

Evidence for widespread cytoplasmic structuring into mesoscale condensates

Received: 31 January 2022

Accepted: 23 January 2024

Published online: 29 February 2024

 Check for updates

Felix C. Keber^{1,2,3}, Thao Nguyen^{2,3}, Andrea Mariossi^{1,3},
Clifford P. Brangwynne^{2,4,5} & Martin Wühr^{1,2,3}

Compartmentalization is an essential feature of eukaryotic life and is achieved both via membrane-bound organelles, such as mitochondria, and membrane-less biomolecular condensates, such as the nucleolus. Known biomolecular condensates typically exhibit liquid-like properties and are visualized by microscopy on the scale of $\sim 1 \mu\text{m}$ (refs. 1,2). They have been studied mostly by microscopy, examining select individual proteins. So far, several dozen biomolecular condensates are known, serving a multitude of functions, for example, in the regulation of transcription³, RNA processing⁴ or signalling^{5,6}, and their malfunction can cause diseases^{7,8}. However, it remains unclear to what extent biomolecular condensates are utilized in cellular organization and at what length scale they typically form. Here we examine native cytoplasm from *Xenopus* egg extract on a global scale with quantitative proteomics, filtration, size exclusion and dilution experiments. These assays reveal that at least 18% of the proteome is organized into mesoscale biomolecular condensates at the scale of $\sim 100 \text{ nm}$ and appear to be stabilized by RNA or gelation. We confirmed mesoscale sizes via imaging below the diffraction limit by investigating protein permeation into porous substrates with defined pore sizes. Our results show that eukaryotic cytoplasm organizes extensively via biomolecular condensates, but at surprisingly short length scales.

Protein components of biomolecular condensates (BMCs) have typically been identified via imaging, co-isolation or proximity labelling^{9–14}. However, these approaches require prior knowledge of at least one constituent of the assembly. Moreover, imaging approaches favour the detection of large ($\sim 1 \mu\text{m}$) assemblies due to the diffraction limit of light microscopy, and are often facilitated using overexpression of labelled proteins, with potential impacts on native condensate structure. So far, these approaches have identified ~ 100 ‘scaffold’ proteins suggested to drive liquid–liquid phase separation (LLPS)¹⁵, which is $\sim 0.5\%$ of the human proteome. However, on the basis of sequence similarity to these proteins, it has been speculated that as much as 20% of the proteome is functionally involved in LLPS¹⁶. In this Letter, we developed methods to

measure what part of the proteome is organized in BMCs and at what length scale these typically form.

To assay the physical properties of protein assemblies throughout the native cytoplasm, we sought to combine filtration experiments of undiluted cytoplasm with quantitative proteomics. We reasoned that we could identify BMCs based on their behaviour upon filtration. When encountering a pore, assemblies smaller than the pore diameter should pass freely. However, the permeation of assemblies larger than the pore will depend on their material properties. While large rigid bodies will not pass, deformable assemblies can squeeze through pores. Thus, size and viscoelastic properties will determine the passage time, establishing a chromatography-like process (Fig. 1a). At the earliest time, the

¹Department of Molecular Biology, Princeton University, Princeton, NJ, USA. ²Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, USA. ³Lewis–Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. ⁴Howard Hughes Medical Institute, Princeton University, Princeton, NJ, USA. ⁵Omnenn–Darling Bioengineering Institute, Princeton University, Princeton, NJ, USA.

✉ e-mail: cbrangwy@princeton.edu; wuhr@princeton.edu

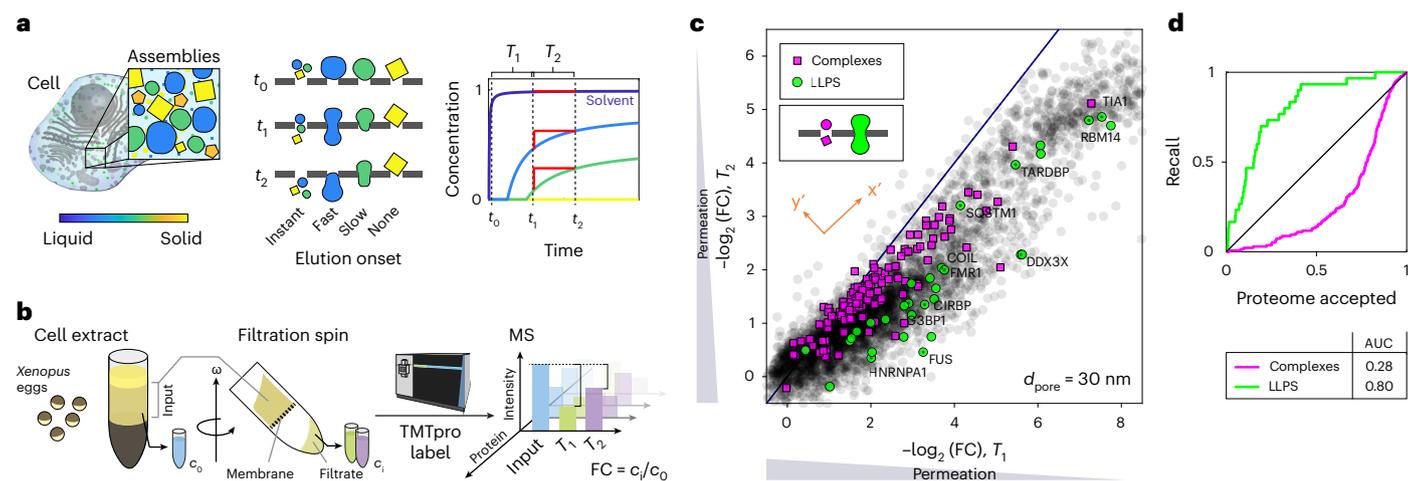


Fig. 1 | Proteomics of filtrated cytoplasm reveals organization in liquid-like assemblies. **a**, An illustration of the filtration process. For larger assemblies, squeezing through pores enables passage, where more liquid assemblies pass faster. Right: sampling protein concentrations throughout the process (red lines) informs about the different elution behaviours. **b**, Experimental outline. We spun undiluted cytoplasm from frog eggs through filters and analysed the filtrate in comparison to the input by mass spectrometry (MS). For each protein, we obtain the concentration fold change (FC) between input and filtrate (FC of c/c_0) via quantitative proteomics. T_i denotes the time interval of the sampling (t_{i-1} , t_i). FCs were normalized separately for each experiment to their 0.95 quantile, representing freely passing proteins. **c**, A scatter plot of protein permeation later (T_2) versus earlier (T_1) in the process. Close to the identity line (blue), the permeation is almost unchanged, corresponding to the

flat part of the curves in **a**. Below the identity line permeation is increased in T_2 , as expected for a steep curve section in **a**. Proteins established to form LLPS (green) or large complexes (magenta) exhibit distinguishable behaviours, matching the expectation (inset schematic). $N=1$ biological sample. **d**, We rank order the proteome along an axis orthogonal to the upper edge of the data close to the identity line (y' in the orange inset in **c**). The ROC plots the true positive rate (recall) against the false positive rate (proteome accepted). A perfect classifier curve would pass through the top-left corner, whereas a diagonal line indicates randomness. LLPS proteins are recalled against y' preferentially, while complexes are underrepresented. This distinction is quantified by the AUC. Source numerical and proteomics data are provided in Source data and Supplementary Table 3. $N=1$ biological sample.

filtrate should contain only solvent and freely passing proteins, since assemblies that must squeeze through pores spend more time in the porous medium and their elution onset is later; such behaviour can be described in a simplistic model (Supplementary Figs. 1–3). Thus, by comparing filtrates at different times, we sought to identify assemblies that exhibit squeezing behaviour. We performed microscopy experiments as proof of principle (Supplementary Figs. 4 and 5).

To study the organization of near-native cytoplasm, we chose lysate from eggs of the frog *Xenopus laevis*, which provides easy access to large amounts of undiluted cytoplasm in a near-native state. The eggs are naturally arrested in metaphase; thus, the nuclear proteins are in the cytoplasm. *X. laevis* extracts have been a powerful model for studying biochemistry in the cytoplasm, which allows observing protein interactions in a close-to-native environment, including complex processes such as the formation of nuclei or spindles^{17,18}. Eggs are crushed in an extract preparation spin. We verified that this spin did not sediment proteins known to be involved in BMCs (Extended Data Fig. 1). Cytoplasmic extract was then centrifuged against polyethersulfone filter membranes with a defined particle size cut-off diameter (d_{pore}). We analysed the filtrates of various experimental conditions by multiplexed proteomics, quantifying each protein's concentration relative to the input^{19–21}, where proteins with the earliest onset served for normalization (Fig. 1b).

We compared the permeation of all proteins at a fixed pore size ($d_{\text{pore}} = 30$ nm) at exemplary stages in the elution process (t_1 and t_2). At both timepoints, the protein concentrations relative to the input display a broad spectrum, ranging from free passage to heavy retention (Fig. 1c). The data display a sharp top edge slightly below the identity line that originates from the systematically higher elution at t_2 . We observe two major regimes: (1) a large part of the proteome shows approximately the same concentration at t_1 and t_2 (following close to the identity line). This includes proteins that are organized smaller than the pores (close to the origin). In our passage-time model,

these protein species' elution has started long before t_1 , thus t_1 and t_2 both sample the flat part of the curve (close to the dark-blue solvent curve, Fig. 1a). (2) For many other proteins, elution is increased at t_2 (notably below the identity line). This corresponds to sampling the steep part of the curves in our model (blue or green curves, Fig. 1a), indicating a longer passage time. We find that annotated canonical large protein complexes, which we do not expect to exhibit 'squeezing' behaviour, exhibit a preponderance along the top edge of the data, closer to the identity line (magenta symbols, Fig. 1c); the deviation from the identity line probably reflects the influence of their size alone. By contrast, we find that proteins known to be associated with LLPS are enriched farther from the identity line (green symbols, Fig. 1c and Supplementary Table 1)^{22–26}, consistent with their presence within liquid-like assemblies that slowly squeeze through pores. Executing the experiment with a larger pore size ($d_{\text{pore}} = 100$ nm) yields qualitatively similar results (Extended Data Fig. 2a). To characterize the squeezing behaviour, we measure each protein's distance from the top edge close to the identity line. This 'squeezing score' displays pronounced differences between known LLPS proteins and the whole proteome, as reflected in the receiver operating characteristic (ROC; Fig. 1d and Extended Data Fig. 2b for $d_{\text{pore}} = 100$ nm). These findings suggest that filtration chromatography probes the physical properties of the cytoplasm on the mesoscale and could be used to identify novel BMCs and their components.

A characteristic property of phase-separated BMCs is that they typically form via dynamic, multivalent interactions, and disassemble below a particular saturation concentration^{1,2,27,28}. To test whether altering concentration impacts the apparent mesoscale cytoplasmic organization, we diluted cell extracts to various extents and examined their filtration behaviour at a fixed pore size ($d_{\text{pore}} = 100$ nm) (Fig. 2a). Remarkably, at a dilution of only 1.4-fold, known LLPS proteins show vastly different permeation behaviour compared with canonical larger complexes or the entire proteome (Fig. 2b,c). LLPS proteins exhibit

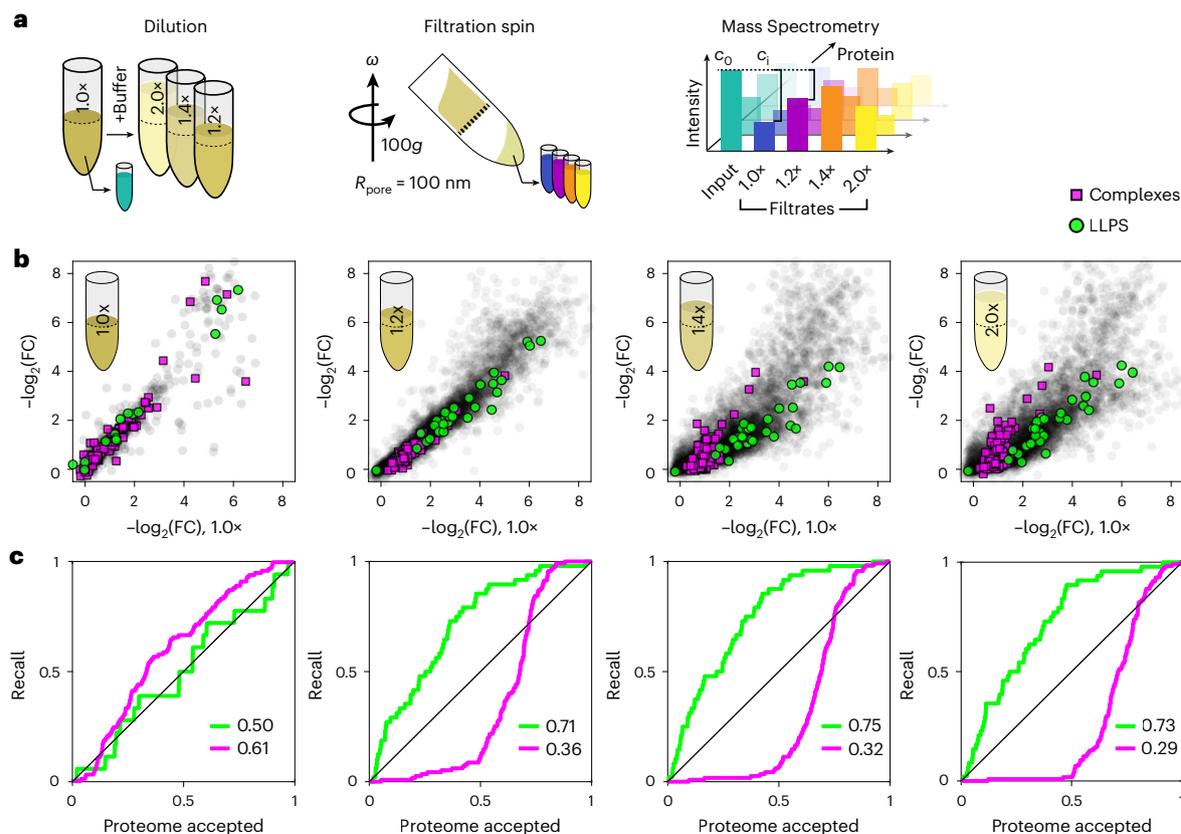


Fig. 2 | Liquid-like assemblies are sensitive to moderate dilution. **a**, We performed the cytoplasmic filtration experiment with 100-nm pores and compared protein retention with lysate diluted by various amounts. The concentrations of the filtrates relative to the input are measured by MS. **b**, An experiment of filtration of the diluted lysate reveals that LLPS proteins (green) seem to disassemble at dilutions larger than 1.2 \times . FCs were normalized

separately for each experiment to their 0.95 quantile, representing freely passing proteins. $N = 1$ biological sample. **c**, The ROCs of the scatter plots in **b** along an axis orthogonal to the upper edge of the data, with the AUCs displayed at the legends. Source numerical and proteomics data are provided in Source data and Supplementary Table 3.

a progressive increase in permeation with dilution, which probably originates from shrinking BMC sizes as concentrations approach and cross below different saturation thresholds. Interestingly, however, even at high dilution factors, we do not observe unhindered permeation, as may be expected upon full dissolution of assemblies. Similarly, a complementary assay applying hard spins to cytoplasm detected sedimentation of LLPS proteins as expected for large assemblies; however, dilution did not abolish this sedimentation (Extended Data Fig. 3). These data suggest that, while these mesoscale condensates can be partially dissolved upon dilution, they also exhibit partially solid-like characteristics that may indicate stable cores, potentially formed by specific protein–protein interactions, gelation or binding to RNA molecules^{29,30}.

RNAs are long polymers that are usually coated in RNA-binding proteins (RBPs). Together with their protein-binding partners, RNAs often help drive formation condensates, which can thereby impact associated functions, such as messenger RNA (mRNA) translation efficiency and stability. Many of the proteins that exhibit differential filtration behaviour in our assay are also RBPs. Indeed, RBPs were preferentially retained and squeezable, especially if they were annotated to be involved in LLPS (Extended Data Fig. 4a,b). We thus sought to further examine the role of RNA in this emergent cytoplasmic organization. When we repeated the filtration experiment after treating the extract with RNase, the elution of RBPs was increased, with the effect strongest for LLPS-associated proteins (Extended Data Fig. 4c). These findings suggest that RNA–RBP interactions play a major role in the time-dependent squeezing behaviour and

are consistent with the widespread presence of RNA in many or all known condensates.

To further examine which RNAs contribute to squeezability, we performed RNA sequencing (RNA-seq) of filtrates at different elution times. We detect RNA exhibiting comparable behaviour to what we observed on the protein level (Fig. 1c and Extended Data Fig. 5a). Remarkably, more than 15% of all RNAs showed behaviour consistent with organization via BMCs. Intriguingly, these RNAs were enriched for developmental regulations, while RNAs of housekeeping genes³¹ rarely showed high retention (Extended Data Fig. 5b–d). This suggests that mRNAs that undergo transient translation, such as germ granules^{32,33}, might be particularly prone to organization via BMCs.

Cellular BMCs so far have been reported mainly on the micrometre length scale. However, there is growing evidence for much smaller structures^{34,35}. To investigate the length scale of the liquid-like organization in our system, we compared the elution behaviour at different pore diameters (30, 100 and 200 nm). At the largest pore size of 200 nm, we observe an overall higher permeation with very few proteins retained (Fig. 3a). Notable exceptions include retention of mitochondrial proteins, which is not surprising (Fig. 3a). The strikingly higher retention as the pore size is decreased to 100 nm suggests this is a characteristic length scale of liquid-like organization in the cytoplasm. This is further supported by filtration experiments in a complementary setup, using polymer mesh filters and gravity flow to avoid force-induced squeezing (Extended Data Fig. 6). These data suggest the widespread presence of liquid-like cytoplasmic assemblies on the submicrometre scale, involving a broad swath of the proteome.

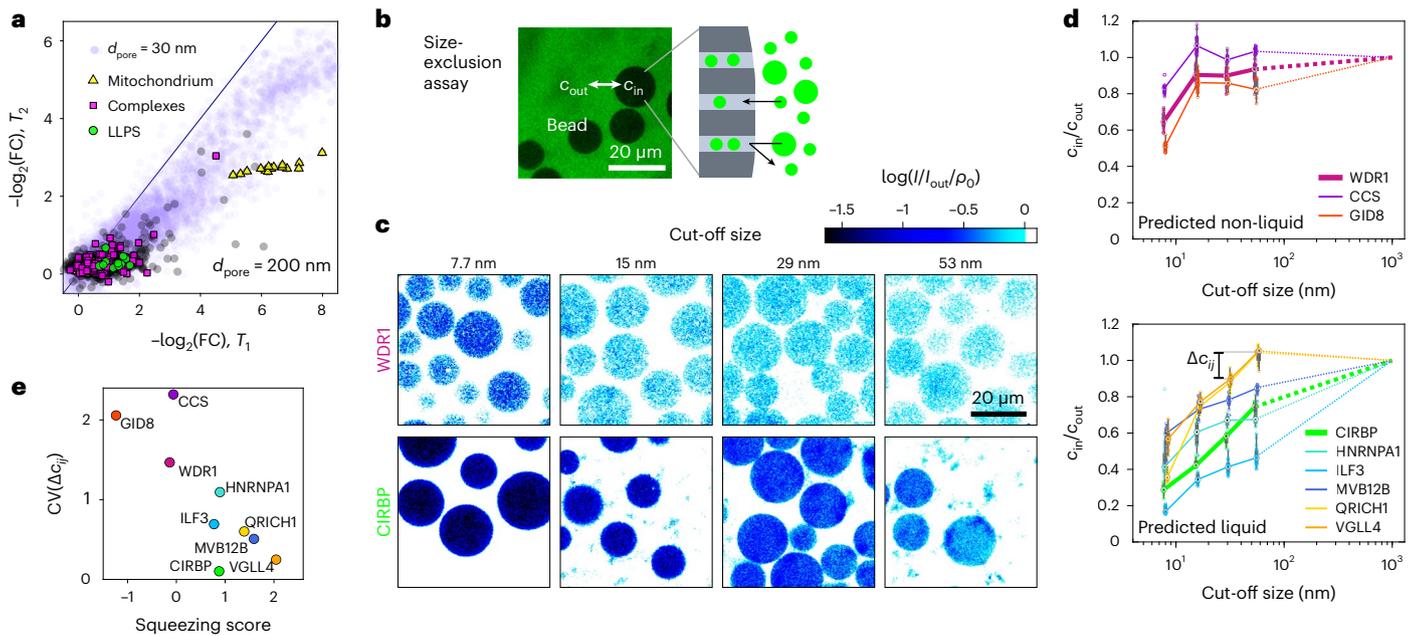


Fig. 3 | Proteomics and imaging assays indicate that BMCs predominantly organize at the ~100 nm length scale. **a**, The filtration experiment with 200 nm pore size indicates that few proteins besides mitochondrial ones (yellow) are retained (30-nm data for comparison (lilac)). $N = 1$ biological sample. **b**, Microscopy of GFP-labelled proteins with size-exclusion beads to assess the protein assembly sizes. The concentration inside a bead's polymer matrix c_{in} compared with outside solution c_{out} provides information about the protein's assembly size. The example image shows beads with 400 kDa size cut-off in an extract expressing CCS-GFP. $N = 1$ biological sample. **c**, Examples of the size-exclusion assay for the proteins WDR1-GFP and CIRBP-GFP (examples of proteins in small assemblies and LLPS, respectively) with four different bead cut-off sizes. Confocal image intensities I are normalized to their respective outside solution intensity I_{out} and each column is scaled by the accessible volume for monomers ρ_0 of the used bead. Thus, the white bead colour stands for $c_{in}/c_{out} \approx 1$ and darker blue tones report higher exclusions. **d**, Size-exclusion plots. Top: proteins devoid of squeezing behaviour show a characteristic

jump at short length scale. Bottom: in contrast, proteins filtrations suggested to be in BMCs (ILF3, MVB12B, QRICH1 and VGLL4) including well-established LLPS (CIRBP and HNRNPA1) show organization throughout the mesoscale. The dotted lines extrapolate data (c_{in}/c_{out}) to 1 at 1 μm based on a lack of visible organization via microscopy. The box plots denote the median (circle), 0.25/0.75 quantiles (box) and 0.05/0.95 quantiles (whiskers). $N = 3$ biologically independent samples examined over n_1, n_2, n_3 and n_4 measurements: WDR1: (21, 21, 21 and 19); GID8: (20, 30, 9 and 18); CCS: (20, 21, 11 and 23); CIRBP: (8, 7, 7 and 7); HNRNPA1: (19, 9, 7 and 8); ILF3: (15, 11, 17 and 20); MVB12B: (16, 10, 20 and 16) and VGLL4: (15, 13, 13 and 4). **e**, The coefficient of variation (CV) of subsequent steps Δc_{ij} of the curves in **d** serves as a metric for the organization patterns. Proteins with a high squeezing score (Fig. 1d) that are predicted to be in liquid-like assemblies exhibit a low CV, suggesting an assembly size distribution spanning the mesoscale. Source numerical and proteomics data are provided in Source data and Supplementary Table 3.

To further validate our findings of liquid-like organization on surprisingly short length scales, we employed an experimentally orthogonal light microscopy assay that investigates the size of cytoplasmic assemblies below the diffraction limit. We selected proteins exhibiting different filtration behaviour and expressed green fluorescent protein (GFP)-tagged versions by doping the cytoplasm with the corresponding mRNA. Interestingly, at our low expression levels of only a few tens of nanometres, we can detect fluorescence, but none of the investigated proteins—including established LLPS proteins such as HNRNPA1 (ref. 36) or CIRBP³⁷—show any signs of assemblies on the micron scale (Extended Data Fig. 7a). However, diffusion constant measurements by fluorescence correlation spectroscopy suggested organization into small assemblies (Extended Data Fig. 7b). To interrogate potential assemblies on smaller length scales, we equilibrated the lysates with size-exclusion beads (Fig. 3b). The size exclusion of biomolecular assemblies by a bead's polymer matrix reduces the concentration inside the bead c_{in} compared with the outside c_{out} . The concentration ratio c_{in}/c_{out} at different cut-off sizes (~7.7, 15, 29 and 53 nm) serves as a proxy for the cumulative assembly size distribution. We measure c_{in} and c_{out} by their GFP intensity and correct c_{in} for the excluded volume in the beads, which we determined with a dextran–rhodamine solution (hydrodynamic radius of ~5 nm) (Fig. 3c and Extended Data Fig. 8). Proteins that eluted early in the filtration assay (WDR1, GID8 and CCS) mainly organized on the 10 nm scale. By contrast, proteins that we found in the squeezing regime (MVB12B, VGLL4, QRICH1 and ILF3) exhibit c_{in}/c_{out}

increasing with the cut-off size, similar to established LLPS proteins (CIRBP and HNRNPA1) (Fig. 3d,e). This behaviour suggests that the size of these assemblies is not sharply defined but spans across the sampled scale, matching the expectation for phase-separated droplets, which appear to assemble typically on the scale of ~100 nm.

We next sought to integrate our multiple proteomics experiments to improve predictions on which proteins are associated with phase-separated condensates. Existing predictors for proteins driving LLPS analyse published data to extract sequence features but suffer from a lack of comprehensive reference data. To this end, we trained a classifier, which learns to identify LLPS proteins by bagging an ensemble of linear discriminators and decision trees (Fig. 4a)^{38–40}. We used the results from our filtration chromatography and dilution experiments as features. We additionally included coarsened sequence information on intrinsic disorder⁴¹, nucleic acid binding^{42–44} and amino acid (a.a.) composition⁴³. The performance of our predictor is assessed by the recall of known LLPS proteins with fivefold cross-validation. While prediction using experiments alone already has an area under curve (AUC) of 0.86, our final predictor, including all features, reaches an AUC of 0.93 (Supplementary Table 2). This approach establishes an improvement over the state-of-the-art predictions of LLPS proteins by catGRANULE⁴⁵ (AUC of 0.84), Pscore⁴⁶ (AUC of 0.85) or phase separation analysis and prediction (PSAP)²³ (AUC of 0.88). Notably, we gain sensitivity by a steeper early increase, which is arguably the most relevant regime for the prediction (Fig. 4b).

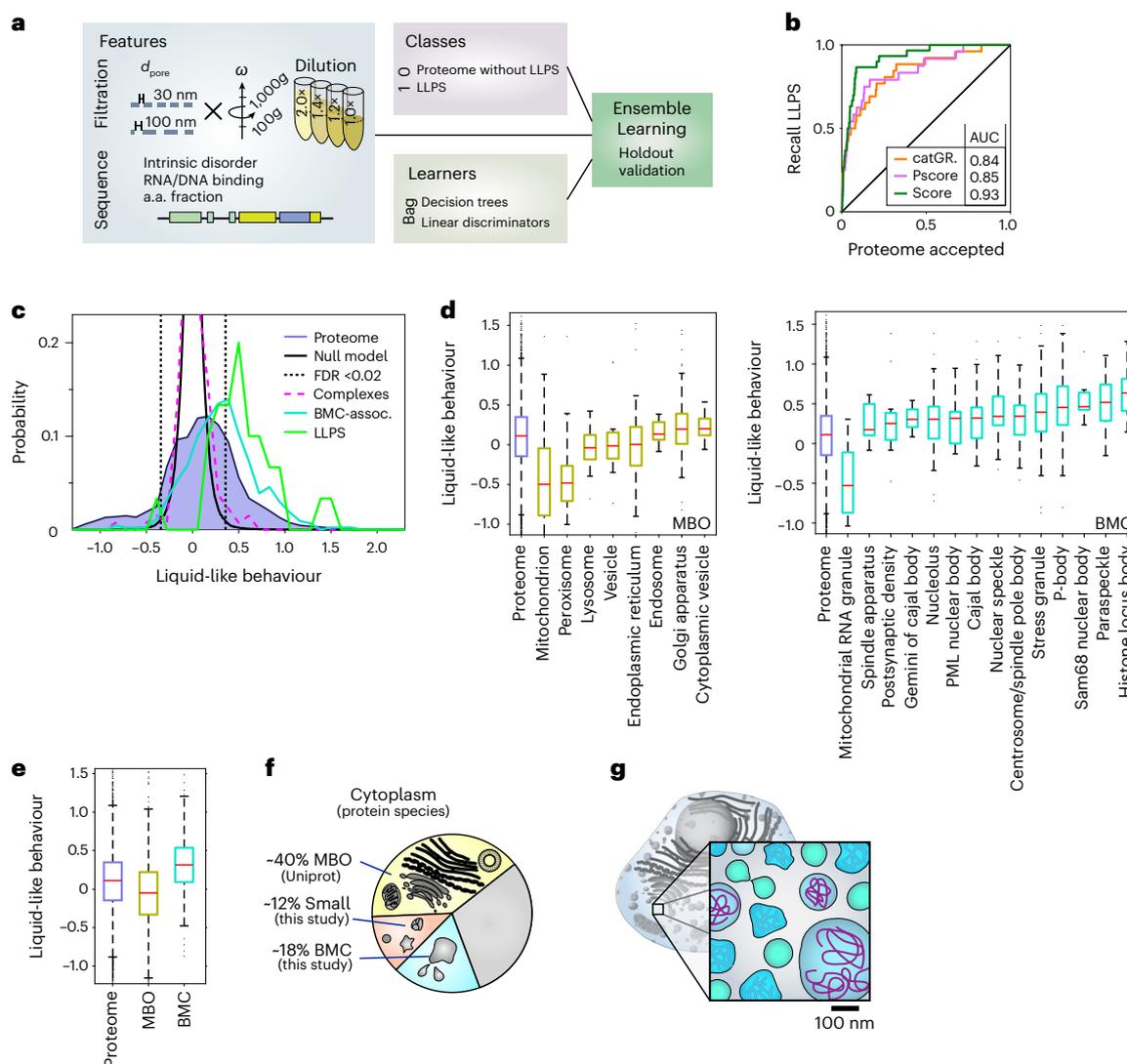


Fig. 4 | Integrating proteomics experiments enhances LLPS prediction and classifies at least 18% of the cytoplasm as BMC. **a**, Classification ensemble learning to distinguish LLPS proteins from the bulk proteome. We use our deep proteomics data of filtration chromatography (Fig. 1, $N = 1$ biological sample) and diluted filtration chromatography (Fig. 2, $N = 1$ biological sample) and additional sequence annotations as features. LLPS proteins serve as the positive training class against the rest of the proteome. Bagging of decision trees and linear discriminators returns the classification score. **b**, ROC displaying the recall of LLPS proteins by the respective scoring. The AUC of our prediction score (dark green) is higher than that of the established predictors catGRANULE (catGR)⁴⁵ (orange) and Pscore⁴⁶ (purple) and shows a steeper increase. **c**, We define liquid-like behaviour as the accumulative shift in dilution and filtration experiments ($n = 7$ measurements pooled from $N = 2$ biological independent samples), where the null model (black) is the replicate noise centred around strongly retained assemblies (complexes (magenta) and transmembrane domain proteins). A quarter of the proteome (blue, filled) and 68% of the LLPS references (green) are beyond a 2% FDR (black dotted lines). BMC-associated, supposedly not LLPS driving proteins (teal) are also shifted towards this regime. **d**, Distributions of

liquid-like behaviour for canonical MBOs (left) and BMCs (right). The different liquid-like behaviours of MBOs agree with the picture that mitochondria are large and stable and by contrast, the Golgi apparatus' intricate structures partially squeeze through filters. BMCs in general exhibit higher liquid-like behaviour, but also show broad distributions. **e**, Accumulatively, BMCs and MBOs exhibit distributions distinct from the proteome. The box plots (**d** and **e**) display data distribution with the centre as the median, the box limits as quartiles, the grey dots as outliers and whiskers as non-outlier extremes. **f**, Eukaryotic cytoplasmic organization, to a large extent, is achieved by MBOs (~40% of our sample, yellow, UniProt). In our filtration experiments, only around 12% of protein species stay unaffected (salmon). On the basis of the 2% FDR and correcting for MBOs, 18% of the cytoplasm is organized in BMCs (cyan). We cannot confidently assign any of these three organization modalities for the remaining third of the proteome (grey). **g**, Our results suggest that BMCs frequently have stable cores, contain RNAs and contribute markedly to cytoplasmic organization at the ~100 nm scale. This indicates that cytoplasm is widely structured at the mesoscale. Source numerical and proteomics data are provided in Source data and Supplementary Table 3.

Finally, we sought to estimate what fraction of the proteome is organized into these liquid-like assemblies. To this end, we integrated our filtration chromatography and dilution experiments and developed a noise model based on replicates (Fig. 4c). BMCs exhibit a liquid-like behaviour, different from membrane-bound organelles (MBOs) (Fig. 4d,e and Extended Data Fig. 9). About one-quarter of the protein species exhibit behaviour consistent with BMCs and is shifted beyond a 2% false discovery rate (FDR), including two-thirds of the LLPS

references (Fig. 4c). Additionally, only 12% of the detected proteins do not respond to the filtration and appear assembled smaller than the assayed scale. Correcting for MBOs, we conclude that at least 18% of the detected protein species are in BMCs (Fig. 4f). However, based on the spread of known BMCs into our noise model, we consider this a very conservative lower bound (Fig. 4c). We expect a substantial fraction of the remaining third of the protein species to be also organized via BMCs (Fig. 4f).

Previous studies have focused primarily on large ($\geq 1 \mu\text{m}$) BMCs, which are easily observed by microscopy. Nevertheless, recent studies have increasingly demonstrated smaller structures^{34,35}. Our differential filtration and size-exclusion studies on intact cytoplasm reveal that phase-separation-prone proteins ubiquitously organize cytoplasm into mesoscale assemblies, which exhibit liquid-like deformability (Fig. 4g). Notably, these assemblies are more stable upon dilution than expected for assemblies formed through LLPS alone, and probably have less dynamic, potentially solid-like core structures. Besides gelation or specific binding interactions, RNA may play a key role in this stabilization, as our data suggest that many RNAs are contained in BMCs. Furthermore, our findings illustrate the potential of proteomics data to enhance the prediction of LLPS proteins. Presumably, this can be effectively integrated with orthogonal approaches, such as those previously mentioned or those that analyse disordered sequences⁴⁷.

It remains an exciting question how these tiny BMCs can exist without ripening into larger condensates, via coalescence or Ostwald ripening (the growth of large condensates at the expense of smaller ones). We speculate that the mesoscale organization we have uncovered is highly dynamic, reflecting continuous assembly and disassembly. Such behaviour can originate from associative polymers and their percolation^{29,30}, or chemical activity⁴⁸, but is also reminiscent of phase-separating systems in the vicinity of a critical point, as has been suggested for two-dimensional phase separation in the plasma membrane⁴⁹.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-024-01363-5>.

References

- Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285 (2017).
- Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, aaf4382 (2017).
- Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
- Ivanov, P., Kedersha, N. & Anderson, P. Stress granules and processing bodies in translational control. *Cold Spring Harb. Perspect. Biol.* **11**, a032813 (2019).
- Wippich, F. et al. Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. *Cell* **152**, 791–805 (2013).
- Chong, P. A. & Forman-Kay, J. D. Liquid–liquid phase separation in cellular signaling systems. *Curr. Opin. Struct. Biol.* **41**, 180–186 (2016).
- Alberti, S. & Dormann, D. Liquid–liquid phase separation in disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).
- Tsang, B., Pritisanac, I., Scherer, S. W., Moses, A. M. & Forman-Kay, J. D. Phase separation as a missing mechanism for interpretation of disease mutations. *Cell* **183**, 1742–1756 (2020).
- Fei, J. et al. Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.* **130**, 4180–4192 (2017).
- Go, C. D. et al. A proximity-dependent biotinylation map of a human cell. *Nature* **595**, 120–124 (2021).
- Hubstenberger, A. et al. P-body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell* **68**, 144–157 (2017).
- Jain, S. et al. ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell* **164**, 487–498 (2016).
- Markmiller, S. et al. Context-dependent and disease-specific diversity in protein interactions within stress granules. *Cell* **172**, 590–604 (2018).
- Youn, J. Y. et al. High-density proximity mapping reveals the subcellular organization of mRNA-associated granules and bodies. *Mol. Cell* **69**, 517–532 (2018).
- Panca, R., Vranken, W. & Meszaros, B. Computational resources for identifying and describing proteins driving liquid–liquid phase separation. *Brief. Bioinform.* **22**, bbaa408 (2021).
- Vernon, R. M. & Forman-Kay, J. D. First-generation predictors of biological protein phase separation. *Curr. Opin. Struct. Biol.* **58**, 88–96 (2019).
- Lohka, M. J. & Maller, J. L. Induction of nuclear envelope breakdown, chromosome condensation, and spindle formation in cell-free extracts. *J. Cell Biol.* **101**, 518–523 (1985).
- Hannak, E. & Heald, R. Investigating mitotic spindle assembly and function in vitro using *Xenopus laevis* egg extracts. *Nat. Protoc.* **1**, 2305–2314 (2006).
- Sonnett, M., Yeung, E. & Wuhr, M. Accurate, sensitive, and precise multiplexed proteomics using the complement reporter ion cluster. *Anal. Chem.* **90**, 5032–5039 (2018).
- Johnson, A., Stadlmeier, M. & Wuhr, M. TMTpro complementary ion quantification increases plexing and sensitivity for accurate multiplexed proteomics at the MS2 level. *J. Proteome Res.* **20**, 3043–3052 (2021).
- McAlister, G. C. et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
- Meszaros, B. et al. PhaSePro: the database of proteins driving liquid–liquid phase separation. *Nucleic Acids Res.* **48**, D360–D367 (2020).
- van Mierlo, G. et al. Predicting protein condensate formation using machine learning. *Cell Rep.* **34**, 108705 (2021).
- Ning, W. et al. DrLLPS: a data resource of liquid–liquid phase separation in eukaryotes. *Nucleic Acids Res.* **48**, D288–D295 (2020).
- You, K. et al. PhaSepDB: a database of liquid–liquid phase separation related proteins. *Nucleic Acids Res.* **48**, D354–D359 (2020).
- Li, Q. et al. LLPSDB: a database of proteins undergoing liquid–liquid phase separation in vitro. *Nucleic Acids Res.* **48**, D320–D327 (2019).
- Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
- Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
- Kar, M. et al. Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc. Natl Acad. Sci. USA* **119**, e2202222119 (2022).
- Mittag, T. & Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* **82**, 2201–2214 (2022).
- Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).
- Thomas, L., Putnam, A. & Folkmann, A. Germ granules in development. *Development* **150**, dev.201037 (2023).
- Neil, C. R. et al. L-bodies are RNA–protein condensates driving RNA localization in *Xenopus* oocytes. *Mol. Biol. Cell* **32**, ar37 (2021).
- Ma, W. & Mayr, C. A membraneless organelle associated with the endoplasmic reticulum enables 3' UTR-mediated protein–protein interactions. *Cell* **175**, 1492–1506 (2018).

35. Lee, D. S. et al. Size distributions of intracellular condensates reflect competition between coalescence and nucleation. *Nat. Phys.* **19**, 586–596 (2023).
36. Mollieux, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
37. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
38. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
39. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
40. Mardia, K. V. Fisher's pioneering work on discriminant analysis and its impact on AI. Preprint at *arXiv* 2309.04774 (2023).
41. Walsh, I., Martin, A. J., Di Domenico, T. & Tosatto, S. C. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509 (2012).
42. Castello, A. et al. Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell* **63**, 696–710 (2016).
43. UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
44. Huntley, R. P. et al. The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res.* **43**, D1057–1063 (2015).
45. Bolognesi, B. et al. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.* **16**, 222–231 (2016).
46. Vernon, R. M. et al. Pi–Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7**, e31486 (2018).
47. Zarin, T. et al. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *eLife* **8**, e46883 (2019).
48. Zwicker, D., Seyboldt, R., Weber, C. A., Hyman, A. A. & Jülicher, F. Growth and division of active droplets provides a model for protocells. *Nat. Phys.* **13**, 408–413 (2017).
49. Honerkamp-Smith, A. R., Veatch, S. L. & Keller, S. L. An introduction to critical points for biophysicists; observations of compositional heterogeneity in lipid membranes. *Biochim. Biophys. Acta* **1788**, 53–63 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Methods

Xenopus laevis egg extracts

Frog husbandry. Mature *Xenopus laevis* females were purchased from Nasco/Xenopus1 and maintained by the laboratory animal resources at Princeton University. All animal procedures are approved under the Institutional Animal Care and Use Committee (IACUC) protocol 2070, reviewed in March 2023. Ovulation was induced with at least 6 month rest intervals.

Egg collection. *X. laevis* eggs were collected as previously described⁵⁰ from wild-type females aged between 1 and 5 years old. Frogs were primed with pregnant mare serum gonadotropin (HOR-272, ProSpec-Tany TechnoGene Ltd.) within 2 months before the experiment. At 16 h before egg collection, frogs were injected with 500 U of human chorionic gonadotropin (Sigma CG10) and kept at 16 °C in Marc's Modified Ringer's⁵¹ (MMR) solution (50 mM HEPES pH 7.8, 100 mM NaCl, 2 mM KCl, 1 mM MgCl₂ and 2 mM CaCl₂). We collected eggs the next day in MMR buffer and sorted out pre-activated ones for further use.

Extract preparation. *X. laevis* egg extracts were prepared as previously described^{52,53}. The eggs were dejellied in MMR with added L-cysteine (2 wt%, pH 7.8) and washed in CSF-XB buffer (100 mM KCl, 20 mM HEPES, 2 mM MgCl₂, 0.1 mM CaCl₂ and 4 mM ethylene glycol tetraacetic acid, pH 7.7). Eggs were collected into 3.5 ml centrifuge tubes (Beckmann) in the presence of cytochalasin D (Sigma C8273) and LPC (leupeptin, pepstatin, chymostatin) protease inhibitor premix (Sigma L2884), Sigma P5318 and Sigma C7268. The surplus buffer was removed after a soft spin at 500g for 1 min. The eggs were crushed and fractionated in a spin at 14,400g for 15 min. The cytoplasmic fraction was extracted using an 18G gauge needle. The extract was supplemented with 10 µg ml⁻¹ cytochalasin D, 10 µg ml⁻¹ LPC, 1 µM nocodazole (Sigma M1404) and 50 mM sucrose. All used drugs were dissolved in dimethylsulfoxide (DMSO), resulting in a total DMSO concentration of <2.5%. We pre-filtered extracts through a 6 µm poly-ether mesh filter to remove residual debris and stored them on ice for further use.

Preparation spin control experiment

Immediately after the preparation spin, the centrifuge tube containing the spin-crushed and sedimented eggs was shock frozen in liquid nitrogen at 77 K. The extract section, excluding lipid and yolk/debris sections, was cut out and cut in halves using a razor blade.

Filtration experiments

Filtrations were performed in 2 ml tubes in a tabletop centrifuge, using three-dimensional printed filter holders⁵⁴. The design files are available on our GitHub page⁵⁵.

We used the Hubs platform (<https://www.hubs.com/>) to print the holders by stereolithography (SLA) using either standard resin or dental resin materials at 20% infill rate and 50 µm layer height. We used polyethersulfone membranes (Sterlitech, PES00347100, PES0147100, PES0247100 and PES0847100) in the spin filtration setup and the in vitro assay, and cellulose acetate membranes (Sterlitech, C080A047A, C300A047A, CA1247100 and CA0247100) in the gravity flow setup. Membranes were wetted with methanol and washed with CSF-XB buffer. After excess buffer removal, the sample chamber was flushed twice with 50 µl cell extract.

Filtration chromatography experiments. For filtration chromatography, 50 µl (167 µl) of the cytoplasmic extract was loaded and spun at the respective speed (at 100g (30g) for early timepoints and at 1,000g (300g) for late timepoints) in turns of 4 min until approximately 15 µl of filtrate was collected. To avoid material build up or cake formation, the samples were spun at a fixed angle (45°), and the sample chamber

was stirred after each turn. The input sample (-10 µl) was collected before the spin.

Dilution experiments. At 30 min before the spin filtration process, the extracts were diluted in CSF-XB buffer in a dilution series to a factor of 1.2, 1.44 and 2.

RNase experiments. A total of 1 U µl⁻¹ of RNase I (Ambion, ThermoFisher AM2294) was added to the solutions 45 min before the filtration. One unit is the amount of enzyme required to produce 1 µg of acid-soluble material from mouse liver RNA in 30 min at 37 °C (ThermoFisher).

Gravity flow experiments. The samples (200 µl) were loaded and placed in a box providing a humid atmosphere until sufficient flow through accumulated. The input chamber was stirred occasionally to avoid clogging.

Sedimentation assay

Then, 200 µl of extract was spun in 5 × 20mm polypropylene tubes (Beckman Coulter) at 200 krcf for 15 min (30 min), using a TLS55 swinging bucket with tube adaptors in an Optima TLX ultracentrifuge (Beckman Coulter). The samples were flash frozen in liquid nitrogen and cut into halves.

Expression of GFP fusion proteins

The gateway entry plasmids of desired proteins were retrieved from the *Xenopus laevis* ORFeome⁵⁶. The destination vector carrying an enhanced GFP sequence-TEV site-S-tag was purchased from Addgene (pCSF107mT-GATEWAY-3'-LAP tag, plasmid #67618). For the gateway LR (left and right recombination) cloning reaction (which involves the recombination of attL and attR sites), the entry plasmid, the destination plasmid and the gateway LR clonase II enzyme mix (Invitrogen I1791) were combined at the ratios recommended in the manufacturer's protocol. After the reaction, the expression cloned vector was purified then linearized using restriction enzymes, which were chosen so that the region of protein of interest was protected. The linearized plasmids were in vitro transcribed using the mMMESSAGE mMACHINE SP6 kit (Invitrogen AM1340) supplemented with a 7-methyl guanosine cap protected on the 5' end terminal and a poly(A) tail (NEB M0276). Finally, RNA products were purified using Trizol LS reagent (Invitrogen 10296010), then resuspended in nuclease-free water at -1 µg µl⁻¹ as the final RNA concentration. The RNA solution was added in a volume ratio of 1:100 to the extracts.

Size-exclusion assay

Size-exclusion chromatography beads (GE, Sephacryl High Resolution, S-200, 300, 400 and 500, 17-584-10, 17-599-99, 17060999 and 170613-10; molecular size cut-offs 400 kDa, 2 MDa, 9 MDa and 100 MDa) were washed in CSF-XB buffer and equilibrated in three rounds of sedimentation, supernatant removal, 1:5 add-up in the plain cell extract and waiting times of about 10 min. After being finally added to the labelled cell extracts, we waited at least 15 min before imaging. The samples were enclosed in mineral oil to prevent evaporation (Sigma M5904). For calibration, we added the beads to a 70 kDa dextran-rhodamine-B-isothiocyanate (Sigma, R9379, approximately 5 nm) solution in CSF-XB buffer. The sizes were estimated from the molecular weights using Zetasizer (Malvern Panalytical).

Image analysis. Line profiles of beads and the surrounding solution were measured manually in ImageJ/Fiji^{57,58}. Raw image intensities were corrected for the detector background to make them proportional to concentrations. An estimation of the accessible volume for each bead type was measured by the intensity ratio for the dextran solution.

The GFP intensities measured inside the bead were adjusted by this factor to calculate the concentration ratio.

Microscopy

Confocal images of labelled cytoplasmic extracts were taken in glass-bottom well plates (Cellvis) on a Nikon A1 laser scanning confocal microscope using 60× and 20× oil immersion objectives. Images of the *in vitro* assay were taken on a Nikon spinning disc confocal microscope with a 100× oil immersion objective. Fluorescence correlation microscopy (FCS) was performed with an oil immersion objective (Plan Apo 60×/1.4 numerical aperture, Nikon) using an FCS Upgrade Kit for Laser Scanning Microscopes (PicoQuant). FCS measurements were performed using the SymPhoTime Software (PicoQuant).

MS sample preparation and analysis

Low-complexity samples. Samples for mass spectrometry (MS) underwent preparation spin control, sedimentation assay, RNase treatment, gravity flow, filtration of data used for fitting the model and filtration at 200 nm. Samples were labelled using tandem mass tag (TMT)-10plex and TMTpro-16plex (Thermo Fisher Scientific) and analysed by TMT-MS3 (ref. 21) and TMTproC (RNase treatment). For detailed tag assignment to channels, see Supplementary Table 3.

High-complexity samples. Samples underwent filtration chromatography at 30 nm and 100 nm, then were labelled using TMT-10plex and analysed by TMTc¹⁹. The diluted cytoplasm was filtered, labelled using TMT-10plex and TMTpro-16plex (Thermo Fisher Scientific) and then analysed by TMTproC²⁰. For detailed tag assignment to channels, see Supplementary Table 3.

Sample preparation. Samples were prepared mostly as previously described⁵⁹. Lysates were collected in 100 mM HEPES pH 7.2 and proteins were denatured by adding 2% sodium dodecyl-sulfate at volumes of -100 μ l. The concentrations were determined by the bicinchoninic acid assay (Pierce) and similar amounts underwent further processing. All conditions for a multiplex were prepared in the same batch. To reduce disulfides, dithiothreitol (DTT) (500 mM in water) was added to a final concentration of 5 mM (20 min at 60 °C). Samples were cooled to room temperature (RT), and cysteines were alkylated by the addition of *N*-ethyl maleimide (NEM, 1 M in acetonitrile) to a final concentration of 20 mM followed by incubation for 20 min at RT. Then, 10 mM dithiothreitol (500 mM stock in water) was added at RT for 10 min to quench any remaining NEM. A methanol-chloroform precipitation was performed for protein cleanup, and the collected protein pellets were allowed to air dry. Samples were taken up in 6 M guanidine chloride in 200 mM 3-(4-(2-hydroxyethyl)piperazin-1-yl)propanesulfonic acid (EPPS) pH 8.5. Subsequently, the samples were diluted to 2 M guanidine chloride in 200 mM EPPS pH 8.5 for overnight digestion with 20 ng μ l⁻¹ Lys-C (Wako) at RT. The samples were further diluted to 0.5 M guanidine chloride in 200 mM EPPS pH 8.5 and then digested with 20 ng μ l⁻¹ Lys-C and 10 ng μ l⁻¹ trypsin (Promega) at 37 °C overnight. For the samples of the sedimentation assay and the model fit, methanol-chloroform precipitation was replaced by SP3 magnetic bead (SpeedBead Magnetic Carboxylate, Thermo Scientific -45/651521050250) cleanup⁶⁰.

The digested samples were dried using a vacuum evaporator at RT and taken up in 200 mM EPPS pH 8.0. To equalize channel loading, the same protein masses for each condition in a multiplexed sample were labelled with TNTs. The total mass per sample was -20 μ g and -200 μ g for low-complexity and high-complexity samples, respectively. TMT/TMTpro samples were labelled for 2 h at RT. Labelled samples were quenched by adding 0.5% hydroxylamine to the solution. Samples from all conditions were combined into one tube, acidified to pH <2 with phosphoric acid (high-performance liquid chromatography (HPLC) grade, Sigma) and cleared by ultracentrifugation at

100,000g at 4 °C for 1 h in polycarbonate tubes (Beckman Coulter, 343775) in a TLA-100 rotor. The supernatants were dried using a vacuum evaporator at RT. For a low-complexity sample, dry samples were taken up in HPLC-grade water, stage-tipped for desalting⁶¹ and resuspended in 1% formic acid (FA) to 1 μ g μ l⁻¹ for MS analysis. For high-complexity samples, the supernatant was sonicated for 10 min and then fractionated by medium pH reverse-phase HPLC (Zorbax 300Extend C18, 4.6 × 250 mm column, Agilent) with 10 mM ammonium bicarbonate pH 8.0, using 5% acetonitrile for 17 min followed by an acetonitrile gradient from 5% to 30%. Fractions were collected starting at minute 17 with a flow rate of 0.5 ml min⁻¹ into a 96-well plate every 38 s. These fractions were pooled into 24 fractions by alternating the wells in the plate⁶². Each fraction was dried and resuspended in 100 μ l of HPLC water. Fractions were acidified to pH <2 with HPLC-grade trifluoroacetic acid and stage-tipping was performed to desalt the samples. For liquid chromatography (LC)-MS analysis, samples were resuspended to 1 μ g μ l⁻¹ in 1% FA and HPLC-grade water and -1 μ g of peptides were analysed per 1 h run time. The quality of the sample preparation was controlled by checking the labelling degree, channel loading, content of cysteine-containing peptides and missed cleavages in a single-shot MS3 analysis.

MS analysis. Approximately 1–3 μ g of the sample was analysed by LC-MS. LC-MS experiments were performed with an nLC-1200 HPLC (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Lumos (Thermo Fisher Scientific). Injected volumes were in the range of 1–4 μ l. For each run, peptides were separated on an Aurora Series emitter column (25 cm × 75 μ m ID and 1.6 μ m C18) (IonOpticks), held at 60 °C during separation by an in-house built column oven. Separation was achieved by applying a 12–35% acetonitrile gradient in 0.125% FA and 2% DMSO over 90 min for fractionated samples and 180 min for unfractionated samples at 350 nl min⁻¹ at 60 °C. Electrospray ionization was enabled by applying a voltage of 2.6 kV through a MicroTee at the inlet of the microcapillary column. As indicated in each proteomics experiment, we used the Orbitrap Fusion Lumos with a TMT-MS3 (ref. 21), TMTc¹⁹ or TMTproC²⁰, as previously described.

Reference databases

If not stated otherwise, all annotations are from Uniprot⁴³.

Complexes and organelles. Our reference group for large ‘complexes’ includes proteins from ribosome, proteasome, the vault complex, Arp2/3, RNA polymerase II core complex, proliferating cell nuclear antigen-DNA polymerase delta complex and gamma-tubulin ring complex (the latter four from CORUM⁶³). In the group ‘Mitochondrion’, we exclude proteins with promiscuous subcellular location annotations. The reference group ‘BMC-associated proteins’ is the union of the categories ‘client’ and ‘regulator’ in DrLLPS²⁴. Organelles in DrLLPS that had an abundance of client and regulator annotations were further refined using the following additional databases: ‘Stress granule’ and ‘P-body’ by the RNPgranule database⁶⁴ and ‘Nucleolus’, ‘Postsynaptic density’, ‘Centrosome/Spindle pole body’ and ‘promyelocytic leukemia (PML) nuclear body’ by UniProt subcellular location. For the estimation of the fraction of proteins in MBOs, we used QuickGO⁴⁴ and restrict to ‘UniProtKB’ (swiss-prot) and ‘located in’ MBOs. ‘Small’ proteins were located around the origin in our filtration experiments within the bounds of our null model. Transmembrane helices were predicted by Krogh’s algorithm⁶⁵.

LLPS database. To create a comprehensive, high-confidence database of known LLPS proteins with minimal personal curation bias, we follow the approach of merging several published databases¹⁵: PhasePro²², DrLLPS²⁴, PhaSepDB²⁵ and LLPSDB²⁶, and extend this set by the reference list for the PSAP predictor²³. We also include proteins assigned ‘candidate’ in PhasePro and the updated annotation ‘PS-self’

in PhaSepDB version 2 that indicates de novo phase separation without partners. We define the 'consensus level' as the number of databases in which a protein is found; throughout the manuscript we use consensus level 4, except Fig. 2 (level 3) and Extended Data Fig. 4 (level 0). Overall, LLPS proteins with higher consensus level showed larger shifts. This may reflect the increasing curation quality but may also be due to system-specific or partner-dependent phase separation.

Predictor learning

We train our predictor for phase separating proteins in MATLAB (MathWorks), using the function `fitensemble` for ensemble classification. Class one is proteins from our LLPS database and class two is the rest of the proteome. We use only proteins identified in both the filtration chromatography and the dilution experiments ($N \sim 4,000$) and thus have no missing values. The features are the filtrate concentrations relative to the input in the different conditions (filtration chromatography: 30-nm and 100-nm at T_1 and T_2 and diluted filtration: 100 nm diluted $1\times$ (undiluted), $1.2\times$, $1.4\times$ and $2\times$). We include features for intrinsically disordered regions (IDRs) determined by Espritz⁴¹ (the fraction of a.a. in IDRs, the number of IDRs of >50 a.a./ >30 a.a./any length and setting: `disprot`, `BestSw`). We include features on DNA binding⁴³, RNA binding (QuickGo⁴⁴) and RNA-binding domains⁴². Following van Mierlo et al.²³, we also include the sequence fractions of glycine, cysteine, leucine and isoleucine, as well as the content of aliphatic and aromatic residues. We use the method 'Bag' to train ensembles with 500 learners of decision trees and linear discriminators^{38–40}. The trees are restricted to a minimum number of 32 proteins per leaf and have a maximum number of splits equalling three-quarters the number of used features. We train $N = 600$ ensembles on partitions of the data using 80% of class 1 ($N_1 \sim 40$) and 10% of class 2 ($N_2 \sim 400$). Thus, we obtain each protein's cross-validated score by the median score of these ensembles, excluding any runs where it was used in the training. When trained separately on experiment (s_{exp}) and sequence features (s_{seq}) and combined by the Euclidean distance, $s = \sqrt{(\max(s_{\text{exp}}) - s_{\text{exp}})^2 + (\max(s_{\text{seq}}) - s_{\text{seq}})^2}$, the

final score s shows a slightly better performance (AUC for recall LLPS 0.93 versus 0.91). Probably, this can be accounted to the relatively small training sets and that there are more sequence features than experimental features. The PSAP predictor could not be plotted as its published output does not contain cross-validated results of the training set.

Data processing

Data exclusion. For all the experiments reported in the manuscript, no data were excluded.

Normalization. Filtration data were compared using their fold changes (FC of c/c_0), where the concentrations c are assumed proportional to the MS signal sum of peptides' reporter ions (MS3 method) or complementary ions (TMTc method). The FCs were normalized separately for each experiment to the 0.95 quantile of each experiment to account for loading imbalance in the TMT channels. Proteins with the highest FCs pass the filter medium fastest and represent unhindered flow through.

Squeezing score. The upper edge close to the identity line of the scatter data is determined by fitting a line and taking the top four per cent of points in bins along the line. Another line is fit through these points and the squeezing score is the orthogonal distance to it.

Null model. We created an empirical null model for the measurement noise. We matched replicates with a Pearson correlation larger than 0.8 from 36 early elution conditions and fit a line through each of the resulting 359 sample pairs in \log_2 space. The resulting histogram of residues ($N = 393,000$) serves as an estimate for the errors.

Fraction of BMCs. To estimate the fraction of the proteome that is organized in liquid assemblies, we further constrain this by fitting the line through large complexes and proteins with transmembrane domains^{65,66}. We average the shifts in the 30-nm and 100-nm filtration chromatography and the $1.2\times$ and $1.44\times$ dilution experiments to quantify the 'liquid-like behaviour'. We identified proteins in BMCs beyond a 2% FDR of the null model, omitting any proteins with membrane annotation.

MS data analysis

MS data analysis was performed essentially as previously described⁶⁷, using the Gygi Lab software platform (GFY Core Version 3.8) licensed through Harvard University. The MS data in the Thermo RAW format were converted to mzXML format and erroneous assignments of peptide ion charge state and monoisotopic m/z were corrected⁶⁸. Monoisotopic mass detection was supported by Monocle⁶⁹. ReAdW.exe was modified to include signal-to-noise ratios for each peak during file format conversion⁷⁰. Assignments of MS2 spectra were performed by the SEQUEST algorithm⁷¹, searching against a combined database made of (1) the *X. laevis* v9.2 genome assembly (Xenbase⁷² (RRID:SCR_003280)), (2) common contaminants such as human keratins and digestion enzymes, and (3) the reverse protein sequences of the target ((1) and (2)) as a decoy. Searches were performed using a precursor ion tolerance of 20 ppm and a product ion tolerance of 1 Da or 0.02 Da for MS3 or TMT(pro)c methods, respectively. Both peptide termini were required to be consistent with Lys-C/Trypsin digest specificity, allowing one missed cleavage. Static modifications included TMT/TMTpro tags on lysine residues and peptide N-termini (+229.162932 Da/+304.2071 Da) and NEM on cysteine residues (+125.047679 Da). Up to three differential modifications included oxidating of methionine residues (+15.99492 Da) and water addition to NEM on cysteines (+18.0105 Da). An MS2 spectral assignment FDR of less than 1% was achieved by applying the target–decoy database search strategy⁷³. Filtering was performed using a linear discrimination analysis method to create one combined filter parameter from the following z-scored peptide ion and MS2 spectra properties: SEQUEST parameters XCorr and Diff_Seq_dCN, missed cleavages, adjusted ppm, peptide length, fraction of ions matched and charge state. Forward peptides within three standard deviations of the theoretical m/z of the precursor were used as a positive training set. All reverse peptides were used as negative training set. Linear discrimination scores were used to sort peptides with at least seven residues and to filter with a cut-off of 1% FDR based on the decoy database⁶⁸. Each search was software recalibrated to alleviate any systematic mass error dependent on peptide elution time or observed m/z . All ions in the full MS1 spectra were first adjusted. A representative subset of peptides was selected using those above the median XCorr and within one standard deviation of the global mass error. The mass errors of this subset were then fit to each parameter using locally estimated scatter plot smoothing (LOESS) regression. The m/z of every ion in MS1 spectra was then adjusted by the error predicted by interpolating the values of the nearest data points in the regression model. Adjustments for each of the two parameters were done iteratively. MS2 spectra were then calibrated in a similar manner. Mass errors were calculated from matched peptide fragment ions within two standard deviations of the global mass error and above the upper quartile of intensity. Mass errors were fitted to each parameter using locally estimated scatter plot smoothing regression and the m/z for every ion in MS2 spectra was adjusted as above. Peptides that matched multiple proteins were assigned to the proteins with the greatest number of unique peptides. TMT-MS3 (ref. 21), TMTc¹⁹ or TMTproC²⁰ data were analysed as previously described (the MATLAB module available on github⁷⁴).

The MS proteomics data have been deposited to the ProteomeX-change Consortium via the PRIDE⁷⁵ partner repository with the dataset identifier [PXD029879](https://doi.org/10.1038/s41556-024-01363-5).

RNA-seq library preparation and sequencing

Procedures for RNA extraction from input, and filtered samples T_1 and T_2 were similar to protein samples preprocessing for MS except that the samples were immediately lysed in TRIZOL, frozen in liquid nitrogen and stored at -80°C . Three biological replicates were collected in total. RNA extraction was performed using phenol–chloroform precipitation and purified with the RNA Clean & Concentrator kit (Zymo) following the manufacturer's instructions, including a 10 min DNase treatment in column. The quality of the RNA was assessed using a NanoDrop (Thermo Fisher Scientific), Qubit (Invitrogen) and TapeStation system (Agilent Biotechnologies), and only samples with high quality (RNA integrity number ≥ 7.0) were used to prepare libraries on an Apollo 324 system with the PrepX RNA-seq protocol (Takara Bio) and a ribo-depletion step (RiboCop rRNA, Lexogen). Paired-end (300 cycles total) sequencing was performed on an NovaSeq SP 100nt Flowcell v1.5 (Illumina) at the Genomics Core Facility of Princeton University with a read depth of 30–90 million reads per sample.

Bioinformatic analysis of sequencing results

The RNA-seq reads were first assessed using FastQC (0.10.0)⁷⁶ and TrimGalore (0.6.10)⁷⁷ for adaptor removal and quality control. The processed reads were then aligned to the *X. laevis* v10.1 reference genome (Xenbase)⁷⁸ using STAR (v2.7.10a) with the option ‘-quant-Mode GeneCounts’⁷⁹. The output files were imported into R (v3.5.1). FCS between T_1 and T_2 filtration versus input samples were calculated after performing the median of ratios normalization and regularized log transformation using DESeq2 (v1.32.0) (ref. 80). Sample similarity was assessed using hierarchical clustering with both Euclidean and Pearson distances. Functional enrichment analysis was done using the STRING database⁸¹ and its reported FDRs determined by Bonferroni-corrected Kolmogorov–Smirnov tests. The category ‘Development-related’ was built by text mining all enrichments for the terms ‘morphogenesis’, ‘development’, ‘differentiation’, ‘fate’ or ‘growth’. All RNA-seq datasets are deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (GEO) with the GEO accession number [GSE232651](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232651).

Randomization and blinding

No randomization was performed. Randomization is not relevant to this study for comparisons within one set of isobaric tags in multiplexed proteomics studies. Data collection and analysis were not performed blind to the conditions of the experiments. Proteomics or transcriptomics samples underwent the same workflow during which they were indistinguishable.

Statistics and reproducibility

Proteomics experiments on filtration chromatography and dilution shown in Figs. 1 and 2 and Extended Data Figs. 1–4, 6 and 9 were performed on $n = 1$ biological sample each. This way, the number of conditions compared in a multiplex could be maximized. Proteomics replicates were performed in shallow samples. Each experiment was measured on at least $n = 2$ biologically independent samples. Samples showed good agreement to an initial screen for experimental conditions, among similar conditions and to a broad variation of conditions shown in Supplementary Fig. 2 ($n = 89$ experiments pooling $N = 18$ biologically independent samples).

Transcriptomics experiments were conducted using $N = 3$ biologically independent samples. In total, $n = 9$ independent experiments were performed, that is, a biological triplicate for each experiment. The triplicate, as well as its accompanying proteomics data, showed high reproducibility.

Confocal microscopy micrographs in Extended Data Fig. 7 are representative of ten micrographs taken on $n = 1$ biological sample for each protein. The observation of lack of micron-scale structure was confirmed throughout the other replicates in the FCS and size-exclusion

bead assays. FCS spectra were collected from N biologically independent samples, measured by n FCS traces across the extract, N/n : Dex70: 2/36; ILF3: 5/36; G3BP2: 1/12; HNRNPA1: 4/26; CIRBP: 4/68; GID8: 4/28; WDRI: 4/23; AP2S1: 3/13; PCF11: 4/23; and CCS: 3/33. Confocal micrographs in Fig. 4 and Extended Data Fig. 8 are representative for n_1 , n_2 , n_3 and n_4 micrographs for each bead category of $n = 1$ biological sample, WDRI: 4, 4, 4 and 5; GID8: 4, 5, 5 and 5; CCS: 4, 5, 5 and 4; CIRBP: 4, 6, 6 and 5; HNRNPA1: 4, 4, 4 and 5; ILF3: 4, 4, 5 and 5; MVB12B: 5, 4, 6 and 6; QRICH1: 4, 2, 3 and 4; and VGLL4: 5, 4, 4 and 6. Size-exclusion assays were performed in $n = 3$ biologically independent replicates showing similar results. The confocal micrographs in Supplementary Fig. 4 are representative of five micrographs of an in vitro experiment performed in duplicate. The confocal micrographs in Supplementary Fig. 5 are representative of four micrographs of $n = 2$ biologically independent samples.

Data distribution was assumed to be normal when performing t -tests, but this was not formally tested.

No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications^{54,82}.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Assignments of MS spectra were searched against the *X. laevis* v9.2 genome assembly (Xenbase (RRID:SCR_003280))⁷². The LLPS database was sourced from PhasePro²², PhaSepDB²⁵, DrLLPS²⁴ and LLPSDB²⁶, and the references from the PSAP predictor²³, and protein complex data from CORUM⁶³. Protein nucleic acid binding was sourced from Uniprot⁴³, QuickGO⁴⁴ and Castello et al.⁴². Organelle data were sourced from DrLLPS²⁴, Uniprot⁴³ and RNPgranuleDB⁶⁴. Predictions of catGRANULE⁴⁵, Pscore⁴⁶ or PSAP²³ were re-evaluated and plotted. IDRs, transmembrane helix annotations and enrichment analyses were generated from Espritz⁴¹, Krogh et al.⁶⁵ and STRING⁸¹, respectively. Housekeeping gene and L-body annotations were derived from Eisenberg et al.³¹ and Neil et al.³³. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXD029879](https://www.ebi.ac.uk/pride/archive/study/PXD029879). The raw sequencing data and gene expression matrices have been deposited to the National Center for Biotechnology Information's GEO with the GEO series accession number [GSE232651](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232651). All other data are available from the corresponding authors upon reasonable request. Source data are provided with this study in Source data and Supplementary Table 3. Source data are provided with this paper.

Code availability

Custom code is available from the corresponding authors upon reasonable request. Code for the analysis of TMTproC data is available on GitHub⁷⁴.

References

50. Wlzl, M., McNamara, S. & Horb, M. E. Generation and care of *Xenopus laevis* and *Xenopus tropicalis* embryos. *Methods Mol. Biol.* **1865**, 19–32 (2018).
51. Ubbels, G. A., Hara, K., Koster, C. H. & Kirschner, M. W. Evidence for a functional role of the cytoskeleton in determination of the dorsoventral axis in *Xenopus laevis* eggs. *J. Embryol. Exp. Morphol.* **77**, 15–37 (1983).
52. Good, M. C. & Heald, R. Preparation of cellular extracts from xenopus eggs and embryos. *Cold Spring Harb. Protoc.* **2018**, 097055 (2018).
53. Sawin, K. E. & Mitchison, T. J. Mitotic spindle assembly by two different pathways in vitro. *J. Cell Biol.* **112**, 925–940 (1991).

54. Nguyen, T. et al. Differential nuclear import sets the timing of protein access to the embryonic genome. *Nat. Commun.* **13**, 5887 (2022).
55. 3DFilterHolderDesigns. *GitHub* <https://github.com/wuhrlab/3DFilterHolderDesigns> (2021).
56. Grant, I. M. et al. The *Xenopus* ORFeome: a resource that enables functional genomics. *Dev. Biol.* **408**, 345–357 (2015).
57. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
58. Abramoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with ImageJ. *Biophoton. Int.* **11**, 36–42 (2004).
59. Gupta, M., Sonnett, M., Ryazanova, L., Presler, M. & Wühr, M. Quantitative proteomics of *Xenopus* embryos I, sample preparation. *Methods Mol. Biol.* **1865**, 175–194 (2018).
60. Hughes, C. S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).
61. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
62. Edwards, A. & Haas, W. Multiplexed quantitative proteomics for high-throughput comprehensive proteome comparisons of human cell lines. *Methods Mol. Biol.* **1394**, 1–13 (2016).
63. Ruepp, A. et al. CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* **38**, D497–D501 (2010).
64. Youn, J. Y. et al. Properties of stress granule and p-body proteomes. *Mol. Cell* **76**, 286–294 (2019).
65. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
66. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182 (1998).
67. Sonnett, M., Gupta, M., Nguyen, T. & Wühr, M. Quantitative proteomics for *Xenopus* embryos II, data analysis. *Methods Mol. Biol.* **1865**, 195–215 (2018).
68. Huttlin, E. L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
69. Rad, R. et al. Improved monoisotopic mass estimation for deeper proteome coverage. *J. Proteome Res.* **20**, 591–598 (2021).
70. Sashimi. *Source Forge* <http://sashimi.svn.sourceforge.net/viewvc/sashimi/> (2023).
71. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* **5**, 976–989 (1994).
72. Fortriede, J. D. et al. Xenbase: deep integration of GEO and SRA RNA-seq and ChIP-seq data in a model organism database. *Nucleic Acids Res.* **48**, D776–D782 (2020).
73. Elias, J. E. & Gygi, S. P. Target–decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
74. TMTProC. *GitHub* <https://github.com/wuhrlab/TMTProC> (2021).
75. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2018).
76. Andrews, S. FastQC. *Babraham Bioinformatics, Babraham Institute* <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
77. Krueger, F. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files *Babraham Bioinformatics, Babraham Institute* http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (2015).
78. <http://www.xenbase.org/entry/>
79. Dobin, A. & Gingeras, T. R. Optimizing RNA-seq mapping with STAR. *Methods Mol. Biol.* **1415**, 245–262 (2016).
80. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
81. Szklarczyk, D. et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–d646 (2023).
82. Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301–304 (2014).

Acknowledgements

We thank L. Ryazanova for help with sample preparation and members of the Wühr and the Brangwynne laboratories for useful discussions. We thank D. Hill for access to the *Xenopus* ORFeome and J. Pelletier for help with designing the filter holders. This work was supported by European Molecular Biology Organization ALTF 601-2018 (F.C.K.), the Lewis–Sigler Institute (M.W. and C.P.B.), National Institutes of Health grant R35GM128813 (M.W.), American Heart Association predoctoral fellowship 20PRE35220061 (T.N.), Princeton Catalysis Initiative (A.M. and M.W.), Eric and Wendy Schmidt Transformative Technology Fund (M.W. and C.P.B.), the Howard Hughes Medical Institute (C.P.B.), the Princeton Biomolecular Condensate Program (C.P.B.), the Princeton Center for Complex Materials, the Materials Research Science and Engineering Center (NSF DMR-2011750) (C.P.B.) and the Air Force Office of Scientific Research Multi-University Research Initiative (FA9550-20-1-0241) (C.P.B.).

Author contributions

F.C.K., C.P.B. and M.W. designed the research. F.C.K. conducted the experiments and analysed the data. F.C.K. and T.N. developed and performed in vitro protein expression. F.C.K. and A.M. analysed the RNA-seq data. M.W. and C.P.B. provided funding and supervised the study. F.C.K., C.P.B. and M.W. wrote the manuscript, and all authors helped edit the manuscript.

Competing interests

C.P.B. is a founder and consultant of Nereid Therapeutics. F.C.K., T.N., C.P.B. and M.W. are inventors and applicants in the provisional patent application US 63/433,243 on the filtration chromatography method. A.M. declares no competing interests.

Additional information

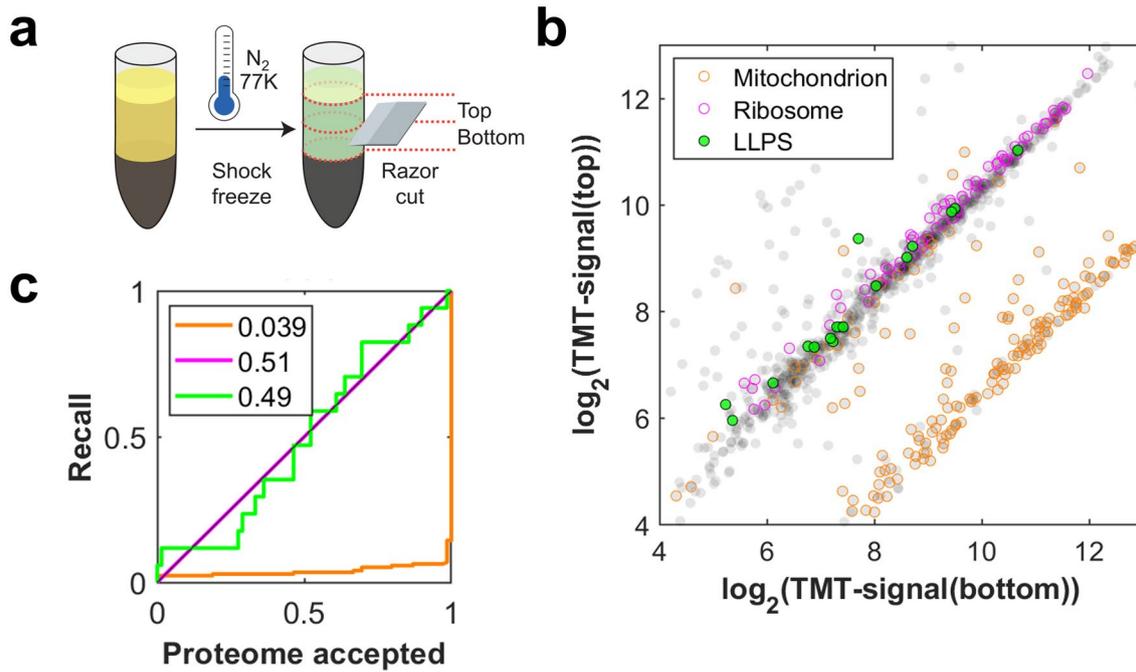
Extended data is available for this paper at <https://doi.org/10.1038/s41556-024-01363-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41556-024-01363-5>.

Correspondence and requests for materials should be addressed to Clifford P. Brangwynne or Martin Wühr.

Peer review information *Nature Cell Biology* thanks Jonathon Ditlev and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

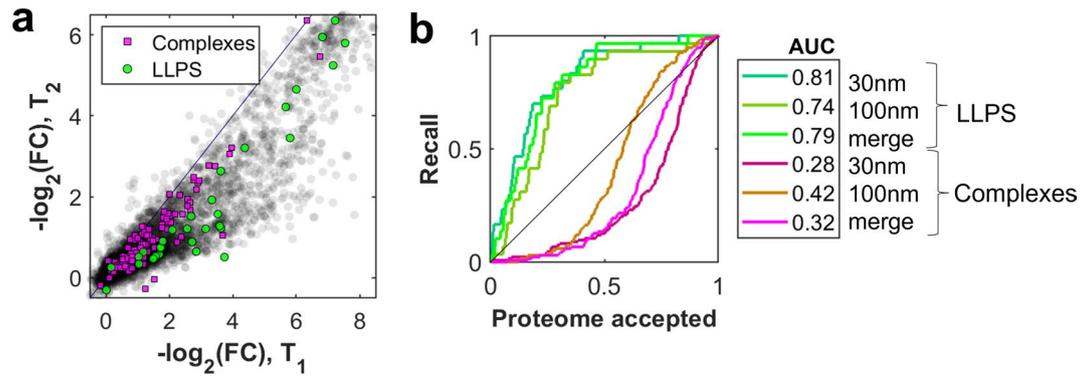
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Sedimentation in extract preparation spin is negligible.

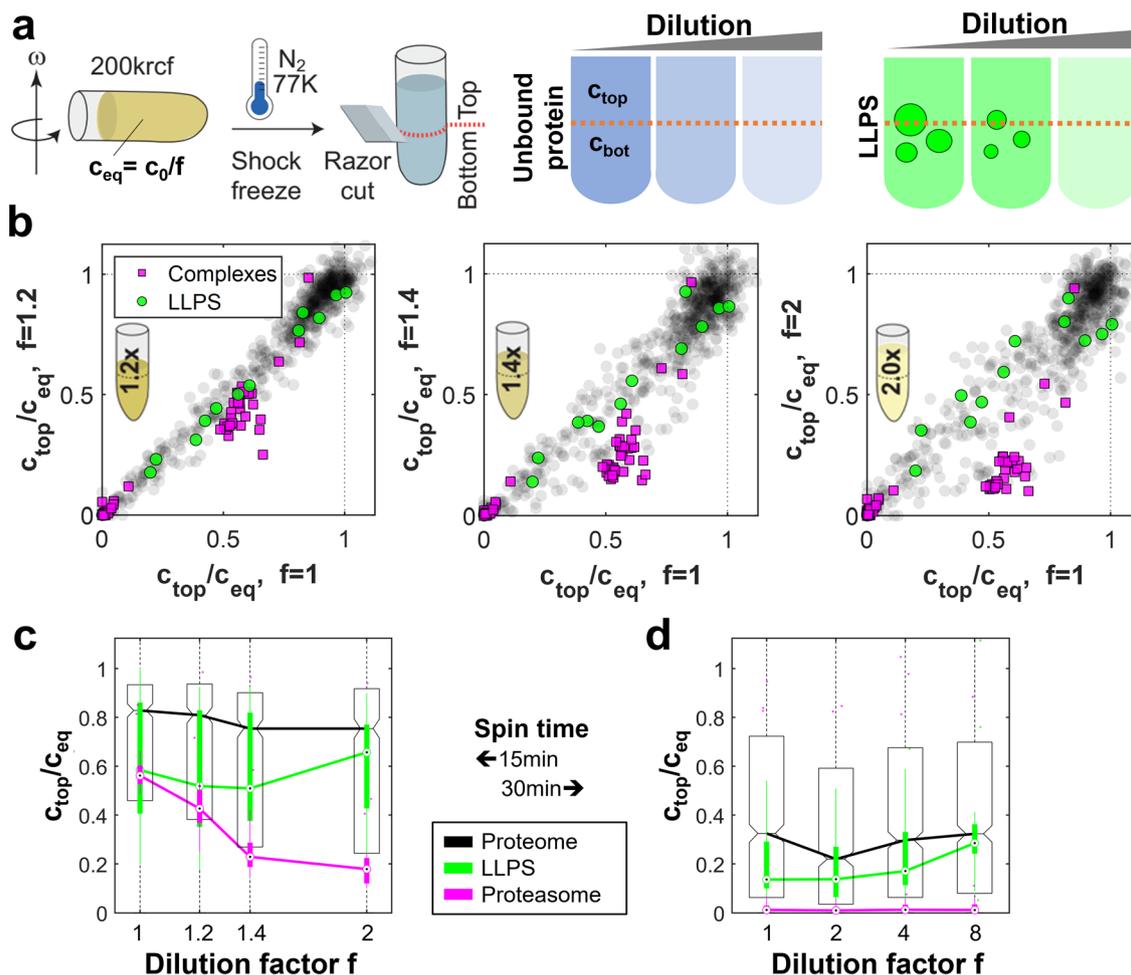
a, Schematic illustrating the sample collection. After 16 min of the preparation spin (14400 krcf), the tube is shock frozen in liquid nitrogen and subsequently cut into a top and a bottom section with a razor blade. **b**, Comparison of the two halves of the extract. While mitochondrial proteins (orange) are shifted to the

bottom, indicating sedimentation, the bulk part of the proteome, including both ribosomal (magenta) and LLPS proteins (green), stays unchanged. Scatter plot of raw TMT signals. $N = 1$ biological sample. **c**, Receiver operator characteristics on the signal ratio top to bottom of the data in (b). Source numerical data and proteomics data are provided in Source Data and Table 3.



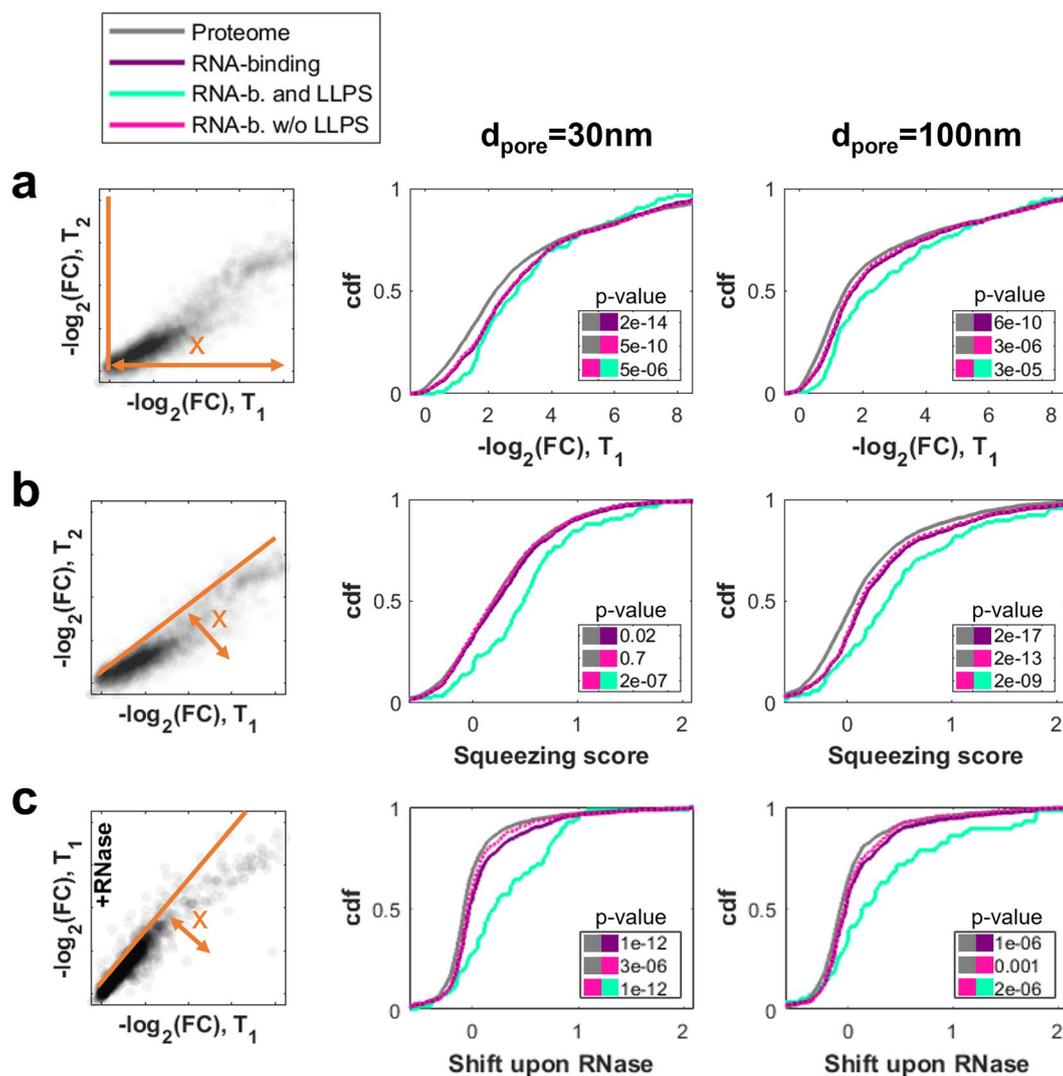
Extended Data Fig. 2 | Filtration at pore size 100 nm. **a**, Executing the experiment presented in Fig. 1c with a larger pore size ($d_{\text{pore}} = 100 \text{ nm}$) yields qualitatively similar results. However, there are quantitative changes, and, as expected, the overall permeation is higher. FCs were normalized separately for each experiment to their 0.95 quantile, representing freely passing proteins.

$N = 1$ biological sample. **b**, The receiver operating characteristic for LLPS proteins is best for the 30 nm filtration (AUC = 0.81). Notably, the separation of the LLPS and complexes groups is weaker in the 100 nm condition. Source numerical data and proteomics data are provided in Source Data and Supplementary Table 3.



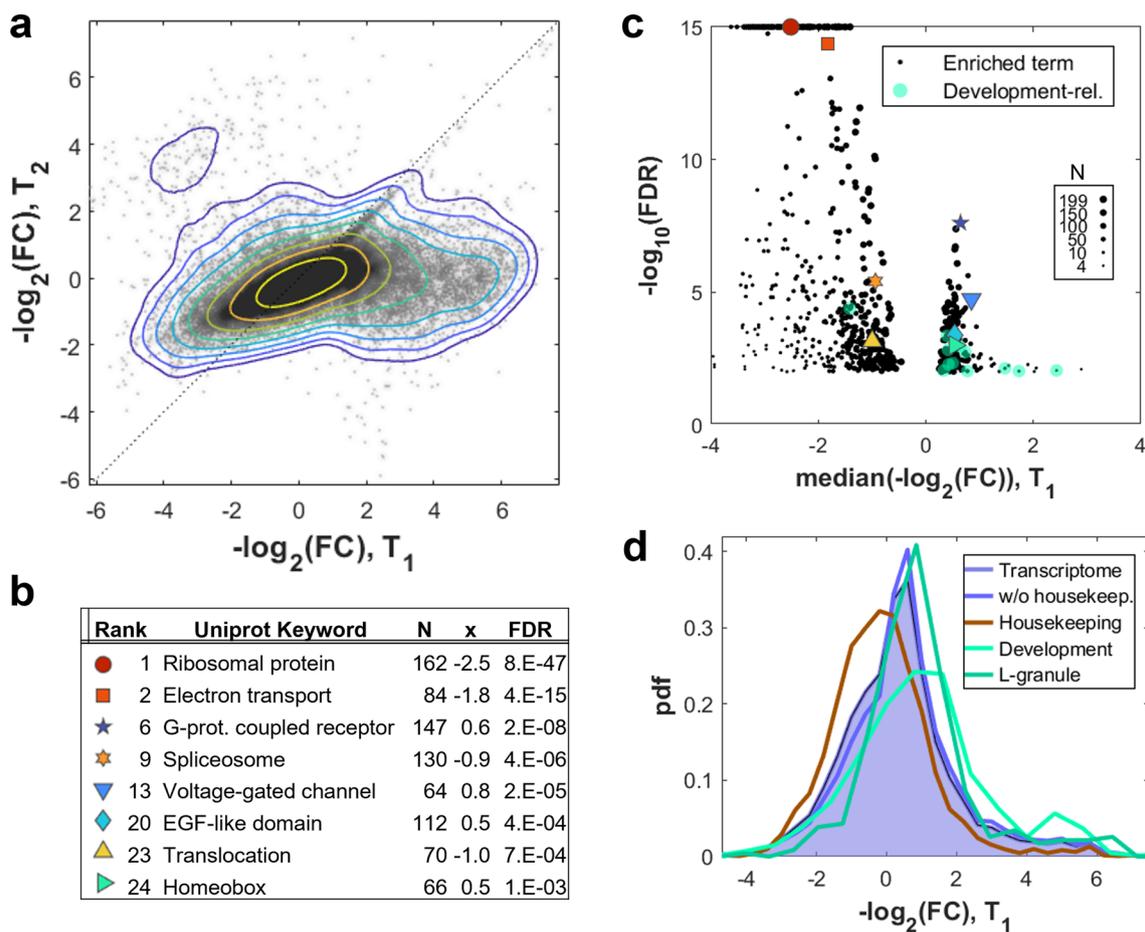
Extended Data Fig. 3 | Sedimentation of diluted cytoplasm suggests partial dissolution of condensates. **a**, Schematic of sedimentation assay. Extract is diluted by a factor f and centrifuged in a hard spin; the top and bottom part are analyzed by mass-spectrometry. For an unbound protein, the top and bottom part have equal concentrations independent of dilution, $c_{top} = c_{bot} = \frac{c_0}{f} := c_{eq}$. By contrast, phase separated proteins sediment in assemblies and the concentration in the bottom part is higher unless they fully dissolve upon dilution. **b**, Scatter plots of sedimentation data for different dilution conditions ($f = 1.2/1.4/2.0$) against the undiluted case ($f=1$). Concentrations are normalized to c_{eq} , derived from the 20% least sedimenting proteins. At the chosen timepoint (200 krcf, 15 min), the ribosome is fully sedimented (magenta, bottom left corner), while the proteasome (magenta cluster, middle) is not. The proteasome indicates facilitated sedimentation for the diluted conditions, however LLPS proteins (green) still exhibit similar c_{top}/c_{eq} . $N = 1$ biological sample. **c**, Box-plot

representation of the data in (b). In all tested dilutions, most LLPS proteins are observed to sediment. While their median starting point suggests sedimentation similar to the proteasome, c_{top}/c_{eq} remains on a similar level, whereas the proteasome sediments further. This (together with (d)) may be explained by the counteracting effects of sedimentation and dissolution. $N = 1$ biological sample, number of proteins per group: 788 (proteome), 35 (proteasome), 11 (LLPS). **d**, At longer centrifugation times (200 krcf, 30 min), an overall higher sedimentation is observed, along with a trend towards equilibration upon higher dilution. Importantly, LLPS proteins are sedimenting even up to 8-fold dilution. $N = 1$ biological sample, number of proteins per group: 844 (proteome), 19 (proteasome), 11 (LLPS). Boxplots (c, d) display data distribution with the center as the median, box limits as quartiles, and whiskers to non-outlier extremes. Source numerical data and proteomics data are provided in Source Data and Supplementary Table 3.



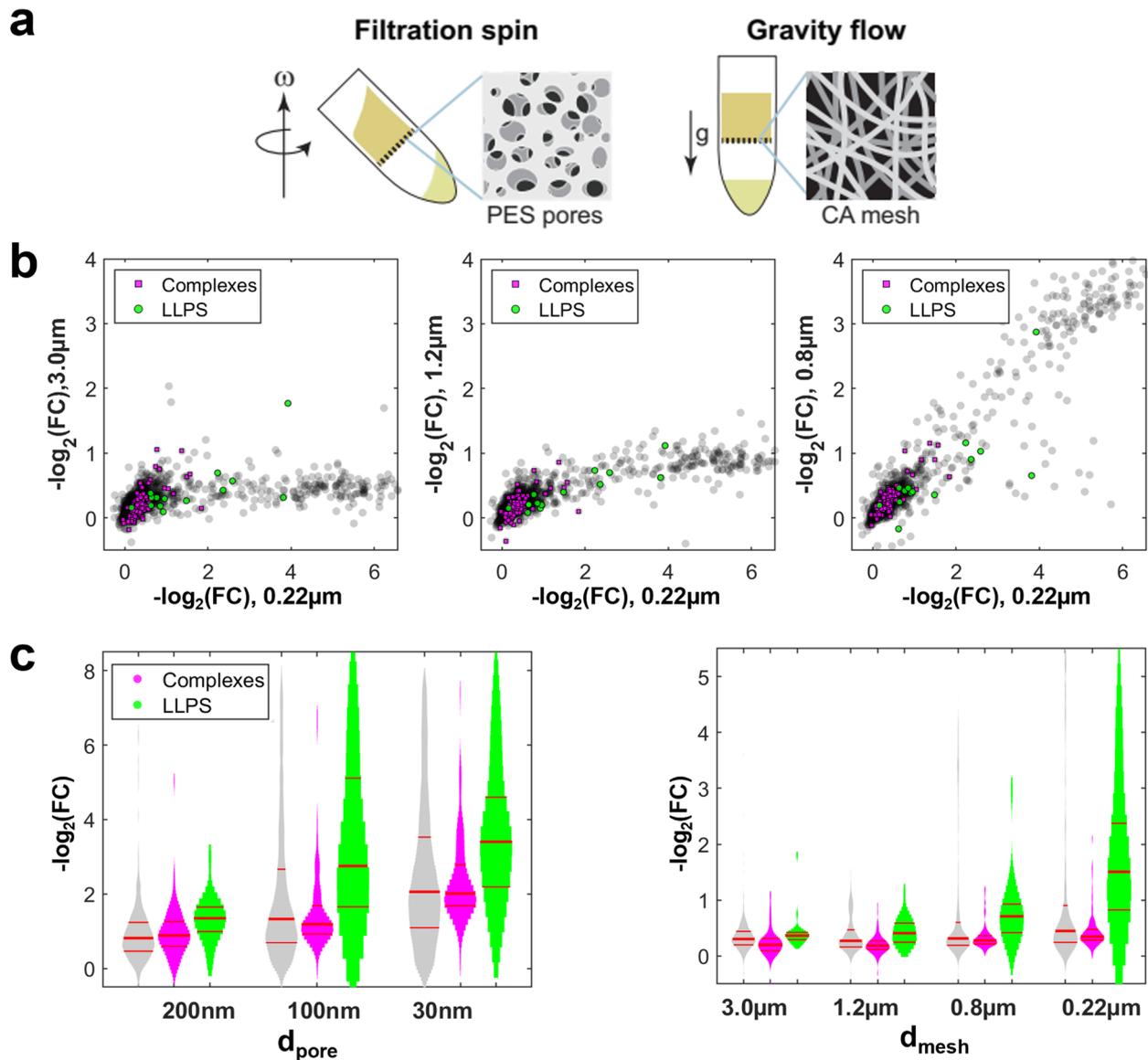
Extended Data Fig. 4 | RNA binding proteins elute later, especially when part of liquid assemblies. Comparison of the flowthrough behavior of the proteome (gray) with RBPs (purple), grouped by their LLPS annotation (w: aqua, w/o: pink). Left column depicts which metric is used in the cumulative histograms for 30 nm (middle) and 100 nm (right) pore diameters. The insets list the p-values (two-sided ks-test) between the groups color-coded by the marker. **a**, In the filtration experiment (see Fig. 1), RBPs exhibit higher retention than the bulk proteome. Notably, only few RBPs pass the filters unhindered. These observations are more pronounced for the LLPS subgroup. FCs were normalized separately for each experiment to their 0.95 quantile. N = 1 biological sample. Proteins per group

3922 (proteome), 941 (RNA-binding), 116 (RNA-b. and LLPS), 825 (RNA-b. w/o LLPS). **b**, The squeezing behavior of most RBPs is slightly greater than that of the average protein. However, the LLPS subgroup shows a much greater shift. N = 1 biological sample. Proteins per group 3922 (proteome), 941 (RNA-binding), 116 (RNA-b. and LLPS), 825 (RNA-b. w/o LLPS). **c**, Filtration of RNase-treated cytoplasm facilitated the overall flowthrough of RBPs. Again, the effect was much stronger on the LLPS subgroup, suggesting disruption of their assemblies. N = 1 biological sample. Proteins per group 1802 (proteome), 489 (RNA-binding), 69 (RNA-b. and LLPS), 420 (RNA-b. w/o LLPS). Source numerical data and proteomics data are provided in Source Data and Supplementary Table 3.



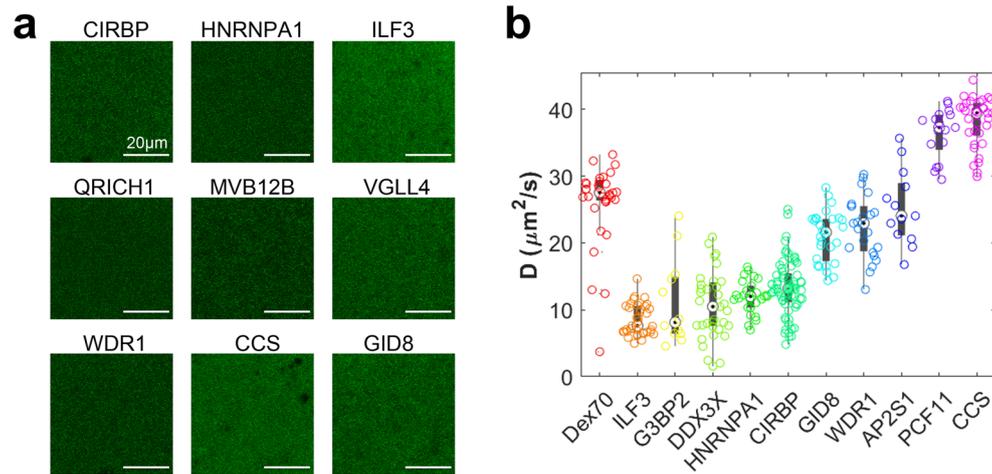
Extended Data Fig. 5 | Transcriptomics of filtrated cytoplasm suggests organization for accessibility. **a**, Scatter plot of the transcriptome for a filtration experiment as in Fig. 1c, comparing the foldchanges to the input of an early and a late time T_1 and T_2 . The majority of RNAs exhibit lightly enhanced flowthrough later on, while about 15% of RNAs are strongly retained at T_1 but become much more abundant at T_2 . The contour denotes the marker density, each line marking a two-fold increase. FC is the ratio between RNA abundance in transcripts per million reads at T_1 and T_2 normalized by the unfiltered condition. $N = 3$ biologically independent replicates. **b**, **c**, Enrichment analysis of gene database terms (STRING⁸¹) for the FC at T_1 . The selection of top UniProt Keywords (b) suggests that constitutively translated mRNAs pass the pores

easily, while mRNAs for transiently translated are more retained. Highlighting development-related terms in the volcano plot of all enriched terms (c) supports this picture. False discovery rates (FDR) calculated by two sided ks-test⁸¹. $N = 3$ biologically independent replicates. **d**, Histogram of filtration retention values for the transcriptome. Housekeeping gene transcripts³¹ are shifted towards easier flowthrough ($p = 3.3e-78$), while the cumulative of all development-related terms from (c) have mRNAs in the retained cluster ($p = 3.9e-29$). Similarly enriched are transcripts which are contained in L-bodies³³ ($p = 2.4e-11$). P-values determined by two-sided t-test. $N = 3$ biologically independent replicates. Source numerical data and transcriptomics data are provided in Source Data and Supplementary Table 3.



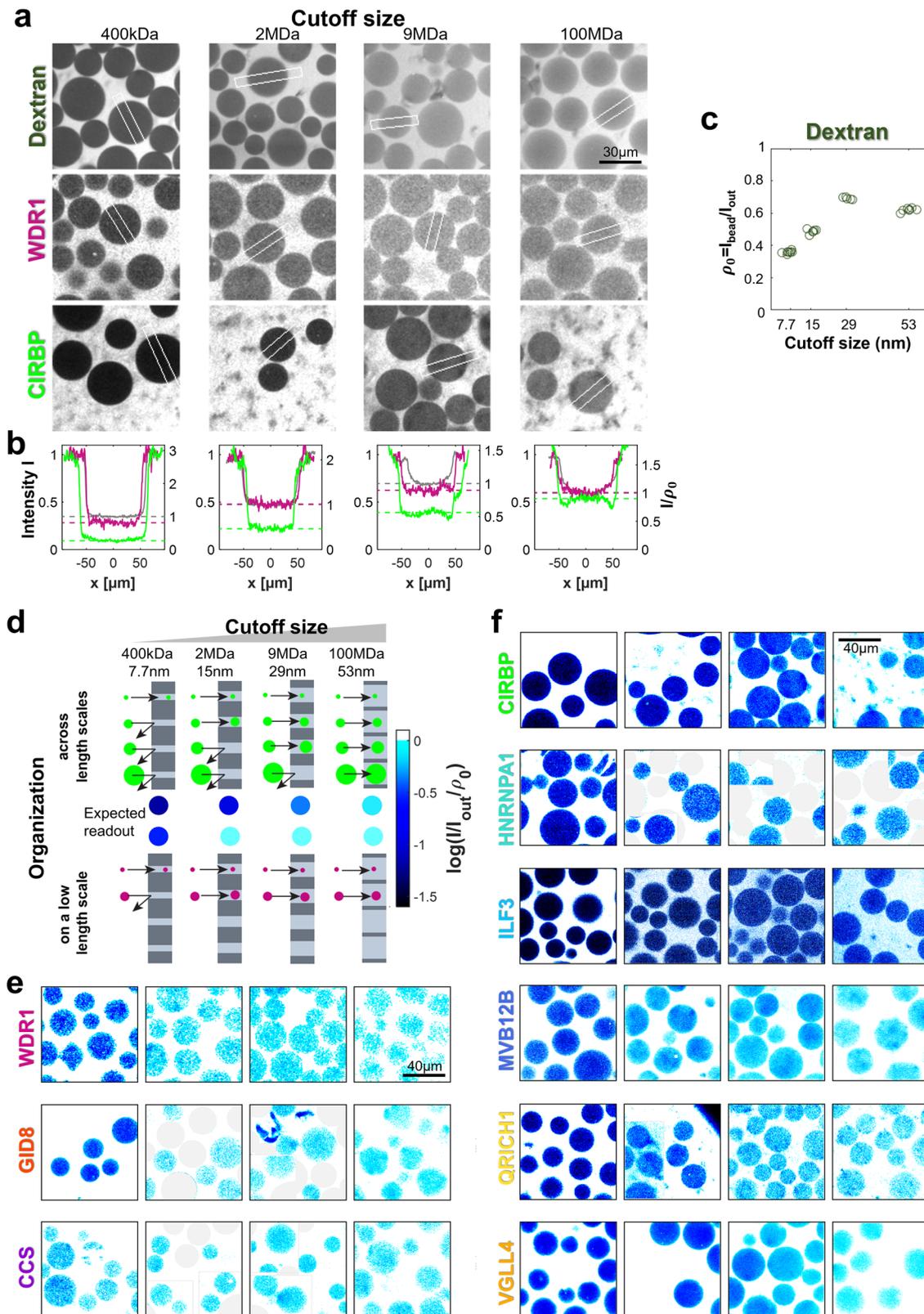
Extended Data Fig. 6 | Pore size dependence of filtration. **a**, Schematics of the spin filtration setup, using polyethersulfone (PES) membranes and the alternative setup gravity flow setup, using cellulose acetate (CA)-mesh filters. The large open area of the CA mesh enables flow through at 1 g force. **b**, Scatter plots of the fold-changes $FC = c/c_0$ in CA gravity flow experiments. Larger meshes (3 μm , 1.2 μm , 0.8 μm ; left, mid, right panel on y-axis) can only resolve few structures compared to a 0.22 μm mesh (x-axis). LLPS proteins are shifted to less permeation. $N = 1$ biological sample. **c**, Permeation histograms of the PES filters at the early elution (from main text) (left) and the CA mesh filters (right). The retention increases with smaller pores or meshes, suggesting assemblies

on the sub-micrometer length scale. This behavior is pronounced for LLPS proteins. Note that pore and mesh sizes d_{pore} and d_{mesh} are stated as the filter cutoff, i.e., particles larger than this size are confidently retained, and thus most pores or meshes are smaller than this size. However, due to the squeezing behavior of assemblies, we cannot determine a precise size from the cutoff. Red lines in the violin plots denote the 0.25, 0.5, 0.75 quantiles. FCs were normalized separately for each experiment to their 0.95 quantile. $N = 1$ biological sample. Source numerical data and proteomics data are provided in Source Data and Supplementary Table 3.



Extended Data Fig. 7 | Exemplary proteins organize below the micrometer scale. **a**, Confocal micrographs of GFP-fused proteins expressed from mRNA in the cell extract. Solutions appear relatively homogenous, and we did not detect structures on the micrometer scale. Lookup tables adjusted individually to 0.35% saturated pixels to enhance contrast. **b**, Diffusion constants D of GFP-fused proteins in extract measured by fluorescence correlation spectroscopy. A 70 kDa dextran-rhodamine serves as a reference. Phase separating proteins (ILF3, G3BP2, DDX3X, HNRNPA1, CIRBP) exhibit low diffusion constants, suggesting the presence of assemblies. The measured values would correspond to assemblies

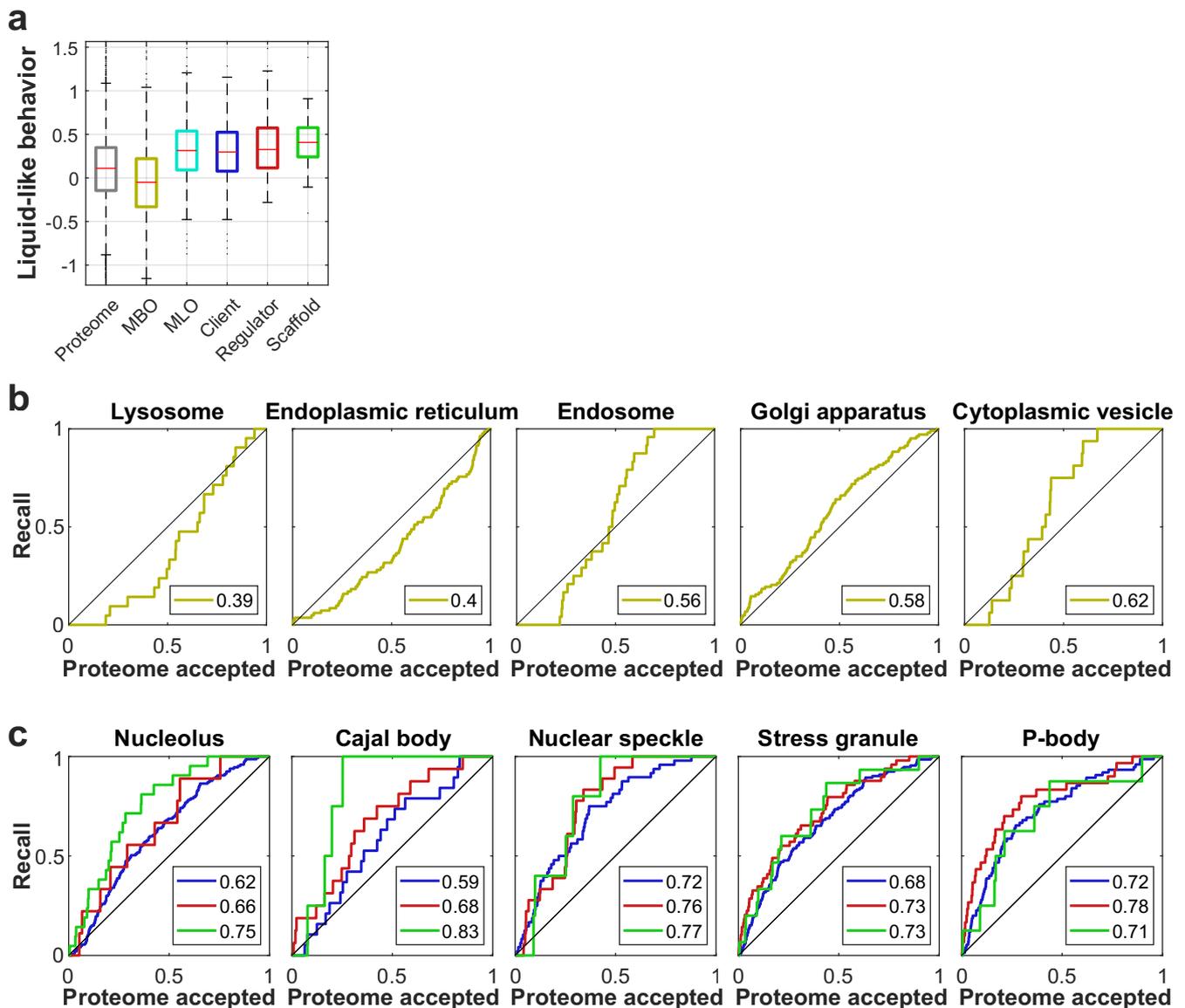
with at least tens to hundreds of monomers, based on a rough estimation by the Einstein-Stokes equation and the dextran reference. Boxplots display data distribution with the center as the median, box limits as quartiles, gray dots as outliers, and whiskers to non-outlier extremes. Dex70 ($N = 2$, $n = 36$), ILF3 ($N = 5$, $n = 36$), G3BP2 ($N = 1$, $n = 12$), HNRNPA1 ($N = 4$, $n = 26$), CIRBP ($N = 4$, $n = 68$), GID8 ($N = 4$, $n = 28$), WDR1 ($N = 4$, $n = 23$), AP2S1 ($N = 3$, $n = 13$), PCF11 ($N = 4$, $n = 23$), CCS ($N = 3$, $n = 33$); where N denotes the number of biological samples, and n denotes the number of FCS measurements. Source numerical data are provided in Source Data and Supplementary Table 3.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Size exclusion assay. **a**, Confocal micrographs of the assay described in the main text illustrating the measurement. Chromatography beads of different cutoff sizes are placed in extracts with GFP-labeled proteins (WDRI and CIRBP) and Dextran 70 kDa-rhodamine solution for the calibration measurement. Intensities are normalized to the outside solution. White rectangles indicate regions of the line plots in panel (b). Intensities are normalized to the outside solutions I_{out} for comparability. **b**, Density comparison of the solutions above. Left y-axis displays the normalized, background corrected fluorescence intensity I . Right y-axis displays the intensity normalized to the dextran density I/ρ_0 , as done to correct for accessible volume. **c**, Measurement of dextran intensity ratios that serve as calibration of the assay. The bright,

homogenous solutions allow for a precise determination of ρ_0 . **d**, Schematic illustrating the estimate of the size distribution from the exclusion of assemblies from beads with different cutoff sizes. If all assemblies can enter a bead, a fill fraction $f = I/(I_{out}\rho_0)$ of 1 (cyan) is expected. Exclusion of assemblies means lower f . We observe organization happening either on the 10 nm scale (jump of f) or spanning across scales (gradual increase f). **e**, **f**, Bead assays for the proteins analyzed in Fig. 4d. Color scale as in panel (d). In case of multiplexed assays, beads with different pore size are masked by the gray circles. Insets are provided to show more beads per image. Source numerical data are provided in Source Data and Supplementary Table 3.



Extended Data Fig. 9 | Membrane-bound organelles do not exhibit the typical filtration behavior of BMCs. **a**, Summary of liquid-like behavior, the integrated result of filtration and dilution experiments ($n = 7$ measurements pooled from $N = 2$ biological independent samples). While both distributions for MBOs and BMCs are wide, they are strongly centered on opposite sides of the scale. Scaffold proteins of BMCs are typically more shifted than co-proteins, termed clients and regulators²⁴. Boxplots display data distribution with the center as the

median, box limits as quartiles, gray dots as outliers, and whiskers to non-outlier extremes. **b, c**, ROCs of organelles. Legends display the AUC, color-code as in (a). (c) Most MBOs -with exemption of the Golgi- are depleted from the most liquid-like region of the proteome. (d) BMCs show a good recall characteristic. Typically, co-proteins have lower AUC, as expected from the picture that scaffolds get populated by them depending on the context. Source numerical data and proteomics data are provided in Source Data and Supplementary Table 3.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Fluorescence microscopy data was acquired on a Nikon A1 confocal microscope using the Nikon Elements software (Nikon). Fluorescence correlation spectroscopy data was acquired using an FCS Upgrade Kit for Laser Scanning Microscopes and the SymPhoTime software (2.0) (PicoQuant).
Proteomics data was collected using Orbitrap Fusion Lumos (ThermoFisher Scientific).
All codes and software used for data collection were described in the Methods section.

Data analysis

Proteomics data analysis was performed using GFY Core version 3.8 (licensed through Harvard Medical School) and Matlab (R2021a). Searches were performed using the SEQUEST algorithm and the *X. laevis* v9.2 genome assembly (Xenbase). RNA-Seq reads were assessed using FastQC (0.10.0), TrimGalore (0.6.10), and STAR (v2.7.10a); and further analyzed using R (v3.5.1) and Matlab (R2021a, Mathworks). Machine learning was performed using Matlab. Imaging data was analyzed using Fiji/ImageJ (NIH) and Matlab.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Mass spectrometry raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029879. Prediction data and used positive training class are provided in the Supplementary Tables. All RNA-seq datasets are deposited in the National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) with the GEO accession number GSE232651.

Assignments of mass spectrometry spectra were searched against the *X. laevis* v9.2 genome assembly (Xenbase (RRID:SCR_003280)1. LLPS database was sourced from PhasePro2, PhaSepDB3, DrLLPS4, LLPSDB5, and the references from the PSAP predictor6; protein complex data from CORUM7. Protein nucleic acid binding was sourced from Uniprot8, QuickGO9, and Castello et. al.10. Organelle data was sourced from DrLLPS4, Uniprot8, and RNPgranuleDB11. Predictions of catGRANULE12, Pscore13 or PSAP6 were reevaluated and plotted. Intrinsically disordered regions, transmembrane helix annotations, and enrichment analysis were generated from Espritz14, Krogh et.al.15, and STRING16 respectively. Housekeeping gene and L-body annotations were derived from Eisenberg et. al.17 and Neil et. al.18. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029879. The raw sequencing data and gene expression matrices have been deposited to the National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) with the GEO Series accession number GSE232651. All other data are available from the corresponding authors upon reasonable request. Source data are provided with this study in Source Data and Supplementary table S3.

- 1 Fortriede, J. D. et al. Xenbase: deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database. *Nucleic Acids Research* 48, D776-D782 (2020).
- 2 Meszaros, B. et al. PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res* 48, D360-D367, doi:10.1093/nar/gkz848 (2020).
- 3 You, K. et al. PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res* 48, D354-D359, doi:10.1093/nar/gkz847 (2020).
- 4 Ning, W. et al. DrLLPS: a data resource of liquid-liquid phase separation in eukaryotes. *Nucleic Acids Res* 48, D288-D295, doi:10.1093/nar/gkz1027 (2020).
- 5 Li, Q. et al. LLPSDB: a database of proteins undergoing liquid-liquid phase separation in vitro. *Nucleic Acids Research* 48, D320-D327, doi:10.1093/nar/gkz778 (2019).
- 6 van Mierlo, G. et al. Predicting protein condensate formation using machine learning. *Cell Rep* 34, 108705, doi:10.1016/j.celrep.2021.108705 (2021).
- 7 Ruepp, A. et al. CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic acids research* 38, D497-501, doi:10.1093/nar/gkp914 (2010).
- 8 UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49, D480-D489, doi:10.1093/nar/gkaa1100 (2021).
- 9 Huntley, R. P. et al. The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* 43, D1057-1063, doi:10.1093/nar/gku1113 (2015).
- 10 Castello, A. et al. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* 63, 696-710, doi:10.1016/j.molcel.2016.06.029 (2016).
- 11 Youn, J. Y. et al. Properties of Stress Granule and P-Body Proteomes. *Mol Cell* 76, 286-294, doi:10.1016/j.molcel.2019.09.014 (2019).
- 12 Bolognesi, B. et al. A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Rep* 16, 222-231, doi:10.1016/j.celrep.2016.05.076 (2016).
- 13 Vernon, R. M. et al. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* 7, doi:10.7554/eLife.31486 (2018).
- 14 Walsh, I., Martin, A. J., Di Domenico, T. & Tosatto, S. C. Espritz: accurate and fast prediction of protein disorder. *Bioinformatics* 28, 503-509, doi:10.1093/bioinformatics/btr682 (2012).
- 15 Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305, 567-580, doi:10.1006/jmbi.2000.4315 (2001).
- 16 Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 51, D638-d646, doi:10.1093/nar/gkac1000 (2023).
- 17 Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends in Genetics* 29, 569-574, doi:https://doi.org/10.1016/j.tig.2013.05.010 (2013).
- 18 Neil, C. R. et al. L-bodies are RNA-protein condensates driving RNA localization in *Xenopus* oocytes. *Molecular Biology of the Cell* 32, ar37 (2021).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications (Liu et al. Bioinformatics, 2014; Nguyen et al., Nature communications, 2022). Replicate samples showed good agreement to an initial screen for experimental conditions, among similar conditions, and to a broad variation of conditions presented in Supplementary Figure 2 (n=18 biologically independent samples over n=89 independent experiments).
Data exclusions	For all the experiments reported in the manuscript, no data was excluded.
Replication	All experiments were independently biological replicated. Proteomics experiments n>=2. Transcriptomics experiments n=3. Microscopy experiments n=3.
Randomization	No randomization was performed. Randomization is not relevant to this study for comparisons within one set of isobaric tags in multiplexed proteomics studies.
Blinding	Data collection and analysis were not performed blind to the conditions of the experiments. Proteomics or transcriptomics samples underwent the same workflow during which they were indistinguishable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involvement in the study	n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants		

Antibodies

Antibodies used	N/A
Validation	N/A

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	N/A
Authentication	N/A
Mycoplasma contamination	N/A
Commonly misidentified lines (See ICLAC register)	N/A

Palaeontology and Archaeology

Specimen provenance	<input type="text" value="N/A"/>
Specimen deposition	<input type="text" value="N/A"/>
Dating methods	<input type="text" value="N/A"/>
<input type="checkbox"/> Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	
Ethics oversight	<input type="text" value="N/A"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	<input type="text" value="Adult Xenopus laevis wild-type females obtained from Nasco/Xenopus1 were used for egg collection to generate extracts. The age range of used laboratory animals is estimated to be between one and five years."/>
Wild animals	<input type="text" value="The study did not involve wild animals."/>
Reporting on sex	<input type="text" value="The study used unfertilized frog eggs."/>
Field-collected samples	<input type="text" value="The study did not involve samples collected from the field."/>
Ethics oversight	<input type="text" value="The study protocol was approved by Princeton University Institutional Animal Care and Use Committee; IACUC 2070, reviewed in March 2023."/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
<input type="checkbox"/>	<input type="checkbox"/>	Public health
<input type="checkbox"/>	<input type="checkbox"/>	National security
<input type="checkbox"/>	<input type="checkbox"/>	Crops and/or livestock
<input type="checkbox"/>	<input type="checkbox"/>	Ecosystems
<input type="checkbox"/>	<input type="checkbox"/>	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
<input type="checkbox"/>	<input type="checkbox"/>	Demonstrate how to render a vaccine ineffective
<input type="checkbox"/>	<input type="checkbox"/>	Confer resistance to therapeutically useful antibiotics or antiviral agents
<input type="checkbox"/>	<input type="checkbox"/>	Enhance the virulence of a pathogen or render a nonpathogen virulent
<input type="checkbox"/>	<input type="checkbox"/>	Increase transmissibility of a pathogen
<input type="checkbox"/>	<input type="checkbox"/>	Alter the host range of a pathogen
<input type="checkbox"/>	<input type="checkbox"/>	Enable evasion of diagnostic/detection modalities
<input type="checkbox"/>	<input type="checkbox"/>	Enable the weaponization of a biological agent or toxin
<input type="checkbox"/>	<input type="checkbox"/>	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.
Files in database submission	Provide a list of all files available in the database submission.
Genome browser session (e.g. UCSC)	Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.
Software	Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

- Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.
Instrument	Identify the instrument used for data collection, specifying make and model number.
Software	Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI

 Used

 Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis:

Whole brain

ROI-based

Both

Statistic type for inference

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See [Eklund et al. 2016](#))

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

- n/a | Involved in the study
- Functional and/or effective connectivity
 - Graph analysis
 - Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.