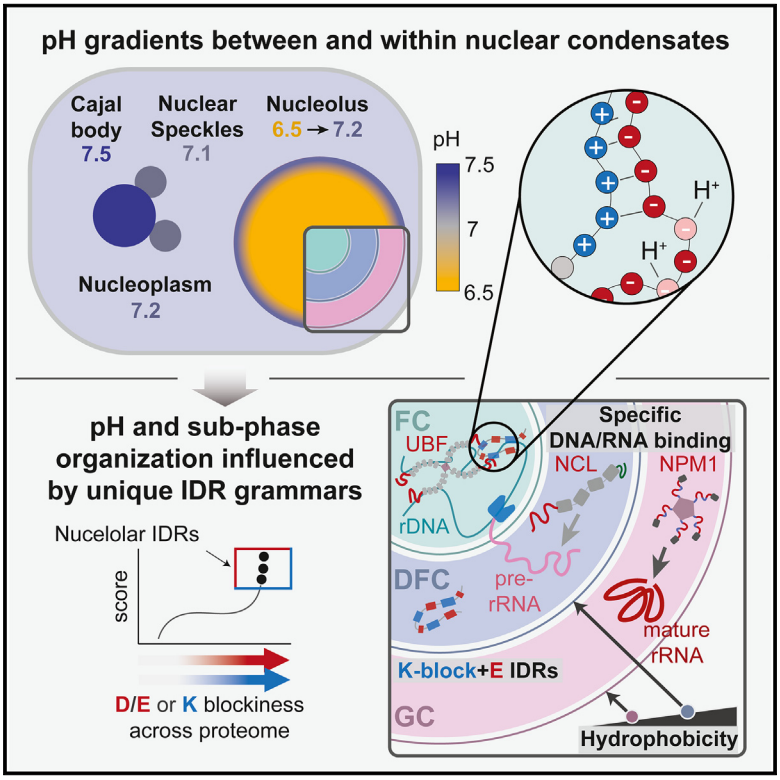
Cell

Macromolecular condensation organizes nucleolar sub-phases to set up a pH gradient

Graphical abstract



Highlights

- Nucleolar IDRs are enriched in D/E tracts and K-blocks interspersed by E-rich regions
- Specificity of complex coacervation and hydrophobicity organize nucleolar layers
- Binding of protons to D/E tracts sets up pH gradients between nucleoli and nucleoplasm
- Condensate-specific pH values in the nucleus correlate with mean net charge of IDRs

Authors

Matthew R. King, Kiersten M. Ruff, Andrew Z. Lin, ..., Emma Lundberg, Michael D. Vahey, Rohit V. Pappu

Correspondence

pappu@wustl.edu

In brief

Proteins containing blocks of charged amino acids in intrinsically disordered regions contribute to the layered organization of nucleoli and establish distinct pH environments in nuclear condensates.

King et al., 2024, Cell *187*, 1889–1906
April 11, 2024
https://doi.org/10.1016/j.cell.2024.02.029

CellPress

Cell

CellPress
OPEN ACCESS

Article

Macromolecular condensation organizes nucleolar sub-phases to set up a pH gradient

Matthew R. King,[1,2] Kiersten M. Ruff,[1,2] Andrew Z. Lin,[1,2] Avnika Pant,[1,2] Mina Farag,[1,2] Jared M. Lalmansingh,[1,2] Tingting Wu,[2,3] Martin J. Fossat,[1,2] Wei Ouyang,[4,5,6] Matthew D. Lew,[2,3] Emma Lundberg,[4,5,6] Michael D. Vahey,[1,2] and Rohit V. Pappu[1,2,7,*]

[1]Department of Biomedical Engineering, James McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO, USA
[2]Center for Biomolecular Condensates, James McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO, USA
[3]Department of Electrical and Systems Engineering, James F. McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA
[4]Department of Bioengineering, Schools of Engineering and Medicine, Stanford University, Stanford, CA, USA
[5]Department of Pathology, School of Medicine, Stanford University, Stanford, CA, USA
[6]Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH—Royal Institute of Technology, Stockholm, Sweden
[7]Lead contact
*Correspondence: pappu@wustl.edu
https://doi.org/10.1016/j.cell.2024.02.029

SUMMARY

Nucleoli are multicomponent condensates defined by coexisting sub-phases. We identified distinct intrinsically disordered regions (IDRs), including acidic (D/E) tracts and K-blocks interspersed by E-rich regions, as defining features of nucleolar proteins. We show that the localization preferences of nucleolar proteins are determined by their IDRs and the types of RNA or DNA binding domains they encompass. *In vitro* reconstitutions and studies in cells showed how condensation, which combines binding and complex coacervation of nucleolar components, contributes to nucleolar organization. D/E tracts of nucleolar proteins contribute to lowering the pH of co-condensates formed with nucleolar RNAs *in vitro*. In cells, this sets up a pH gradient between nucleoli and the nucleoplasm. By contrast, juxta-nucleolar bodies, which have different macromolecular compositions, featuring protein IDRs with very different charge profiles, have pH values that are equivalent to or higher than the nucleoplasm. Our findings show that distinct compositional specificities generate distinct physicochemical properties for condensates.

INTRODUCTION

The nucleolus is the largest membraneless body in the cell.[1] Widely acknowledged as the site of ribosomal biogenesis,[2,3] the nucleolus is "plurifunctional"[1,4] being involved in a range of cellular functions that include synthesis of signal recognition particles,[5–7] DNA replication and repair,[8,9] stress responses,[9,10] viral replication,[11–14] and stem cell differentiation.[15,16]

In higher metazoans, the nucleolus contains at least three nested layers.[17] The layers, which we refer to as sub-phases[18] enable spatial organization of nucleolar components. Ribosomal DNA (rDNA) and the transcriptional machinery are housed in the fibrillar center (FC). The transcribed premature ribosomal RNA (pre-rRNA) is folded and matured in the dense fibrillar component (DFC) and complexed with ribosomal proteins in the granular component (GC). New sub-phases beyond the FC, DFC, and GC have also been described. A recent study identified 12 proteins including unhealthy ribosome biogenesis 1 (URB1) that localize to the DFC periphery (PDFC).[3] A fifth sub-phase refers to the enrichment of key components at the GC-nucleoplasm interface is the nucleolar rim or NR.[19] The GC itself is defined by rRNA- and protein-rich territories that do not overlap with one another[15] giving rise to GC-R and GC-P regions.

Nucleolar sub-phases have different macromolecular densities.[20] Sub-phases of different coexisting densities most likely result from a combination of active processes and spontaneous phase transitions.[18,21–24] These are processes such as phase separation coupled to percolation or complex coacervation.[25,26] The case for spontaneous phase transitions was made by Feric et al.[18] They showed that a ternary mixture comprising the GC protein nucleophosmin 1 (NPM1), a monomeric version of the FC protein fibrillarin (mFBL), and a generic rRNA was sufficient for generation of an inner phase rich in FBL that coexists with an outer phase rich in NPM1.[18]

Nucleolar components are also expected to engage in a combination of site-specific, saturable, stoichiometric binding,[27–30] non-saturable, associative multivalent interactions that give rise to percolation also known as gelation that generates viscoelasticity,[18,23–25,31–33] and complementary electrostatic interactions that drive complex coacervation.[34–37] Therefore, spontaneous

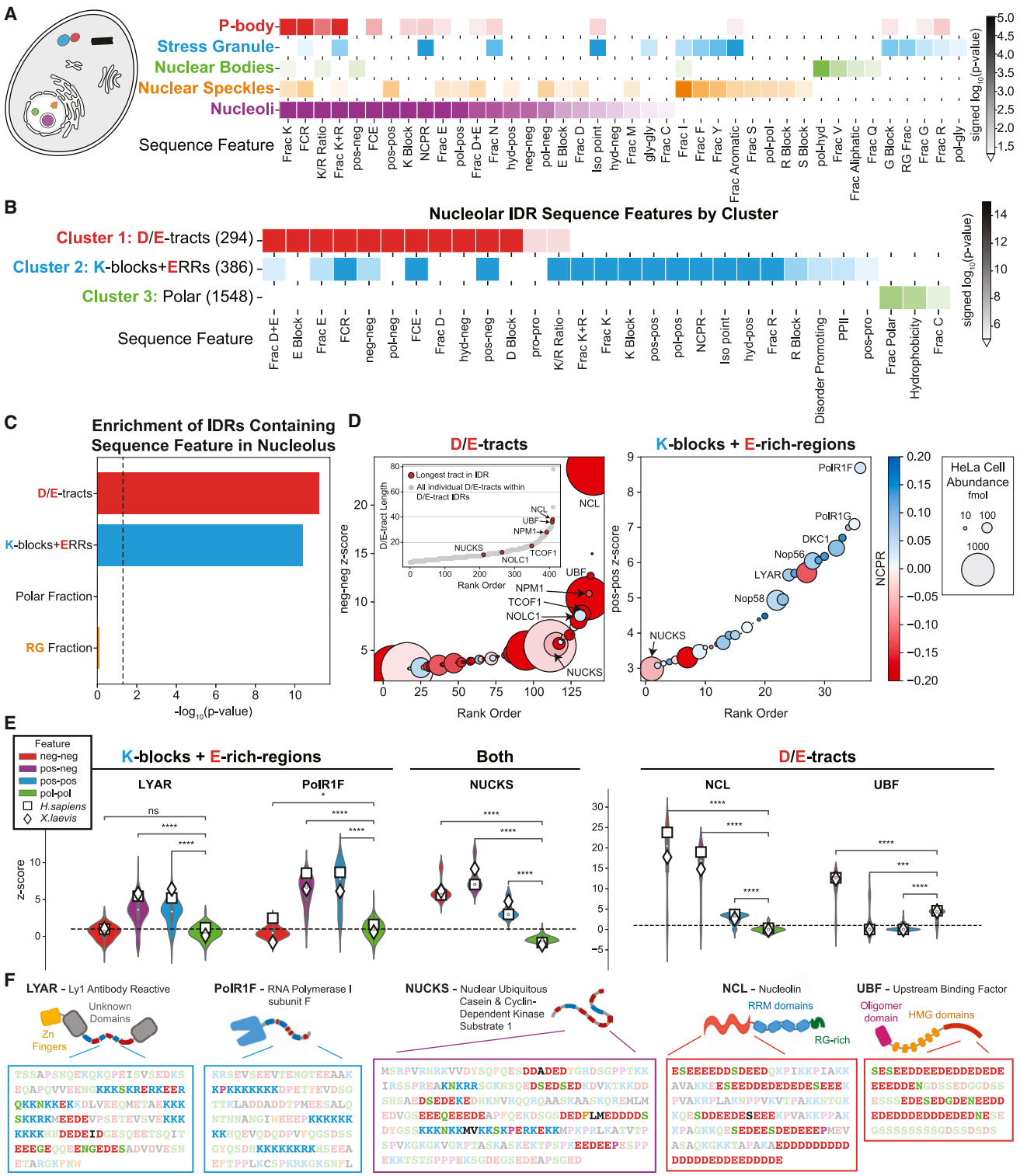


Figure 1. IDRs found in nucleoli are enriched in K-blocks + E-rich regions and D/E tracts

(A) Summary of compositional and patterning features that are enriched in the IDRs across different condensates. Patterning parameters are designated as x-z, where x and z refer to the residue types whose binary pattern is being quantified. p values were calculated using the two-sample Kolmogorov-Smirnov test comparing the distribution of *Z* scores for IDRs in the given condensate versus the remaining IDRs of the human proteome.

(legend continued on next page)

processes that contribute to nucleolar organization are likely to involve phase separation that accounts for solvent-mediated effects, saturable site-specific binding interactions, and non-saturable associative phase transitions such as percolation. We refer to the totality of these processes as *condensation*.

Here, we focus on uncovering the physical driving forces that underlie the generation of nucleolar sub-phases, specifically coexisting FC, and DFC layers. Our work is motivated by differences between pre-mature and mature rRNA (pre-rRNA and mat-rRNA), the distinct substrate binding domains that are present in different nucleolar layers, and the hypothesis that the specificity of intrinsically disordered regions (IDRs) contribute distinct types of multivalent interactions to condensation. Our findings highlight the importance of processes that come under the rubric of condensation. Further, compositional specificities lead to proton gradients making nucleoli more acidic than the surrounding nucleoplasm.

RESULTS

Nucleolar proteins harbor D/E tracts and K-blocks + E-rich regions

Molecular grammars of IDRs refer to distinct non-random compositional biases and sequence patterns.[38,39] These grammars contribute to condensation and compositional specificities of condensates.[40–48] We deployed a computational pipeline to enable unbiased, comprehensive, and comparative assessments of IDR grammars across proteins that localize to distinct condensates. Each IDR in the human proteome was annotated using a 90-component feature vector.[43] Components of feature vectors are *Z* scores that quantify the extents of enrichment of specific amino acid types[39,42,49–51] and extents of non-randomness associated with specific binary sequence patterns.[38,41,51–56]

The computational analysis, referred to as NARDINI+[43], recovered known grammars of IDRs of proteins from cytoplasmic and nuclear condensates[57–61] (Figure 1A). Nucleoli exhibit the highest number of significantly enriched features of the five condensates studied[62] (Figure 1A). Features that are highly enriched in IDRs of nucleolar proteins compared with the remaining human IDRome include a high fraction of charged residues that are organized into linear blocks (Figure 1A). Additionally, the nucleolus is highly enriched in IDRs with high Z scores ($\geq$ 3) for blocky patterns of Lys versus acidic residues (Figure S1B).

Unbiased hierarchical clustering[64] of feature vectors across IDRs from nucleolar proteins yielded three main clusters (Figures 1B and S1A). Of these, the standout clusters refer to D/E tracts defined as tracts of aspartic acid (D) and glutamic acid (E) residues and K-blocks + E-rich regions defined by blocks of lysine (K) residues interspersed by E-rich regions (Figure S1C). A third cluster corresponds to high fractions of polar and/or hydrophobic residues. Neither the frequency of nucleolar IDRs with high polar fraction nor IDRs with high RG contents are statistically enriched when compared with the human IDRome (Figure 1C). Instead, proteins with high-scoring RG/RGG-rich IDRs (*Z* scores > 9) contain either D/E tracts or K-blocks + E-rich regions, or they are in complexes with proteins with these features (Figure S1D). Three such proteins include nucleolin (NCL), FBL, and Dead-box Helicase 21 (DDX21).[65]

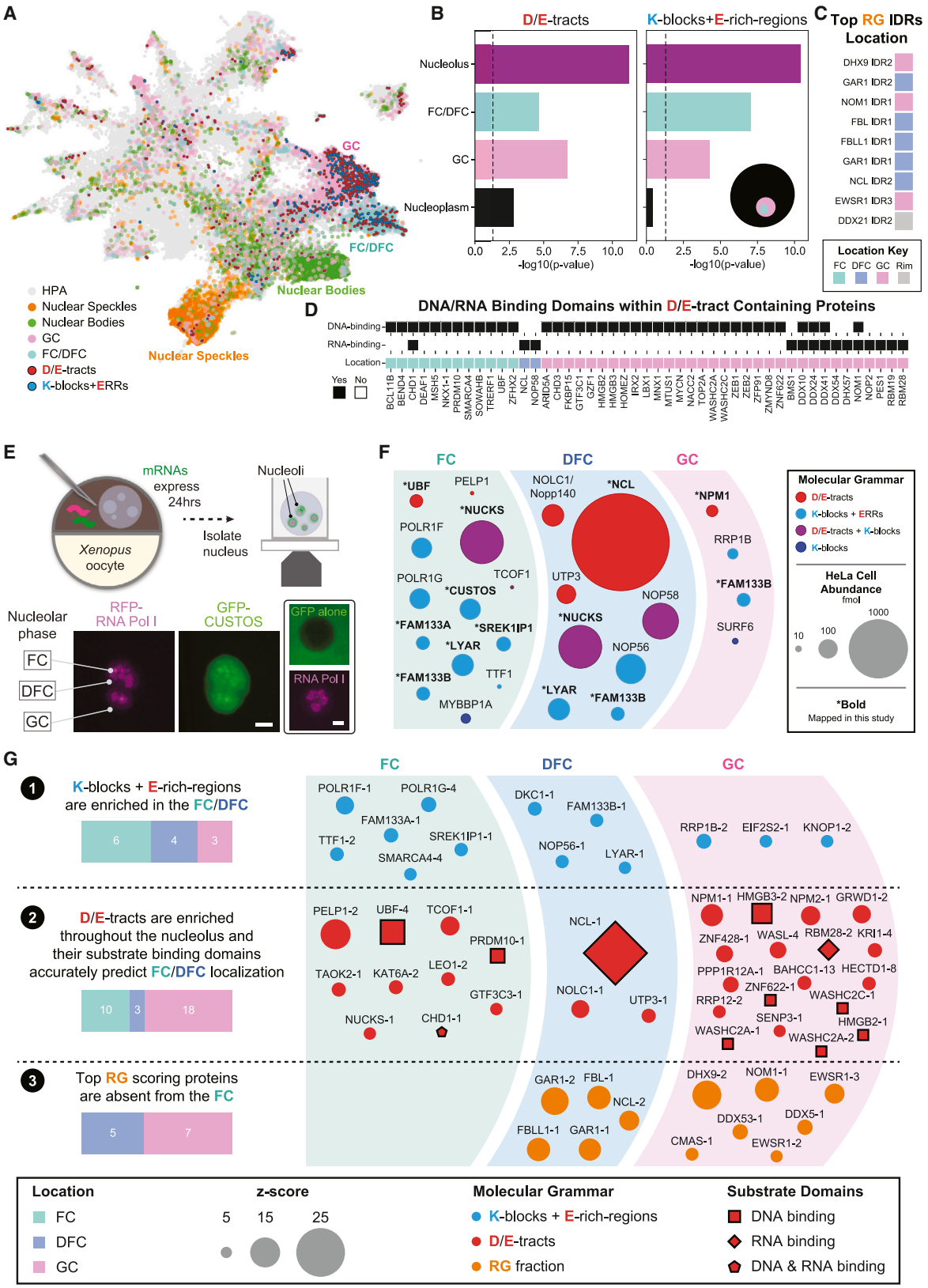
Next, we sorted IDRs from nucleolar proteins in ascending orders of the *Z* scores that quantify non-random segregation of acidic residues into D/E tracts (neg-neg *Z* score) and the non-random segregation of lysine residues into K-blocks (pos-pos *Z* score) (Figures 1D and S1C). Of the 140 D/E tracts across the nucleolar IDRome, three that have the highest neg-neg *Z* scores are found in upstream binding factor (UBF), NCL, and NPM1, which are known markers of the FC, DFC, and GC, respectively.[18,29,66,67] Their D/E tracts are among the longest in nucleolar proteins (Figure 1D inset). Additional proteins with prominent D/E tracts include the fully disordered FC and DFC phosphoprotein (NOLC1/Nopp140), and treacle ribosome biogenesis factor 1 (TCOF1)[49,68,69] (Figures 1D and S1C).

Among the 36 K-blocks + E-rich regions are those found in RNA polymerase I subunits F and G (PolR1F and PolR1G), components of essential nucleolar snoRNPs (DKC1, Nop56, and Nop58),[70] and the NCL-chaperone Ly-1 antibody reactive (LYAR)[71,72] (Figures 1D and S1C). Nucleolar casein kinase substrate (NUCKS)[73] is predicted to be fully disordered, and it is an abundant protein that harbors both D/E tracts and K-blocks + E-rich regions (Figures 1D, S1C, and S1E). Importantly, the patterning *Z* score values are highly conserved (Figures 1E and 1F).[74–77] Next, we sought to map the spatial localizations of proteins within nucleolar sub-phases and identify contributions of the IDR grammars of nucleolar proteins.

Molecular grammars of IDRs and the substrate binding domains to which they are tethered determine intra-nucleolar localization

We analyzed the uniform manifold approximation and projection for dimension reduction (UMAP) representation of the Subcellular Atlas of the Human Protein Atlas.[19,78–82] These maps distinguish proteins that localize to the GC from those that localize to the FC and DFC (termed FC/DFC) (Figure 2A). The GC shows a higher

(B) Significantly enriched compositions and blocky patterns of each cluster of nucleolar IDRs compared with the human IDRome. Each cluster is defined by a distinct set of sequence features which are: (1) D/E tracts, (2) K-blocks + E-rich regions, and (3) high fractions of polar residues.
(C) Fisher exact test of the frequencies of D/E-tracts, K-blocks+ERRs, IDRs with fraction polar *Z* scores $\geq$ 3, and RG-containing IDRs (*Z* score $\geq$ 3) in the nucleolus compared with the rest of the human IDRome. The dashed vertical line corresponds to a p value of 0.05.
(D) D/E-tract and K-blocks + E-rich-regions IDRs sorted in descending orders of their neg-neg and pos-pos *Z* score, respectively. Circle size denotes abundance in HeLa cells[63] and circle color refers to the net charge per residue (NCPR); the names of key nucleolar proteins are indicated. Inset shows the lengths of individual D/E-tracts within nucleolar proteins.
(E) Conservation of top-scoring D/E-tract and K-block + E-rich-regions IDRs across Metazoan species. Violin plots show the range of *Z* scores across Metazoan species. Polar-polar patterning is used as a control since this feature was not found to be a distinct feature of nucleolar IDRs; *Z* scores of the human and frog IDRs are shown explicitly to emphasize the conservation across these two species. Significance was determined using the Wilcoxon test with *: 1.00e−02 < p $\leq$ 5.00e−02, **: 1.00e−03 < p $\leq$ 1.00e−02, ***: 1.00e−04 < p $\leq$ 1.00e−03, and ****: p $\leq$ 1.00e−04.
(F) Schematic of top scoring D/E-tract and K-block + E-rich-region containing proteins and their IDR sequences.
See also Figure S1.



(legend on next page)

enrichment of D/E tracts when compared with the FC/DFC. K-blocks + E-rich regions are more enriched in the FC/DFC when compared with the GC (Figure 2B), and high-scoring RG-rich IDRs are excluded from the FC (Figure 2C). Next, we examined all nucleolar proteins with DNA- or RNA-binding domains (DBDs/RBDs) that contain a D/E tract.[83] Of these proteins, the ones that localize to the FC contain at least one DBD, whereas those that localize to the DFC contain RBDs (Figure 2D). Interestingly, proteins with D/E tracts that localize to the GC feature either DBDs and RBDs and sometimes both.

The extracted grammars provide a first-order predictor of localization preferences of nucleolar proteins. We tested these using direct mapping in cells (Figure S2A). Leveraging conserved grammars across metazoans, we mapped locations of proteins with high-scoring D/E tracts and K-blocks + E-rich regions using live *Xenopus laevis* germinal vesicles (GVs) obtained from oocytes expressing green or red fluorescent protein (FP)-tagged proteins of interest from micro-injected mRNAs. Although FPs alone are mostly excluded from nucleoli, all FP-tagged proteins with high-scoring K-blocks + E-rich regions, and D/E tracts readily localized to nucleoli (Figures 2E, 2F, and S2B–S2G). By contrast, FP-tagged Huntingtin (Httex1) was excluded to the same extent as free FPs (Figures S2H–S2K).

The bioinformatics and *in vivo* analyses point to key molecular grammars that determine sub-phase localization of nucleolar proteins (Figure 2G). First, proteins harboring high-scoring K-blocks + E-rich regions are enriched in the FC/DFC followed by the GC. Second, proteins containing D/E tracts are enriched throughout the nucleolus, and their FC versus DFC versus GC localization is governed by whether RBDs or DBDs are tethered to the D/E tracts. Specifically, proteins with D/E tracts and RBDs are excluded from the FC. For example, NCL features a prominent D/E tract, four RNA recognition motifs (RRMs), and an RG-rich C-terminal IDR. UBF also features a prominent D/E tract with six DBDs, in the form of high mobility groups (HMGs), tethered N-terminally to the D/E tract. These differences in grammars lead to UBF localizing to the FC, whereas NCL is excluded from the FC.

UBF, rDNA, and proteins containing K-blocks + E-rich regions drive FC condensation

We investigated eight FC-localized proteins for their ability to promote condensation in cells. If the FC is a *bona fide* condensate, then its size should increase as the copy numbers of scaffold molecules are increased.[18,84] FC enlargement resulted from increased copy numbers of UBF mRNA, and this was dependent on mRNA dosage (Figures 3A, 3B, S3A, and S3B). Of the other seven proteins, only increases in TCOF1 and LYAR resulted in FC enlargement, and the enlargement was modest when compared with what was realized upon increasing UBF levels (Figures 3A–3D). Increased copy numbers of LYAR, CUSTOS, or SREK1IP1 mRNA led to micron-sized puncta that do not overlap with nucleoli (Figures 3C, S3C, and S3D). We found significant FC enlargement (p value $< 10^{-4}$) in *Xenopus* oocytes injected with 10 ng UBF mRNA or 60 ng TCOF mRNA (Figures S3E and S3F). In contrast to UBF, which localizes to the FC regardless of mRNA load, TCOF1 localized exclusively to the FC only when 60 ng of mRNA was injected. At lower mRNA loads, TCOF1 localized to both the FC and GC (Figure S3G). Overall, the results point to UBF as the key scaffold of the FC. Proteins such as LYAR, with their K-blocks + E-rich regions, contribute as co-factors to the stabilization of FCs. We tested these inferences via *in vitro* reconstitutions using purified protein and nucleic acid components (Figure S3H).

Full-length UBF forms tetramers in solution (Figures 3E, S3I, and S3J). Tetramers and higher-order oligomers generate multivalency. The numbers of DNA binding HMGs and D/E tracts increase multiplicatively by factors of 6n and n, respectively, where n is the number of UBF molecules per oligomer. Next, we investigated the condensation ability of UBF alone and in two-component mixtures with rDNA (Figure 3F). Similar measurements of phase boundaries were performed for LYAR ± rDNA and NUCKS ± rDNA (Figures S3K and S3L). In all measurements of condensation with just proteins or binary mixtures with rDNA, the threshold concentrations of the relevant proteins required for forming condensates were above their endogenous levels (Figure 3G). Furthermore, in binary mixtures with UBF, rDNA is interoperable with an orthogonal, non-ribosomal DNA (Figure S3M).

By contrast, to one- and two-component systems, the threshold concentrations of each of rDNA, UBF, and LYAR required to form *in vitro* condensates in three-component mixtures are below their endogenous levels (Figures 3I and 3J). However, if we fix the rDNA and UBF levels to be at endogenous levels, the concentration of NUCKS required for condensation in ternary mixtures is higher than their endogeneous concentration (Figures 3J and S3L). These results suggest that the higher valence of K-blocks and E-rich regions in chaperoning proteins such as LYAR enables it to stabilize condensate formation with UBF and rDNA more efficiently than NUCKS. Based on measured threshold concentrations, the ability

Figure 2. Proteins with K-block + E-rich regions preferentially localize to the FC, whereas the top RG-rich IDRs are depleted from the FC

(A) UMAP representation of the Human Protein Atlas (HPA) with nuclear speckles, nuclear bodies, GC, and FC/DFC highlighted. Proteins containing the 140 distinct D/E tracts and 36 K-blocks + E-rich regions (ERRs) are shown in red and blue, respectively.

(B) Fisher's exact test of the frequencies of D/E tracts and K-blocks + E-rich-regions IDRs in the nucleolus, FC/DFC, GC, and nucleoplasm compared with the remaining human IDRome. The vertical line corresponds to a p value of 0.05.

(C) Localization (from HPA) of the top nine scoring RG-rich IDRs in the nucleolus.

(D) Examination of all DNA- or RNA-binding proteins that also contain D/E tract IDRs within the FC/DFC/GC. Both location and whether the protein contains a DNA- or RNA-binding domain are shown.

(E) Schematic of method for obtaining live GVs from *Xenopus* oocytes and imaging nucleoli therein (see STAR Methods for details). Representative image of a nucleolus co-expressing GFP-tagged CUSTOS/c12orf43 or GFP with RFP-tagged PolR1E (subunit of RNA polymerase I with known preferential FC localization).

(F) Spatial map of proteins with high-scoring K-blocks + E-rich regions and D/E tracts IDRs in the FC/DFC/GC phases of the nucleolus.

(G) Summary schematic showing localization of K-blocks + E-rich regions, D/E tracts, and RG-rich IDRs with Z score ≥ 5.5.

See also Figure S2.

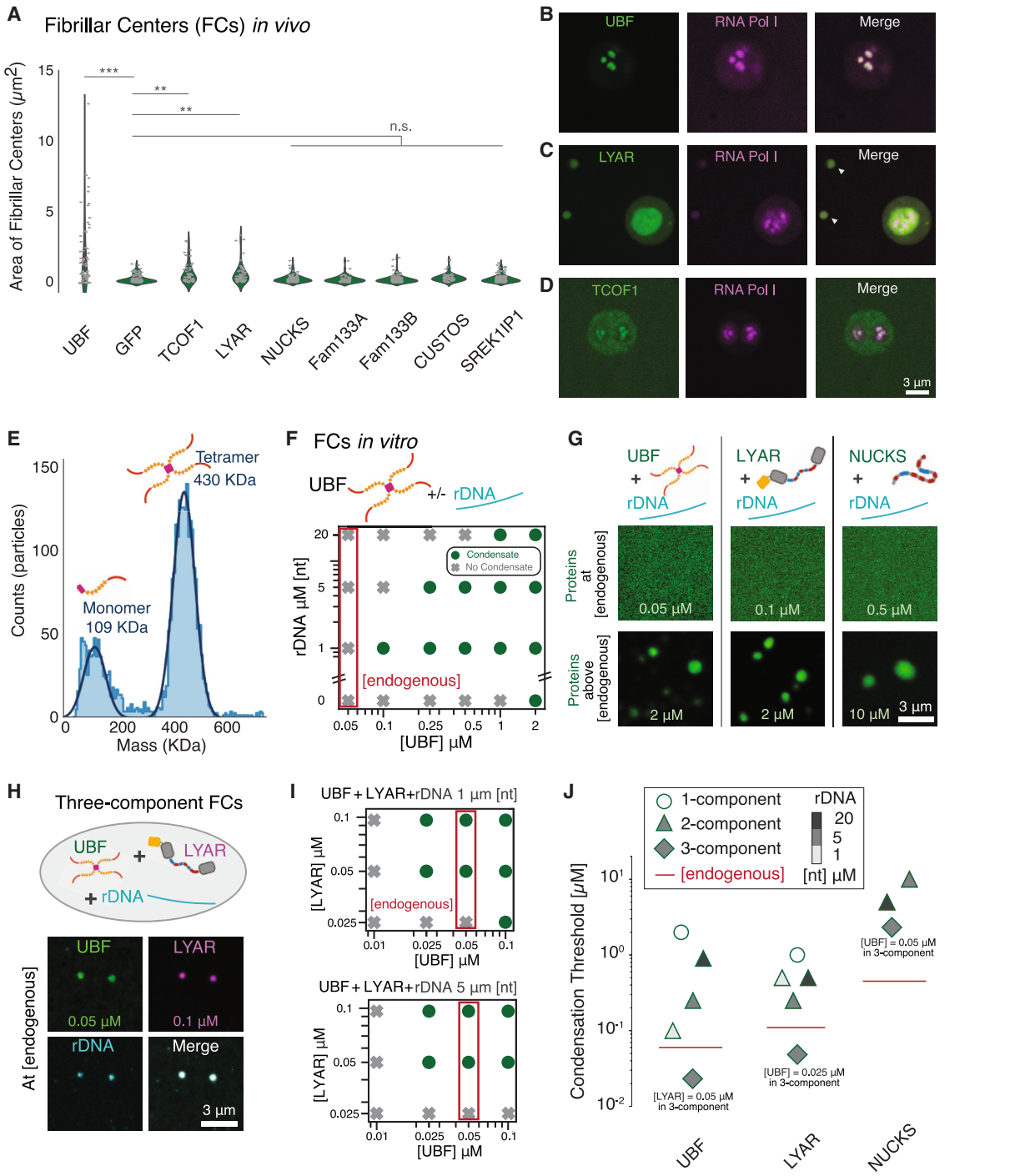


Figure 3. UBF, rDNA, and proteins containing K-blocks + E-rich regions proteins drive the formation of FC-like condensates

(A) Violin plots of FC size from *Xenopus* oocytes monitored upon injection of 20 ng of mRNA for each of the proteins. The sizes were monitored following protein expression and leveraging the GFP tags of each of the expressed proteins; n.s. (not significant) denotes p values > 5.00e−02; **p values < 1.00e−03, ***p value < 1.00e−04.

(B–D) (B) Representative images of oocyte nucleoli expressing 20 ng mRNA of GFP-tagged UBF, (C) RFP-tagged TCOF1, (D) GFP-tagged LYAR; arrowheads indicate puncta or incidental condensates[85] of GFP-LYAR that form at high copy numbers.

(E) Mass photometry trace of purified UBF showing raw counts (light blue histogram) and calculated distribution (dark blue line); the peak mass values indicate two populations corresponding to the masses of UBF monomers and tetramers.

(legend continued on next page)

of FC proteins to co-condense with rDNA follows the hierarchy of UBF > LYAR >> NUCKS.

FC/DFC-like organization is reproduced *in vitro* with quinary mixtures comprising UBF, rDNA, NCL, LYAR, and pre-rRNA

We sought to determine how a DFC-like sub-phase might organize spontaneously around an FC-like condensate. We focused on NCL as a dominant DFC scaffold.[66,67] NCL is abundant, harbors multiple RRMs, and contains a high-scoring D/E tract (Figure 4A). Our mapping in cells showed that NCL is excluded from the FC and localizes preferentially to the DFC (Figure 4B), which is consistent with its role in processing pre-rRNA.

Recombinant NCL forms well-mixed condensates with pre-rRNA and LYAR *in vitro* (Figure 4C). We included LYAR because it has an NCL-chaperoning role.[86] In quinary mixtures with the UBF:rDNA:LYAR:NCL:pre-rRNA ratio set at endogenous levels,[63] we observed two coexisting phases. The "core" comprises rDNA and UBF, and the "shell" is enriched in NCL and pre-rRNA (Figures 4D, 4E, S4A, and S4B). This shows that *in vitro* facsimiles of coexisting FCs and DFCs can be reconstituted with a minimum of three proteins, UBF, LYAR, and NCL, and two native nucleic acids, namely, rDNA and pre-rRNA. Inhibiting pre-rRNA production with drugs that target RNA Pol I causes FCs to extrude out into the nucleoplasm while partially wetting a DFC-like region[17,87] (Figure 4F). The inside-out extrusion phenomenon is reproduced *in vitro* by excluding pre-rRNA from the quinary mixture (Figures 4G, 4H, and S4C).

Driving forces for co-condensation of NCL versus NPM1 are specific for the type of rRNA

Unlike pre-rRNA, mat-rRNA is folded and matured via a series of modifications that include 2′ methylation and pseudouridinylation.[88] We investigated the relative abilities of NCL to co-condense with purified versions of 18S pre-rRNA versus mat-rRNA, which was a mixture of 18S and 28S molecules. NCL co-condenses at endogeneous concentrations with pre-rRNA, but it does not co-condense with mat-rRNA (Figures 4I and 4J). In contrast to NCL, purified NPM1 weakly co-condenses with pre-rRNA, but it readily co-condenses with mat-rRNA (Figures 4K and 4L).

Condensates formed by mixtures of NCL and pre-rRNA exhibited similar extents of wetting on hydrophobic and hydrophilic surfaces[18,89] (Figures 4M, 4N, 4SD, and 4SE). By contrast, condensates formed by mixtures of NPM1 and mat-rRNA adsorb onto and wet hydrophilic surfaces, while they are repelled by hydrophobic surfaces (Figures 4M, 4N, 4SD, and 4SE). Therefore, condensates formed by mixtures of NPM1 and mat-rRNA are hydrophilic whereas condensates formed by mixtures of NCL and pre-rRNA are amphiphilic. Taken together with the work of Feric et al.[18] and Li et al.,[90] our findings highlight the hydrophilicity of NPM1-based condensates.

Complex coacervation contributes to nucleolar stability and condensation

Complex coacervation is driven by a combination of complexation of oppositely charged macromolecules,[26,34–37,91–93] release of counterions,[94–97] networking of macromolecules enabled by multivalency,[35,98] and diminished solubilities of higher-order complexes.[99] Salting-in behavior is a signature of complex coacervation. This refers to the dissolution of condensates as salt concentrations increase. To test if nucleoli in *Xenopus* oocyte GVs dissolve at high salt concentrations, we injected *Xenopus* GVs with either an isotonic or high salt solution. Isotonically injected GVs were indistinguishable from non-injected GVs, whereas in GVs subjected to high salt injections that brought the final KCl concentration to 200 mM, the core nucleolar proteins UBF and NCL no longer showed partitioning into nucleoli after 5 min (Figures 5A and 5B). This is suggestive of complex coacervation playing a role in stabilizing nucleoli in cells. We also titrated the concentrations of monovalent salts *in vitro* and discovered system-specific thresholds for salt concentrations above which the different FC and DFC components do not form condensates (Figures 5C–5F).

D/E tracts enhance asymmetrical complex coacervation of nucleolar components

We performed separate *in vitro* experiments using full-length wild-type (WT) UBF and a construct lacking the D/E tract (ΔD/E-tract UBF) as well as full-length WT NCL and a construct lacking its D/E tract (ΔD/E-tract NCL). For symmetrical complex coacervation, the expectation would be that the driving forces for condensation of UBF with rDNA in two- and three-component mixtures would be enhanced upon deletion of the D/E tract. Instead, the ability of ΔD/E-tract UBF to co-condense with LYAR and rDNA, as measured by the threshold concentration required for condensation, was weakened by 100-fold (Figures 5G, S5A, and S5B). In binary mixtures with equivalent concentrations of rDNA, the threshold concentration of ΔD/E-tract UBF required to drive condensation was 20-fold higher than the threshold concentration of the full-length UBF (Figures 5H,

(F) Two-component phase boundaries of a mixture comprising rDNA and UBF. Here, green circles indicate the presence of condensates, and gray x symbols indicate a one-phase regime.

(G) Confocal images of two-component mixtures consisting of rDNA (5 μM [nt]) (untagged) and UBF, LYAR, or NUCKS (fluorescently tagged) at their endogenous concentrations or at high concentrations (indicated).

(H) Representative confocal image of condensates formed in three-component mixtures of rDNA (5 μM [nt]), UBF (0.05 μM), and LYAR (0.1 μM). The proteins are at their endogenous concentrations.[63]

(I) Phase boundaries of three-component mixtures consisting of LYAR and UBF at concentrations corresponding to their endogenous concentration that contain either 1 μM rDNA (top) or 5 μM rDNA (bottom).

(J) Threshold concentrations for condensation of recombinant UBF, LYAR, and NUCKS either alone (one-component, unfilled circle), with increasing concentrations of rDNA (two-component, filled triangles), or with three components (filled diamonds). In three-component mixtures, the threshold concentrations of UBF and LYAR lie below the endogenous levels of the relevant proteins (red, horizontal bar).[63] Condensation was not observed for NUCKS alone or with 1 μM [nt] rDNA. Scale bars for all images, 3 μm.

See also Figure S3.

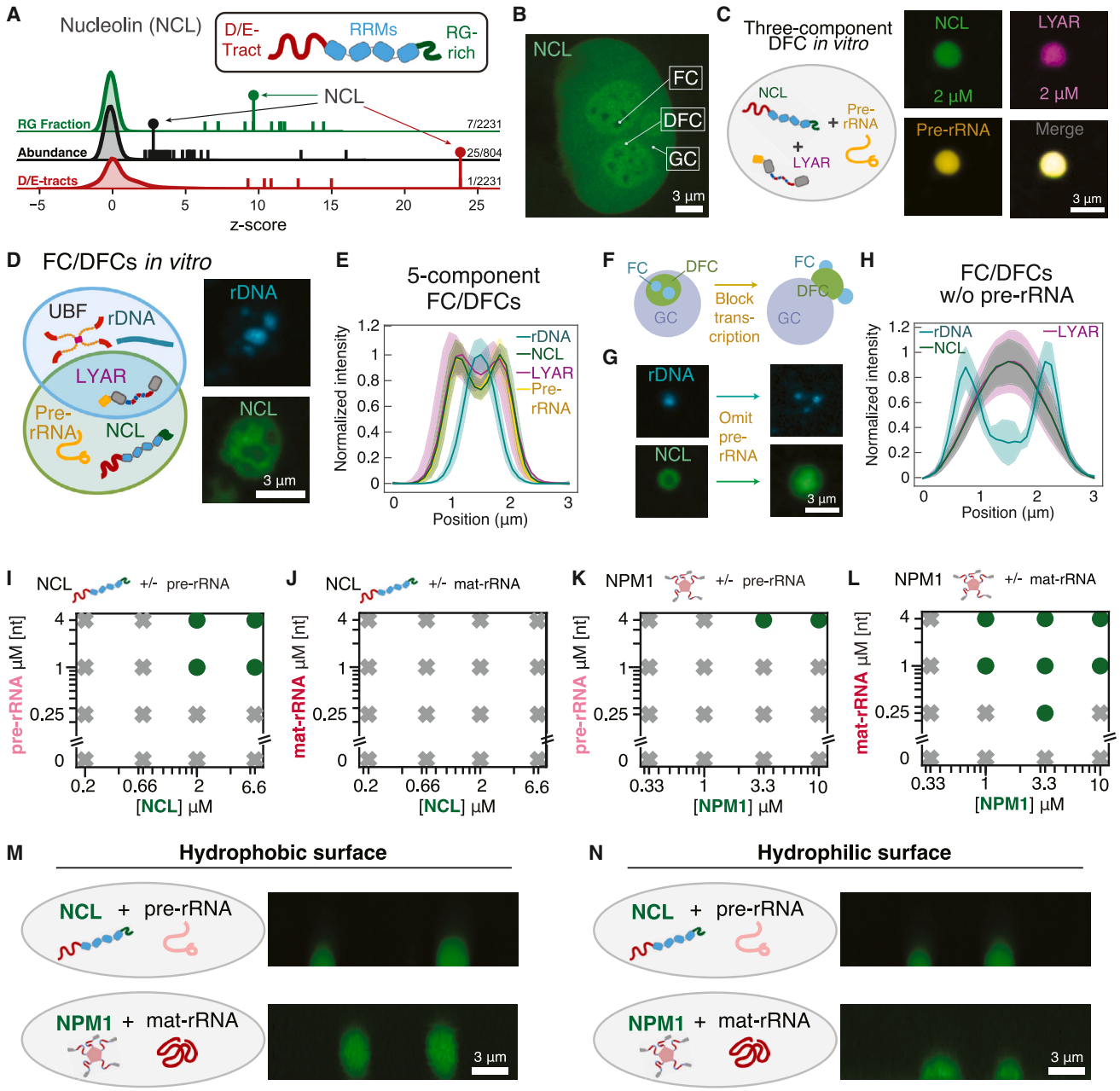


Figure 4. Substrate-specific condensation and hydrophobicity/hydrophilicity of condensates contribute to organization of nucleolar subphases

(A) Schematic of NCL and density distribution graphs showing the relative rank of NCL for fraction of sequence that contains an RG block among all nucleolar IDRs (7th out of 2,231), abundance among all nucleolar proteins (25th out of 804), and D/E tract (neg-neg patterning) among all nucleolar IDRs (1st out of 2,231). The tall/capped marker indicates NCL; vertical stripes indicate all proteins ranking above NCL and the five closest proteins ranking below it.

(B) Representative images of nucleoli from a living *Xenopus laevis* oocyte expressing GFP-tagged NCL. This highlights the exclusion of NCL from FCs and localization to the DFC and GC.

(C) Schematic and images of three-component mixtures of NCL, LYAR, and pre-rRNA as minimal *in vitro* reconstitutions of the DFC.

(D) Schematic of the quinary mixture and images of multiphase condensates reconstituted *in vitro*. UBF always colocalizes with rDNA (see Figure 3G).

(E) Line scan analysis of condensates formed by the quinary mixture shown in (C). Solid curves correspond to median values, and shaded envelopes correspond to the 95% confidence intervals.

(F) Schematic of nucleolar re-organization resulting from transcription inhibition.

(G) Images of rDNA and NCL in condensates formed by quinary mixture and a ternary mixture where pre-rRNA is omitted.

(H) Line scan analysis of condensates in the ternary mixture. Curves and shadings are similar to panel (E).

(I–L) Two-component phase diagrams of NCL or NPM1 with either pre-rRNA or mat-rRNA.

(legend continued on next page)

S5A, and S5B). Similarly, when compared with an equivalent ternary mixture with NCL, the threshold concentration of ΔD/E-tract NCL that is required for condensation increases 15-fold in three-component mixtures comprising ΔD/E-tract NCL, LYAR, and pre-rRNA. In binary mixtures, the threshold concentration of ΔD/E-tract NCL increased by 7-fold in binary mixtures with pre-rRNA when compared with mixtures of WT NCL and pre-rRNA (Figures 5I, 5J, S5C, and S5D). These data show that D/E-tracts contribute to asymmetrical complex coacervation.

K-blocks contribute cohesive interactions to condensation of FC and DFC components

Next, we tested if K-blocks contribute to cohesive interactions with proteins that feature D/E tracts. We chose NUCKS because it is a fully disordered, polyampholytic protein with K-blocks + E-rich regions that lacks either RBDs or DBDs. This allowed us to titrate the contributions of charge blocks without being concerned about specific binding domains. Adding 2 μM NUCKS to the binary mixture of UBF and rDNA leads to co-condensation in the ternary system. By contrast, we did not observe condensate formation in the ternary system when we added up to 20 μM of ΔK-rich NUCKS to the UBF + rDNA mixture (Figures 5K and S5E). Here, ΔK-rich NUCKS refers to a variant of NUCKS where the larger of two regions containing K-blocks has been deleted. Addition of 1 μM of NUCKS to a mixture of NCL and pre-rRNA drives condensate formation in this ternary system. By contrast, the ternary mixture of NCL, pre-rRNA, and ΔK-rich NUCKS does not form condensates under the range of conditions we investigated (Figures 5L and S5F). These data suggest that proteins with K-blocks + E-rich regions enable condensation through their K-blocks when added to low-concentration mixtures of UBF and rDNA or NCL and pre-rRNA.

To uncover the molecular-scale contributions of K-blocks + E-rich regions, we used a combination of atomistic and coarse-grained simulations and experiments that were performed *in vitro* and in cells. We chose WT NUCKS and ΔK-rich NUCKS as well as the C-terminal IDR in the PolR1F for our investigations. Polymerase I is an essential FC factor, and the IDR on subunit F has the highest pos-pos *Z* score in the human proteome. We performed atomistic simulations of the PolR1F IDR, WT NUCKS, and ΔK-rich NUCKS using the ABSINTH implicit solvation model and forcefield paradigm.[100–103] Complementary electrostatic interactions between oppositely charged blocks drive intramolecular compaction in each of these systems[53] (Figures 6A–6C, S6A, and S6B). The degree of compaction is an indicator of the strengths of interactions between K-blocks and E-rich regions. The PolR1F IDR features the strongest cohesive and complementary interactions between oppositely charged blocks. This is followed by WT NUCKS and ΔK-rich NUCKS.

Next, we performed lattice-based coarse-grained simulations[104] of ternary mixtures comprising UBF, rDNA, and one of PolR1F IDR, WT NUCKS, or ΔK-rich NUCKS (Figures 6D, S6C, and S6D). From the simulations, we quantified partition coefficients (PCs) for PolR1F IDR, WT NUCKS, or ΔK-rich NUCKS into condensates formed by models that mimic mixtures of UBF and rDNA (Figures 6E and S6E–S6J). The computed PCs are largest for PolR1F IDR and smallest for ΔK-rich NUCKS. This hierarchy of PCs was also found *in vitro* (Figures 6F and 6G) and in cells (Figures 6H and 6I). The PCs *in silico* correlate positively with the *Z* scores for K-blocks in the different constructs (Figure 6J). Likewise, we observe positive correlations between (1) PCs measured *in vitro* and those computed *in silico* and (2) the PCs measured in cells and *in vitro* (Figures 6K–6M). These results provide a physicochemical rationale for K-rich nucleolar localization signals.[42,105]

Nucleolar molecular grammars set up a pH gradient with respect to the nucleoplasm

To understand the contributions of the 140 nucleolar D/E-tract IDRs found in nucleoli, we drew motivation from studies showing that poly-acids such as D/E tracts are likely to be carriers of protons.[106–108] Binding of protons to Asp/Glu residues lowers the net negative charge, thus reducing electrostatic repulsions among D/E tracts. It also enables a lowering of entropic penalties associated with organizing solvent molecules around charged residues. These considerations cause upshifted pK_a values for Asp/Glu residues in D/E tracts.[106,109–111] If this is prevalent in FC/DFC-like condensates, then an emergent consequence of condensation would be a lowering of the pH inside condensates with respect to the coexisting dilute phase, which is the definition of an interphase electrochemical potential.

We measured pH values inside and outside *in vitro* condensates using ratiometric, pH-sensitive fluorescent dyes based on Seminaphtharhodafluor (SNARF),[112] specifically SNARF-4 and SNARF-1 dyes, and cross-validated our results using pHluorin2.[113] We performed a scan of the emission spectrum and tabulated the ratio of emission intensities (r_{GR}) in the green-to-red regions of the emission spectrum as a function of pH (Figures S7A–S7D). These ratiometric calibrations allowed us to use SNARF dyes to quantify differences in pH between coexisting dense and dilute phases (Figure S7E). To benchmark inferences from ratiometric calibrations, we used an electrochemical probe and measured the pH in the undiluted nucleoplasm of *Xenopus* oocytes to be 7.2 (Figure S7F). The ratiometric emission intensity $r_{579:646}$ for SNARF-4 in the *Xenopus* nucleoplasm matches the value extracted using the direct pH measurement (Figure S7G), lending confidence in the use of SNARF-4 for measurements of pH. SNARF-1 is also robust in terms of the relative pH values it yields (Figure S7H).

Using the SNARF-4 dye we find that condensates formed by NCL and pre-rRNA have median pH values of ~6.8, which are ~0.25 units lower than the dilute phase (Figure 7A). Likewise,

(M–N) Sagittal slices of three-dimensional (3D) confocal images of condensates formed by mixtures of NCL and pre-rRNA (top row) or NPM1 and mat-rRNA (bottom row) resting on coverslips with (M) hydrophobic or (N) hydrophilic surface treatments. For all quantifications, n = 25 condensates. Scale bars for all images, 3 μm. Concentrations of UBF, NCL, and NPM1 were set at endogenous levels. The concentration of LYAR was 2 μM in mixtures shown in (C) and (D) and endogenous levels for those shown in (G);
See STAR Methods for additional details.
See also Figure S4.

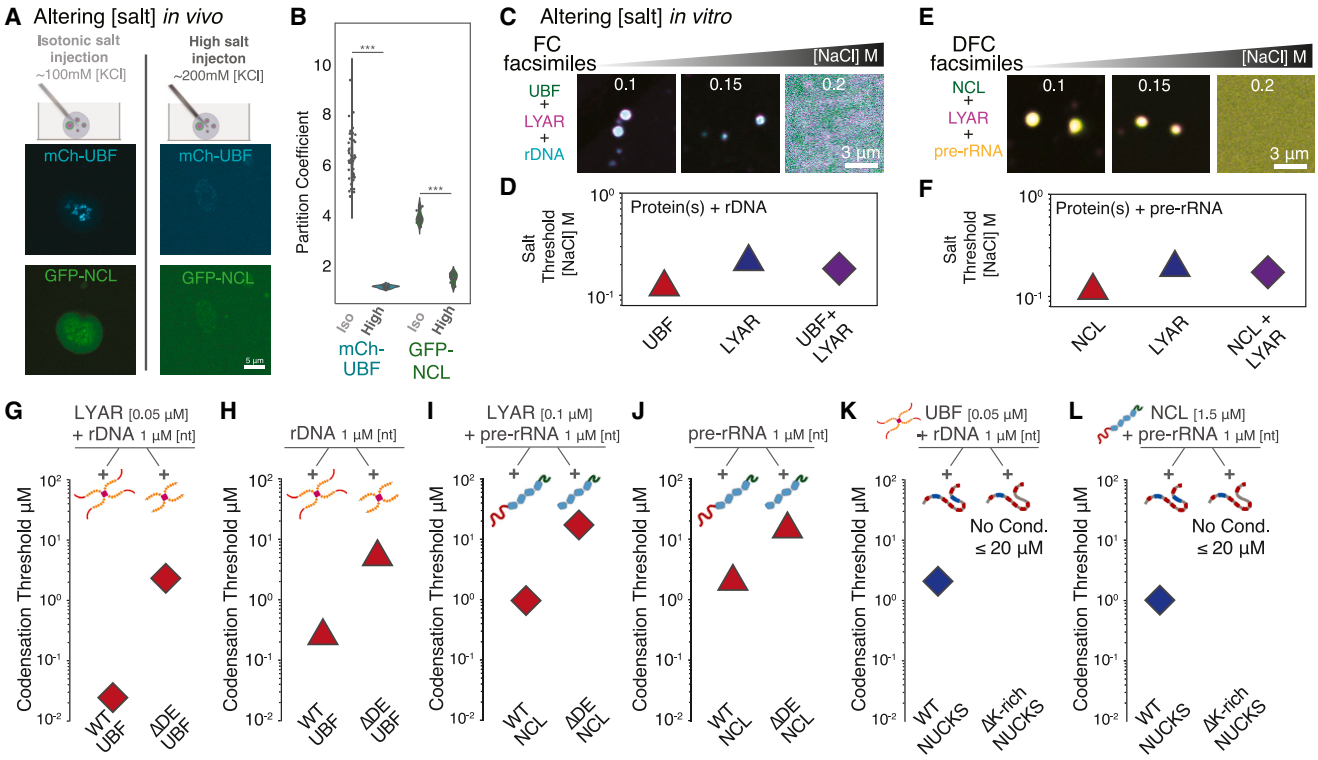


Figure 5. Complex coacervation underlies condensation of nucleoli in cells and nucleolar components *in vitro*

(A) Schematic of injections of isotonic (light gray) or high salt (dark gray) solutions into isolated *Xenopus* GVs expressing mCherry-UBF and GFP-NCL; resultant KCl concentrations in GVs are predicted to be 100 and 200 mM, respectively. Confocal images shown are taken 5 min after injections.

(B) Quantification of partition coefficients of mCherry-UBF in FCs and GFP-NCL in FC/DFCs relative to the surrounding nucleoplasm ($n > 20$); isotonic and high salt treatments are compared, ***p value < 1.00e−05.

(C) Salting-in behaviors of three-component condensates comprising UBF (0.1 μM), LYAR (0.25 μM), and rDNA (5 μM [nt])) measured at the indicated salt concentrations. Representative confocal images of all three components (merge) are shown.

(D) Threshold salt concentration for condensation of indicated two- and three-component mixtures.

(E) Salting-in behaviors of three-component condensates comprising NCL (2 μM), LYAR (0.25 μM), and pre-rRNA (5 μM [nt])) measured at the indicated salt concentrations. Representative confocal images of all three components (merge) are shown.

(F) Threshold salt concentration for salting-in of condensates indicated for two- and three-component mixtures.

(G and H) (G) Condensate thresholds of UBF and ΔD/E-UBF in three-component condensates with rDNA (5 μM [nt]) and LYAR (0.1 μM) or (H) two-component condensates with rDNA (5 μM [nt]).

(I and J) (I) Condensate thresholds of NCL and ΔD/E-NCL in three-component condensates with pre-rRNA (5 μM [nt]) and LYAR (0.1 μM) or (J) two-component condensates with pre-rRNA (5 μM [nt]).

(K) Condensation thresholds of WT NUCKS and ΔK-rich NUCKS in three-component condensates with rDNA (5 μM [nt]) and UBF (0.05 μM) or (L) in three-component condensates with pre-rRNA (5 μM [nt]) and NCL (1.5 μM).

See also Figure S5.

the condensates formed by NPM1 have median internal pH values of ~6.5, and this is ~0.6 units lower than the median pH of the coexisting dilute phase (Figure 7B).

In condensates formed by ternary mixtures with LYAR, the median values of measured internal pH in NCL + pre-rRNA + LYAR condensates increases as the dose of LYAR increases, becoming equivalent to the buffer pH of 7.2 at a LYAR concentration of 5 μM (Figure S7I). For a direct assessment of the contributions of D/E tracts, we used the ΔD/E-tract NCL construct. In ternary mixtures containing NCL or ΔD/E-tract NCL with pre-rRNA and LYAR, the median pH within condensate interiors containing NCL was 6.86, whereas the median pH of condensates containing ΔD/E-tract NCL was larger than 7.0 (Figures 7C and 7D), and the gap between the median pH for dense and dilute phases was not statistically significant. These data suggest that D/E tracts are the carriers of protons into condensates formed by nucleolar components.

Next, we measured the pH of nucleoli in GVs from *Xenopus* oocytes using SNARF-4 dyes. Scanning the emission spectrum of micro-injected SNARF dyes (Figures 7E, 7F, and S7J), we find a pH gradient that increases within the GC, from values of ~6.5 within the FC and DFC to a value of 7.2 in the nucleoplasm (Figures 7E–7H, S7K, and S7L). The FC/DFC have apparent pH values of 6.6 ± 0.1 (Figure S7K). The pH of inner half of the GC is akin to that of the FC and DFC, whereas a sharp gradient is set up within outer half of the GC such that the pH increases to converge to the value measured in the nucleoplasm (Figure S7L). Overall, the FC/DFC, the internal half of the GC, the external

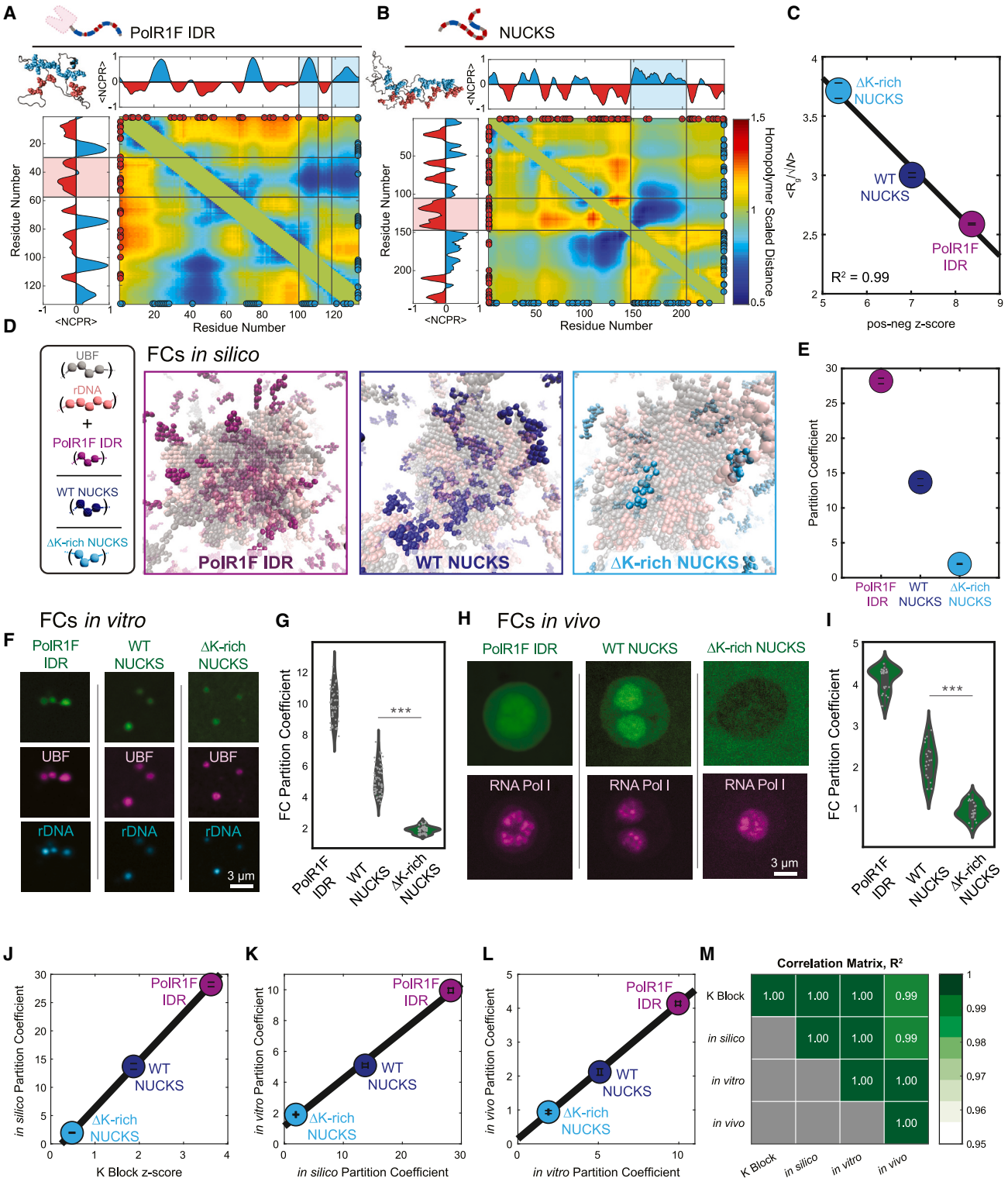


Figure 6. K-blocks within FC/DFC proteins drive enrichment and partitioning into FC in cells and their facsimiles *in vitro*

(A and B) Sequence-specific inter-residue distance maps from atomistic simulations shown as normalized quantities, with respect to the best fit homopolymer model for PolR1F IDR (A) and WT NUCKS (B). Positive and negative residues are shown as circles in blue and red, respectively.

(C) Comparison of positive-negative (pos-neg) patterning *Z* score and sequence length (N) normalized radius of gyration (R_g) (R^2 = 0.99 for linear regression). Error bars indicate standard errors from the mean across five independent simulations.

(legend continued on next page)

half of the GC, and the nucleoplasm have median pH values of 6.5 ± 0.1, 6.6 ± 0.1, 6.8 ± 0.1, and 7.2 ± 0.1, respectively (Figure 7H).

The *in vitro* data suggest that increasing the concentration of LYAR diminishes the proton-carrying potential of the D/E tracts (Figure S7I). However, in cells, data show that increasing the concentrations of proteins such as LYAR leads to the formation of condensates that are non-overlapping with nucleoli. Therefore, concentrations of proteins such as LYAR need to be maintained below levels that drive condensation via homotypic interactions, lest they form "incidental condensates.[85]" This apparent asymmetry in concentrations of K-blocks and D/E-tracts leaves uncompensated charges on D/E tracts, which appears to be neutralized via proton binding, leading to nucleoli being acidic compartments.

Across a phase boundary, the chemical potentials of all solution components will be equalized when phase equilibrium is established. We designate the chemical potentials of protons in condensates (C) versus dilute (D) phases as $\mu_{H^+,C}$ and $\mu_{H^+,D}$, respectively. Each of these chemical potentials is written as: $\mu_{H^+,D} = \mu^0_{H^+,D} + RTln a_{H^+,D}$ and $\mu_{H^+,C} = \mu^0_{H^+,C} + RTln a_{H^+,C}$. Here, $\mu^0_{H^+,D}$ and $\mu^0_{H^+,C}$ are the standard state chemical potentials, whereas $a_{H^+,D}$ and $a_{H^+,C}$ are the dilute (D) and condensed phase (C) activities of protons. For protons, the standard state chemical potentials are free energies associated with transferring protons from a reference state, such as an aqueous solution comprising 1 M HCl, to either the dilute phase ($\mu^0_{H^+,D}$) or condensate ($\mu^0_{H^+,C}$). Note that $a_{H^+,D} = \lambda_{H^+,D}[H^+_D]$ and $a_{H^+,C} = \lambda_{H^+,C}[H^+_C]$ where the activity coefficients $\lambda_{H^+,D}$ and $\lambda_{H^+,C}$ quantify non-idealities due to buffering, and this will modulate the concentration of free protons in the dilute phase and condensate.

In a system lacking energy sinks or sources, as would be the case *in vitro*, the chemical potentials of the proton will be equalized across the dilute phase and condensate. Equalizing chemical potentials of protons across the dilute phase and coexisting condensate leads to: $\Delta\mu^0_{H^+} = (\mu^0_{H^+,C} - \mu^0_{H^+,D}) = -RT \ln\left(\frac{a_{H^+,C}}{a_{H^+,D}}\right)$. Therefore, any difference in the pH across the phase boundary results in an interphase pH gradient that sets up an interphase proton motive force, and this will be the result of differential binding of protons in condensates versus the dilute phases.

D/E tracts appear to create a microenvironment that is favorable for the accumulation of protons. Previous studies uncovered asymmetries between proton association with Asp/Glu rich regions compared with proton dissociation from Lys-rich regions.[106,107] This implies an asymmetry in contributions to charge regulation from D/E tracts versus K-rich regions. Asymmetries arise from differences in free energies of hydration of Asp/Glu vis-à-vis Lys, the lower charge density of the basic groups due to longer sidechains, and the local effects this creates around D/E tracts versus K-rich regions.[109]

The internal pH of nuclear bodies correlates with net charge of constituent IDRs

We hypothesized that if IDRs can influence nucleolar pH, then they also influence other nuclear bodies. Nuclear speckles[114] and Cajal bodies[115] house IDRs with a bias for positive net charge (Figure 7I). We compared the pH of the nucleolus to those of the nuclear speckle and Cajal body using the fluorescence emission of a freely diffusing pHluorin2 protein (Figure 7J). These measurements show that the median pH of nucleoli, nuclear speckles, and Cajal bodies are, respectively, 0.7 units lower, equivalent to, and 0.3 units higher than the nucleoplasm, which has a pH of 7.2. We observe a positive correlation between the measured pH values and mean net charge per IDR (Figure 7M). Negatively charged GFP and its derivative proteins (RFP and pHluorin2) do not partition strongly into acidic nucleoli as autonomous units. Conversely, we observe a strong positive correlation between the PCs of pHluorin2 into nuclear bodies and the r_{510} measurement of pH. This demonstrates increased partitioning of negatively charged proteins into more basic environments created by speckles and Cajal bodies (Figure S7O).

DISCUSSION

We discovered two non-random, highly enriched, and conserved IDR grammars in nucleoli. These are D/E tracts and K-blocks + E-rich regions (Figure 1). Key scaffolds of the FC, DFC, and GC, namely, UBF, NCL, and NPM1 feature D/E tracts. FC proteins such as LYAR and PolR1F/G feature K-blocks + E-rich regions. Proteins with high-scoring RG-rich IDRs and RBDs are excluded from the FC, and overall localization preferences are governed, in part, by the numbers (valence) and types of nucleic acid binding domains that are tethered to D/E tracts.

In vitro condensation in quinary mixtures comprising endogenous amounts of UBF, rDNA, LYAR, NCL, and pre-rRNA reproduces the spatial organization observed in nucleoli whereby DFC-like condensates adsorb onto FC-like condensates. Leaving pre-rRNA out of the quinary mixture leads to inside-out

(D) Representative snapshots from LaSSI simulations of simulated FCs involving UBF with rDNA and either PolR1F IDR, WT NUCKS, or ΔK-rich NUCKS; snapshots are a zoom-in of the entire simulation box, wherein spherical condensates are formed.
(E) Calculated partition coefficients (PCs) of PolR1F IDR, WT NUCKS, or ΔK-rich NUCKS into simulated FCs. Error bars indicate standard errors from the mean across ten replicates.
(F) Confocal images of three-component FC facsimiles comprising UBF and rDNA with either Pol1RF IDR, WT NUCKS, or ΔK-rich NUCKS.
(G) Violin plots of PCs into the *in vitro* FC-like condensates.
(H) Confocal images of nucleoli from live *Xenopus* oocytes expressing GFP-tagged PolR1F IDR, WT NUCKS, or ΔK-rich NUCKS; RFP-PolRIe co-expressed to highlight FC.
(I) Violin plots of PCs into FC phases. For all violin plots (G,I) *** denotes p value < 1.00e−05. For all microscopy images (F, H), scale bars correspond to 3 μm.
(J) Comparison of K-block *Z* score and mean *in silico* partition coefficient.
(K) Comparison of mean *in silico* and mean *in vitro* PCs.
(L) Comparison of mean *in vitro* and mean in cells PCs. Black lines denote the linear regression fit, and error bars indicate standard errors from the mean. (M) R^2 values for linear regression comparing all combinations of K-block *Z* score, mean *in silico* PCs, mean *in vitro* PC coefficient, and mean PC in cells.
See also Figure S6.

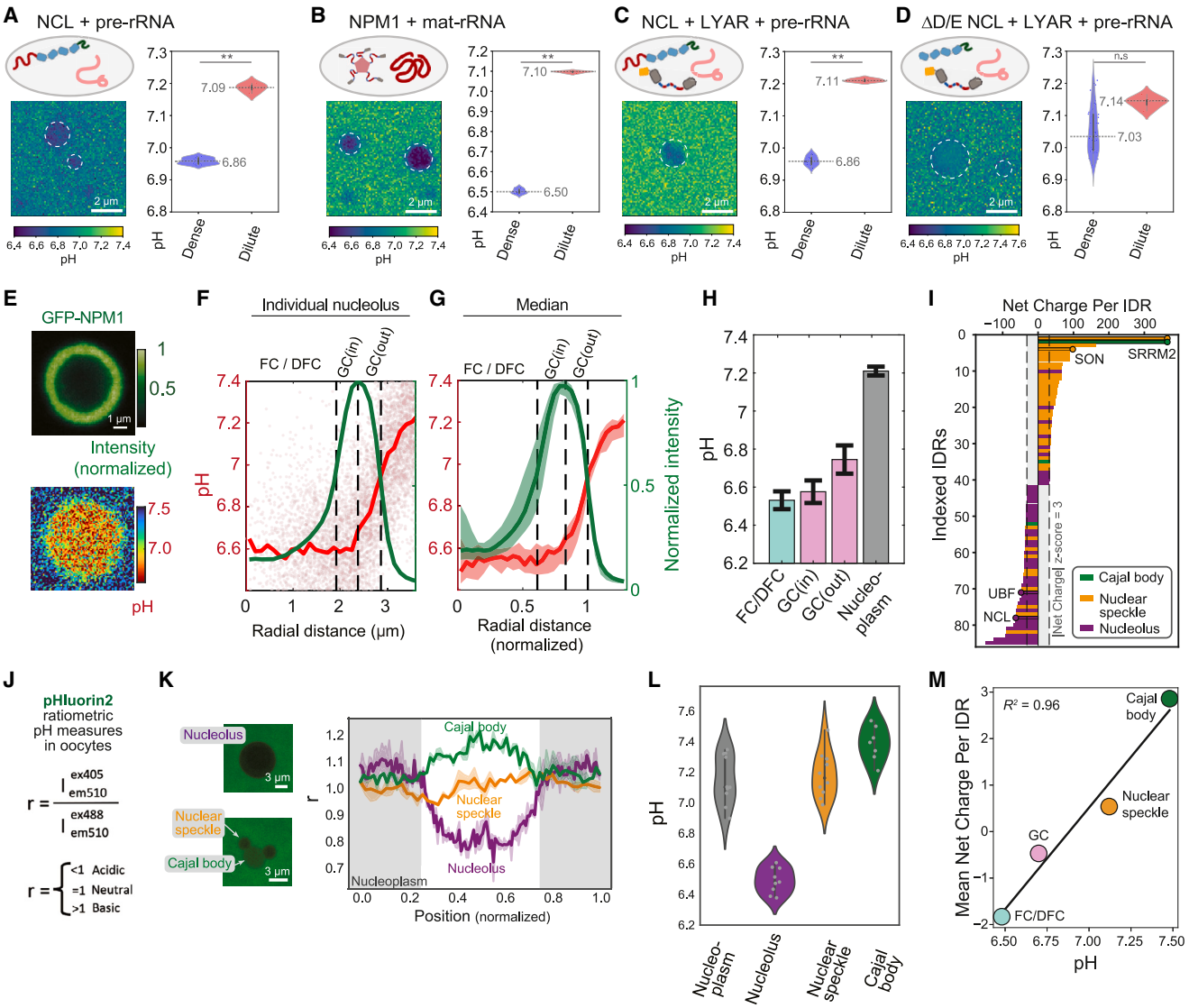


Figure 7. Distinctive compositional biases give rise to acidic nucleoli and non-acidic nuclear speckles and Cajal bodies

(A–D) Images of *in vitro* condensates and violin plots of measured pH values for dense phase condensates and surrounding dilute phase for the following mixtures: (A) NCL (5 μM) and pre-rRNA (1 μM [nt]), (B) NPM1 (5 μM) and mat-rRNA (2 μM [nt]), (C) wild-type (WT) NCL (15 μM), LYAR (0.2 μM), and pre-rRNA (1 μM [nt]), and (D) ΔD/E NCL (15 μM), LYAR (0.2 μM), and pre-rRNA (1 μM [nt]). Images show the measured pH on a per-pixel basis. Median pH values are indicated in violin plots; n.s. (not significant) denotes p values > 5.00e−02, ** denotes p value < 1.00e−03.

(E) Confocal images of a nucleolus in a living *Xenopus* oocyte GV expressing GFP-NPM1 (shown in upper panel) with freely diffusing SNARF-4 dye, whose ratiometric fluorescence is used quantify pH on a per-pixel basis (shown in lower panel); scale bars, 1 μm.

(F and G) Values of pH and normalized GFP-NPM1 intensity plotted against radial distance from the center of a nucleolus for the individual nucleolus (F) and median values (G) (x axis = actual distance [μm]) from five separate nucleoli (x axis = normalized distance). Solid curves correspond to median values, and dashed vertical lines demarcate boundaries of the GC (full-width half max of GFP-NCL signal) and its peak-intensity point. Points in the individual nucleolus plot (F) correspond to quantified pH values for individual pixels, and shaded regions in the median plot (G) correspond to the 95th percentile confidence interval of all values.

(H) Bar plot of median pH values across five nucleoli binned by FC/DFC, GC (inner half), GC (outer half), and nucleoplasm; error bars are standard deviation of measurements.

(I) Sorted net charge per IDR for all IDRs in the nucleolus, nuclear speckles, and Cajal body with a |net charge| *Z* score of ≥3 when compared with the human IDRome.

(J) The pHluorin2 r value is obtained by dividing the two emission intensities (each at 510 nm) from the same pHluorin2-containing sample excited (sequentially) by a 405 nm laser and a 488 nm laser; relationship of r to pH is indicated.

(K) Representative images and line scan profiles of pHluorin2 r values for nucleoli, nuclear speckles, and Cajal bodies in living *Xenopus* oocyte GVs expressing freely diffusing pHluorin2. Line scans through nuclear bodies are represented as normalized distance; solid curves and surrounding shaded region correspond to median values and 95th percentile confidence intervals, respectively (n ≥ 7 for each nuclear body). Gray shaded region corresponds to nucleoplasm.

(L) Violin plots of computed pH based on pHluorin2 r values for each nuclear body and nucleoplasm.

(M) Plot of mean net charge per IDR against measured pH for each nuclear body (R^2 = 0.96 for linear regression).

See also Figure S7.

morphologies *in vitro* that is reminiscent of observations in cells when transcription is inhibited.[17] The inside-out morphology observed in the absence of pre-rRNA, suggests that in its absence, the UBF-rDNA complex is preferentially solvated when compared with the combination of LYAR and NCL. This finding provides a thermodynamic rationale for why the intra-nucleolar organization is governed by pre-rRNA levels.[17,87,116,117]

The work of Feric et al., together with our results, and explorations of the effects of deleting the RG-rich IDR of NCL in *Caenorhabditis elegans* (*C. elegans*)[22] appear to explain the layering of the DFC versus GC. The DFC is scaffolded by proteins with RG-rich IDRs, and the GC is scaffolded by NPM1. RG-rich regions likely contribute hydrophobicity to condensates whereas highly charged IDRs like D/E tracts contribute to hydrophilicity. Proteins like NCL that contain both types of IDRs generate condensates that are amphiphilic. FBL in its natural context (snoRNP) is bound to proteins with K-block E-rich region IDRs; thus, *in vivo* DFCs may be less hydrophobic than suggested by condensates composed of mFBL.[23,118–121] Our work and that of others show that NPM1-based condensates are hydrophilic relative to condensates containing RG-rich.

In vivo FCs and DFCs, as well as *in vitro* reconstitutions thereof, exhibit salting-in behavior. Furthermore, K-block IDRs preferentially partition into nucleoli through electrostatic interactions that promote condensation. These findings are consistent with a mechanism of nucleolar condensation via complex coacervation, which leads us to designate the nucleolus as being a *complex coacervatome*. We expected D/E-Tracts to reduce driving forces for condensation through electrostatic repulsion, but instead found the opposite to be true. Our discovery that the relative acidity of nucleoli is due, in part, to uncompensated D/E-tracts favoring nucleolar proton accumulation provides one potential explanation. Protonation of acidic residues in D/E-tracts could both reduce repulsion and promote homotypic interactions among D/E-tracts.

As for the functional role of the interphase proton gradient, we propose that this might generate proton-motive forces[122,123] to enable directional transport within nucleoli.[24] A proton gradient is also likely to enable specificity of enzyme-catalyzed biochemical reactions within nucleoli by altering local microenvironments of enzymes and generating distinct conformational signatures[124,125] and electrochemical potentials at internal interfaces within nucleoli.[126,127]

Limitations of the study

Our method for mapping sub-phase localizations of proteins in living *Xenopus* nucleoli is based on a stepwise decrease in PolR1E intensity. We cross-validated that the highest PolR1E signal overlaps with the known FC localizer UBF. However, additional validation with more FC, DFC, and GC markers will be useful to strengthen our sub-phase assignments. We have not yet reconstituted a full nucleolar facsimile with all six sub-phases using endogenous levels of the relevant proteins and nucleic acid substrates. Such bottom-up reconstitutions will likely require the inclusion of small nucleolar ribonucleoprotein complexes and key RG-rich IDRs from the DFC, PDFC proteins such as URB1, mixtures of ribosomal proteins from the GC,[5] and proteins from the NR. It might also require the correct balance of pre-rRNA versus mat-rRNA and a functional mimic of RNA Pol I to mimic active processes *in vitro*.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Generation of DNA constructs
  - Preparation of DNA and RNA reagents for assays
  - Protein expression and purification
  - Tagging proteins with fluorescent reporters
  - Analysis of IDRs using NARDINI+
  - Extraction of condensate-enriched sequence features
  - Extraction of IDR Clusters within Nucleolar Proteins
  - Defining D/E-tracts and K-blocks+ERRs IDRs
  - Determining enrichment of IDR sequence features in different condensates
  - Evolutionary Analysis of Nucleolar IDRs
  - Calculating Sequence Block Profiles of IDRs
  - Analysis of Human Proteome Atlas (HPA) dataset
  - Identification of DNA / RNA-binding Domains
  - Selection criteria for summary of enriched nucleolar proteins
  - Bioinformatics analysis of NCL sequence features
  - Atomistic simulations of K-Block ERR IDRs
  - LaSSI simulations of condensates formed by components of fibrillar centers
  - Image collection and analysis
  - Isolation and imaging analysis of nucleoli from *Xenopus* oocytes
  - *In vitro* condensate / phase separation assays
  - Analysis of monomer-tetramer equilibrium of UBF
  - Wetting assays *in vitro*
  - Standard curves for measurements of pH
  - Direct electrochemical probe measure of pH
  - pHluorin2-based measurements of pH
  - Data visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.02.029.

ACKNOWLEDGMENTS

This work was funded by the US Air Force Office of Scientific Research grant (FA9550-20-1-0241 to R.V.P.), the St. Jude Research Collaborative on the Biology and Biophysics of RNP granules (to R.V.P.), the US National Science

Foundation (MCB-2227268 to R.V.P.), the US National Institutes of Health (F32GM146418-01A1 to M.R.K., R21AI163985 to M.D.V., R35GM124858 to M.D.L., and R01NS121114 to R.V.P.), and the Knut and Alice Wallenberg Foundation (grant 2021.0346 to E.L.).

AUTHOR CONTRIBUTIONS

Conceptualization, M.R.K. and R.V.P.; methodology, M.R.K., A.Z.L., K.M.R., J.M.L., T.W., W.O., M.D.V., M.D.L., and R.V.P.; investigation, M.R.K., K.M.R., A.P., M.F., and R.V.P.; funding acquisition, R.V.P., M.D.V., M.D.L., M.R.K., and E.L.; project administration, R.V.P.; supervision, M.D.V., E.L., M.D.L., and R.V.P.; writing, M.R.K. and R.V.P.; review & editing, all authors.

DECLARATION OF INTERESTS

R.V.P. is a member of the scientific advisory board for and shareholder in Dewpoint Therapeutics Inc.

Received: July 18, 2023
Revised: January 2, 2024
Accepted: February 22, 2024
Published: March 18, 2024

REFERENCES

1. Pederson, T. (2011). The Nucleolus. Cold Spring Harbo. Perspect. Biol. *3*, a000638.
2. Shaw, P.J., and Jordan, E.G. (1995). The Nucleolus. Annu. Rev. Cell Dev. Biol. *11*, 93–121.
3. Shan, L., Xu, G., Yao, R.-W., Luan, P.-F., Huang, Y., Zhang, P.-H., Pan, Y.-H., Zhang, L., Gao, X., Li, Y., et al. (2023). Nucleolar URB1 ensures 3′ ETS rRNA removal to prevent exosome surveillance. Nature *615*, 526–534.
4. Boisvert, F.M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat. Rev. Mol. Cell Biol. *8*, 574–585.
5. Kraushar, M.L., Krupp, F., Harnett, D., Turko, P., Ambrozkiewicz, M.C., Sprink, T., Imami, K., Günnigmann, M., Zinnall, U., Vieira-Vieira, C.H., et al. (2021). Protein Synthesis in the Developing Neocortex at Near-Atomic Resolution Reveals Ebp1-Mediated Neuronal Proteostasis at the 60S Tunnel Exit. Mol. Cell *81*, 304–322.e16.
6. Politz, J.C., Yarovoi, S., Kilroy, S.M., Gowda, K., Zwieb, C., and Pederson, T. (2000). Signal recognition particle components in the nucleolus. Proc. Natl. Acad. Sci. USA *97*, 55–60.
7. Massenet, S. (2019). In vivo assembly of eukaryotic signal recognition particle: A still enigmatic process involving the SMN complex. Biochimie *164*, 99–104.
8. Boulon, S., Westman, B.J., Hutten, S., Boisvert, F.-M., and Lamond, A.I. (2010). The Nucleolus under Stress. Mol. Cell *40*, 216–227.
9. Nishimura, K., Kumazawa, T., Kuroda, T., Katagiri, N., Tsuchiya, M., Goto, N., Furumai, R., Murayama, A., Yanagisawa, J., and Kimura, K. (2015). Perturbation of Ribosome Biogenesis Drives Cells into Senescence through 5S RNP-Mediated p53 Activation. Cell Rep. *10*, 1310–1323.
10. Frottin, F., Schueder, F., Tiwary, S., Gupta, R., Körner, R., Schlichthaerle, T., Cox, J., Jungmann, R., Hartl, F.U., and Hipp, M.S. (2019). The nucleolus functions as a phase-separated protein quality control compartment. Science *365*, 342–347.
11. Salvetti, A., and Greco, A. (2014). Viruses and the nucleolus: The fatal attraction. Biochim. Biophys. Acta *1842*, 840–847.
12. Hiscox, J.A. (2007). RNA viruses: hijacking the dynamic nucleolus. Nat. Rev. Microbiol. *5*, 119–127.
13. Rawlinson, S.M., and Moseley, G.W. (2015). The nucleolar interface of RNA viruses. Cell. Microbiol. *17*, 1108–1120.
14. Hiscox, J.A., Wurm, T., Wilson, L., Britton, P., Cavanagh, D., and Brooks, G. (2001). The coronavirus Infectious Bronchitis Virus Nucleoprotein Localizes to the Nucleolus. J. Virol. *75*, 506–512.
15. Politz, J.C.R., Polena, I., Trask, I., Bazett-Jones, D.P., and Pederson, T. (2005). A Nonribosomal Landscape in the Nucleolus Revealed by the Stem Cell Protein Nucleostemin. Mol. Biol. Cell *16*, 3401–3410.
16. Ma, H., and Pederson, T. (2008). Nucleophosmin Is a Binding Partner of Nucleostemin in Human Osteosarcoma Cells. Mol. Biol. Cell *19*, 2870–2875.
17. Lafontaine, D.L.J., Riback, J.A., Bascetin, R., and Brangwynne, C.P. (2021). The nucleolus as a multiphase liquid condensate. Nat. Rev. Mol. Cell Biol. *22*, 165–182.
18. Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R.V., and Brangwynne, C.P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell *165*, 1686–1697.
19. Stenström, L., Mahdessian, D., Gnann, C., Cesnik, A.J., Ouyang, W., Leonetti, M.D., Uhlén, M., Cuylen-Haering, S., Thul, P.J., and Lundberg, E. (2020). Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. Mol. Syst. Biol. *16*, e9469.
20. Handwerger, K.E., Cordero, J.A., and Gall, J.G. (2005). Cajal Bodies, Nucleoli, and Speckles in the Xenopus Oocyte Nucleus Have a Low-Density, Sponge-like Structure. Mol. Biol. Cell *16*, 202–211.
21. Falahati, H., Pelham-Webb, B., Blythe, S., and Wieschaus, E. (2016). Nucleation by rRNA Dictates the Precision of Nucleolus Assembly. Curr. Biol. *26*, 277–285.
22. Spaulding, E.L., Feidler, A.M., Cook, L.A., and Updike, D.L. (2022). RG/RGG repeats in the C. elegans homologs of nucleolin and GAR1 contribute to sub-nucleolar phase separation. Nat. Commun. *13*, 6585.
23. Riback, J.A., Zhu, L., Ferrolino, M.C., Tolbert, M., Mitrea, D.M., Sanders, D.W., Wei, M.T., Kriwacki, R.W., and Brangwynne, C.P. (2020). Composition-dependent thermodynamics of intracellular phase separation. Nature *581*, 209–214.
24. Riback, J.A., Eeftens, J.M., Lee, D.S.W., Quinodoz, S.A., Donlic, A., Orlovsky, N., Beckers, L., Becker, L.A., Strom, A.R., Rana, U., et al. (2023). Viscoelastic RNA entanglement and advective flow underlie nucleolar form and function. Mol. Cell *82*, 3095–3107.e3099.
25. Pappu, R.V., Cohen, S.R., Dar, F., Farag, M., and Kar, M. (2023). Phase Transitions of Associative Biomacromolecules. Chem. Rev. *123*, 8945–8987.
26. Brangwynne, C.P., Tompa, P., and Pappu, R.V. (2015). Polymer physics of intracellular phase transitions. Nature. Phys. *11*, 899–904.
27. Mais, C., Wright, J.E., Prieto, J.L., Raggett, S.L., and McStay, B. (2005). UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery. Genes Dev. *19*, 50–64.
28. Prieto, J.L., and McStay, B. (2007). Recruitment of factors linking transcription and processing of pre-rRNA to NOR chromatin is UBF-dependent and occurs independent of transcription in human cells. Genes Dev. *21*, 2041–2054.
29. Grob, A., Colleran, C., and McStay, B. (2014). Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. Genes Dev. *28*, 220–230.
30. Tartakoff, A., DiMario, P., Hurt, E., McStay, B., Panse, V.G., and Tollervey, D. (2022). The dual nature of the nucleolus. Genes Dev. *36*, 765–769.
31. Mittag, T., and Pappu, R.V. (2022). A conceptual framework for understanding phase separation and addressing open questions and challenges. Mol. Cell *82*, 2201–2214.
32. Bergeron-Sandoval, L.P., Kumar, S., Heris, H.K., Chang, C.L.A., Cornell, C.E., Keller, S.L., Francois, P., Hendricks, A.G., Ehrlicher, A.J., Pappu, R.V., and Michnick, S.W. (2021). Endocytic proteins with prion-like domains form viscoelastic condensates that enable membrane remodeling. Proc. Natl. Acad. Sci. USA *118*, e2113789118.

33. Feric, M., Sarfallah, A., Dar, F., Temiakov, D., Pappu, R.V., and Misteli, T. (2022). Mesoscale structure–function relationships in mitochondrial transcriptional condensates. Proc. Natl. Acad. Sci. USA *119*, e2207303119.

34. Neitzel, A.E., Fang, Y.N., Yu, B., Rumyantsev, A.M., de Pablo, J.J., and Tirrell, M.V. (2021). Polyelectrolyte Complex Coacervation across a Broad Range of Charge Densities. Macromolecules *54*, 6878–6890.

35. Pak, C.W., Kosno, M., Holehouse, A.S., Padrick, S.B., Mittal, A., Ali, R., Yunus, A.A., Liu, D.R., Pappu, R.V., and Rosen, M.K. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. Mol. Cell *63*, 72–85.

36. Sing, C.E., and Perry, S.L. (2020). Recent progress in the science of complex coacervation. Soft Matter *16*, 2885–2914.

37. Adhikari, S., Leaf, M.A., and Muthukumar, M. (2018). Polyelectrolyte complex coacervation by electrostatic dipolar interactions. J. Chem. Phys. *149*, 163308.

38. Cohan, M.C., Shinn, M.K., Lalmansingh, J.M., and Pappu, R.V. (2022). Uncovering Non-random Binary Patterns Within Sequences of Intrinsically Disordered Proteins. J. Mol. Biol. *434*, 167373.

39. Zarin, T., Strome, B., Peng, G., Pritišanac, I., Forman-Kay, J.D., and Moses, A.M. (2021). Identifying molecular features that are associated with biological function of intrinsically disordered protein regions. eLife *10*, e60220.

40. Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., et al. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell *174*, 688–699.e16.

41. Lyons, H., Veettil, R.T., Pradhan, P., Fornero, C., De La Cruz, N., Ito, K., Eppert, M., Roeder, R.G., and Sabari, B.R. (2023). Functional partitioning of transcriptional regulators by patterned charge blocks. Cell *186*, 327–345.e28.

42. Greig, J.A., Nguyen, T.A., Lee, M., Holehouse, A.S., Posey, A.E., Pappu, R.V., and Jedd, G. (2020). Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. Mol. Cell *77*, 1237–1250.e4.

43. Patil, A., Strom, A.R., Paulo, J.A., Collings, C.K., Ruff, K.M., Shinn, M.K., Sankar, A., Cervantes, K.S., Wauer, T., St.Laurent, J.D., et al. (2023). A disordered region controls cBAF activity via condensation and partner recruitment. Cell *186*, 4936–4955.e26.

44. Mensah, M.A., Niskanen, H., Magalhaes, A.P., Basu, S., Kircher, M., Sczakiel, H.L., Reiter, A.M.V., Elsner, J., Meinecke, P., Biskup, S., et al. (2023). Aberrant phase separation and nucleolar dysfunction in rare genetic diseases. Nature *614*, 564–571.

45. Kilgore, H.R., and Young, R.A. (2022). Learning the chemical grammar of biomolecular condensates. Nat. Chem. Biol. *18*, 1298–1306.

46. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842–1855.e16.

47. Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., and Baldwin, A. (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. Mol. Cell *57*, 936–947.

48. Ng, S.C., and Görlich, D. (2022). A simple thermodynamic description of phase separation of Nup98 FG domains. Nat. Commun. *13*, 6172.

49. Lee, B., Jaberi-Lashkari, N., and Calo, E. (2022). A unified view of low complexity regions (LCRs) across species. eLife *11*, e77058.

50. Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O.G., and Pappu, R.V. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys. J. *112*, 16–21.

51. Ruff, K.M., Choi, Y.H., Cox, D., Ormsby, A.R., Myung, Y., Ascher, D.B., Radford, S.E., Pappu, R.V., and Hatters, D.M. (2022). Sequence grammar underlying the unfolding and phase separation of globular proteins. Mol. Cell *82*, 3193–3208.e8.

52. Shinn, M.K., Cohan, M.C., Bullock, J.L., Ruff, K.M., Levin, P.A., and Pappu, R.V. (2022). Connecting sequence features within the disordered C-terminal linker of Bacillus subtilis FtsZ to functions and bacterial cell division. Proc. Natl. Acad. Sci. USA *119*, e2211178119.

53. Das, R.K., and Pappu, R.V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc. Natl. Acad. Sci. USA *110*, 13392–13397.

54. Martin, E.W., Holehouse, A.S., Grace, C.R., Hughes, A., Pappu, R.V., and Mittag, T. (2016). Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. J. Am. Chem. Soc. *138*, 15323–15335.

55. Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R.V., and Mittag, T. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science *367*, 694–699.

56. Zeng, X., Ruff, K.M., and Pappu, R.V. (2022). Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA *119*, e2200559119.

57. Consortium, T.U. (2020). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research *49*, D480–D489.

58. Piovesan, D., Necci, M., Escobedo, N., Monzon, A.M., Hatos, A., Mičetić, I., Quaglia, F., Paladin, L., Ramasamy, P., Dosztányi, Z., et al. (2021). MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Res. *49*, D361–D367.

59. Zarin, T., Strome, B., Nguyen Ba, A.N., Alberti, S., Forman-Kay, J.D., and Moses, A.M. (2019). Proteome-wide signatures of function in highly diverged intrinsically disordered regions. eLife *8*, e46883.

60. Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.-B., Munier, A., Fradet, M., et al. (2017). P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. Mol. Cell *68*, 144–157.e5.

61. Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487–498.

62. Marsaglia, G., Tsang, W.W., and Wang, J. (2003). Evaluating Kolmogorov's Distribution. J. Stat. Soft. *8*, 1–4.

63. Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. Mol. Syst. Biol. *7*, 548.

64. Ward, J.H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc. *58*, 236–244.

65. Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249–253.

66. Ma, N., Matsunaga, S., Takata, H., Ono-Maniwa, R., Uchiyama, S., and Fukui, K. (2007). Nucleolin functions in nucleolus formation and chromosome congression. J. Cell Sci. *120*, 2091–2105.

67. Ugrinova, I., Monier, K., Ivaldi, C., Thiry, M., Storck, S., Mongelard, F., and Bouvet, P. (2007). Inactivation of nucleolin leads to nucleolar disruption, cell cycle arrest and defects in centrosome duplication. BMC Mol. Biol. *8*, 66.

68. Valdez, B.C., Henning, D., So, R.B., Dixon, J., and Dixon, M.J. (2004). The Treacher Collins syndrome TCOF1 gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. Proc. Natl. Acad. Sci. USA *101*, 10709–10714.

69. Jaberi-Lashkari, N., Lee, B., Aryan, F., and Calo, E. (2023). An evolutionarily nascent architecture underlying the formation and emergence of biomolecular condensates. Cell Rep. *42*, 112955.

70. Gerlach, P., Schuller, J.M., Bonneau, F., Basquin, J., Reichelt, P., Falk, S., and Conti, E. (2018). Distinct and evolutionary conserved structural features of the human nuclear exosome complex. eLife *7*, e38686.

71. Cortez, K.J., Roilides, E., Quiroz-Telles, F., Meletiadis, J., Antachopoulos, C., Knudsen, T., Buchanan, W., Milanovich, J., Sutton, D.A., Fothergill, A., et al. (2008). Infections Caused by Scedosporium spp. Clin. Microbiol. Rev. *21*, 157–197.

72. Abdelmohsen, K., and Gorospe, M. (2012). RNA-binding protein nucleolin in disease. RNA Biol. *9*, 799–808.

73. Pfaff, M., and Anderer, F.A. (1988). Casein kinase II accumulation in the nucleolus and its role in nucleolar phosphorylation. Biochim. Biophys. Acta *969*, 100–109.

74. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314.

75. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

76. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview, Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

77. Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biom. Bull. *1*, 80–83.

78. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. eLife *6*, e27041.

79. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321..

80. Ouyang, W., Winsnes, C.F., Hjelmare, M., Cesnik, A.J., Åkesson, L., Xu, H., Sullivan, D.P., Dai, S., Lan, J., Jinmo, P., et al. (2019). Analysis of the Human Protein Atlas Image Classification competition. Nat. Methods *16*, 1254–1261.

81. Le, T., Winsnes, C.F., Axelsson, U., Xu, H., Mohanakrishnan Kaimal, J., Mahdessian, D., Dai, S., Makarov, I.S., Ostankovich, V., Xu, Y., et al. (2022). Analysis of the Human Protein Atlas Weakly Supervised Single-Cell Classification competition. Nat. Methods *19*, 1221–1229.

82. Ouyang, W., Mueller, F., Hjelmare, M., Lundberg, E., and Zimmer, C. (2019). ImJoy: an open-source computational platform for the deep learning era. Nat. Methods *16*, 1199–1200.

83. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419.

84. Weber, S.C., and Brangwynne, C.P. (2015). Inverse size scaling of the nucleolus by a concentration-dependent phase transition. Curr. Biol. *25*, 641–646.

85. Putnam, A., Thomas, L., and Seydoux, G. (2023). RNA granules: functional compartments or incidental condensates? Genes Dev. *37*, 354–376.

86. Li, H., Wang, B., Yang, A., Lu, R., Wang, W., Zhou, Y., Shi, G., Kwon, S.W., Zhao, Y., and Jin, Y. (2009). Ly-1 Antibody Reactive Clone Is an Important Nucleolar Protein for Control of Self-Renewal and Differentiation in Embryonic Stem Cells. Stem Cells *27*, 1244–1254.

87. Burger, K., Mühl, B., Harasim, T., Rohrmoser, M., Malamoussi, A., Orban, M., Kellner, M., Gruber-Eber, A., Kremmer, E., Hölzel, M., and Eick, D. (2010). Chemotherapeutic Drugs Inhibit Ribosome Biogenesis at Various Levels. J. Biol. Chem. *285*, 12416–12425.

88. Henras, A.K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip. Rev. RNA *6*, 225–242.

89. Piehler, J., Brecht, A., Valiokas, R., Liedberg, B., and Gauglitz, G. (2000). A high-density poly(ethylene glycol) polymer brush for immobilization on glass-type surfaces. Biosens. Bioelectron. *15*, 473–481.

90. Li, D., Cao, R., Li, Q., Yang, Y., Tang, A., Zhang, J., and Liu, Q. (2023). Nucleolus assembly impairment leads to two-cell transcriptional repression via NPM1-mediated PRC2 recruitment. Nat. Struct. Mol. Biol. *30*, 914–925.

91. Margossian, K.O., Brown, M.U., Emrick, T., and Muthukumar, M. (2022). Coacervation in polyzwitterion-polyelectrolyte systems and their potential applications for gastrointestinal drug delivery platforms. Nat. Commun. *13*, 2250.

92. Overbeek, J.T.G., and Voorn, M.J. (1957). Phase separation in polyelectrolyte solutions; Theory of complex coacervation. J. Cell. Physiol. Suppl. *49* (*Supplement 1*), 7–26.

93. Rumyantsev, A.M., Jackson, N.E., and de Pablo, J.J.d. (2021). Polyelectrolyte Complex Coacervates: Recent Developments and New Frontiers. Annu. Rev. Condens. Matter Phys. *12*, 155–176.

94. Mascotti, D.P., and Lohman, T.M. (1990). Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids. Proc. Natl. Acad. Sci. USA *87*, 3142–3146.

95. Record, M.T., Lohman, T.M., and Haseth, P.d. (1976). Ion effects on ligand-nucleic acid interactions. J. Mol. Biol. *107*, 145–158.

96. Chang, L.-W., Lytle, T.K., Radhakrishna, M., Madinya, J.J., Vélez, J., Sing, C.E., and Perry, S.L. (2017). Sequence and entropy-based control of complex coacervates. Nat. Commun. *8*, 1273.

97. Heidarsson, P.O., Mercadante, D., Sottini, A., Nettels, D., Borgia, M.B., Borgia, A., Kilic, S., Fierz, B., Best, R.B., and Schuler, B. (2022). Release of linker histone from the nucleosome driven by polyelectrolyte competition with a disordered protein. Nat. Chem. *14*, 224–231.

98. Galvanetto, N., Ivanović, M.T., Chowdhury, A., Sottini, A., Nüesch, M.F., Nettels, D., Best, R.B., and Schuler, B. (2023). Extreme dynamics in a biomolecular condensate. Nature *619*, 876–883.

99. Ogston, A.G. (1970). On the interaction of solute molecules with porous networks. J. Phys. Chem. *74*, 668–669.

100. Vitalis, A., and Pappu, R.V. (2009). ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. J. Comput. Chem. *30*, 673–699.

101. Radhakrishnan, A., Vitalis, A., Mao, A.H., Steffen, A.T., and Pappu, R.V. (2012). Improved Atomistic Monte Carlo Simulations Demonstrate That Poly-l-Proline Adopts Heterogeneous Ensembles of Conformations of Semi-Rigid Segments Interrupted by Kinks. J. Phys. Chem. B *116*, 6862–6871.

102. Mao, A.H., and Pappu, R.V. (2012). Crystal lattice properties fully determine short-range interaction parameters for alkali and halide ions. J. Chem. Phys. *137*, 64104.

103. Lalmansingh, J.M., Keeley, A.T., Ruff, K.M., Pappu, R.V., and Holehouse, A.S. (2023). SOURSOP: A Python package for the analysis of simulations of intrinsically disordered proteins. J. Chem. Theory. Comput. *19*, 5609–5620.

104. Choi, J.M., Dar, F., and Pappu, R.V. (2019). LASSI: A lattice model for simulating phase transitions of multivalent proteins. PLoS Comput. Biol. *15*, e1007028.

105. Scott, M.S., Boisvert, F.M., McDowall, M.D., Lamond, A.I., and Barton, G.J. (2010). Characterization and prediction of protein nucleolar localization sequences. Nucleic Acids Res. *38*, 7388–7399.

106. Fossat, M.J., Posey, A.E., and Pappu, R.V. (2021). Quantifying charge state heterogeneity for proteins with multiple ionizable residues. Biophys. J. *120*, 5438–5453.

107. Fossat, M.J., Posey, A.E., and Pappu, R.V. (2023). Uncovering the Contributions of Charge Regulation to the Stability of Single Alpha Helices. ChemPhysChem *24*, e202200746.

108. Fossat, M.J., and Pappu, R.V. (2019). q-Canonical Monte Carlo Sampling for Modeling the Linkage between Charge Regulation and Conformational Equilibria of Peptides. J. Phys. Chem. B *123*, 6952–6967.

109. Fossat, M.J., Zeng, X., and Pappu, R.V. (2021). Uncovering Differences in Hydration Free Energies and Structures for Model Compound Mimics of Charged Side Chains of Amino Acids. J. Phys. Chem. B *125*, 4148–4161.

110. Franzmann, T.M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A.S., Nuske, E., Richter, D., Baumeister, W., Grill, S.W., Pappu, R.V., et al. (2018). Phase separation of a yeast prion protein promotes cellular fitness. Science *359*, eaao5654.

111. Guillen-Boixet, J., Kopach, A., Holehouse, A.S., Wittmann, S., Jahnel, M., Schlüßle, R., Kim, K., Trussina, I., Wang, J., Mateju, D., et al. (2020). RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. Cell *181*, 346–361.e17.

112. Salvi, A., Quillan, J.M., and Sadée, W. (2015). Monitoring intracellular pH changes in response to osmotic stress and membrane transport activity using 5-chloromethylfluorescein. AAPS PharmSci *4*, 21.

113. Reifenrath, M., and Boles, E. (2018). A superfolder variant of pH-sensitive pHluorin for in vivo pH measurements in the endoplasmic reticulum. Sci. Rep. *8*, 11985.

114. Spector, D.L., and Lamond, A.I. (2011). Nuclear Speckles. Cold Spring Harb. Perspect. Biol. *3*.

115. Gall, J.G., Bellini, M., Wu, Z.a., and Murphy, C. (1999). Assembly of the Nuclear Transcription and Processing Machinery: Cajal Bodies (Coiled Bodies) and Transcriptosomes. Mol. Biol. Cell *10*, 4385–4402.

116. Falahati, H., and Wieschaus, E. (2017). Independent active and thermodynamic processes govern the nucleolus assembly in vivo. Proc. Natl. Acad. Sci. USA *114*, 1335–1340.

117. Schede, H.H., Natarajan, P., Chakraborty, A.K., and Shrinivas, K. (2023). A model for organization and regulation of nuclear condensates by gene activity. Nat. Commun. *14*, 4152.

118. Yu, G., Zhao, Y., and Li, H. (2018). The multistructural forms of box C/D ribonucleoprotein particles. RNA *24*, 1625–1633.

119. Mitrea, D.M., Cika, J.A., Guy, C.S., Ban, D., Banerjee, P.R., Stanley, C.B., Nourse, A., Deniz, A.A., and Kriwacki, R.W. (2016). Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. eLife *5*, e13571.

120. Ferrolino, M.C., Mitrea, D.M., Michael, J.R., and Kriwacki, R.W. (2018). Compositional adaptability in NPM1-SURF6 scaffolding networks enabled by dynamic switching of phase separation mechanisms. Nat. Commun. *9*, 5064.

121. Mitrea, D.M., Cika, J.A., Stanley, C.B., Nourse, A., Onuchic, P.L., Banerjee, P.R., Phillips, A.H., Park, C.-G., Deniz, A.A., and Kriwacki, R.W. (2018). Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. Nat. Commun. *9*, 842.

122. Abad, J.P. (2011). Proton Motive Force. In Encyclopedia of Astrobiology, M. Gargaud, R. Amils, J.C. Quintanilla, H.J. Cleaves, W.M. Irvine, D.L. Pinti, and M. Viso, eds. (Springer), pp. 1355–1356.

123. Simon, J., van Spanning, R.J.M., and Richardson, D.J. (2008). The organisation of proton motive and non-proton motive redox loops in prokaryotic respiratory systems. Biochim. Biophys. Acta *1777*, 1480–1490.

124. Farag, M., Cohen, S.R., Borcherds, W.M., Bremer, A., Mittag, T., and Pappu, R.V. (2022). Condensates formed by prion-like low-complexity domains have small-world network structures and interfaces defined by expanded conformations. Nat. Commun. *13*, 7722.

125. Farag, M., Borcherds, W.M., Bremer, A., Mittag, T., and Pappu, R.V. (2023). Phase separation of protein mixtures is driven by the interplay of homotypic and heterotypic interactions. Nat. Commun. *14*, 5527.

126. Dai, Y., Chamberlayne, C.F., Messina, M.S., Chang, C.J., Zare, R.N., You, L., and Chilkoti, A. (2023). Interface of biomolecular condensates modulates redox reactions. Chem *9*, 1594–1609.

127. Hoffmann, C., Murastov, G., Tromm, J.V., Moog, J.-B., Aslam, M.A., Matkovic, A., and Milovanovic, D. (2023). Electric Potential at the Interface of Membraneless Organelles Gauged by Graphene. Nano Lett. *23*, 10796–10801.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
DH5α strain *E. coli* cells	New England Biolabs	Cat#C2987I
Bl21 strain *E. coli* cells	New England Biolabs	Cat#C2530H
Biological samples		
Stage VI *Xenopus leavis* oocytes	John Silva Laboratory at Washington University in St. Louis	N/A
Chemicals, peptides, and recombinant proteins		
CF®405M dUTP for labeling DNA reagents	Biotium	Cat#40100-T
CF®640R dUTP for labeling DNA reagents	Biotium	Cat#40007-T
Cy®3-UTP Cytiva for labeling RNA reagents	Sigma	Cat#GEPA53026
Cy®5-UTP Cytiva for labeling RNA reagents	Sigma	Cat#GEPA55026
AlexaFluorTM405 NHS Ester for labeling proteins	Fisher	Cat#A30000
AlexaFluorTM488 NHS Ester for labeling proteins	Fisher	Cat#A20100
AlexaFluorTM555 NHS Ester for labeling proteins	Fisher	Cat#A20009
AlexaFluorTM647 NHS Ester for labeling proteins	Fisher	Cat#A20008
UBF: Upstream Binding Transcription factor	This paper - Source plasmid- pMK137	Uniprot: P17480 (Human)
UBF ΔD/E-tract: Upstream Binding Transcription factor with C-terminal D/E-tract removed	This paper - Source plasmid pMK152	Uniprot: P17480 (Human)
LYAR: Ly1-antibody reactive	This paper - Source plasmid- pMK39	Uniprot- Q6GP45 (Xenopus laevis)
NUCKS: Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	This paper - Source plasmid- pMK106	Uniprot- Q9H1E3 (Human)
NUCKS ΔK-rich: Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1 with K-rich region removed	This paper - Source plasmid- pMK141	Uniprot- Q9H1E3 (Human)
PolRIF_IDR: RNA polymerase I subunit F amino acids 204-338 (End)	This paper - Source plasmid- pMK140	Uniprot- Q3B726 (Human)
NCL: Nucleolin	This paper - Source plasmids: pMK55 and pMK78	Uniprot- Q06459 (Xenopus laevis)
NCLΔD/E: Nucelolin with D/E-tract IDR removed (a.a. 1-257)	This paper – Source plasmid pMK147	Uniprot- Q06459 (Xenopus laevis)
NPM1: Nucleophosmin	This paper- Source plasmids: pMK59 and pMK61	Uniprot - P07222 (Xenopus laevis); Nuclephosmin;
pHluorin2	This paper	Synthetic; generated via directed evolution in Mahon et al.[128] Source plasmid pMK169
Critical commercial assays		
Fisher Pure link quick plasmid miniprep kit	Fisher	Cat#K210011
Q5® High-Fidelity 2X Master Mix	NEB	Cat#M0492L
PCR and Gel clean up Kit	IBI scientific	Cat#IB47020
SP6 mMESSAGE mMACHINE Transcription Kit	Fisher	Cat#AM1340
T7 mMESSAGE mMACHINE Transcription Kit	Fisher	Cat#M0255A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Monarch RNA clean up	NEB	Cat#T2040L
TRIZOL reagent extraction	Fisher	Cat#15596026
Deposited data		
Raw and analyzed data	This paper	https://zenodo.org/doi/10.5281/zenodo.10661405
Swissprot Homo sapiens database	Consortium[57]	https://www.uniprot.org/proteomes/UP000005640
MobiDB	Piovesan et al.[58]	https://mobidb.bio.unipd.it/
Human Protein Atlas (HPA)	Thul et al.[79]	https://www.proteinatlas.org/
P-Body proteome	Hubstenberger et al.[60]	N/A
Stress granule proteome	Jain et al.[61]	N/A
EggNOG (protein orthologues)	Huerta-Cepas et al.[74]	http://eggnog5.embl.de/#/app/home
pFam (hosted b InterPro)	Mistry et al.[83]	http://pfam.xfam.org/
HeLa cell protein abundances	Nagaraj et al.[63]	N/A
Experimental models: Organisms/strains		
Mature female *Xenopus laevis* frogs (3-7 years old)	Xenopus1	Cat#4800
Recombinant DNA		
GFP_PolR1e	Feric et al.[18]	MK16
mCherry_PolR1e	Feric et al.[18]	MK17
NPM1_GFP	Feric et al.[18]	MK20
NPM1_RFP	Feric et al.[18]	MK21
His-LYAR (Xenopus)	This paper	MK39
GFP-xlNCL (mRNA)	This paper	MK49
mCherry-xlNCL (mRNA)	This paper	MK50
GFP_LYAR (mRNA)	This paper	MK51
mCherry-LYAR (mRNA)	This paper	MK52
Str-His-GFP-Tev-NCL	This paper	MK55
His-MBP-Tev-NCL	This paper	MK78
His-mCherry-Tev-NPM1	This paper	MK59
His-MBP-Tev-NPM1	This paper	MK61
SP6_5'ETS_rDNA	This paper	MK89
SP6_18s_rDNA	This paper	MK90
GFP_FAM133A (mRNA)	This paper	MK93
GFP_FAM133B (mRNA)	This paper	MK94
GFP_C12orf43 (mRNA)	This paper	MK95
GFP_NUCKS (mRNA)	This paper	MK97
GFP_SREK1IP1 (mRNA)	This paper	MK98
His-SUMO-NUCKS-Tev-SUMO	This paper	MK106
GFP_NUCKS_ΔKrich (mRNA)	This paper	MK117
Huntington_exon1_WT_RFP (mRNA)	This paper	MK131
Huntington_exon1_ΔN17_RFP (mRNA)	This paper	MK132
His-SUMO-UBF-Tev-MBP	This paper	MK137
His-SUMO-PolR1FIDR-Tev-GFP	This paper	MK140
His-SUMO-NUCKSΔKRich-Tev-SUMO	This paper	MK141
GFP alone (mRNA)	This paper	MK145
RFP alone (mRNA)	This paper	MK146
Str-His-GFP-Tev-NCLΔDE	This paper	MK147

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
His-SUMO-UBFΔDE-Tev-MBP	This paper	MK152
His-TEV-pHluorin2	This paper	MK169
pHluorin (mRNA)	This paper	MK170
TCOF1_RFP (mRNA)	This paper	MK208
GFP-PolR1FIDR (mRNA)	This paper	AP1
T7_Cln3	This paper	pUC19_T7_Cln3
Software and algorithms		
NARDINI	Cohan et al.[38] and Shinn et al.[52]	https://github.com/mshinn23/nardini
localCIDER	Holehouse et al.[50]	http://pappulab.github.io/localCIDER/
MUSCLE multiple sequence alignment	Edgar et al.[103]	https://www.ebi.ac.uk/Tools/msa/muscle/
ImJoy HPA-UMAP plugin	Ouyang et al.[82]	https://github.com/imjoy-team/imjoy-core
SOURSOP	Lalmansingh et al.[126]	https://github.com/holehouse-lab/soursop
LaSSI	Choi et al.[104]	https://github.com/Pappulab/LASSI
Custom Python 3 scripts for plot generation	This paper	https://zenodo.org/doi/10.5281/zenodo.10661405

RESOURCE AVAILABILITY

Lead contact

Please direct requests for resources and reagents to Rohit Pappu (pappu@wustl.edu).

Materials availability

Plasmids generated in this study are available upon request to Rohit Pappu (pappu@wustl.edu).

Data and code availability

- Original gel images, details of all quantitative measurements (as examples: threshold concentrations and pH measurements) as part of this work are publicly available via the Pappu lab Github repository: https://github.com/Pappulab/King-etal-Nucleoli and via zenodo https://zenodo.org/doi/10.5281/zenodo.10661405. We analyzed existing, publicly available data from the Swissprot Homo sapiens database, the Human Protein Atlas (HPA), the EggNOG repository of protein orthologues, and various mass spectrometry databases. The accession numbers for the datasets are listed in the key resources table.
- All original code that was developed as part of this work has been deposited to the Pappu lab Github repository and is publicly available via https://github.com/Pappulab/King-etal-Nucleoli and via zenodo at https://zenodo.org/doi/10.5281/zenodo.10661405.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Xenopus laevis (frog) oocytes were used for live cell analysis of protein localization in nucleoli and as source material for biochemical purification of mature rRNA. Mature female *Xenopus laevis* frogs (3-7 years old) are housed in a professional animal facility located in Whitaker Hall, within the McKelvey School of Engineering at Washington University in St. Louis. Oocyte harvesting is conducted by trained personnel, within the animal preparation room of this animal facility. All procedures were performed under the supervision of the Washington University Institutional Animal Care and Use Committee Office. All procedures were approved by the Washington University Institutional Animal Care and Use Committee (Animal Welfare Assurance #: A-3381-01). The animal care facility provided additional technical assistance where needed.

Two strains of *Escherichia coli* (*E. coli*) were used in this study. DH5α *E. coli* cells (New England Biolabs (NEB) - C2987I) were used to amplify DNA and BL21 *E. coli* cells (NEB - C2530H) were used to express protein. DH5α *E. coli* cells were grown in Luria-Bertani (LB) broth (Sigma - L3522) at 37°C in an orbital shaker incubator, operating at 220 revolutions per minute (RPM). LB broth was prepared at a concentration of 20 g/L. BL21 *E. coli* cells in either LB Broth or Terrific broth (TB) (Sigma: T9179) at 37°C (pre-induction) or 26°C and 18°C (post-induction) in an orbital shaker incubator, operating at 220 revolutions per minute (RPM). TB broth was prepared at a concentration of 48 g/L and supplemented with 8 mL/L (v/v) glycerol. All cell growths contained the necessary antibiotics that correspond to the plasmid's resistance gene, these are Carbenicillin (50 μg/mL) (GoldBio - C-103-25) and Kanamycin (34 μg/mL) (Sigma-Aldrich - 60615-5G).

METHOD DETAILS

Generation of DNA constructs

Detailed information about each DNA construct used in this study can be found in the key resources table provided as an addendum to this document. DH5α strain *E. coli* (NEB - C2987I) cells were used for all subcloning steps and for permanent storage of transfected strains. All DNA constructs were validated using Sanger sequencing (Azenta). Methods for generating DNA and RNA reagents used in assays, such as reconstitution experiments, are described in the section "Preparation of DNA and RNA reagents for assays" below. All plasmid DNA constructs were prepared from 3mL of saturated *E. coli* cells using the Fisher Pure link quick plasmid miniprep kit (K210011). All buffers used were those supplied in the kit (RNase A and ethanol were added to the resuspension buffers and wash buffers, respectively). The "Miniprep protocol" provided in this kit was adhered to for preparing plasmid DNA, which was eluted in 50 μL elution buffer and stored on ice or at 20°C. Polymerase chain reaction (PCR) products were generated using Q5® High-Fidelity 2X Master Mix (NEB - M0492L). PCR reaction mixtures contained 10 μM each forward and reverse primers, 25 ng template DNA and 1 volume Q5® High-Fidelity 2X Master Mix. The final volumes of these mixtures ranged from 10 to 50 μL. DNA amplicons were generated using the following thermocycling regime: initial denaturation: 95°C for 10 seconds; 30 Amplification cycles of the 98°C for 5–10 seconds, 50 to 72°C* for 20 seconds, 72°C for 30 seconds per kb**; final extension: 72°C for 3 minutes (*temperature used were -5, 0, or +5 degrees from primer annealing temperature predicted by NEB; **amplicons <1kB were amplified for 30 seconds). Amplicons were purified using PCR clean up (IBI Scientific - IB47020), wherein all buffers used were those supplied in the kit, and were supplemented with RNase A or ethanol, as needed. The "PCR Clean-up protocol" provided in this kit was adhered to and the optional wash step with "W1" buffer was omitted. PCR products were eluted in 30 μL elution buffer and stored on ice or at 20°C. Sub-cloning was carried out using via Gibson assembly. For Gibson assembly reactions, the insert DNA and plasmid DNA were incubated at a 3:1 molar ratio with the latter fixed at 50 ng total mass and were combined with 2X Gibson Master Mix (NEB - E2611L) (final volume: 20 μL) and incubated at 50°C for one hour then cooled on ice. This mixture, which contained the assembled plasmid, was used directly to transform into DH5α *E. coli* (NEB - C2987I). In all cases, DNA concentration was measured by absorbance at 260 nm on a Nanodrop 2000.

Preparation of DNA and RNA reagents for assays

To generate tagged DNA, 1% of 405 or 640-labeled dUTP (key resources table) was incorporated into a PCR reaction and amplified DNA was generated using the same procedure, reagents (Q5® High-Fidelity 2X Master Mix (NEB - M0492L)), and thermocycling regime described above. Amplified DNA was run on a 1% agarose gel and purified using IBI scientific Gel Clean up (IB47020). All buffers used were those supplied in the kit, and were supplemented, as needed, with either RNase A or ethanol. The "Gel Clean-up protocol" provided in this kit was adhered to except for the elation step. Labeled DNA was eluted in 30 μL ddH_2O buffer and aliquoted into single use volumes (typically 5 μL), flash-frozen in liquid N_2 and stored at -80°C. For the 5′ ETS region of rDNA we used DNA derived from PCR amplification of a 1622 bp region corresponding to the annotated 5′ ETS region of the Human rDNA loci (Gene ID: 6052). This construct was used to generate rDNA reagents via PCR amplification and pre-rRNA reagents using in vitro transcription. For the 18S region we used DNA derived from PCR amplification of an 1844 bp region corresponding to the annotated 18S region of the Human rDNA loci (Gene ID: 6052). This construct was used to generate rDNA reagents via PCR amplification and pre-rRNA reagents using in vitro transcription.

For all RNA reagents other than mat-rRNA, plasmids containing the desired RNA transcript were linearized via incubation with 5% v/v NotI restriction enzyme and 10% v/v CutSmart Buffer (both are provided in the same NEB product - R3189) at 37°C for 4 hours. Linearization was confirmed by gel electrophoresis of DNA samples in a 1% agarose gel (included were control digests where enzyme was omitted). Linearized plasmids were purified via PCR clean up (IBI Scientific IB47020) as described above. RNA was transcribed in vitro (IVT) using the ThermoFisher mMESSAGE mMACHINE Transcription Kit (using either SP6 polymerase – AM1340 or T7 polymerase – M0255A). For all IVT reactions, 100 ng or 1 μg of DNA was used, depending on length of the reaction; 100 ng DNA was used for overnight reactions, whereas or 1 μg of DNA was used for 2–4-hour reactions. The buffers used were those provided in the kit and the mixture was generated using the following reagents and volumes: 2 μL enzyme mix, 10μl 2X NTP/CAP mix, 2 μL 10X buffer, and 6 μL DNA. To generate tagged RNA, 1% of the total rNTPs were supplemented with either Cy3- or Cy5-UTP (key resources table) was incorporated into the IVT reaction. Transcribed RNAs were purified using the NEB Monarch RNA clean up (T2040L). The purity and weight of pure RNA was checked using gel electrophoresis, then aliquoted into single use volumes (typically 2 μL), which were flash frozen in liquid N_2 and stored at -80°C.

Mature rRNA (mat-rRNA) is a mixture of rRNAs (18S and 28S). Crude rRNA was purified endogenously from *Xenopus* oocytes via TRIzol reagent extraction (Fisher - 15596026). A 50 μL packed volume of stage VI *Xenopus* oocytes was incubated with 1 mL TRIzol for 5 minutes at room temperature then lysed via homogenization and pulse vortexing. A pellet of precipitated RNA was obtained by adding 0.5 mL isopropanol, incubating at -80°C for 45 minutes and centrifuging the reaction at 4°C for 10 minutes at 14,000 g. RNA pellets were desiccated using a gentle flow of nitrogen and reconstituted in 4°C ddH_2O. Concentration was measured by absorbance at 260 nm on a Nanodrop 2000. From this point on, all procedures took place at 4°C. RNA was either immediately aliquoted into single use volumes (typically 10 μL), flash-frozen in liquid N_2 and stored at -80°C or used to generate tagged mat-rRNA. To obtain pure mixtures or labeled 18S and 28S rRNA, freshly prepared or thawed crude RNA of a concentration of at least 7.5 μM was added to freshly prepared sodium periodate and sodium acetate (pH 5). The final concentrations of sodium periodate and sodium acetate are each

100 mM. DEPC-treated ddH_2O was used to bring the volume of the reaction to 100 μL. The reaction was then incubated for 90 minutes at room temperature in the absence of light. The reaction was then added to an Amicon Ultra-0.5 centrifugal filter (Sigma - UFC5050) with a molecular weight cutoff of 50 kDa, spun for 30 minutes at 14,000 g, and recovered the oxidized RNA. We then added sodium acetate (pH 5) to a final concentration of 100 mM, and a hydrazide fluorescent dye solution to a final concentration of 1.7 mM. We then added DEPC-treated water to a final volume of 30 μL. The reaction was incubated for 4 hours in the absence of light under gentle agitation. The mixture containing tagged RNA was then run over a Superose 6 3.2/300 column using DEPC-treated ddH20 at 4°C. The fractions containing mixtures of 18S and 28S rRNA (as assessed by gel electrophoresis) were recovered with Amicon Ultra-0.5 centrifugal filters with a cutoff of 50 kDa. The concentration of recovered mat-rRNA was measured by absorbance at 260 nm on a Nanodrop 2000, then the material was aliquoted into single use volumes (typically 2 μL), flash-frozen in liquid N_2 and stored at -80°C.

Protein expression and purification

All proteins were expressed in BL21 *E. coli* cells (NEB - C2530H) grown in either in Terrific broth (TB) (Sigma - T9179) (UBF, NUCKS, and PolR1F IDR) or LB broth (Sigma: L3522). For LYAR and NCL we used Erlenmeyer flasks with greater than or equal to five-fold head volume. Cultures were grown at 37°C in an orbital shaker incubator operating at 220 RPM until OD_{600} ~0.6 was reached. Cultures were then chilled for 10 minutes in an ice bath before being induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.35 mM). Expression-induced cultures in TB were grown for 14-18 hours at 18°C; those in LB broth were grown for 6-8 hours at 26°C. Cells were harvested by centrifugation, washed of residual media, and stored as pellets in 50 mL Falcon tubes at -80°C. Descriptions of the purification procedure of each protein are provided below. Details of the proteins, including their amino acid sequence and species origin, can be found in the key resources table and Table S1.

All protein purifications were carried out at 4°C. At the start of a purification a cell pellet was gently resuspended to homogeneity in 35 mL supplemented lysis buffer (buffer is protein-specific and provided below) then lysed via sonication on a Branson 550 with an L102C horn attachment using five series of the following 20-round cycle: 1 second on / 2 second off at 30% power. Lysis buffer supplements are: 500 U DNase I (Sigma - 4536282001), 500 U RNase A (Sigma - 10109169001), 5 mg lysozyme (Sigma - 62971-10G-F), and one protease inhibitor tablet (Sigma - 40694200). After each stage of purification (affinity, ion exchange, and / or size exclusion), protein concentration and nucleotide contamination was assessed on a Nanodrop 2000 by measuring absorbance at 280 nm and measurements of the ratio A260 / A280. Similarly, after each purification stage the inputs, flow-throughs, washes and elutes were examined using SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) (Gel: BioRad mini-PROTEAN TGX AnyKD - 4569036; MW latter: BioRad Precession Plus Protein Standard Unstained – 1610363), stained with EZblue Coomassie stain (10% (v/v) phosphoric acid, 10% (w/v) Ammonium sulfate, 20% (v/v) Methanol, 1.2% (w/v) Coomassie Blue) and de-stained via serial washes in ddH_2O. The amino acid sequences for all protein constructs that were expressed and purified as part of this work may be found in Table S1.

UBF (all variants):

Cells were lysed in supplemented lysis buffer (20 mM sodium phosphate, 75 mM NaCl, 20 mM imidazole, 14.3 mM BME (β-mercaptoethanol), 200 μM PMSF (phenylmethylsulphonyl fluoride) pH 7.5). Supernatant was recovered from a 25-minute spin at 38,000 g and bound to an equilibrated HisTrap FF Crude 5 mL column (Cytiva – 11000458) using an ÄKTA Pure fast protein liquid chromatography (FPLC) module. The Ni-NTA column was washed in 75 mL lysis buffer then the protein eluted in elution buffer (20 mM sodium phosphate, 500 mM NaCl, 350 mM imidazole, 14.3 mM BME, 200 μM PMSF; pH 7.5). Peak fractions from the affinity purification were pooled and diluted five-fold in dilution buffer (20 mM sodium phosphate, 14.3 mM BME; pH 7.5). This solution was further purified via ion exchange chromatography using a continuous gradient purification protocol with a HiTrap Heparin HP 5 mL column (Cytiva – 17040703), Buffer A (20 mM sodium phosphate, 100 mM NaCl, 14.3 mM BME; pH 7.5), and Buffer B (20 mM sodium phosphate, 100 mM NaCl, 14.3 mM BME; pH 7.5) on the ÄKTA Pure FPLC module. Peak fractions containing His-SUMO-UBF-tev-MBP were pooled and cleaved of SUMO and MBP tags during an overnight dialysis in the presence of 0.02X (molar amount) ULP1 protease and 0.1X (molar amount) TEV protease in cleavage buffer (20 mM sodium phosphate, 300 mM NaCl, 1 mM DTT (dithiothreitol), pH 7.5). UBF was purified to ≥95% using size exclusion chromatography - HiLoad 16/600 Superdex 200 pg column (Cytiva - 28989335) on the ÄKTA Pure FPLC module in storage buffer (22 mM sodium phosphate, 1.1 M NaCl, 14.3 mM BME, pH 7.5). The UBF solution was supplemented with 10% (v/v) glycerol and concentrated in Amicon Ultra 3 MWCO (molecular weight cut-off) concentrator columns (Millipore-Sigma UFC 500396 - hereafter called Amicon concentrators). Concentrated protein was aliquoted into single use volumes (typically 10 μL), flash frozen in liquid N_2 and stored at -80°C.

LYAR

Cells were lysed in supplemented lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, 14.3 mM BME, 200 μM PMSF, pH 7). Supernatant was recovered from a 25-minute spin at 38,000 g and bound to a 10 mL volume of equilibrated HisPur Ni-NTA resin (Fisher - 88222) in a gravity column. The Ni-NTA column was washed in lysis buffer until no contaminant protein was detected by a Bradford assay. Protein eluted in 20 mL elution buffer (20 mM sodium phosphate, 500 mM NaCl, 350 mM imidazole, 14.3 mM BME, 200 μM PMSF; pH 7). Peak fractions from this affinity purification were pooled and diluted five-fold in dilution buffer (20 mM sodium phosphate, 14.3 mM BME; pH 7). This solution was further purified via ion exchange chromatography using a continuous gradient purification protocol with a HiTrap Heparin HP 5 mL column (Cytiva – 17040703), Buffer A (20mM sodium phosphate, 100mM NaCl, 14.3mM BME; pH 7), and Buffer B (20mM sodium phosphate, 100mM NaCl, 14.3mM BME; pH 7) on the ÄKTA Pure

FPLC module. Peak fractions containing His-LYAR were pooled and purified to ≥95% using size exclusion chromatography - HiLoad 16/600 Superdex 75pg column (Cytiva - 28989333) on the ÄKTA Pure FPLC module in storage buffer (22 mM sodium phosphate, 1.1 M NaCl, 14.3 mM BME, pH 7.5). The LYAR solution was supplemented with 10% (v/v) glycerol and concentrated in Amicon concentrators. Concentrated protein was aliquoted into single use volumes (typically 5 μL), flash frozen in liquid N_2 and stored at -80°C.

NUCKS (all variants)

Cells were lysed in supplemented lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, 14.3 mM BME, 200 μM PMSF, pH 7.5). Supernatant was recovered from a 25-minute spin at 38,000 g and bound to an equilibrated HisTrap FF Crude 5 mL column (Cytiva – 11000458) using an ÄKTA Pure FPLC module. The Ni-NTA column was washed in 75 mL lysis buffer then protein eluted in elution buffer (20 mM sodium phosphate, 300 mM NaCl, 350 mM imidazole, 14.3 mM BME, 200 μM PMSF; pH 7.5). Peak fractions from this affinity purification were pooled and diluted five-fold in dilution buffer (20 mM sodium phosphate, 14.3 mM BME; pH 7.5). This solution was further purified via ion exchange chromatography using a contiguous gradient purification protocol with a HiTrap SP 5 mL column (GE – 17115201), Buffer A (20 mM sodium phosphate, 100 mM NaCl, 14.3 mM BME; pH 7.5), and Buffer B (20 mM sodium phosphate, 1.1M NaCl, 14.3 mM BME; pH 7.5) on the ÄKTA Pure FPLC module. Peak fractions containing His-SUMO-NUCKS-tev-SUMO were pooled and cleaved of both SUMO tags during an overnight dialysis in the presence of 0.02X (molar amount) ULP1 protease and 0.1X (molar amount) TEV protease in cleavage buffer (20 mM sodium phosphate, 300 mM NaCl, 1mM DTT, pH 7.5). NUCKS was purified to ≥95% using size exclusion chromatography - HiLoad 16/600 Superdex 75 pg column on the ÄKTA Pure FPLC module in storage buffer (22 mM sodium phosphate, 1.1 M NaCl, 14.3 mM BME, pH 7.5). The NUCKS protein solution was supplemented with 10% (v/v) glycerol and concentrated in Amicon concentrators. Concentrated protein was aliquoted into single use volumes (typically 5 μL), flash frozen in liquid N2 and stored at -80°C.

PolR1F IDR

Cells were lysed in supplemented lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM I midazole, 14.3 mM BME, 200 μM PMSF, pH 7). Supernatant was recovered from a 25-minute spin at 38,000g and bound to an equilibrated HisTrap FF Crude 5 mL column (Cytiva – 11000458) using an ÄKTA Pure FPLC module. The Ni-NTA column was washed in 75 mL lysis buffer then protein eluted in elution buffer (20 mM sodium phosphate, 300 mM NaCl, 350 mM imidazole, 14.3 mM BME, 200 μM PMSF; pH 7). Peak fractions from this affinity purification were pooled and diluted five-fold in dilution buffer (20 mM sodium phosphate, 14.3 mM BME; pH 7). This solution was further purified via ion exchange chromatography using a continuous gradient purification protocol with a HiTrap Heparin HP 5 mL column (Cytiva – 17040703), Buffer A (20 mM sodium phosphate, 100 mM NaCl, 14.3 mM BME; pH 7), and Buffer B (20 mM sodium phosphate, 1 M NaCl, 14.3 mM BME; pH 7) on the ÄKTA Pure FPLC module. Peak fractions containing His-SUMO-PolR1FIDR-tev-GFP were pooled and cleaved SUMO and GFP tags during an overnight dialysis in the presence of 0.02x (molar amount) ULP1 protease and 0.1X (molar amount) TEV protease in cleavage buffer (20 mM Sodium phosphate, 300 mM NaCl, 1 mM DTT, pH 7). PolR1F IDR was purified to ≥95% using size exclusion chromatography - HiLoad 16/600 Superdex 75 pg column on the ÄKTA Pure FPLC module in storage buffer (22 mM sodium phosphate, 1.1 M NaCl, 14.3 mM BME, pH 7.5). The PolR1 F IDR protein solution was supplemented with 10% (v/v) glycerol and concentrated with Amicon concentrators. Concentrated protein was aliquoted into single use volumes (typically 5 μL), flash frozen in liquid N_2 and stored at -80°C.

NCL (all variants)

Cells were lysed in supplemented lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, 14.3 mM BME, 200 μM PMSF, pH 7.5). Supernatant was recovered from a 25-minute spin at 38,000 g and bound to an equilibrated HisTrap FF Crude 5 mL column (Cytiva – 11000458) using an ÄKTA Pure FPLC module. The NiNTA column was washed in 75 mL lysis buffer then protein eluted in elution buffer (20 mM Sodium phosphate, 500 mM NaCl, 350 mM imidazole, 14.3 mM BME, 200 μM PMSF; pH 7.5). Peak fractions from this affinity purification were pooled and diluted five-fold in dilution buffer (20 mM sodium phosphate, 14.3 mM BME; pH 7.5). This solution was further purified via ion exchange chromatography using a continuous gradient purification protocol with a HiTrap Heparin HP 5 mL column (Cytiva – 17040703), Buffer A (20 mM sodium phosphate, 100 mM NaCl, 14.3 mM BME; pH 7.5), and Buffer B (20 mM sodium phosphate, 1 M NaCl, 14.3 mM BME; pH 7.5) on the ÄKTA Pure FPLC module. Peak fractions containing His-GFP-Tev-NCL (or in some cases His-MBP-Tev-NCL) were pooled and cleaved of GFP during an overnight dialysis in the presence of 0.025x TEV protease in cleavage buffer (20 mM Sodium phosphate, 100 mM NaCl, 1 mM DTT, pH 7.5). NCL was purified to ≥95% using size exclusion chromatography - HiLoad 16/600 Superdex 200 pg column (Cytiva - 28989335) on the ÄKTA Pure FPLC module in storage buffer (22 mM Sodium phosphate, 1.1 M NaCl, 14.3 mM BME, pH 7.5). The NCL solution was supplemented with 10% (v/v) glycerol and concentrated in Amicon concentrators. Concentrated protein was aliquoted into single use volumes (typically 10 μL), flash frozen in liquid N2 and stored at -80°C.

pHluorin2

Cells were lysed in supplemented lysis buffer (20 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 14.3 mM BME, 200 μM PMSF, pH 7.5). Supernatant was recovered from a 25-minute spin at 38,000 g and bound to an equilibrated HisTrap FF Crude 5 mL column (Cytiva – 11000458) using an ÄKTA Pure FPLC module. The NiNTA column was washed in 75 mL lysis buffer then protein eluted in elution buffer (20 mM sodium phosphate, 300 mM NaCl, 350 mM imidazole, 14.3 mM BME, 200 μM PMSF; pH 7.5). Peak fractions from this affinity purification were pooled and diluted five-fold in dilution buffer (20 mM sodium phosphate, 14.3 mM BME; pH 7.5). This solution was further purified via ion exchange chromatography using a contiguous gradient purification protocol with a HiTrap Heparin HP 5 mL column (Cytiva – 17040703), Buffer A (20 mM sodium phosphate, 100 mM NaCl, 14.3 mM

BME; pH 7.5), and Buffer B (20 mM sodium phosphate, 1 M NaCl, 14.3 mM BME; pH 7.5) on the ÄKTA Pure FPLC module. Peak fractions containing His-tev-pHluorin (or in some cases His-MBP-Tev-NCL) were pooled and cleaved of His during an overnight dialysis in the presence of 0.025X TEV protease in cleavage buffer (20 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, pH 7.5). pHluorin2 was purified to $\geq$95% using size exclusion chromatography - HiLoad 16/600 Superdex 200pg column (Cytiva - 28989335) on the ÄKTA Pure FPLC module in storage buffer (22 mM sodium phosphate, 1.1 M NaCl, 14.3 mM BME, pH 7.5). The pHluorin2 solution was supplemented with 10% (v/v) glycerol and concentrated in Amicon concentrators. Concentrated protein was aliquoted into single use volumes (typically 10 μL), flash frozen in liquid N_2 and stored at -80°C.

Tagging proteins with fluorescent reporters

A proportion of purified proteins were covalently conjugated using NHS (N-hydroxysuccinimide ester) Alexa Fluor™ dye (see the key resources table for specifics on fluorescent dyes used in this study). After the final stage of purification, ~1 mg of protein was desalted into 100 mM sodium bicarbonate ($NaHCO_3$) buffer at pH 8.3 using a PD10 desalting column (GE Healthcare - 17085101). Desalted protein was concentrated using Amicon concentrators and protein concentration was monitored using 280 nm absorbance on a Nanodrop 2000 (Fisher). Concentrated protein solutions $\geq$5 mg/mL were combined with in NHS-Alexa Fluor™ dye (maintained in DMSO) at a dye:protein ratio of 3:1 molarity. This mixture was incubated under gentle rocking for one hour at room temperature in the absence of light. The mixture was then dialyzed overnight into protein storage buffer (20 mM sodium phosphate, 1 M NaCl, 14.3 mM BME, 10% (v/v) glycerol, pH 7.5) to remove unincorporated dye. Fluorescently tagged protein was concentrated using Amicon 3 kD MWCO columns to $\geq$ 20 μM protein. Concentrations of protein and the covalently attached dye were determined by absorption measurements using nanodrop. Labeling efficiency was calculated as the molar ratio of dye to protein; these values ranged from 0.7-1.2 for all labeled proteins. Labeled proteins were aliquoted into single-use volumes (typically 2 μL), flash frozen in liquid N_2, and stored at -80°C.

Analysis of IDRs using NARDINI+

NARDINI+ is a bioinformatic analysis of the human IDRome that stands for Non-random Arrangement of Residues in Disordered Regions Inferred using Numerical Intermixing. Protein sequences from the human proteome were downloaded using Swissprot *Homo sapiens* database (May 2015, 20882 entries).[57] Each sequence was analyzed using MobiDB to extract regions that are predicted to be disordered.[58] A residue was considered disordered if the consensus prediction labeled it as disordered. All consecutive disordered stretches of length greater than or equal to 30 residues were extracted leading to 24,508 IDRs. We refer to this set of IDRs as the human IDRome. All IDR sequences were analyzed using 90 sequence features suggested to be important for the function and / or phase separation of disordered regions.[43,59] The sequence features were split into patterning and composition categories.[38,39,43,52] Patterning features were analyzed using NARDINI.[38,52] Briefly, each wild-type sequence was compared to 10^5 random sequences of the same composition to determine whether the wild-type sequence has specific binary patterns that are more well-mixed or blockier than expected by chance. For patterning analysis, residues were grouped as follows: pol $\equiv$ (S, T, N, Q, C, H), hyd $\equiv$ (I, L, M, V), pos $\equiv$ (K, R), neg $\equiv$ (E, D), aro $\equiv$ (F, W, Y), ala $\equiv$ (A), pro $\equiv$ (P), and gly $\equiv$ (G). Z-scores greater than zero imply the disordered sequence is blockier than expected and z-scores less than zero imply the disordered sequence if more well-mixed than expected for the given residue groups. For each IDR sequence, we also extracted 54 compositional features. Most compositional features were quantified using localCIDER.[50] These included the fraction of each amino acid type (20 features); the fraction of positive, negative, polar, aliphatic, and aromatic residues (5 features); the ratios of Ks to Rs and Es to Ds (2 features); the fraction promoting **c**hain **e**xpansion (FCE) residues, the fraction of charged residues (FCR), the net charge per residue (NCPR), the fraction of disorder promoting residues, the hydrophobicity, the isoelectric point, and the polyproline-II propensity (7 features). We also included "blocky" features defined as the fraction of the IDR sequence that is made up of blocks of a given residue type or Arginine-Glycine (RG) pair. Here, a block had to have at least four occurrences of the given residue or two occurrences of RG and could not extend past two interruptions. We examined 20 block features as no W block was found in the human IDRome. Then z-scores for each individual sequence in the human IDRome were calculated using the mean and standard deviation of the entire IDRome.

Extraction of condensate-enriched sequence features

Protein sequences from nucleoli, nuclear speckles, and nuclear bodies were extracted from the Human Protein Atlas (HPA).[79] Cross-referencing with the human IDRome led to 2231, 911, and 1143 IDRs located in nucleoli, nuclear speckles, and nuclear bodies, respectively. The datasets from Hubstenberger et al., and Jain et al., were used to determine P-body and stress granule proteins.[60,61] Cross-referencing with the human IDRome led to 209 and 639 IDRs located in P-bodies and stress granules, respectively. To extract sequence features that are enriched and condensate-specific, the human IDRome was split into the condensate-specific IDRome and the remaining IDRome. Then, the distribution of z-scores for a given sequence feature in the two sets was examined and a p-value was calculated using the two-sample Kolmogorov-Smirnov (KS) test[62] to determine whether the two distributions were identical. If the p-value < 0.05, then the signed $\log_{10}$(p-value) was calculated. Here, the $\log_{10}$(p-value) was positive if the mean z-score was larger in the condensate IDRome and negative if the mean z-score was smaller in the condensate IDRome. Thus, positive $\log_{10}$(p-value) values imply that compositional feature is enriched in the condensate, or the patterning feature is blockier in the condensate, whereas the converse is true for negative $\log_{10}$(p-value) values.

Extraction of IDR Clusters within Nucleolar Proteins

Sequence feature z-scores for all 2231 nucleolar IDRs were hierarchically clustered using the Euclidean distance and Ward's linkage method.[64] Using a cophenetic distance threshold of 90, five clusters were identified. We focused on three of the clusters given that the other two clusters had only one or two IDRs.

Defining D/E-tracts and K-blocks+ERRs IDRs

IDRs containing D/E-tracts were defined by having a negative-negative patterning z-score $\geq$ 3. IDRs containing K-blocks+ERRs were defined by having a positive-positive and positive-negative patterning z-score $\geq$ 3, a K Block z-score $\geq$ 1, a fraction of Es > 0 and greater than or equal to the fraction of Ds, and a negative-negative patterning z-score less than five times the positive-positive patterning z-score. The use of these criteria led to 140 D/E-tract containing IDRs, 36 K-blocks+ERRs containing IDRs, and 9 IDRs that contained both features.

Determining enrichment of IDR sequence features in different condensates

To determine the enrichment of high z-scoring features, the human IDRome was split into the condensate specific IDRome and the remaining IDRome. The number of IDRs with a z-score $\geq$ 3 for each of the 90 sequence features, as well as the number of IDRs containing D/E-tracts and K-blocks+ERRs, for each category was determined. Then enrichment was determined using the one-sided Fisher's Exact Test.

Evolutionary Analysis of Nucleolar IDRs

The list of eukaryotic orthologous proteins was extracted using EggNOG[74] with the full human nucleolar sequence as the input. BLASTp was then used to keep only the most similar ortholog for each species per nucleolar protein. For LYAR (Q9NX58) residues 1 to 52 were used as the query, for POLR1F (Q3B726) residues 1 to 205 were used as the query, for NUCKS (Q9H1E3) residues 1 to 243 were used as the query, for NCL (P19338) residues 307 to 647 were used as the query, and for UBF (P17480) residues 298 to 475 were used as the query. Next, MUSCLE multiple sequence alignment was used to align the orthologous sequences.[75] Then, the alignments were loaded into Jalview[76] and the full protein sequences were cut to only include the region that corresponds to the human IDR of interest. Orthologous IDRs with unnatural amino acids or sequence lengths less than fifteen were then removed and NARDINI analyses were performed on the remaining orthologous IDRs. Distributions of charge patterning z-scores were then compared to the polar-polar patterning z-score distribution to assess significance using the Wilcoxon signed-rank test.[77] The polar-polar patterning z-score distribution was chosen as a control given that polar residues are highly abundant in IDRs and polar-polar blockiness was not found to be an enriched feature in nucleolar IDRs.

Calculating Sequence Block Profiles of IDRs

To calculate sequence block profiles, local asymmetries were calculated across sliding windows of five residues using $\psi_x = \frac{(f_x - f_{x\neq y})^2}{(f_x + f_{x\neq y})}$ for each sliding window. Here, x is either positive residues (K, R) or negative residues (D, E) and f_x is the fraction of those residues within the five-residue sliding window. Additionally, ψ_x was set to be positive if $f_x \geq f_{x\neq y}$, and negative otherwise. Then the mean signed ψ_x was calculated for each residue in the IDR by averaging the signed ψ_x overall all sliding windows containing that residue. Thus, blocks of a given charge should have positive signed ψ_x values. Negative charged blocks were then set to $-\psi_x$ such that positive and negative blocks could be plotted together and identified across zero. Two blocks of like-charge were combined if the number of interruptions was less than or equal to two.

Analysis of Human Proteome Atlas (HPA) dataset

UniProt accessions were used to map proteins to subcellular regions and by whether they contained D/E-tract or K-blocks+ERRs IDRs onto the uniform manifold approximation and projection for dimension reduction (UMAP) coordinates of the HPA dataset (April 2019).[80] For subcellular location, the human protein atlas field category Subcellular location (ICC) was used to extract the accessions of nuclear speckles IDRs, Nuclear bodies IDRs, FC/DFC (Nucleoli fibrillar center) IDRs, and GC (Nucleoli) IDRs. Additionally, D/E-tract or K-blocks+ERRs IDR containing nucleolar proteins were mapped to the UMAP data for visualization of the cellular localization using their UniProt accessions. Visualization was achieved using the ImJoy HPA-UMAP plugin.[82]

Identification of DNA / RNA-binding Domains

The sequences of nucleolar D/E-tract containing proteins were run though the NCBI Batch CD-Search Tool using default parameters and "Concise Results" output. To identify which domains corresponded to DNA/RNA-binding domains cddannot.dat was downloaded from https://ftp.ncbi.nlm.nih.gov/pub/mmdb/cdd/ for CDD domains and Pfam-C from http://ftp.ebi.ac.uk/pub/databases/Pfam/current_release/ for PFAM domains. CDD accessions whose function contained "DNA bind" or "DNA-bind" were classified as DNA-binding domains and CDD accessions whose function contained "RNA bind" or "RNA-bind" were classified as RNA-binding domains. PFAM DNA- and RNA-binding domain accessions[83] were also determined using the same search criteria within the Pfam-C file. The list of CDD and PFAM DNA- and RNA-binding accessions were then used to categorize the D/E-tract containing proteins.

Selection criteria for summary of enriched nucleolar proteins

For the summary analysis in Figure 2G only high scoring D/E-tracts, K-blocks + E-rich-regions, and RG regions containing IDRs were considered. Here, the patterning negative-negative, patterning positive-positive, and RG Fraction z-scores had to be greater than or equal to 5.5, respectively.

Bioinformatics analysis of NCL sequence features

NCL has the 7th highest RG fraction z-score of all 2231 nucleolar IDRs of length $\geq$ 30. HeLa cell proteome abundance data were extracted from Nagaraj et al.[63] The mean and standard deviation of this dataset was used to determine the abundance z-score values for all nucleolar proteins in this dataset. NCL has the 25th highest z-score of all 804 nucleolar proteins in the dataset of Nagaraj et al.[63] NCL has the highest negative-negative z-score of all 2231 nucleolar IDRs.

Atomistic simulations of K-Block ERR IDRs

The ABSINTH implicit solvation model[100] and forcefield paradigm as implemented in the CAMPARI simulation engine (http://campari.sourceforge.net) with appropriate local and non-local Monte Carlo move sets[101] was used to perform atomistic simulations using a parameter set based on abs3.2_opls.prm. The radius of the sodium ions was increased to 1.81 Å to improve sampling of sequences with high negative charge regions. Each simulation was performed at a temperature of 340 K in a spherical droplet of 275 Å with counterions and an excess of 5 mM NaCl, which were modeled explicitly.[102] For each IDR, five independent simulations were conducted with 10^7 equilibration steps and 5.5x10^7 production steps.

Analyses of trajectories were performed using the SOURSOP analysis package.[103] Homopolymer scaled distance maps were computed by first finding the best fit homopolymer model using the function get_scaling_exponent(). This function estimates A_0 and ν using the standard polymer relationship: $\langle r_{ij}^2 \rangle^{1/2} = A_0 |i - j|^{\nu}$. Here, r_{ij} is the distance between residue number i and j, A_0 is a prefactor that provides information on the average chain persistence length and volume, and ν yields information on the solvent quality. Then, A_0 and ν were used as inputs into the function get_polymer_scaled_distance_map() which calculates the average distance between residues from the simulations normalized to the expected distance based on the best fit homopolymer model.

LaSSI simulations of condensates formed by components of fibrillar centers

Simulations were performed using LaSSI, a lattice-based Monte Carlo simulation engine.[104] Monte Carlo moves are accepted or rejected based on the Metropolis-Hastings criterion Total system energies were calculated using a nearest neighbor model with pairwise interaction parameters described in supplemental Table S2. These parameters were chosen to reflect a domain-spacer architecture wherein the domains include HMGs that bind to DNA, as well as oppositely charged blocks of residues that make complementary interactions with one other. On the lattice, these interactions between oppositely charged blocks are short-range interactions, reflecting the high screening length expected at physiologically relevant salinities. Like-charged blocks of residues form very weak interactions with each other, behaving as spacers that contribute a high positive excluded volume. All other interactions are of a medium strength, typical of spacers that contribute a low or negligible excluded volume. All simulations described in this study were performed at a reduced temperature of $T = 55$, written in units of k_B the Boltzmann constant set to unity. Each simulation comprises a total of 4×10^{10} Monte Carlo steps and ten independent simulations were performed for each system and condition to enable statistical rigor. The simulations described in Figures 3F and 3G involved 250 molecules of NUCKS, 100 molecules of UBF tetramers, and 150 molecules mimicking rDNAs. We used a cubic lattice with periodic boundary conditions, and each side in the central cell comprises 140 lattice units. The simulations described in Figure S6C involved 385 molecules of PolR1F IDR, 100 molecules of UBF tetramers, and 150 molecules of rDNAs. Again, we used a cubic lattice with periodic boundary conditions, and each side in the central cell comprises 140 lattice units. Chain architectures were determined by coarse-graining IDRs such that each bead in LaSSI is equivalent to ~10-15 amino acid residues. All simulations were initialized using a smaller cubic lattice to speed up the equilibration into a single condensate. After initialization, the system was allowed to equilibrate fully, as determined by a plateauing of the total system energy. Although this plateauing typically occurs after about 5×10^9 Monte Carlo steps, simulations were only analyzed after 2×10^{10} Monte Carlo steps to ensure a fully equilibrated system. All simulations comprised a single large condensate with a coexisting dilute phase. Partition coefficients (PCs) of different molecules were calculated as follows: (1) We calculated the radial density of each macromolecular species in the system with reference to the condensate center-of-mass. (2) We fit the following logistic function to the radial density data:

$$\log_{10}[\varphi(x)] = \frac{1}{2}[\log_{10}(\varphi'') + \log_{10}(\varphi')] - \frac{1}{2}[\log_{10}(\varphi'') - \log_{10}(\varphi')]\tanh\left[\frac{2(x - x_{mid})}{\delta}\right].$$

Here, $\varphi(x)$ is the calculated radial density, φ'' is the density in the dense phase, φ' is the density in the dilute phase, x_{mid} is the distance between the condensate center-of-mass and the midpoint of the logistic curve, and δ describes the width of the interface between the dense and dilute phases.[124] (3) Lastly, we calculate the PC as φ''/φ'. We followed this routine for all replicates and report the mean value alongside the standard error from the mean for all 10 replicates.

Image collection and analysis

Confocal imaging of all samples except those with SNARF dyes were carried out on a Nikon Eclipse Ti2 microscope with a Yokogawa CSU X1 disk module and a LunF laser launch equipped with 405 nm, 488 nm, 561 nm, and 647 nm lasers. Images shown were taken using a 60X, 1.4 NA Apo oil immersion objective (Nikon) and a Hamamatsu Orca Flash 4.0 CMOS camera. All experiments were carried out at room temperature. NIS-Elements software was used for all image acquisition. Images within a data set were taken with identical imaging parameters ensuring that signal was not saturated (averaging, binning, and projecting were not used). Images were taken and saved as 16-bit '.nd2' files. ImageJ, Python3, and MATLAB were used for all quantitative image analysis (details provided below). All images shown are representative crops of one or a few entities (e.g., a condensate) and the brightness and contrast have been optimized.

Samples that contain SNARF dye were imaged on Zeiss LSM 980 with Airyscan 2 using a 63x 1.4 NA oil immersion objective equipped with manufacturers specific laser lines, multi-element spectral detector, and 16-bit CMOS camera. Zen Blue software was used for all image acquisition. Samples containing GFP were imaged with 488nm laser. All SNARF dyes were imaged with 561nm laser excitation followed by an emission scan from 570-696nm with 8.8nm step sizes.

Isolation and imaging analysis of nucleoli from *Xenopus* oocytes

Harvested oocytes are manually de-flocculated, then subjected to collagenase digestion (gentle rocking) for 2 hours at 18°C. Oocytes are stored in ND96 Buffer (0.005 M HEPES, 0.096 M NaCl, 0.002 M KCl, 0.0018 M $CaCl_2$, 0.001 M $MgCl_2$) that was filter sterilized and supplemented with 2.5 mM sodium pyruvate (Thermo - 11360070) and 1X penicillin-streptomycin (Sigma - P4333) at 18°C. Healthy stage VI oocytes were selected and injected using freshly pulled microneedles (Drummond - 3-000-203-G/X) and a Drummond Nanoinject II (3-000-204). A total of 23 nL of mRNA in ddH_2O (typically at a total mass of 20 ng) were injected into each oocyte. In all cases, oocytes were injected with mCherry-Pol1Re mRNA or GFP-Pol1Re mRNA and an mRNA coding for a protein of interest (GFP or RFP tagged). Injected oocytes were stored individually in wells in a 48-well polystyrene SterileTissue Culture Plates (Fisher - FB012930) in supplemented ND-96 buffer at 18°C for at least 18 hours to allow expression and localization of exogenous protein. Immediately prior to imaging, germinal vesicles were manually dissected in mineral oil and mounted on a glass slide with 6 μL of mineral oil. A 22 x 22mm glass coverslip was gently overlaid onto the sample and nucleoli were then immediately imaged. This procedure was carried out for all proteins of interest for at least two separate harvests of oocytes and similar results were obtained across these biological replicates.

Nucleoli were imaged as a confocal Z-stack. All images shown are of a single Z-slice that is representative and is at least 2 μm above the coverslip; brightness and contrast have been optimized. For more details, see Image collection and analysis section. FC area was calculated as follows: individual crops of an entire nucleolus were Otsu thresholded in the PolR1e channel (a *bona fide* FC-localizing protein) to create a mask corresponding to the location of FCs. Apparent FCs smaller than 3^2 pixels (0.10835 μm/pixel) were omitted due to falling below resolution limits. Using the analyze particle and particle manager tools, the FC mask was used to obtain the area (μm^2) and the mean intensity value (AU) of FCs for both the PolR1e channel and the protein of interest channels. The area obtained using the PolR1e channel was used to report condensate size in the context of expressing various proteins of interest. The mean intensity values of the protein of interest channel for individual FCs was saved as 'FC signal'. To use these values to obtain PCs, we carried out additional analysis as follows: An independent thresholding procedure was applied to duplicated images to obtain the aggregate intensity of all signals that fell below the FC, which we term 'non-FC signal'. Individual PCs are the difference of the FC signal and non-FC signal, corrected for microscope background. PC values are reported as a normalized value where 1 is the non-FC signal.

Enrichment of target proteins in a nucleolar sub-phase was determined by relative RNA Polymerase I subunit E (PolR1e) fluorescence. The intensity of PolR1e is highest in the FC and drops in a stepwise fashion with the following hierarchy: FC > DFC > GC > NP (intensities are normalized to peak value of 1 – shown in Figure S2A). Mapping the sub-phase enrichment hierarchy of target proteins involved co-injection of PolR1e tagged with fluorescent protein – FP-PolR1e and FP-tagged version of the target protein. Differential Otsu thresholding of the hierarchal PolR1e signal and the consequent masks were used to demarcate sub-phases. Demarcated sub-phases were used to identify which sub-phase contains the peak signal of the target protein. For all proteins of interest at least 50 FCs from at least three independent germinal vesicles were analyzed for FC size.

In vitro condensate / phase separation assays

To prepare *in vitro* reconstituted condensates, mixtures of protein(s), DNA, and/or RNA(s) were combined in protein storage buffer (20 mM sodium phosphate, 1 M NaCl, 14.3 mM BME, 10% glycerol at pH 7.5) at 10x the final desired concentration. The mixture was then diluted 10X into a no salt buffer (20 mM sodium phosphate 14.3 mM BME, pH 7.5) to achieve a final [NaCl] of 100 mM (roughly physiological conditions). Ten-fold dilutions into higher [NaCl] buffers were used to map the salt condensation thresholds (Figures 3C and 3D). All dilutions were performed directly in individual wells in a glass-bottomed 384-well plate (Cellvis- P384-1.5H-N). Immediately upon dilution these mixtures were mixed by pipetting up and down three times. Wells were then sealed, and the contents allowed to equilibrate at room temperature for 20 minutes. After equilibration, the contents of these mixtures were imaged as a confocal Z-stack. All images shown are of a single Z-slice taken at least 1 μm above the coverslip and are representative of the contents of the well; brightness and contrast have been optimized. For more details, see Image collection and analysis section.

PC values of *in vitro* condensates were obtained as follows. For an entire field of view, Otsu thresholding in the channel corresponding to the protein of interest was used to create a mask corresponding to individual condensates. Entities smaller than 5^2 pixels (0.10835 μm/pixel ≈ ≤0.5um in diameter) were omitted. Using the analyze particle and particle manager tools, the mask was used to obtain the mean intensity value (AU) of condensates for all channels. The inverse mask was used to obtain the aggregate intensity of 'non-condensate' signal. Individual PCs are the difference of the condensate signal and non-condensate signal, corrected for microscope background. PC values are reported as a normalized value where 1 is the non-condensate signal.

Analysis of monomer-tetramer equilibrium of UBF

For both mass photometry and gel electrophoresis analysis of UBF, the protein mixture in storage buffer (20 mM sodium phosphate, 1 M NaCl, 14.3 mM BME, 10% (v/v) glycerol at pH 7.5) at 500 nM UBF was diluted 10X into a no salt buffer (20 mM sodium phosphate, 14.3 mM BME, pH 7.5) to achieve a final [NaCl] of 100 mM and UBF of 50 nM (endogenous concentration). This mixture was kept at room temperature for 20 minutes, then analyzed. For gel electrophoresis measurements, equivalent dilutions were also performed into 20 mM sodium phosphate, 1 M NaCl, pH 7.5 buffer.

Mass photometry measurements were carried out on a Refyn One Mass Photometer. For a single measurement protein solution (~10 μL) was applied to a pre-focused slide and counts were recorded for 3 minutes (automated quantification of molecular weight based on airy pattern intensity). The resultant frequency distribution histogram distributions had two peaks at predicted MWs of 109 kDa and 443 kDa, representing monomer and tetramer species of UBF, respectively.

For gel electrophoresis measurements, UBF in either 100 mM NaCl or 1 M NaCl was diluted 4X into 4X Laemmli SDS sample buffer (Fisher - AAJ60015AC) then immediately run on SDS-PAGE (no sample heating). Gels were stained using Coomassie stain and destained. Gels were imaged on a flatbed scanner and processed for optimal brightness and contrast in ImageJ. A line scan of the 100 mM NaCl sample was carried out to determine the positional locations and area-under-the-curve (AoC) values for the monomeric and tetrameric species of UBF. AoC values were obtained by determining the area occupied by each species with the MagicWand tool. Relative amounts of each species were determined using the following: Tetramer relative amount = Tetramer AoC/Tetramer AoC + Monomer AoC; Monomer relative amount = Monomer AoC/Tetramer AoC + Monomer AoC.

Wetting assays *in vitro*

Coverslips (size 22x55 mm) for wetting assays with hydrophobic surface coatings were prepared following a PEGylation protocol modified from Piehler et al.[89] Glass coverslips were cleaned with sonication using 1.5 M NaOH in 50% ethanolfor 30 min, followed by two rinses in 1 L beakers of MilliQ water. Coverslips were then cleaned using piranha solution (60 vol% sulfuric acid, 40 vol% H_2O_2) and sonication for 45 min, rinsed, dried, and functionalized with (3-glycidyloxypropyl)trimethoxysilane (GOPTS) for 1 hr at 75°C. Excess GOPTS was rinsed from coverslips using anhydrous acetone, and a mixture of biotin-PEG-amine/methoxy-PEG-amine (Rapp Polymer) was prepared in anhydrous acetone at a ratio of 10 mol% biotinylated PEG. The PEG solution was coupled to coverslips overnight at 75°C, rinsed twice in 1 L beakers of MilliQ water, and stored in MilliQ water at 4°C until use.

Coverslips for wetting assays with hydrophilic surface coatings were prepared in a similar way up until the functionalization step. At this stage coverslips placed in a humidity chamber and 10% solution of Pluronic F-127 and MilliQ (Sigma 540025) was deposited on the coverslip surface.[18] These were allowed to incubate for 1 hr and were prepared the day of use. Prior to use, both hydrophobic surface coated (PEGylated) and hydrophilic surface coated (Pluronic F-127) coverslips were rinsed in ethanol, dried, and silicon well insulators were adhered (Grace Biolabs 665206) on the treated surface. From this point on, phase separation assays were carried out as described above.

Standard curves for measurements of pH

Standard curves for SNARF-1 and SNARF-4 dyes used for analysis of pH were generated by preparing buffers at a range of pH values with SNARF dyes (1:500 dilution) and imaging them in a similar manner as GVs. Buffers for *in vivo* standard curves were as follows: 20 mM MES, 100 mM KCl (buffered to pH values ranging from to 5, 6, or 6.5), 20 mM HEPES, 100 mM KCl (buffered to pH values of 7.0, 7.5, or 8.0), 20 mM CHOPS, 100 mM KCl (buffered to a pH of 9). To prepare buffer samples for imaging, 1 μL of buffer containing either SNARF-4 or SNARF-1 dye was pipetted into the center of a 5 μL drop of mineral oil on a slide, which was then overlaid with a size 22x22 mm coverslip. This procedure matches how GVs were mounted. A standard curve for SNARF-4 dye used for analysis of pH *in vitro* was generated using 20 mM sodium phosphate,100 mM NaCl at pH values of 6.0, 6.5, 7.0, 7.5, and 8.0. The buffer / SNARF-4 mixture was mounted in a coverslip well to match how *in vitro* condensates are imaged.

All buffers were titrated using a Sartorius (PB-11) pH meter with a KCl liquid-filled platinum junction pH probe (Sartorius, PY-11) to be with 0.02 units of their target pH using KOH, NaOH or HCl, as needed. We noticed pH readings of buffers used for the standard curve usually do not deviate beyond 0.05 throughout the day of the experiment, but a deviation of 0.09 was observed once, therefore a conservative error of 0.1 is applied to all measurements reported in Figure SI7. Slides with all pH buffers were prepared one a single imaging day and imaged using the procedure described above. To obtain estimates of error introduced by imaging, this procedure was carried out two additional times on different days and the reported standard curve shows mean and standard error of mean values across the triplicate of samples.

Direct electrochemical probe measure of pH

The pH of the undiluted nucleoplasm was measured using a Sartorius (PB-11) pH meter and a Thermo Scientific™ Orion™ 9810BN Micro pH Electrode (13-299-511) probe. Approximately thirty Germinal Vesicles (GVs) from living *Xenopus leavis* oocytes were harvested via manual dissection in mineral oil. These GVs were collected within ~100 μL of mineral oil and spun at 2000 RPM for 1 minute to allow them to coalesce. The pH of this aggregate GV was measured directly using the Thermo Scientific™ Orion™ 9810BN Micro pH Electrode and pH meter.

pHluorin2-based measurements of pH

A standard curve for pHulorin2 used for analysis of pH was generated by preparing buffers at a range of pH values with recombinant, purified pHluorin2 at 10 μM and imaging them in a similar manner as GVs. Buffers for the standard curve were as follows: 20mM MES, 100mM KCl (buffered to a pH of 5, 6 or 6.5); 20 mM HEPES, 100 mM KCl (buffered to a pH of 7, 7.5, or 8); 20 mM CHOPS, 100 mM KCl (buffered to a pH of 9). To prepare buffer samples for imaging, 1 μL of buffer/pHluorin2 mixture was pipetted into the center of a 5 μL drop of mineral oil on a slide, which was then overlaid with a 22x22 mm coverslip. This procedure matches how GVs are mounted.

Samples (both standard curve samples and *in vivo* samples) were imaged on the spinning disc confocal microscope (described above). Samples were excited in series, first by a 405 nm laser then by a 488 nm laser, both at the same laser power and exposure time. In both cases, the emission parameters are matched for default imaging of GFP samples, namely using a bandpass filter centered at 510 nm. Acidic samples exhibited higher emission intensities when excited with a 405 nm laser, whereas basic samples excited with 488 nm laser emit higher intensity signal in than when excited with a 405 nm laser. Therefore, the ratiometric basis of pHlurorin2 is based on alternate emission intensities arising from different ratios of wavelength dependent excitability that is influenced by pH [113] For samples in cells, the procedure described under "In vivo nucleoli" was used to obtain GVs expressing freely diffusing pHluorin2.

If we denote the emission intensities at 510 nm obtained upon excitation at 405 nm and 488 nm as $I_{ex:405}$ and $I_{ex:488}$, respectively, then r_{510} is the ratio of $I_{ex:488}$ to $I_{ex:405}$. If r_{510} is less than one, then the environment of pHluorin2 is acidic with respect to the buffer; if r_{510} is greater than one, then the environment of pHluorin2 is basic with respect to the buffer; and if r_{510} is equal to one, then the pH of the environment of pHluorin2 is that of the buffer solution. To test pH of nuclear bodies in cells, mRNA encoding pHluorin2 was injected into oocytes. We used purified versions of pHluorin2 to measure its standard curve as a function of pH, thus providing a ratiometric calibration of r_{510}.

Analysis of pH in cells

To measure the pH of nuclei in *Xenopus* oocyte GVs, we first segment the GC, DFC/FC region for nuclei whose GC regions are labeled with GFP-NPM1 (Likewise, we denote FC from the remainder of the nucleolus via GFP-UBF). We image nucleoli that a circular and assumed circular shapes for the boundaries between the GC region and nucleoplasm (NP) as well as between GC and DFC. The center of the GC-NP boundary is determined by fitting the GFP images with a circle. The radii of the GC-NP and GC-DFC boundary circles are determined based on the radius at which the GFP intensity reaches half of the peak intensity, as shown in Figures 7E–7H and S7J–S7L. We then define the DFC/FC regions as the area within the GC-DFC boundary, while the GC regions are the areas outside GC-DFC boundary but inside the GC-NP boundary. The GC region is then further segmented into inner and outer regions, denoted as GC(in) and GC(out), respectively. This segmentation was achieved by using an inner-outer circular boundary where GFP intensity reaches its peak. Additionally, we selected a circular NP region near the current nucleolus.

We quantify the average SNARF intensity within the FC/DFC, GC(in), GC(out), and NP regions at each wavelength. To determine the pH, we use the intensity ratio (R) between 579 nm and 646 nm. The pH value is calculated using the formula pH=(R/82840) – 0.15+C, where C is employed to normalize the pH of NP to 7.2 as determined by direct measurement with an electrochemical probe. The value of C equals the difference between 7.2 and the averaged pH of NP across the six nuclei we analyzed. We also calculate the pixel-wise intensity ratio by dividing the SNARF image at 579 nm by the corresponding image at 646 nm. We then smooth the regions inside the nucleolus using a Gaussian kernel with a standard deviation of 1.5 pixels. To avoid edge effects, smoothing only occurs inside the nucleolus, using the boundary as the point at which NPM1-GFP intensity reaches half of the peak intensity.

For nucleoli with FC regions labeled with GFP, we determined the radius and center of the NP-GC boundary circle by fitting the SNARF image with a circle. The FC regions are defined as regions where GFP intensity exceeds 600 photons. The GC/DFC regions are then defined as regions within NP-GC boundary but outside the FC region. The same pH calculations as described earlier are performed to determine the pH values of the FC, GC/DFC, and NP regions.

Analysis of pH in vitro

For calibrations measurements at a given pH, intensities are obtained for N different wavelength channels corresponding to a spectral emission scan from 570nm to 696nm in roughly 8.8nm increments as described above. In total, 14 intensity measurements were collected spanning the spectrum from 570 nm to 693.2 nm: 570.0 nm, 578.8 nm, 596.4 nm, 605.2 nm, 614.0 nm, 622.8 nm, 631.6 nm, 640.4 nm, 649.2 nm, 658.0 nm, 666.8 nm, 675.6 nm, 684.4 nm, and 693.2 nm. The first intensity measurement at 570.0 nm was not included for the calibration analysis because in the raw intensities for some wavelengths were within 3-fold of microscope background signal. Five different buffers were used, each at a pH from 6 to 8 in 0.5 pH unit increments. A similarity matrix is calculated for all N channels from the ratios of the intensity measurements at a given pH. Next, for a given ratio, $r_{ij} = Em_i / Em_j$, each corresponding *ij* pair is selected for each pH measurement. An exponential curve is fit to these ratios as a function of pH, which has

the form: ae^{bx}. If the ratios correspond to the upper triangle of the similarity matrix, the fitting function is ae^{-bx}. We filter the curves based on the R^2 value as well as the slope of the curves when fit to a linear regression.

To minimize variability, curves that have a $R^2 < 0.97$ are excluded. Furthermore, curves whose slopes are less than 0.5 are also excluded. The 0.5 value is chosen as it also corresponds to our measurement confidence in pH – i.e., the minimal measurement increment in pH is 0.5. The filtered curves are then min-max normalized for uniform comparison, using the expression: $y_{scaled} = \frac{y - y_{min}}{y_{max} - y_{min}}$. Rearranging the fitting equations to use the normalized values, we obtain: $x = \mp\frac{1}{b}\ln\frac{y}{a}$ where $y = y_{scaled} * (y_{max} - y_{min}) + y_{min}$. The sign of the product that is calculated for x is determined by the exponential fitting functions used for the corresponding upper and lower triangles of the similarity matrix, negative for the upper triangle and positive for the lower triangle. With this formula, we can calculate the estimated pH as the corresponding x-coordinate of the point where a horizontal line defined by r_{ij}, intersects with a calibration curve. We use the filtered calibration curves from the upper triangle of the similarity matrix as they produce values that more closely match estimated values of the dilute phase. For images with an unknown experimental value for pH but measured at the same wavelengths as the calibration curves, we calculate the intensity ratios based on the wavelengths that correspond to the r_{ij} ratios of the filtered calibration curves. Using a determined r_{ij} ratio, we can calculate a distribution and its associated median of the estimated pH from the rearranged formulae for all filtered calibration curves.

The above procedure is deployed on a per-pixel basis to obtain an image of condensates and surrounding dilute phase as a function of pH. To obtain median values of pH for entire condensates and mitigate against per pixel measurements that fluctuate across channels, we instead calculate the average value of the regions corresponding to the condensate and dilute phases in each image. In this case, selections, called masks, refer to condensates, and the complementary area refers to the dilute phase. In each case, condensates selected using the thresholding procedure described above (*in vitro* condensate / phase separation assays sections) are averaged across each filtered calibration curve channel. We minimize noise contributions by only selecting objects whose area exceeds 50 squared pixels. The full complement of ratios for each condensate is then calculated, and a corresponding pH can be obtained as described above. For each condensate, there are N calculated values for the pH based on the used filtered calibration curves. We calculate the median of the calculated pH for each condensate and report those values in a violin plot. Similarly, for the dilute phase, 3 square patches with a minimum size of 50x50 are selected and the pH is calculated from the mean intensity values as indicated.

Data visualization

Plots were generated using custom Python 3 scripts; these are available on the Github repository - https://github.com/Pappulab/King-etal-Nucleoli.

QUANTIFICATION AND STATISTICAL ANALYSIS

The distribution of z-scores for a given sequence feature in the two sets was examined and a p-value was calculated using the two-sample Kolmogorov-Smirnov (KS) test[62] to determine whether the two distributions were identical. If the p-value < 0.05, then the signed $\log_{10}$(p-value) was calculated. Here, the $\log_{10}$(p-value) was positive if the mean z-score was larger in the condensate IDRome and negative if the mean z-score was smaller in the condensate IDRome. Signed $\log_{10}$(p-values) are shown in Figures 1A–1C and 2B.

Distributions of charge patterning z-scores were then compared to the polar-polar patterning z-score distribution to assess significance using the Wilcoxon signed-rank test.[77] Significantly different distributions of sizes were those with p-values < 0.05; **denotes p-value 10^{-4}, ***denotes p-value $<10^{-5}$. These data are presented in Figure 1E.

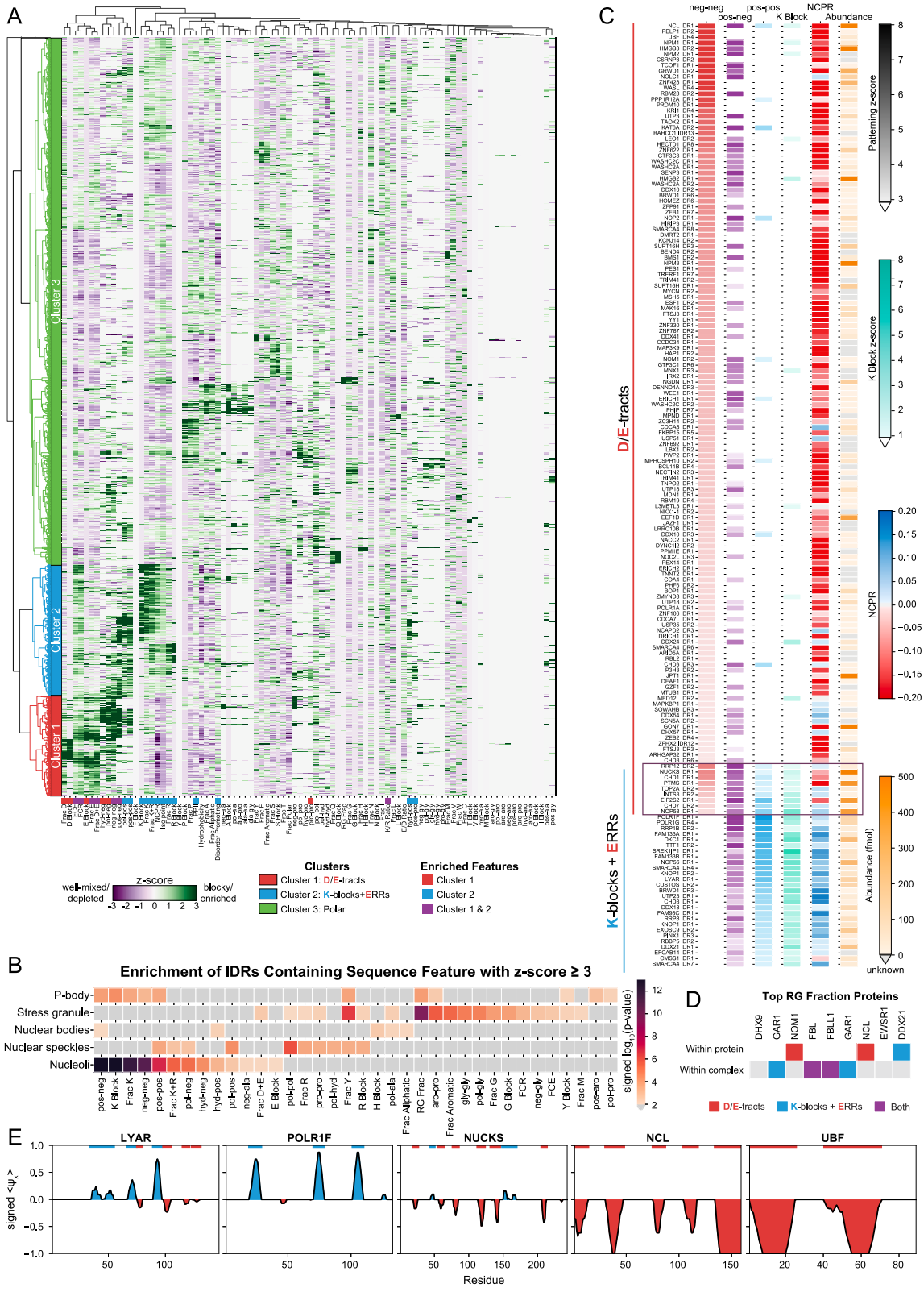
Comparative statistics of FC sizes were performed by extracting p-values using 1-tailed Mann-Whitney-Wilcoxon non-parametric test using the MATLAB function mwwtest. Significantly different distributions of sizes were those with p-values < 0.05; **denotes p-value 10^{-4}, ***denotes p-value $<10^{-5}$. Similarly, comparative statistics of normalized partition coefficient (PC) values of signal within the FC were performed by extracting p-values using 2-tailed Mann-Whitney-Wilcoxon non-parametric test using the same MATLAB function and where ***denotes p-value $<10^{-5}$. For all proteins of interest at least 20 FCs from at least three independent germinal vesicles were analyzed for FC size and their PC values and are plotted in Figure 6I.

Comparative statistics of PC values for protein signal in *in vitro* reconstituted FC-like condensates were performed by extracting p-values using 2-tailed Mann-Whitney-Wilcoxon non-parametric test using the MATLAB function mwwtest. Significantly different distributions of sizes were those with p-values < 0.05, ***denotes p-value $<10^{-5}$. Data from at least 50 *in vitro* condensates, obtained from at least two independent experiments were plotted and used to obtain PC values – shown in Figure 6G.

Comparative statistics of calculated pH values of *in vitro* condensates were performed by extracting p-values using 2-tailed Mann-Whitney-Wilcoxon non-parametric test using the MATLAB function mwwtest. Significantly different distributions of sizes were those with p-values < 0.05, ***denotes p-value $<10^{-5}$. Data from at least 25 condensates (dense phase) and at least 5 control regions (dilute phase) were obtained from at least two independent experiments. These results are shown in Figures 7A–7D.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. Global, unbiased, NARDINI+ analysis of nucleolar IDRs, related to Figure 1

(A) Hierarchical clustering of all nucleolar IDRs yielded three main clusters with distinct sequence features.

(B) Fisher's exact test of the frequencies of IDRs with a Z score ≥ 3 for each of the 90 compositional and patterning features compared with the remaining human IDRome in five different condensates. Features are shown if and only if at least one condensate had a p value < 0.01.

(C) Features of nucleolar IDRs that we classify as D/E tracts and/or K-blocks + E-rich regions. D/E-tract IDRs are those that have a neg-neg Z score ≥ 3. K-blocks + E-rich-regions IDRs have a pos-pos and pos-neg Z score ≥ 3, a K-block Z score ≥ 1, an enrichment of Es compared with Ds, and a neg-neg Z score less than 5 times the pos-pos Z score. These definitions lead to the identification of 140 D/E-tracts, 36 K-blocks + E-rich regions, and nine IDRs that have both features. NCL includes the top scoring D/E-tract IDR and POLR1F encompasses the top scoring K-blocks + E-rich-regions IDR.

(D) Proteins with the top 9 high-scoring RG fraction IDRs (Z score > 9) that are annotated by whether they also have a K-block + E-rich region or D/E-tract IDR in their own protein or in their protein complex.

(E) Sequence block profiles of the LYAR, POLR1F, NUCKS, NCL, and UBF IDRs (see STAR Methods). Blue and red bars along the top of the plots indicate K-blocks and D/E-tracts, respectively. The predicted intrinsic disorder of NUCKS (Q9H1E3) spans the entire protein, whereas the predicted IDRs of the remaining proteins span a portion of the full-length proteins. Here, we analyzed LYAR (Q6GP45) 141–286, POLR1F (Q3B726) 206–338, NCL (Q06459) 95–253, and UBF (P17480) 675–764.

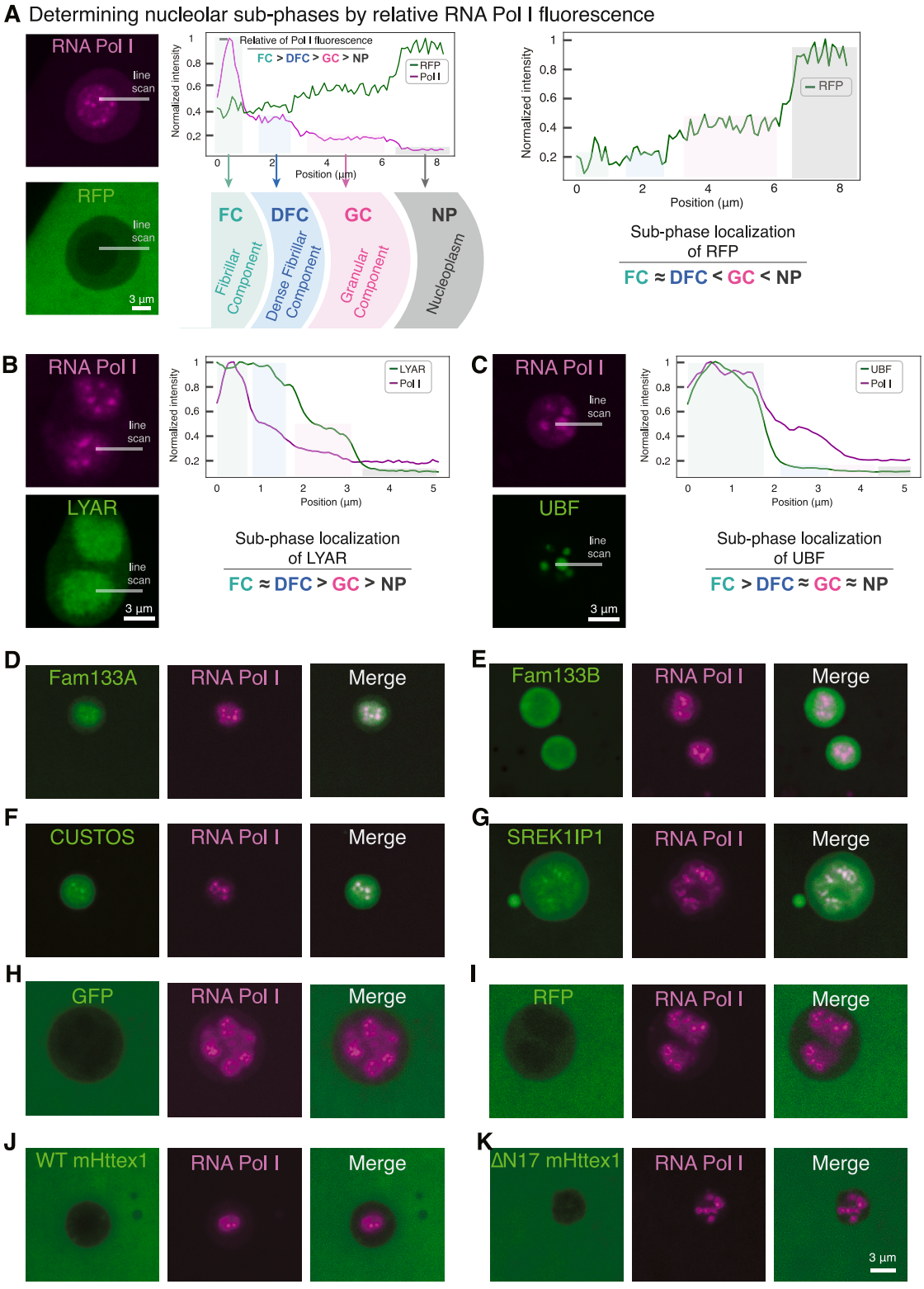


Figure S2. Proteins that localize to the FC and DFC feature K-blocks + E-rich-regions with high pos-pos *Z* scores, related to Figure 2

(A–E) (A) Method for determining nucleolar sub-phase enrichment of target proteins via relative RNA polymerase I fluorescence. On the far left are representative confocal images of a nucleolus in a living *Xenopus laevis* oocyte GV expressing GFP-tagged RNA polymerase I subunit e (PolR1e) and RFP. The resultant line scan begins in a selected FC and ends in the nucleoplasm (NP). The intensity of PolR1e is highest in the FC and drops in a stepwise fashion according to the following

(legend continued on next page)

hierarchy: FC > DFC > GC > NP (intensities were normalized to peak value of 1). Differential thresholding of the hierarchal PolR1e signal was used to demarcate sub-phases, and this was then used to identify the sub-phase that contains the peak signal of the target protein (see STAR Methods for details). Note that the GFP tag on its own has the highest signal in the nucleoplasm. Examples of using this mapping strategy to determine the sub-phase localizations of (B) LYAR and (C) UBF are shown. Representative images of nucleoli in living *Xenopus laevis* oocyte GVs expressing the indicated FP-tagged protein (D) GFP-Fam133A and mCherry-PolR1e, (E) GFP-Fam133B and mCherry-PolR1e, (F) GFP-COSTUS and mCherry-PolR1e, (G) GFP-SREK1IP1 and mCherry-PolR1e, (H) GFP and mCherry-PolR1e, (I) RFP and GFP-PolR1e, (J) RFP-WT_mHttex1 and GFP-PolR1e, and (K) RFP-ΔN17_mHttex1 and GFP-PolR1e. Scale bars for all images, 3 μm.

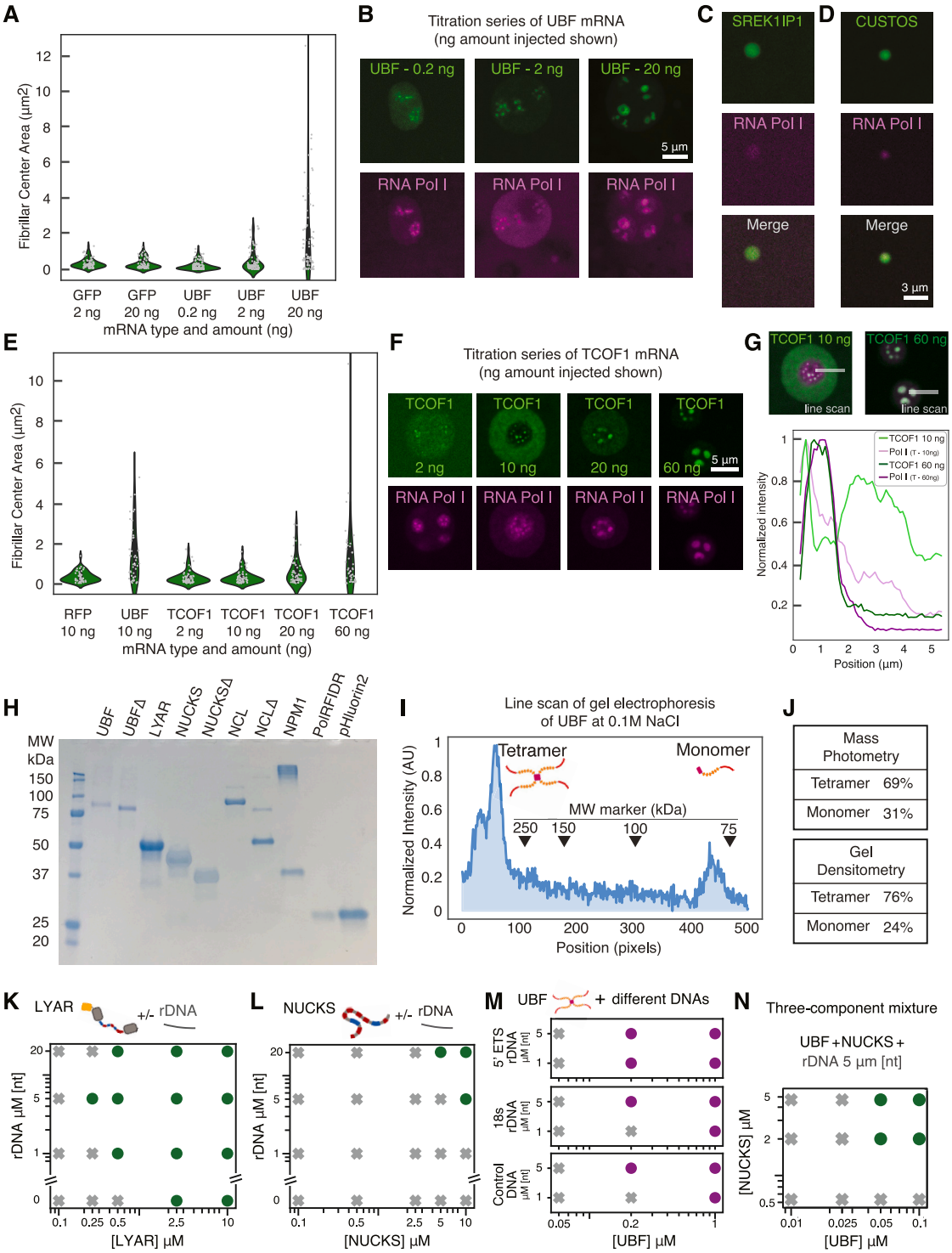


Figure S3. Characterizations of UBF, LYAR, and NUCKS and concentration dependent condensation of these factors *in vitro* and in cells, related to Figure 3

(A) Violin plot of fibrillar center sizes (area, μm^2) for indicated mRNA injection loads of GFP alone or GFP-UBF.

(B) Representative images of nucleoli in living *Xenopus laevis* oocyte GVs expressing GFP-UBF injected at indicated amounts of mRNA and mCherry tagged RNA polymerase I subunit e (PolR1e) (2 ng mRNA).

(legend continued on next page)

(C and D) (C) Representative images of non-nucleolar puncta observed in living *Xenopus laevis* oocytes expressing GFP-SREK1IP1 and mCherry-PolR1e or (D) GFP-CUSTOS and mCherry-PolR1e. GFP-SREK1IP1 and GFP-CUSTOS mRNAs were injected at 20 ng. In all conditions, mCherry-PolR1e mRNA was injected at 2 ng. Scale bars, 3 μm.

(E) Violin plot of FC sizes (area, μm^2) for indicated mRNA injection loads of RFP alone, UBF-RFP, or TCOF1-RFP.

(F) Representative images of nucleoli in living *Xenopus laevis* oocyte GVs expressing TCOF1-RFP injected at indicated amounts of mRNA and GFP-PolR1e (2 ng mRNA).

(G) Line scans of TCOF1-RFP images shown expressing 10 or 60 ng mRNA. Light green and pink curves correspond, respectively, to TCOF1 and PolR1e in the 10 ng TCOF1 mRNA condition. Likewise, dark green and dark magenta curves correspond, respectively, to TCOF1 and PolR1e in the 60 ng TCOF1 mRNA condition.

(H) Coomassie-stained SDS-PAGE of purified recombinant proteins used in this study.

(I) Densitometry profile of the gel lane containing 100 mM NaCl solution UBF showing protein staining intensity in arbitrary units (AUs) normalized to maximum intensity along the gel (position given in pixel units). Arrowheads indicate the position of molecular weight (MW) markers; the expected MW of UBF monomers is 90 kDa, and the expected MW of UBF tetramers is 360 kDa.

(J–L) (J) Summary of area-under-the-curve calculations for UBF species measured using mass photometry and gel electrophoresis with densitometry. Phase boundaries of (K) LYAR, and (L) NUCKS as one- and two-component mixtures without or with rDNA at three concentrations, 1, 5, and 20 μM [nt].

(M) Phase boundaries of three-component mixtures of UBF and LYAR mixed with rDNA at 1 μM [nt] or 5 μM [nt]. Phase boundaries of two-component mixtures of DNA and UBF. UBF is mixed with the 5′-ETS region of the human rDNA gene (1,622 bp), the region of the human rDNA gene corresponding to the final mature 18S rRNA fragment (1,844 bp), or a control DNA fragment corresponding to the Cln3 gene from *Ashbya gossypii* (1,612 bp). UBF is included at the indicated condensations and DNAs are at 1 or 5 μM [nt].

(N) Phase boundaries of three-component mixture of UBF, NUCKS, and rDNA at 5 μM [nt].

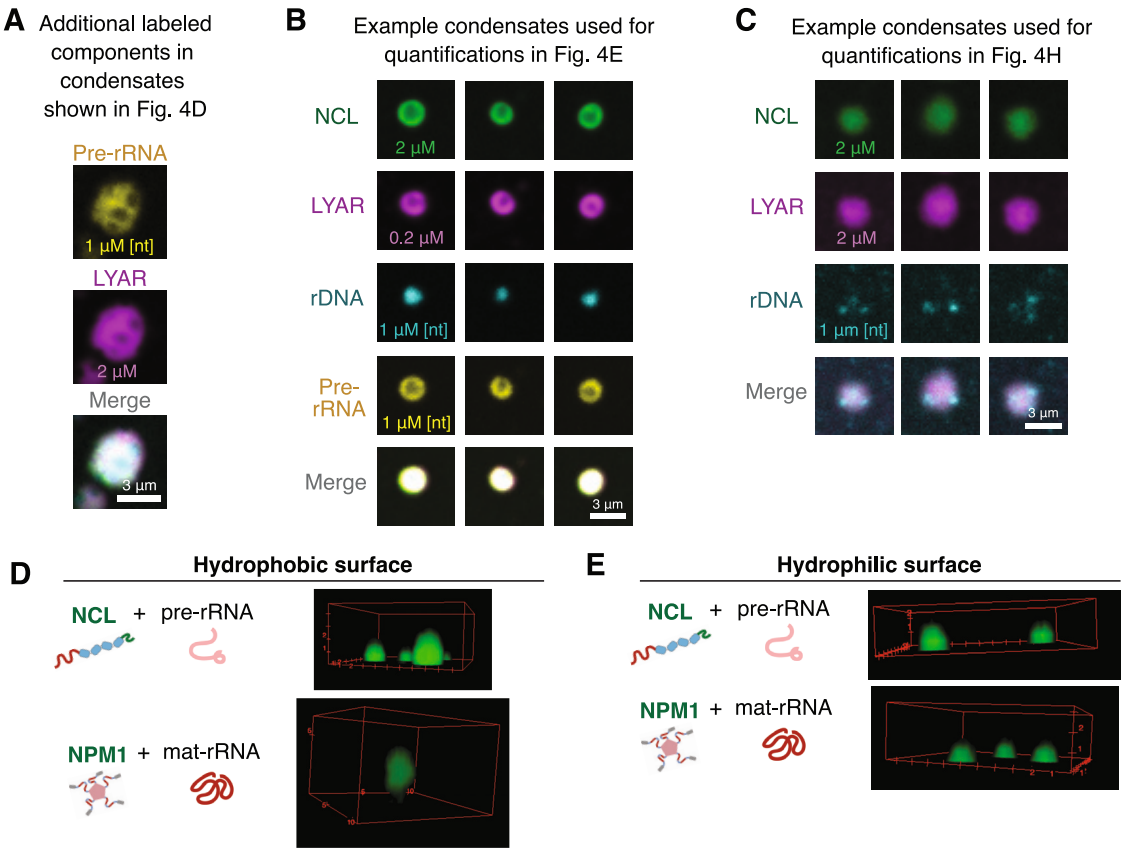


Figure S4. Features of *in vitro* facsimiles of FC, DFC, and GC phases, related to Figure 4

(A) Confocal images of the pre-rRNA and LYAR components of the five-component mixtures displayed in Figure 4C. These mixtures also contain NCL (2 μM), UBF (0.2 μM), and rDNA (1 μM [nt]).

(B) Confocal images of example five-component FC/DFC condensates used to quantify spatial distribution shown in Figure 4D; [UBF] = 0.2 μM, other concentrations are indicated in the figure.

(C) Confocal images of example four-component FC/DFC condensates without pre-rRNA used to quantify spatial distribution shown in Figure 4F; [UBF] = 0.2 μM, other concentrations are indicated in the figure.

(D and E) Rendered 3D volumes of confocal images of condensates from mixtures of NCL and pre-rRNA (top row) or NPM1 and mat-rRNA condensates resting on coverslips with (D) hydrophobic or (E) hydrophilic surface treatments. All scale bars, 3 μm.

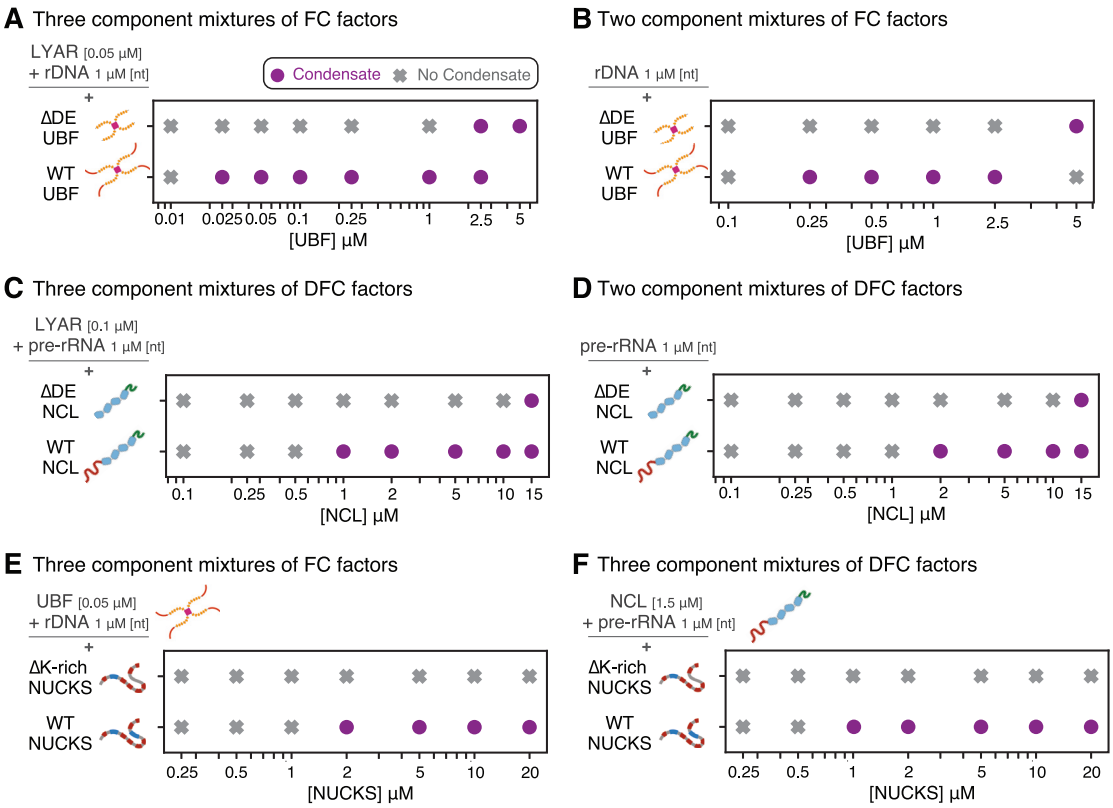


Figure S5. Both D/E-tracts and K-blocks contribute to complex coacervation of nucleolar components, related to Figure 5

(A) Phase boundary of UBF or ΔD/E-tract-UBF in three-component mixtures with LYAR (0.05 μM) and rDNA (1 μM [nt]).
(B) Phase boundary of UBF or ΔD/E-tract-UBF in two-component mixtures with rDNA (1 μM [nt]).
(C) Phase boundary of WT NCL or ΔD/E-tract-NCL in three-component mixtures with LYAR (0.1 μM) and pre-rRNA (1 μM [nt]).
(D) Phase boundary of WT NCL or ΔD/E-tract NCL in two-component mixtures with pre-rRNA (1 μM [nt]).
(E) Phase boundary of WT NUCKS or ΔK-rich NUCKS in three-component mixtures with WT UBF (0.05 μM) and rDNA (1 μM [nt]).
(F) Phase boundary of NUCKS or ΔK-rich NUCKS in three-component mixtures with WT NCL (1.5 μM) and pre-rRNA (1 μM [nt]).

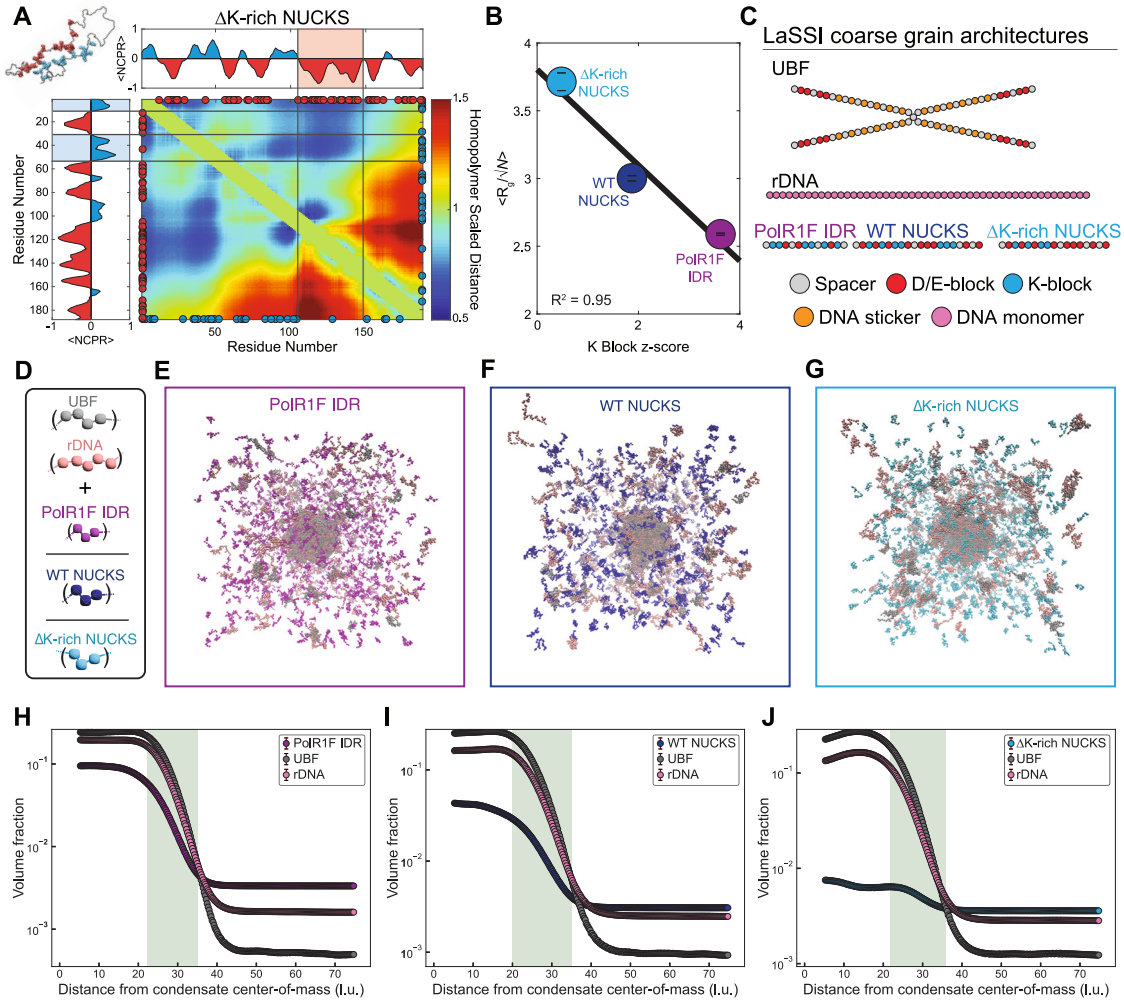


Figure S6. The effect of K-blocks on intra-IDR compaction and partitioning in simulations, related to Figure 6

(A) Inter-residue distance maps normalized by the best-fit homopolymer model for ΔK-rich NUCKS from atomistic simulations. Positive and negative residues are shown as circles in blue and red, respectively.

(B) Comparison of K-block *Z* score and sequence length (N) normalized R_g (R^2 = 0.95 for linear regression). Error bars indicate standard errors from the mean across five independent simulations.

(C) Coarse-grained architectures for each molecule simulated using LaSSI. Here, each circle denotes a coarse-grain bead which represents many residues/nucleotides.

(D–F) Representative snapshots of the full simulation box of simulated FCs involving UBF with rDNA and either PolR1F IDR (D), WT NUCKS (E), or ΔK-rich NUCKS (F).

(G–I) Radial density plots for UBF and rDNA with either PolR1F IDR (G), WT NUCKS (H), or ΔK-rich NUCKS (I). Here, the light green rectangle denotes the condensate interfacial region, and l.u. denotes lattice units.

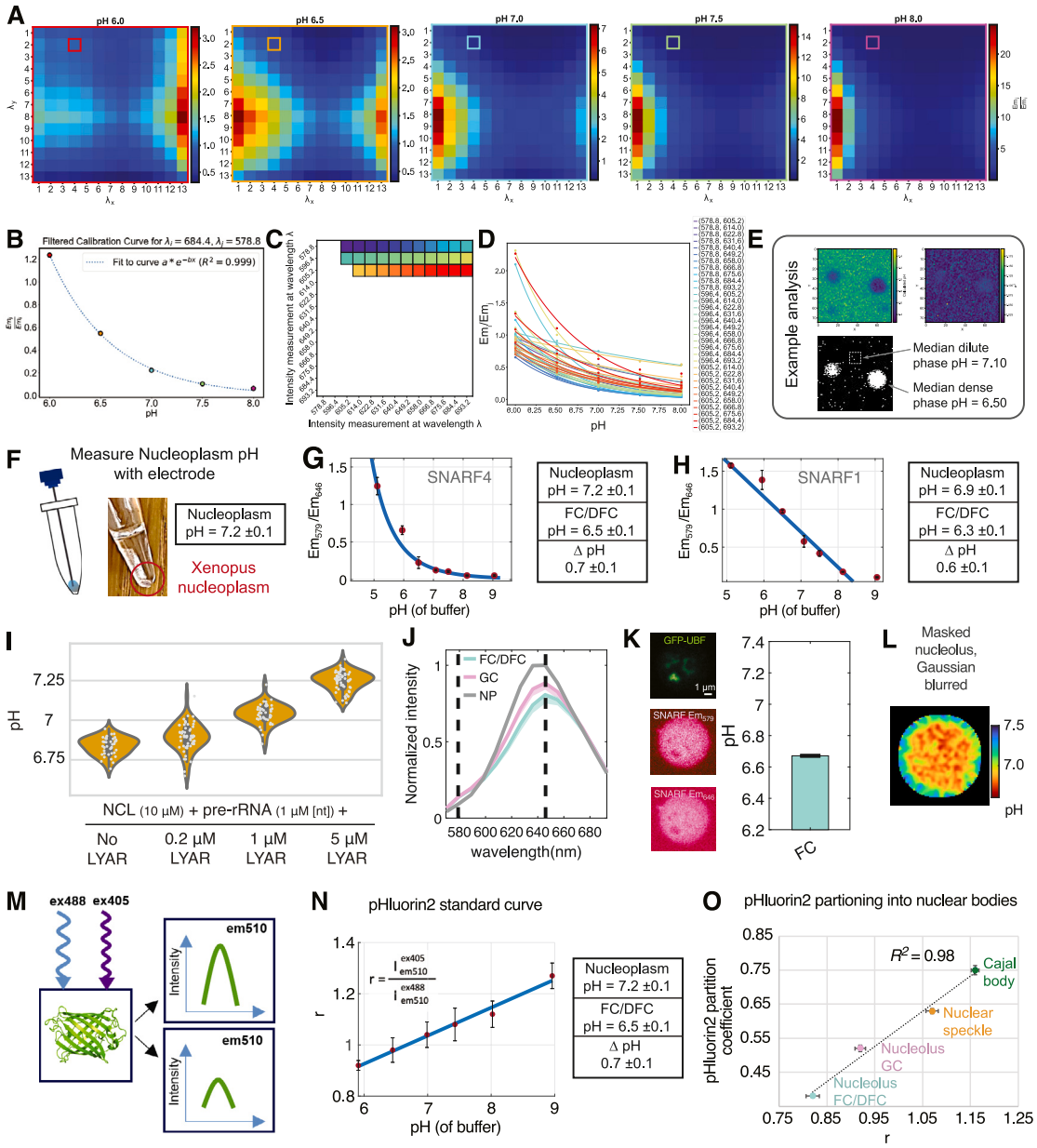


Figure S7. Multiple independent measurements suggest nucleolus is ~0.7 pH units more acidic than nucleoplasm, related to Figure 7

(A) Similarity matrices of the ratios of intensity measurements $r_{ij} = Em_j/Em_i$ obtained at different wavelengths for confocal wavelength scans of five different buffer solutions with SNARF-4 dye (pH 6 to 8 at 0.5 pH increments). Calibration curves were derived from all the similarity matrices by selecting the same intensity measurement ratio for a given pair of wavelengths across all pH values. Shown is an example highlighted cell, corresponding to the ratio of intensity measurements at wavelengths λ_x = 614.0 nm and λ_y = 596.4 nm.

(B) Each calibration curve is fit to an exponential function of the form: $ae^{\mp bx}$ and an R^2 value is calculated. Curves that are within the upper triangle have a negative exponent, and curves in the lower triangle have a positive exponent. Only curves with an R^2 of 0.97 and above were selected. Additionally, curves whose slopes were less than 0.5 are excluded. Shown is a filtered calibration curve for intensity measurements at λ_x = 684.4 nm and λ_y = 578.8 nm.

(C) Shown are the 32 cells corresponding to the intensity measurement ratios whose calibration curves have been filtered as described in (B) (see STAR Methods). This area corresponds to high-quality calibration curves for intensity measurements at larger wavelengths against shorter wavelengths.

(D) A plot of the 32 filtered calibration curves overlain on the same graph. This plot illustrates that although the filtered curves have similar mathematical profiles, the ranges of ratios are variable. We perform min-max normalization of these curves for improved accuracy in calculating the pH from a measurement.

(E) Calculated pH heatmaps for condensate images using the filtered calibration curves on the image shown in Figure 7B (mixture of NPM1 and mat-rRNA). The upper left panel shows a heatmap of the median pH derived from the corresponding intensity measurement ratios of the 32 filtered calibration curves. The upper right panel shows the variance of the calculated pH for each pixel. The lower left panel shows the thresholding mask used to call condensates from surrounding

(legend continued on next page)

dilute phase. Condensates and dilute phase patches are indicated, and their median pH values shown. Only thresholded objects with an area >0.05 μm^2 were considered condensates (see STAR Methods for more details).

(F) Schematic and image of microprobe-based measurement of harvested, undiluted GV nucleoplasm; electrode pH measurement shown (error reported here and in (G), (H), and (N) denotes the maximum daily change in pH of buffer standards, see STAR Methods for details).

(G) Standard curve for SNARF-4 dye generated using a series of buffers at specified pH values (ratiometric intensity of Em_{579}/Em_{646} used). Computed pH of nucleoplasm and nucleolar FC/DFC phases using the SNARF4 standard curve show FC/DFC is ~0.7 units more acidic.

(H) Standard curve for SNARF-1 dye generated using a series of buffers at specified pH values (ratiometric intensity of Em_{579}/Em_{646} used). Computed pH of nucleoplasm and nucleolar FC/DFC phases using the SNARF1 standard curve show FC/DFC is ~0.6 units more acidic.

(I) Violin plot showing pH of condensates in mixtures of NCL and pre-rRNA with or without LYAR at indicated concentrations.

(J) Wavelength scan of intensities (normalized to NP intensity at Em_{646}) of freely diffusing SNARF-4 dye living *Xenopus laevis* oocyte GVs expressing GFP-NPM1 to differentiate nucleoplasm, nucleoli GC, and nucleoli FC/DFC. Solid lines are median values; shaded regions are 95% CI; dashed vertical lines indicate wavelengths used for pH calculation using the standard curve (n = 5 nucleoli).

(K) Confocal images of nucleolus in living *Xenopus* oocyte GV expressing GFP-UBF to demarcate FCs; computed median pH across 3 nucleoli also shown. The error bar is the standard deviation of measurement.

(L) Gaussian-blurred image of the nucleolus shown in Figure 7E. A mask corresponding to the full-width half-maximum intensity of GFP-NPM1 signal surrounding the nucleolus was applied before blurring to display the spatial distributions of pH in the nucleolus specifically.

(M) Schematic of pHluorin2 excitation and emission: pHluorin2-containing samples were excited in series by a 405 nm laser, then a 488 nm laser at matched parameters, and emissions were obtained using the same bandpass filter centered at 510 nm. Acidic samples exhibit higher emission intensities when excited with 405 nm laser, and vice versa for basic samples excited using a 488 nm laser.

(N) Standard Curve of pHluorin2 generated using the imaging scheme in (M) on purified pHluorin2 in a series of buffers at specified pH values. Computed pH of nucleoplasm and nucleolar FC/DFC phases using pHluorin2 standard curve shows FC/DFC is ~0.7 units more acidic.

(O) Correlation of pHluorin2 r value and the extent of partitioning (PC = partition coefficient) of pHluorin2 into nucleolar sub-phases and nuclear bodies (shown).