Linking Biomolecular Condensates to Disease and Therapeutic Development

By

Susana W. Hawken

BA, Computer Science
Williams College, 2018

Submitted to the Program of Computational and Systems Biology in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Computational and Systems Biology

at the

Massachusetts Institute of Technology

September 2023.

Signature of Author……
Computational and Systems Biology Graduate Program
June 21th, 2023

Certified by……
Richard Young
Professor of Biology

Accepted by……...
Christopher Burge
Professor of Biology
Co-Director, Computational and Systems Biology Graduate Program

Linking Biomolecular Condensates to Disease and Therapeutic Development

ABSTRACT

The cell is compartmentalized into membrane-bound and membraneless organelles that organize and regulate key cellular functions. Over the past decade, growing evidence supports the notion that membraneless organelles, called biomolecular condensates, compartmentalize biomolecules – proteins and nucleic acids – involved in shared cellular processes through a biophysical process called phase separation. Biomolecular condensates have distinct physicochemical properties dependent on the molecular features and interactions of constituent biomolecules. Disease-associated mutations in individual biomolecules that compose condensates can alter condensate physicochemical properties. In addition, key drug targets have been identified as components of condensates. This thesis examines biomolecular condensates in disease and therapeutic development. We find that condensate-promoting features in condensate-forming proteins can be mapped and leveraged to build a resource cataloging mutations that likely contribute to condensate dysregulation in human diseases (Banani et al., 2022). Pathogenic mutations in condensate-promoting features span diverse disease classes across both Mendelian diseases and cancers. FDA-approved small molecule therapeutics interact with condensates, selectively partitioning into some condensates and not others (Klein et al., 2020). Selective partitioning of small molecules has broad implications for drug therapeutic activity and resistance. These findings demonstrate the need to integrate condensate-based models in our study of disease and therapeutic development – an effort which will generate novel pathogenic mechanistic hypotheses and improved drug design for human diseases.

I dedicate this thesis to my family,
Laura Wilson, Michael Hawken, and Clara Wilson Hawken

ACKNOWLEDGEMENTS

Scientific discoveries are never the product of one person, so I would like to acknowledge all of the people who have been instrumental in advising, mentoring, contributing and supporting me and the science presented in this thesis.

First, I would like to thank Rick Young, my advisor. Thank you for taking me on as an undergraduate student and for giving me the freedom to grow as a scientist over time. Thank you for being such a great teacher and for always encouraging us to take on new scientific challenges, day in and day out. I learned many lessons in the Young lab – how to be a scientist, a collaborator, and a leader – lessons that I will carry with me throughout my career.

I am forever grateful to the wonderful members of the Young lab who supported me through my PhD and contributed to the science presented in this thesis. Thank you especially to all of my co-authors, specifically Ann Boija, Isaac Klein, Lena Afeyan and Salman Banani. Ann and Isaac, I learned so much from working with both of you – you are always on the cutting edge of science, and you push others to think outside the box. Lena, you are a great scientific teammate. Thank you for your enthusiasm and for your drive to get things done in times of crisis (the pandemic). There are several other members of the Young lab, Alessandra Dall'Agnese, Kalon Overholt, Victoria Clark and Jonathan Henninger, who I have worked with on science not presented in this thesis and whom I would like to acknowledge. Ale, you are truly an inspirational scientist. Your curiosity, innovative thinking, and work ethic has pushed me to want to do better science. Kalon, your structured way of thinking and willingness to take on new ideas transformed the final year of my PhD. Victoria, your versatility as a neurosurgeon, scientist, and mother has changed my thinking as to what is possible for my career. Jon, your passion for basic science has kept me excited about my PhD work throughout my time at MIT. I would like to give a special thank you to Tony Lee and Nancy Hannett who keep the Young lab running. And thank you to the other current and past members of the Young lab who shaped my PhD – Eliot, Alicia, Charlie, Jesse, Ozgur, Ido, Hailey, Catherine, Henry, Sinthu, Hao-Che, Guiseppe, and Ming.

Thank you also to my thesis committee, Laurie Boyer, Seychelle Vos, and Benjamin Sabari. You have provided incredibly valuable insights and guidance throughout my PhD. Thank you to my other scientific mentors, Tony Koleske, Brian Abraham and Daniel Day, who introduced me to research and encouraged me to do my PhD.

I would not have made it through my PhD without my friends. Emma, Kate, Kirsten, and Tiiso, I owe you all so much. Thank you for supporting me.

Thank you, Mom, Dad, and Clara. No need to elaborate.

Hans, I could not have asked for a better partner. Thank you for always encouraging me throughout this process. You make me better, and you make life easier.

STATEMENT OF WORK

Chapter 1

I wrote Chapter 1 myself with suggestions from Rick Young.

Chapter 2

I wrote the manuscript presented in Chapter 2 with my co-authors Salman Banani and Lena Afeyan, as well as Rick Young. In collaboration with Salman Banani and Lena Afeyan, I performed analyses and designed and executed experiments pertaining to all figures in Chapter 2.

Chapter 3

Isaac Klein, Ann Boija, and Rick Young wrote the manuscript presented in Chapter 3. I designed and executed experiments and analyses in Figures 3C, E, and F, Figure 4C, Figure S2, S9A, S13B, C, and D, S16A, B, C and D, and S21A.

Chapter 4

I wrote Chapter 4 myself.

TABLE OF CONTENTS

CHAPTER 1:

Introduction

Overview of thesis

The cell is comprised of billions of biomolecules – carbohydrates, lipids, proteins and nucleic acids – that work together to perform cellular functions. Often, these biomolecules need to be in very close proximity, either in direct physical contact, or in specific cellular locations to properly function. To facilitate this, the cell is organized into compartments, called organelles, that concentrate, compartmentalize, and encapsulate the biomolecules involved in shared functions. Organelles can be membrane-bound, encased by a plasma membrane, or non-membrane-bound (membraneless organelles), lacking a membrane. Membraneless organelles, recently termed biomolecular condensates, have been observed in the cell for almost two centuries, but only in the past two decades have significant strides been made in understanding their assembly and regulation. Studies of biomolecular condensates in pathological cellular states have revealed their dysregulation in disease — galvanizing efforts to rescue these dysregulated states in the hopes of creating new therapeutic approaches. This thesis examines biomolecular condensates in disease and in therapeutic development. In the introduction, I will discuss the history of intracellular organization, and the underlying molecular and physical principles for biomolecular condensate assembly and regulation. I will summarize how the dysregulation of condensates has been heavily implicated in disease, and finally, I will demonstrate how condensate principles can be leveraged for developing better therapeutics.

A physical process called phase separation is the prevailing model by which biomolecular condensates form. In a biological system undergoing phase separation, a well-mixed single-phase system can de-mix to form two separate phases – a dense phase and dilute phase of particular biomolecules. Evidence for phase separation in biology was first described in a study by Brangwynne et al., which showed that P-granules in *Caenohabditis elegans* behave like liquid-like droplets that form in a switch-like fashion, reminiscent of liquid phase separation (Brangwynne et al., 2009). The P granules displayed behaviors that were characteristic of liquids, such as fusing together, dripping off the nucleus, and rapid internal rearrangement. These properties could not have been predicted based solely on studying the individual component biomolecules and their stoichiometric interactions, and thus “emerge” as new properties that are specific to biomolecules as a collection and their super-stoichiometric behaviors. Since 2009, insights from physics, chemistry, computational sciences, microscopy, and molecular biology have contributed to the understanding of the physical properties of condensates – properties with downstream implications for the cellular functions which they regulate (Lyon et al., 2021). In the last decade it has become apparent that biomolecular condensates regulate virtually all cellular processes, including, but not limited to, transcription, RNA synthesis and metabolism, ribosome biogenesis, DNA replication, DNA repair, chromatin organization, protein quality control, innate and adaptive immunity, mitosis, cell-cell adhesions, signaling, and synaptic transmission (Figure 1A) (Banani et al., 2017; Lyon et al., 2021).

Dysregulation of condensates has been associated with multiple diseases, including neurodegenerative diseases, developmental diseases, cancer, viral infections, cardiac, and metabolic diseases (Alberti and Hyman, 2021; Boija et al., 2021; Dall’Agnese et al., 2022; Li et al., 2022; Mensah et al., 2023; Schneider et al., 2020; Spannl et al., 2019). One of the first examples of condensate dysregulation was in amyotrophic lateral sclerosis (ALS), in which an

ALS-associated mutation in the intrinsically disordered region (IDR) of an RNA-binding protein called fused in sarcoma (FUS) was shown to cause an accelerated transition of the condensate from a liquid-like state to a solid-like state (Patel et al., 2015). Following this seminal study, other human genetics studies have linked mutations in condensate-forming proteins to neurodegenerative diseases and several cancers through altered condensate physical properties. However, beyond these case studies, the scope of condensate dysregulation as a pathogenic mechanism in disease was not yet established in early 2022. In Chapter 2, I, along with my co-authors, generate a catalog of pathogenic mutations in condensate-promoting features of putative condensate-forming proteins in disease. We show that these mutations span a broad range of human diseases, biological functions, and protein sequence features. We further validate that the mutations cause condensate dysregulation in cells. This became the first validated map of protein-coding genetic variation associated with condensate dysregulation across human disease.

Condensates have chemical microenvironments that are distinct from the surrounding cellular milieu, allowing for the selective partitioning of biomolecules like proteins and nucleic acids. For small molecule drugs, the physicochemical environment of the drug affects its therapeutic activity. To design drugs that effectively target proteins and nucleic acids in disease, it is important to understand the underlying condensate physicochemical properties in which the drug target resides. In Chapter 3, I and my co-authors investigate how antineoplastic drugs interact with condensates. We find that cancer therapeutics differentially partition in condensates in vitro and in cells and that this behavior affects therapeutic activity. This study paves the way for designing more effective drugs for hard-to-treat diseases.

A history of intracellular organization: From membrane-bound organelles to biomolecular condensates

A cell's organization is critical to its function. Since the 1800s, it has been known that eukaryotic cells are segregated into specific compartments – compartments now known as organelles. The nucleus, which encapsulates and protects the genetic material of the cell, was first observed by Robert Brown in 1833 and further documented by Rudolph Wagner in 1835 (Wagner, 1835; Brown, 1833). These organelles organize and regulate the essential functions of the cell, such as protecting genetic material, protein folding, packaging and trafficking, and generation of metabolic energy. Among the first observations to be made about organelles in eukaryotic cells was that they contain a thin, dense, viscous outer layer surrounding the internal protoplasm, defined by Carl Nägeli in 1855 as the "plasmamembran" (Mullock and Luzio, 2013). This observation was further extended to mammalian cells and by 1925, two Dutch researchers, Evert Gorter and François Grendel described the classical phospholipid bilayer structure of the plasma membrane that we know today (Gorter and Grendel, 1925). As imaging technology advanced from light microscopy to electron microscopy, several more organelles were observed, including mitochondria and the Golgi apparatus, further establishing this class of membrane-bound organelles, in which the function of the plasma membrane is to concentrate, encapsulate, and regulate the biomolecules involved in shared processes (Mullock and Luzio, 2013) (Figure 1).

Many organelles in the cell lack a plasma membrane. Among the earliest organelles in the nucleus to be observed was the nucleolus, which lacks a plasma membrane and is now known to be the site of ribosome biogenesis (Wagner, 1835; Montgomery Jr., 1898; Pederson, 2011). Moreover, the nucleolus has a tripartite structure, and the three components segregate the transcription, processing, and assembly steps of the ribosomal RNAs (rRNAs). Following these early observations, researchers posited that the protoplasm was organized into compartments. E.B Wilson in 1899 described the protoplasm as an "emulsion" containing granules. (Wilson et al., 1899). Throughout the 20th and early 21st century, cytological studies, light microscopy, and

electron microscopy revealed many more examples of non-membrane-bound organelles, including Cajal, Balbiani, and Lewy bodies (Boke and Mitchison, 2016; Cioce and Lamond, 2005; Rodrigues e Silva et al., 2010).

Many of the concepts described in the introduction to this thesis are grounded in colloidal chemistry, which was prevalent throughout the early 20th century, as physicists and biologists alike attempted to understand the structure of the protoplasm, the cytoplasm and nucleus, and its function in the cell. A colloid is a solution that has two phases, a dispersed phase and a continuous phase. The classic example is the difference between a salt water solution and milk. In a salt water solution, the sodium (Na+) and chloride (Cl-) ions are surrounded by water molecules (H_2O) in a dispersed fashion, interacting with each other. In milk, which is a colloid, the colloidal particles are large collections of fat molecules, called globules, rather than individual fat molecules. In 1916, WB Hardy said the following in his paper “Some of the problems of living matter” in reference to the protoplasm: “So we may in general regard the living cell as a mixture of colloidal slimes of varying degrees of fluidity” (Hardy, 1915). He aptly noted that the protoplasm is a combination of fluid and elastic components, alluding to the liquid and gel-like components that we have now dissected in detail (Hardy, 1915). The study of colloidal chemistry and the structure of the protoplasm later fell out of vogue as molecular discoveries in the 1950s and ‘60s shifted biologists’ focus to more mechanistic molecular questions. These observations and theories from scientists in the 1900-1940s inspired biophysicists and molecular biologists to revisit the ideas in the 21st century. However, a general framework for the assembly and regulation of membraneless organelles remained elusive. In 2009, a group at the Max Planck Institute for Molecular Cell Biology led by researcher Anthony Hyman described a new model for membraneless organelles, which we now refer to as biomolecular condensates, that revolutionized our understanding of cellular organization and biological function.

Figure 1

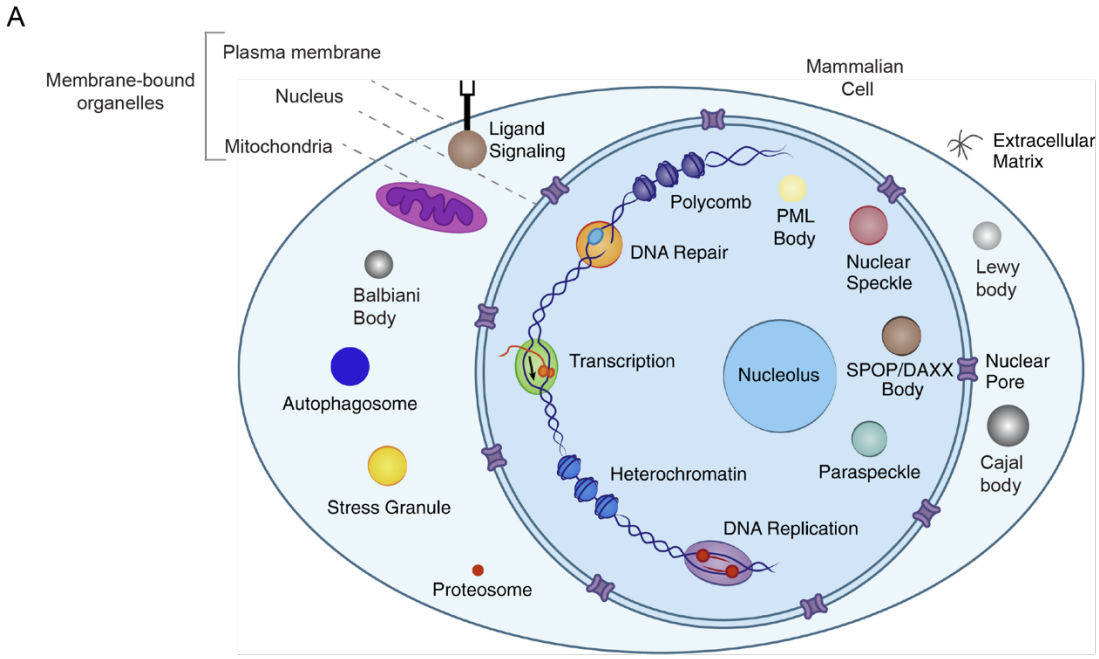


Figure 1. Biomolecular condensates across the cell. (A) Graphic of a mammalian cell with cellular processes that have now been shown to occur within biomolecular condensates labeled across the cell. Adapted from Boija et al, *Cancer Cell* 2020.

Principles of phase separation and biomolecular condensates

The seminal discovery from Cliff Brangwynne in 2009 applied a framework from soft matter and polymer physics to the formation of membraneless organelles. This new framework described membraneless organelles as liquid-like droplets that could form through phase separation. Brangwynne et al. showed that P granules, non-membrane-bound organelles rich with RNA and RNA-binding proteins, behaved like liquid droplets in the cell – rapidly dissolving and condensing to migrate from posterior to anterior of the cell during germ cell specification (Brangwynne et al., 2009). To show this, Brangwynne et al. labeled P granule protein components with green fluorescent protein (GFP) and monitored their movement using three-dimensional (3D) particle tracking. P granules transitioned from a soluble form to a spherical condensed phase and fused to one another, consistent with the behavior of simple liquids.

To further test the hypothesis that P granules' physical characteristics were similar to that of liquids, they applied shear force to the P granules to determine if the stress would induce flows rather than a constant deformed shape, which is characteristic of elastic liquids. Indeed, the granules exhibited flow-like behavior, including dripping, fusing, and wetting. They showed that the liquid-like behavior was in part due to rapid internal molecular rearrangement by photobleaching portions of the P granules and measuring the time it took for the granules to recover. Lastly, they were able to quantify the P granule surface tension, which increased linearly with size, as is expected for liquids. Given these results, they suggested that P granules were liquid-like droplets that could form through liquid phase separation (Figure 2).

Since the study in 2009, many labs and researchers have applied ideas from soft matter physics and polymer/colloidal chemistry to the study of biomolecular condensates in cells, with considerable efforts focused on understanding phase separation in biology. Phase separation offers a thermodynamics-based model by which biomolecular condensates form; as such, phase separation is relevant to this thesis because it has guided the study of the sequence features and molecular interactions that are characteristic of condensate formation – features and interactions that are crucial in linking condensates to disease and therapeutic development.

The prevailing model by which condensates form is through liquid-liquid phase separation (LLPS). Liquid phase separation is the process by which a solution can go from one well-mixed system to a de-mixed system, separating into two phases – a dilute and dense phase (Figure 2A). This mode of formation allows for biomolecular condensates to form in the absence of a membrane.

The theory for biomolecular phase separation comes from polymer physics, as described by Flory-Huggins Theory for Polymer Solution and others (Cohen and Benedek, 1982; Flory and Krigbaum, 1951; Huggins, 1942). At the core of the theory of phase separation is the concept of mixing and de-mixing of solutions, and how thermodynamics can describe behaviors of simple polymer solutions. When two polymers are mixed, they tend to evolve towards states of the highest entropy – a measure of disorder of the system. Typically, the state of highest entropy is the mixed state, however, if one of the polymers favors interactions with itself over interactions with other molecules, the energy reduction of homotypic interactions (enthalpy-driven) outweighs the entropy-driven mixed state of heterotypic interactions (Figure 2B). This is called the de-mixed state. In the case of biological systems, the polymers can be proteins, nucleic acids, or other macromolecules. This type of system is concentration-dependent, meaning that the system will favor a de-mixed state, with a dense and dilute phase, above a critical concentration threshold (C_{sat}). Given the molecular interactions that drive phase separation are also dependent on parameters such as temperature, pH, and salt concentration, varying both concentration and another parameter can produce a binodal phase diagram that describes the conditions in which the system is in the 1-phase (mixed) or 2-phase (de-mixed) regime.

Phase separation models assume that the system of interest is at equilibrium, however, cellular processes that occur within condensates are typically active, non-equilibrium processes, so applying phase separation principles in cells is challenging. Therefore, there is still a question of whether condensates in cells form through super-stoichiometric phase separation process or through a stoichiometric macromolecular assembly, in which biomolecules bind to a static, high valency scaffold molecule that is limited by binding site availability and affinity (Lyon et al., 2021). Some have argued that phase separation is not required to describe the behaviors and functions of condensates in cells (McSwiggen et al., 2019; Trojanowski et al., 2022). It is true that these two models are challenging to deconvolve in a cellular context, however, *in vitro* and cellular studies have aided in providing evidence for the switch-like behavior of phase separation correlated with large changes in functional output of a process associated with condensates. More work will have to be done to distinguish between these two models. However, despite the debate over how condensates are formed, the natural question still becomes – what are the molecular interactions between proteins and nucleic acids that favor condensate formation?

Figure 2

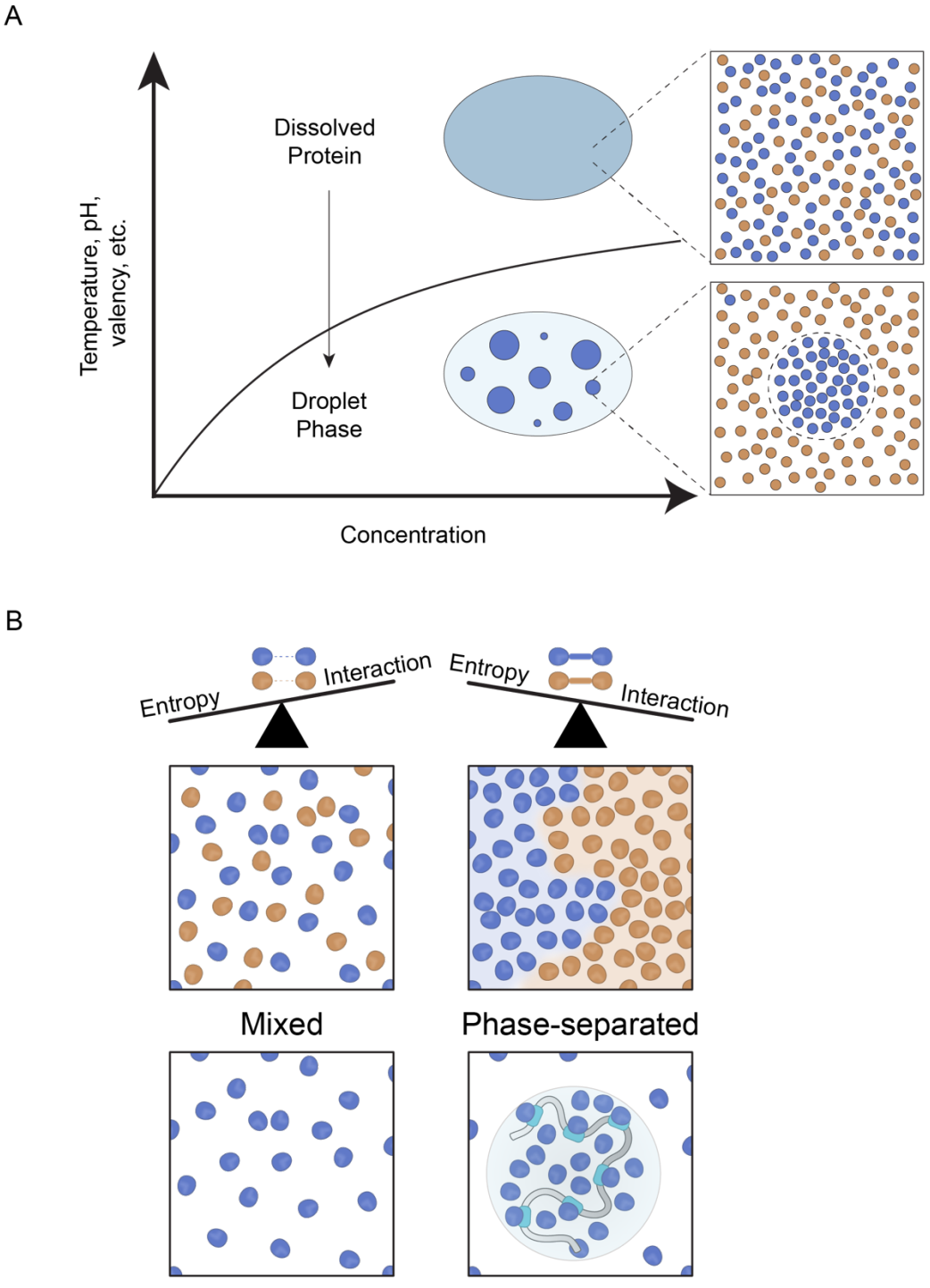


Figure 2. Principles of phase separation. (A) Phase diagram where the two regions of the diagram describe the conditions in which a solution inhabits a dilute or dense phase. Adapted from Brangwynne et al, J Cell Bio 2013. (B) Graphic displaying the thermodynamic forces favored in a mixed (left) and phase-

separated (right) solution. Interaction between biomolecules is energetically favorable in a phase-separated system.

Molecular features involved in condensate assembly: Condensate-promoting features

Molecular-scale interactions are crucial for condensate assembly. Interactions involving specific amino acid types, protein domains, and sequence and structural patterns of proteins have been shown to drive phase separation and condensate formation. In addition, various models for how condensates are nucleated and maintained have been proposed. In this section, I will discuss the features and interactions that have thus far been described for driving condensate assembly, the biophysical frameworks for condensate formation, and the current state of our ability to computationally predict condensate-forming biomolecules. Computational prediction of molecular features involved in condensate formation is important for developing the large-scale approaches to linking condensates and disease described in Chapter 2 of this thesis.

Multivalency: The critical features of biomolecules that drive phase separation are valency, concentration, and affinity as described in *Principles of Polymer Chemistry* by P.J. Flory in 1953 (Figure 3A) (Flory, 1953). The valency of the molecules is directly related to the concentration needed to induce a phase separating system. Multivalent molecules are highly prevalent in diverse cellular processes and drive phase transitions across the cell (Banani et al., 2017, 2016; Banjade and Rosen, 2014; Li et al., 2012; Nott et al., 2015). One of the first descriptions of this principle in a biological system was in an engineered *in vitro* system examining multivalent SH3 domains interacting with disordered proline-rich motifs (PRMs) (Li et al., 2012). Molecules with 2 or more SH3 and PRM domains showed liquid-like droplet behavior at concentrations that were correlated with the valency of the component molecules *in vitro*. Switching one of the multivalent molecules (either SH3 or PRM) with a monovalent version blocked the ability to phase separate. In cells, ectopic expression of multivalent proteins is sufficient for the formation of membraneless bodies and their assembly can drive processes like actin-polymerization—displaying the power of this principle in a functional context (Banjade and Rosen, 2014; Chung et al., 2011; Li et al., 2012).

Condensates contain 100s to 1000s of biomolecules, some of which are multivalent and essential to the condensate integrity, some which are not. From this observation, a framework emerges which entails a multivalent *scaffold* biomolecules (protein, RNA or DNA) that forms and structures the condensates and *client* molecules that can partition in and interact with sites on the scaffold molecules, but are non-essential to the condensate integrity (Banani et al., 2016).

Intrinsically disordered regions (IDRs): Multivalency can be achieved through a series of folded protein domains, such as SH3 domains, or by intrinsically disordered regions (IDRs) – regions of proteins that lack a defined three-dimensional structure (Figure 3A). The classic “lock-and-key”, stoichiometric interactions that are characteristic of many molecular interactions are typically absent in interactions involving IDRs (Wright and Dyson, 2015). Instead, weaker, transient interactions are involved in IDR interactions. IDRs are high valency molecules, in that many regions along the IDR can interact with multiple other biomolecules in a non-stoichiometric fashion. IDRs are therefore drivers of condensate formation through the multivalent interactions of particular amino acids (Banani et al., 2017). Concepts have emerged to describe behaviors of multivalent IDRs in condensates. One is from the study of associative polymers in which the protein polymers themselves contain regions, “stickers”, that specifically interact with binding partners and “spacers”, which are the regions interspersed between the sticker to provide the flexibility for multiple interactions (Figure 3)(Choi et al., 2020; Martin et al., 2020).

Nucleic acids: Nucleic acids are another central component of condensate formation, function, and regulation (Figure 3A) (Roden and Gladfelter, 2021). Several of the well-studied condensates are rich with RNA and RNA binding proteins and some even depend on RNA as the central scaffold of the condensate (Guillén-Boixet et al., 2020; Langdon et al., 2018; Maharana et al., 2018; Zhang et al., 2015). For example, the long non-coding RNA *NEAT1* is the scaffold RNA for paraspeckles – condensate in the nucleus involved in RNA processing (Clemson et al., 2009; Yamazaki et al., 2018). *NEAT1* has multiple binding sites for RNA binding proteins, which allows them to concentrate in regions where *NEAT1* is localized. Properties such as RNA concentration, sequence, length, structure, and modifications are vital for the regulation of condensates (Roden and Gladfelter, 2021; Yamazaki et al., 2018). In transcriptional condensates, nascent RNA production acts as a feedback mechanism for transcriptional output (Henninger et al., 2021). In addition to RNA, DNA also plays a central role in condensate formation and function. One such function is in its role as a crowing agent for protein factors that bind DNA and are involved in gene activation, called transcription factors (TFs). Transcription factors have structured DNA binding domains that recognize specific DNA sequence motifs, and these motifs are clustered in larger regulatory regions called enhancers. Above a specific threshold number of TF DNA binding sites on a piece of DNA, TFs undergo condensate formation and drive gene activation (Sabari, 2020; Shrinivas et al., 2019). Below the threshold number of TF DNA binding sites, TFs remain largely soluble. Therefore, multivalent DNA binding sites act as an agent to crowd TF molecules in close proximity in order to drive condensate formation.

Molecular grammar: Specific chemical interactions are characteristic of condensate-forming biomolecules. Interactions involving aromatic, hydrophobic, and electrostatic amino acids can undergo hydrophobic, π-π, π-cation, polar, and electrostatic interactions (Figure 3B) (Wang et al., 2018). These amino acid subtypes are enriched in IDRs and drive what is called selective partitioning (van der Lee et al., 2014). I will touch more on selective partitioning in the section *Physical properties of condensates*, however the basic concept is that specific condensates selectively partition some biomolecules while excluding others. For example, heterochromatin condensates, associated with inactive genes, partition proteins that regulate heterochromatin, whereas transcriptional condensates associated with active genes selectively partition positive regulators of transcription. In the cell, these condensates are spatially separated, and when mixing key scaffold proteins for these condensates *in vitro*, MeCp2 and BRD4, these proteins form de-mixed droplets driven, presumably, by the immiscibility of their chemical features (Li et al., 2020).

A long-standing question in the field surrounds what specific features of biomolecules provide this sort of specificity for selective partitioning. It has been proposed, and now shown, that blocks of alternating positive and negative charged residues, called charged blocks, can confer specificity for proteins to favor one condensate over another (Figure 3C) (Lyons et al., 2023). Specifically, for transcriptional condensates at active genes, positive regulators of transcription, such as coactivators, positive elongation factors, and transcription factors have been shown to concentrate together, whereas negative regulators of transcription are excluded (Boija et al., 2018; Cho et al., 2018; Sabari et al., 2018). Lyons et al. showed that positive regulators RNA Polymerase II (RNA Pol II) and components of the elongation complex, CTR9 and SPT6, partition into coactivator condensates, as marked by a subunit of the Mediator complex, MED1, *in vitro* and in cells. In contrast, negative elongation factor, NELF, and heterochromatic protein, HP1α, were excluded from MED1 condensates *in vitro* and in cells. This selective partitioning behavior was mediated by patches of positive and negatively charged amino acids that were present in RNA Pol II, CTR9, and SPT6 IDRs, but absent in NELF and HP1a IDRs. Charge blocks are therefore another molecular feature that allows for specific, functional condensate assembly.

More work will need to be done to determine other general rules that govern condensate specificity and selective partitioning.

Figure 3

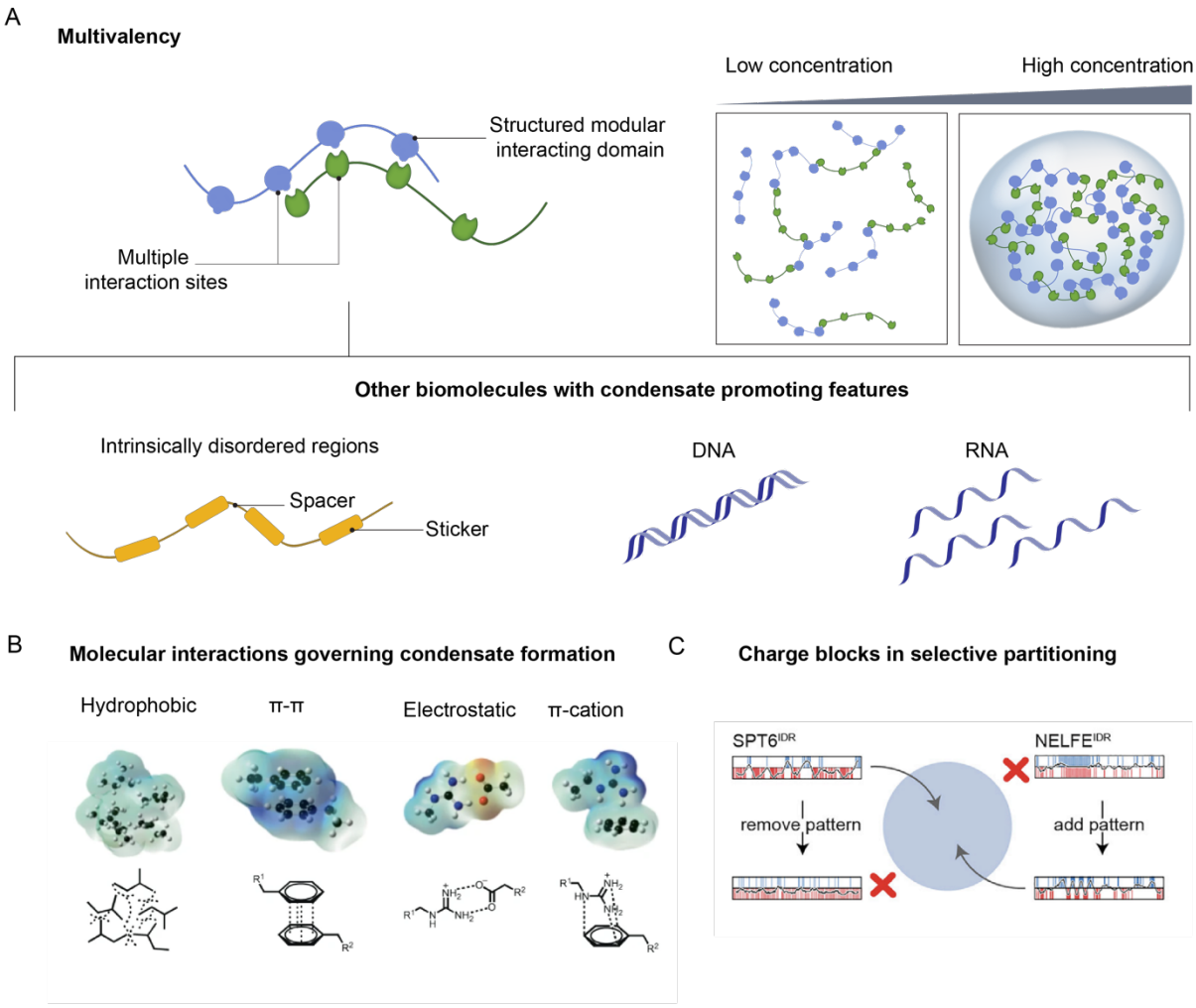


Figure 3: Condensate promoting-features. (A) Multivalency of biomolecules drives condensate formation. Structured domain, IDRs, DNA, and RNA can all participate in multivalent interactions. Adapted from Mitrea et al., *Nature Reviews Drug Discovery* 2022. (B) Interaction types shown to drive condensate formation of biomolecules. Adapted from Boija et al, *Cancer Cell* 2020. (C) Specific patterns of charged amino acid patches drive selective partitioning and downstream functions of transcriptional condensates. Figure 5U, Lyons et al, *Cell* 2023.

Computational approaches for identifying molecular features involved in condensate formation

Experimental studies are vital for uncovering the molecular sequence features of proteins involved in condensate formation, however, many of these studies are limited to individual cases of proteins or condensates. There are ~20,000 proteins in the human proteome that are associated with diverse cellular processes – processes that have now been shown to occur within condensates. How do we identify the proteins and protein features involved in condensate formation and regulation? Several computational tools have been developed to identify features involved in condensate formation and to test the strength of various features in accurately predicting condensate-forming proteins.

Intrinsic disorder and IDR predictors: One such approach is simply testing the amount of intrinsic protein disorder, which in the past has been a useful predictor of the phase separation capacity of a protein. IDR predictors include DisProt, Espritz, and IUPred (Erdős et al., 2021; Vucetic et al., 2005). These predictors combine experimental evidence from various structural techniques, including X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), circular dichroism (CD), and others, and manually curate annotations of intrinsically disordered regions. There are also meta-IDR predictors, D2P2 and MobiDB, which combine the predictions from multiple manually curated databases and neural network predictors (Oates et al., 2013; Piovesan et al., 2022). Neural network predictors, such as the PONDR algorithms, rely on chemical characteristics of amino acids in the primary sequences of proteins to make predictions of which regions are likely disordered. To do this, the PONDR algorithms use feedforward neural networks that identify the sequence attributes within 9-21 amino acids stretches in the protein primary amino acid sequence. The stretches of amino acids are then broken down into parameters such as proportion of specific amino acids, net charge, and sequence complexity. The values of these features are used to train the neural network during predictor construction and used as input during testing.

Tools for identifying molecule features involved in condensate formation: As discussed in the *Molecular Features Involved in Condensate Assembly* section of the introduction, specific amino acid subtypes – aromatic, hydrophobic, and electrostatic amino acids – can undergo interactions that underlie condensate formation and function – hydrophobic, π-π, π-cation, polar, and electrostatic interactions. Computational tools, PLAAC, PScore and CIDER, have been developed to identify a subset of these features. PLAAC identifies prion-like domains, low complexity sequences first characterized in RNA binding proteins. *PScore* identifies π-interacting residues on proteins. CIDER can be used to measure the propensity of a protein to partake in electrostatic interactions by measuring the net charge per residue of a protein using a sliding window approach (Bolognesi et al., 2016; Lancaster et al., 2014; Paiz et al., 2021; van Mierlo et al., 2021; Vernon et al., 2018). As more condensate-promoting features are uncovered through experimental and analytical studies, computational tools such as these can be leveraged for their proteome-wide identification.

In the work presented in Chapter 2, I and my co-authors use a combination of IDR predictors and tools for identifying molecules features to map the regions across the proteome that are likely involved in condensate formation and in which disease mutations therein may result in condensate dysregulation.

Physical properties of condensates

Based on the principles described in the theory of phase separation, the molecular interactions amongst the biomolecules that compose condensates dictate condensate physical properties. All

of the molecular features described in previous sections of this introduction have an influence over how condensates behave and how this behavior affects the downstream condensate-associated cellular processes. In this section, I will describe the physical properties of liquids, gels, and solids that have been shown to be important for our understanding of condensate physical properties thus far. In the following section, I will discuss the emergent behaviors of condensates that arise from the underlying physical principles and how these properties/behaviors can regulate certain biological functions.

Condensates were described in 2009 as protein droplets in the cell with specific physical properties that are characteristic of liquids. Liquids are a subclass of fluids and typically have a measurable surface tension and viscosity. Surface tension is the tendency of a liquid at rest to shrink to minimum surface area and is a result of the attractive forces between component molecules. Viscosity is the measure of a liquid's resistance to flow and is a result of the internal diffusion of molecules. For liquids made up of proteins and nucleic acids, like biomolecular condensates, the surface tension and viscosity are controlled by the sequence features and interactions between molecules.

Gels are also fluids, but have slightly different physical properties than liquids, and one of their defining properties is elasticity. Liquids, even those that are highly viscous, have no memory of their previous shape, meaning that when forces are applied to the liquid it is easily deformed and does not return to its original conformation. A gel, on the other hand, has elastic properties that allows it to have a memory of its previous shape. Many condensates are not pure liquids, but instead have both viscous and elastic (viscoelastic) properties.

Since surface tension, elasticity, and rapid internal exchange of molecules are physical properties of liquids or other fluids, these properties do not apply to solids in the same manner as liquids or gels. However, in a biological context, condensates have been shown to be more "solid-like" meaning they may have measurable surface tension or viscoelasticity, but that in relation to other condensates they tend to behave more like solids.

Experimentally testing physical properties of condensates

One widely used technique to measure the physical properties of condensates, such as viscosity, is fluorescence recovery after photobleaching (FRAP). FRAP is a microscopy technique that involves fluorescently tagging proteins or nucleic acids and measuring the exchange rate between the labeled molecules. This can be done by photo-bleaching the fluorescent molecules in a region of interest (ROI) inside the condensate using a high-power laser and measuring the time it takes for fluorescence signal to recover within the region. More viscous condensates will have a slower recovery than less viscous condensates. The diffusivity of the molecules can be gleaned via the Stokes-Einstein relation, where viscosity is inferred from the Brownian motion of the particles within it (Einstein, 1905). Therefore, FRAP is a powerful tool for measuring physical properties of liquid condensates and the dynamic behavior of the molecules that comprise them. However, extracting the diffusivity of the component molecules relies on the assumption that the condensate of interest is a pure liquid and that the diffusing particle is a rigid sphere, and this method can be more complicated given condensates are often not pure liquids but also exhibit an elastic component (Taylor et al., 2019). Other methods, such as single particle tracking (SPT) can take into account both the viscous and elastic component – methods which will not be discussed in this thesis (Mason et al., 1997).

Emergent behaviors of condensates

The behaviors of condensates that arise from the underlying physical properties can be considered “emergent” properties given the behaviors of the condensates cannot be gleaned from the behaviors of the individual component molecules. Properties of biomolecules have often been studied at the stoichiometric level, however, condensates allow for thousands of molecules to be interacting with each other, resulting in emergent physicochemical behaviors. In this section, I will discuss the behaviors of condensates that emerge through their formation via phase separation and physical properties as liquid-like, gel-like, and solid-like entities.

Selective partitioning: A key emergent behavior of condensates is the ability to selectively partition biomolecules. Selective partitioning behavior is dependent on the physicochemical properties of the biomolecules as well as the physicochemical microenvironment of the condensate (Figure 5A). Changes to the condensate microenvironment, either via post-translational modifications (PTMs) or mutation of the condensate components, can alter selective partitioning behavior.

Coalescence/fusion: Another key behavior of liquid droplets is their ability to fuse or coalesce. Coalescence occurs when two condensates come in close contact with each other, form a capillary bridge, thus joining the condensates into one liquid droplet (Figure 4C). The time it takes for the droplets to coalesce is correlated with the droplet surface tension as well as the viscosity (Alshareedah et al., 2021).

Coarsening: Coarsening is yet another behavior of liquids and is the process by which condensates merge together. Specifically, Oswald ripening is the phenomena in which smaller condensates in a system start to dissolve and deposit material into larger condensates – causing them to grow (Figure 4D).

Interfacial behaviors: Multiple liquids driven by different intramolecular interactions can exhibit specific behaviors when they are mixed in solution. Two liquids can be miscible, meaning that their components can mix in solution, allowing for fusion or coarsening behaviors between separate liquid droplets, however some liquids are immiscible, and this immiscibility results in behaviors at the intersection of where the liquids interact, called interfacial behaviors (Figure 4B). In the cell, condensates exhibit interfacial behaviors such as complete and partial wetting, dripping, and pickering (Gouveia et al., 2022).

Metastability: Liquids can be metastable, meaning they can inhabit intermediate energy states that are not the lowest energy state. Given the dynamic nature of condensates, they can also rapidly inhabit multiple intermediate energy states, transitioning from soluble molecules to liquid-like states to solid-like states under different conditions (Figure 4E).

Figure 4

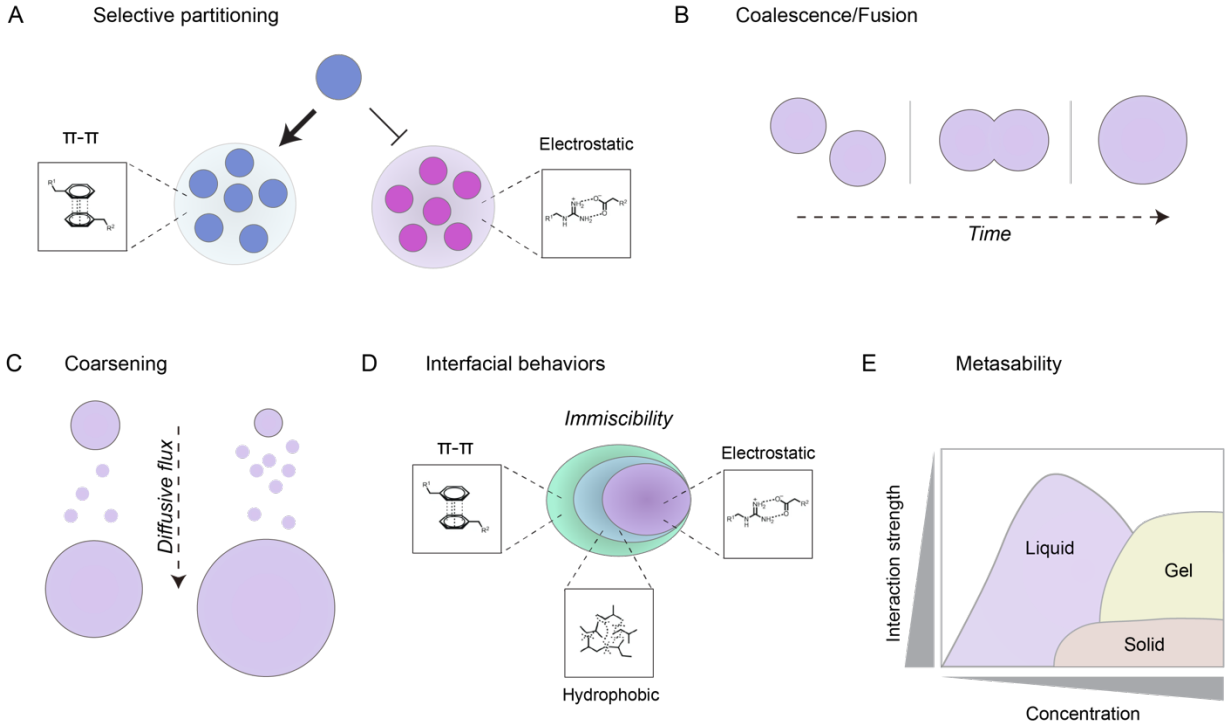


Figure 4. Emergent behaviors of condensates. (A) Biomolecules selectively partition into certain condensates based on their physicochemical properties. Blue protein favors Π-Π interactions. (B) Graphical depiction of condensates fusing over time. (C) Graphical depiction of Oswald ripening. Larger condensates can steal material from smaller condensates over time. (D) Immiscible liquids can form interfaces at the interaction where the liquids interact. (E) Plot depicting metastability. Conditions can dictate the physical state of condensates.

Condensate functions

Emergent behaviors of condensates have implications for the cellular functions that condensates regulate.

One of the emergent behaviors of condensates is selective partitioning, the inclusion or exclusion of specific biomolecules from condensates based on their physicochemical properties. For enzymatic reactions, condensates can aid in increasing reaction rates via mass action by concentrating enzymes with their substrates, excluding enzymatic inhibitors, and limiting enzyme/substrate diffusion space (Figure 5A) (Laflamme and Mekhail, 2020; Lyons et al., 2023; O'Flynn and Mittag, 2021; Prouteau and Loewith, 2018; Smith et al., 1992; Su et al., 2016). For example, in innate immunity, the DNA sensor, cyclic guanosine monophosphate adenosine-monophosphate (cGAS), forms cytoplasmic condensates. DNA induces the phase separation

capacity of cGAS, driven by the multivalency of cGAS IDR interacting with DNA. Longer, higher valency DNA is better able to induce phase separation. Loss of phase separation correlates with loss of the cGAS activity in converting ATP to GTP. Moreover, the condensation capacity of cGAS and the specific partitioning of its substrate, DNA, is important for its catalytic activity – a principle which has been ascribed to other enzyme-substrate interactions (Du and Chen, 2018).

Recent studies show that coalescence, interfacial behaviors and surface tension of condensates can exert force to drive mechanical work in the cell (Figure 5B) (Bergeron-Sandoval et al., 2021; Shin et al., 2018). Chromatin organization requires shaping the chromatin polymer to bring genomic loci into close proximity and exclude other genomic loci. Using an optogenetic system to induce condensate formation on the genome, condensates can restructure the genome by separating euchromatin and heterochromatin condensates through immiscibility, and can pull together certain genomic loci through condensate coalescence (Shin et al., 2018).

Coarsening is an emergent behavior of condensates and can regulate important functions in the cell. For example, the stochastic nature of gene expression causes noisy protein expression, which can significantly alter downstream processes that are dependent on protein concentration (Klosin et al., 2020). However, despite this noise, cellular processes are both precise and robust in time and space. Theoretical proposals have been made that phase-separated condensates are a mechanism utilized by the cell to buffer against noise in protein expression levels, doing so by concentrating superfluous protein material into the condensed fraction, while leaving the concentration in the surrounding milieu (dilute fraction) the same (Figure 5C) (Klosin et al., 2020; Stoeger et al., 2016). This allows for continued precision and robustness of cellular processes.

Interfacial behaviors of condensates drive multi-compartmental organization of processes such as ribosome biogenesis and genome architecture (Figure 5D). The nucleolus is a multiphase liquid condensate organized into distinct sub-compartments that correspond to the steps of ribosome biogenesis: the fibrillar center (FC), the dense fibrillar center (DFC), and the granular component (GC) (Feric et al., 2016; Lafontaine et al., 2021). The FC is responsible for rRNA transcription, the DFC for initial ribosomal subunit maturation steps, and the GC is the site of the final steps of ribosomal subunit maturation. Evidence suggests that these functional compartments are in fact coexisting immiscible condensates. Moreover, their immiscibility is driven by the chemical makeup of the specific proteins that scaffold the compartments. Fibrillarin (FIB1) is a scaffold protein of the DFC and nucleophosmin (NPM1) is a scaffold for the granular component. When the two wildtype proteins are combined *in vitro* they recapitulate the structure of the nucleolus, with NPM1 condensates encapsulating FIB1, akin to the GC encapsulating the DFC and FC. Further, the arginine/glycine (R/G) regions of FIB1 and the acidic regions of NPM1 were necessary for this immiscibility behavior. This showed for the first time that the phase separation capacity of immiscible liquids could drive the functional organization of membraneless organelles in cells. A similar principle has been shown for genome organization, in which euchromatin and heterochromatin proteins have chemical features that drive their liquid immiscibility (Misteli, 2020).

The sensitivity of phase separating systems to transition from soluble to liquid-like to solid-like in response to parameters such as concentration, interaction strength, temperature, pH, and salt concentration has led to the proposal that condensates allow for cells to rapidly respond to environmental changes (Figure 5E) (Yoo et al., 2019). One such example is the cells' observed response to non-lethal changes in temperature. Heat-shock proteins form cytoplasmic granules in response to heat that are thought to have regulatory and protective functions. A study showed the protein poly(A) binding protein (PABPC1) undergoes phase separation in cells and *in vitro* in response to thermal stress (Riback et al., 2017). In yeast, this protein formed condensates at

heat-shock temperatures, a process driven by the Pab1 RNA binding domain. Pab1's rate of condensate formation was correlated with an increase in temperature. Taken together, Pab1's thermosensitive phase separation capacity is linked to its function to sense and rapidly respond to thermal stress.

Figure 5

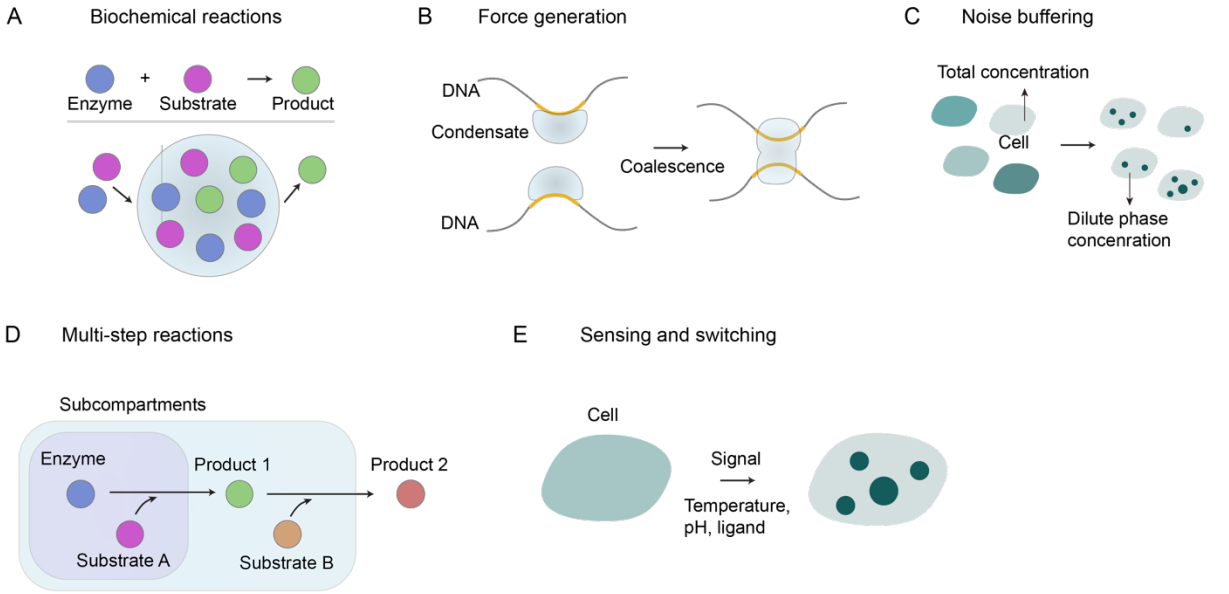


Figure 5. Physical properties and condensate functions. (A) Selective partitioning of biomolecules enhances enzymatic reactions. (B) Condensate coalescence/fusion brings together genomic loci (C) Condensate coarsening aids in noisy protein expression fluctuations. (D) Interfacial behaviors between immiscible liquids aids in regulating multi-step interactions. (E) Metastable liquids allow for rapid responses to cellular signals.

Condensate regulation

Cellular environment, post translational modifications, condensate compositional changes, and protein structural changes can all regulate condensate formation and function.

Chemical changes to condensate components shift the underlying interactions involved in condensate formation and can dissolve condensates, alter condensate composition, and initiate condensate formation. Post translational modifications, such as phosphorylation, methylation, acetylation, glycosylation, palmitoylation, SUMOylation, PARylation (poly-ADP-ribosylation), and ubiquitination are one way the cell regulates condensates through modification of individual components (Snead and Gladfelter, 2019). For example, during transcriptional initiation, RNA Polymerase II forms transcriptional initiation condensates with DNA, RNA, and other transcriptional regulators to initiate transcription. The unstructured C terminal domain of RNA Pol II is initially in a hypo-phosphorylated state, and upon phosphorylation, RNA Pol II favors partitioning into splicing condensates and interacts with elongation factors and the pre-mRNA to promote elongation. The negatively charged phosphate group on Pol II alters the underlying intermolecular interactions that promotes selective partitioning of Pol II into splicing condensates

over initiation condensates (Guo et al., 2019). This is just one example of how post-translational modifications can alter condensate composition. Post-translational modifications have also been shown to dissolve condensates during mitosis (Rai et al., 2018).

Protein, RNA, and metabolite composition of condensates alters their physical properties (Elbaum-Garfinkle et al., 2015; Henninger et al., 2021; Langdon et al., 2018; Woodruff et al., 2017; Zhang et al., 2015). For example, ATP acts as a biological hydrotrope in cells, maintaining the solubility of hydrophobic protein species in liquid droplets both *in vitro* and in cells (Patel et al., 2017). Changes to ATP concentration may contribute to protein aggregation in neurodegenerative diseases.

Ligand-binding that results in a conformational change in proteins can trigger condensate formation for protein scaffolds (Guillén-Boixet et al., 2020; Mitrea et al., 2018). In particular, the stress granule protein G3BP1 binds RNA via an arginine-rich region, competing for binding with its own acidic region, resulting in its ability multivalently interact with more protein and RNA molecules to form stress granules (Guillén-Boixet et al., 2020). Other conformational changes are possible within the chemical environment of a condensate. For example, IDRs can fold upon binding to a protein partner when in close proximity. These higher affinity binding events can alter the physical properties of the condensate, such as viscosity or surface tension (Conicella et al., 2020; Murray et al., 2017).

Condensate dysregulation in disease

Condensate alterations are an important aspect of normal regulation of cellular functions, however they are also involved in disease. In pathological cellular states associated with disease, condensate dysregulation through mutations in component molecules can result in downstream disruption of normal cellular functions. Condensate dysregulation is associated with neurodegeneration, developmental diseases, cancer, cardiac and metabolic diseases, and viral infections (Figure 6A). In this section, I will describe the studies and models by which pathological condensates are implicated in these disease classes. Further, it has been proposed that condensate dysregulation is a potential missing mechanism for interpretation of mutations in a broad range of diseases, spanning thousands of disease mutations, proteins, and cellular functions (Tsang et al., 2020). However, the full scope of genetic variation that is potentially associated with condensate dysregulation was not yet known at the start of my Ph.D. In Chapter 2, I will show that regions of proteins that are involved in condensate formation, termed condensate-promoting features, can serve as a link between pathogenic genetic variation and condensate dysregulation across human disease.

Figure 6

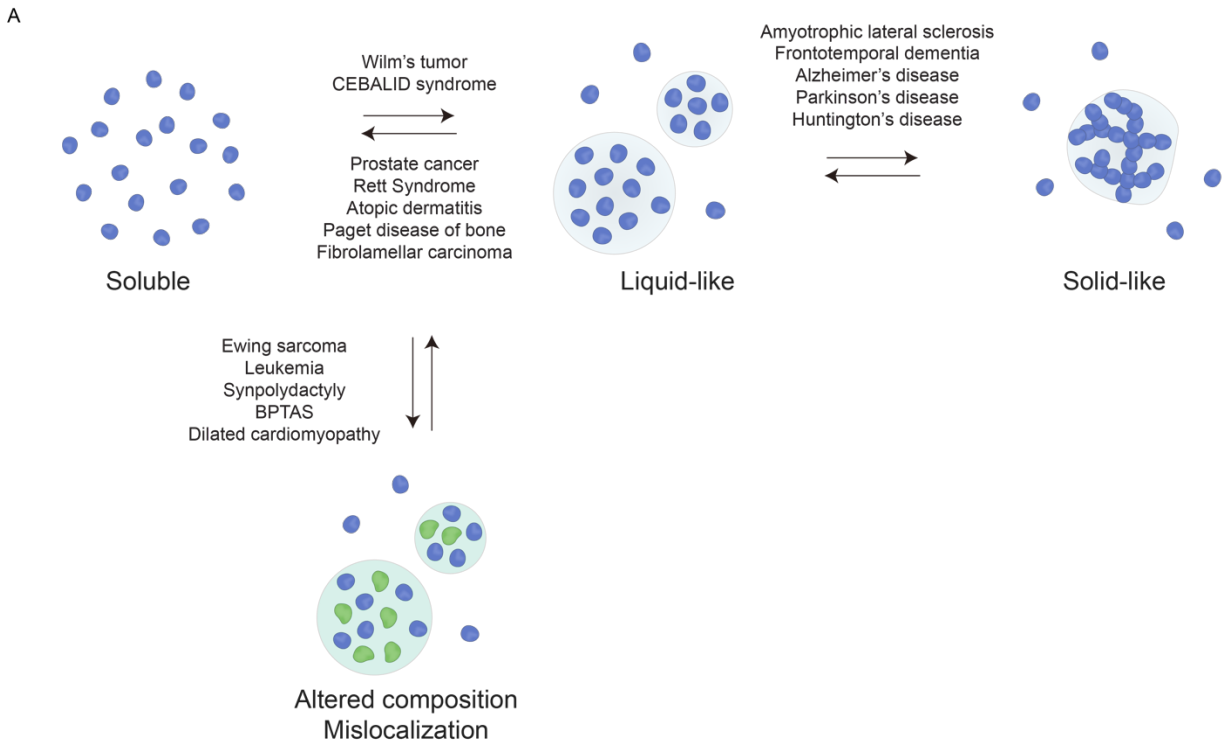


Figure 6. Condensate dysregulation in disease. (A) Changes to condensate physical properties are associated with myriad diseases. Adapted from Alberti et al, *Nat Rev Mol Cell Bio* 2021.

Neurodegenerative diseases

Neurodegeneration was one of the first classes of diseases to be associated with pathological condensates. Aggregates of cytoplasmic RNA binding proteins are a hallmark of neurodegenerative diseases such as frontotemporal dementia (FTD), amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), Alzheimer's disease (AD), and Huntington's disease (HD). Proteins found within these protein aggregates, such as FUS, TDP-43, tau, and α-synuclein, and Huntingtin, respectively, undergo phase separation *in vitro* and condensate formation in cells, dependent on specific amino acid sequence features. These sequence features are targeted by genetic mutations associated with the diseases. This led to the hypothesis that aberrant phase separation results in protein aggregates, and that this could contribute to our understanding of the underlying pathology of neurodegenerative diseases – diseases that still lack a therapeutic treatment. *In vitro*, the liquid-like droplets formed by wildtype versions of the proteins FUS, TDP43, tau and a-synuclein "age" over time and form fibrils that lack liquid-like behavior and are instead more solid-like (Conicella et al., 2020; Hofweber et al., 2018; Murray et al., 2017; Ray et al., 2020; Wegmann et al., 2018). Disease associated mutations in these proteins accelerate the transition from liquid-like to solid-like structures, called a liquid-to-solid transition.

Developmental diseases

Rare, inherited mutations can cause developmental disorders, many of which have no known treatments. Condensate dysregulation is now associated with several developmental disorders that span mutation types and affected tissues. For example, a nucleotide repeat expansion mutation in a transcription factor (TF), HOXD13, which causes hereditary synpolydactily, disrupts the ability for HOXD13 to concentrate in transcriptional condensates with co-activators to drive normal transcriptional programs (Basu et al., 2020). In an X-linked neurodevelopmental disorder called Rett syndrome, a nonsense mutation that truncates the IDR of a heterochromatin regulator, MeCp2, causes increased solubility of MeCp2 and loss of incorporation in heterochromatin condensates (Li et al., 2020). De novo frameshift mutations in HMGB1 that cause brachyphalangy, polydactyly and tibial aplasia syndrome (BPTAS), result in loss of HMGB1 nucleolar partitioning and concurrent effects on rRNA biogenesis, linking nucleolar dysfunction broadly to this rare disease (Mensah et al., 2023). Several other case studies have linked disorders such as CEBALID syndrome and Kabuki syndrome to condensate dysregulation as well (Fasciani et al., 2020; Miyake et al., 2020).

Cancer

Cancer associated mutations can prevent normal condensation behavior of proteins causing aberrant protein function. For example, the tumor suppressor Speckle-type POZ protein (SPOP) is an E3 ubiquitin ligase that is responsible for maintaining normal protein levels in cells through classical degradation pathways. One of its substrates is the protein DAXX. SPOP-DAXX binding was shown to promote condensation in healthy cells, but cancer-associated mutations in SPOP outside of the DAXX binding domain limited its condensate capacity and also its ability to bind DAXX (Bouchard et al., 2018).

Translocations resulting in gene-gene fusion mutations have been associated with condensate formation in areas of the cell they do not normally form. For example, the gene-gene fusion EWS-FLI associated with Ewing Sarcoma, which is a result of a translocation resulting in the DNA binding domain of the transcription factor FLI1 being fused to the unstructured region of EWSR1 containing prion-like domains (PrLD) that have the ability to phase separate *in vitro*. This gene-gene fusion protein is then able to bind acquired microsatellites in cancer cells to drive inappropriate oncogene expression (Boulay et al., 2017; Chong et al., 2018). Other fusions, specifically with the NUP98 IDR which contains phenylalanine-glycine (GC) motifs have also been shown to drive ectopic formation of condensates in pediatric leukemias (Chandra et al., 2022; Michmerhuizen et al., 2020).

In Wilm's tumor, a gain-of-function insertion mutation in a chromatin binding protein, ENL, can enhance condensate formation and dwell time on chromatin to drive oncogenic transcriptional programs, leading to increased oncogene expression and a tendency towards a premalignant cell fate (Wan et al., 2020).

Cardiac diseases

Ribonucleoprotein (RNP) granules are condensates comprised of RNA molecules and RNA binding proteins that regulate various processes such as splicing, RNA metabolism, and stress response. As described in previous sections, mutations in RNA binding proteins that comprise RNP granules have been associated with neurodegenerative diseases. Dysregulation of RNP granules through mutation in RNA binding proteins has also been associated with congenital

dilated cardiomyopathy (DCM). Specifically, a pathogenic point mutation, R636S, in an RNA-binding motif protein-20 (RBM20) has been associated with severe DCM (Brauch et al., 2009). Wild-type RBM20 is involved in splicing regulation and is typically localized in the nuclear splicing speckles in cardiomyocytes, however, the mutant RBM20 cause mislocalization of RMB20 RNP granules to the cytoplasm which was observed in both gene-edited pigs for the mutations as well as human DCM patient samples (Schneider et al., 2020). Mislocalized RNP granules may therefore be a pathogenic mechanism in DCM.

Metabolic diseases

Metabolic diseases stem from cells being unable to metabolize biomolecules properly. These diseases continue to be a growing public health issue, as the prevalence of diabetes and obesity continue to rise a staggering rate. Type 2 diabetes manifests when the body is unable to process insulin properly – resulting in a resistance to insulin signaling. The insulin receptor (IR) translocates into the cytoplasm and nucleus to promote insulin signaling. Recent evidence has shown that in normal, insulin sensitive hepatocytes, IR forms dynamic cytoplasmic and nuclear condensates in response to insulin treatment that aid in its signaling functions (Dall'Agnese et al., 2022). In insulin resistant cells, the IR condensate formation is attenuated, and the condensates that do form are less dynamic – IR molecules within condensates lose mobility and kinase function (Dall'Agnese et al., 2022). These findings implicate the loss of liquid-like properties of IR signaling condensates in insulin resistance and Type 2 diabetes.

Viral infections

Some of the world's largest public health issues include infectious diseases, specifically viruses. For instance, viral outbreaks such as HIV and SARS-CoV-2 have killed millions of people and disrupted society for years to come. Understanding the underlying mechanisms of viral infection and lifecycle are critical for developing therapies to treat these diseases. Many viral proteins that are shared amongst viral species and are critical for replication in host cells, contain intrinsically disordered regions that have implicated phase separation in the regulation of the viral life cycle. Indeed, the Sars-CoV-2 nucleocapsid protein has been shown to undergo phase separation *in vitro* and form dynamic condensates in cells that aid in transcription of its genome (Cascarina and Ross, 2022; Wang et al., 2021).

Targeting condensates in therapeutic development

Given the association of condensate dysregulation and disease and that key drug targets have been shown to reside in condensates, targeting condensates has been proposed as a new mechanism for therapeutic intervention (Kilgore and Young, 2022; Kilgore et al., 2023, Mitrea et al., 2022). In this section, I will cover a case study in which reversing altered condensates in acute promyelocytic leukemia (APL) has been proposed as the one of the therapeutic mechanisms for an FDA-approved curative treatment. I will also cover the academic efforts to modulate condensates as a therapeutic strategy. In Chapter 3, I will show that clinically important cancer therapeutics selectively concentrate into some condensates and not others *in vitro* and in cells, and that this selective partitioning behavior has implications for drug activity, efficacy and drug resistance.

Dysregulated PML bodies, APL, and Retinoic acid/Arsenic treatment

Acute promyelocytic leukemia (APL) is a cancer associated with dysregulated condensates – specifically PML bodies. PML bodies are nuclear condensates formed by the tumor suppressor promyelocytic leukemia protein (PML) and condensates formed by PML are involved in p53 activation and promoting cell senescence. APL is caused by a translocation mutation that results in a fusion protein consisting of PML and retinoic acid receptor alpha (RARA) (de Thé et al., 1990). PML/RARA fusions drive leukemogenesis is two key ways, by dysregulating transcription of RARA and PML target genes and by dysregulation of PML nuclear condensates (de Thé and Chen, 2010; dos Santos et al., 2013). Retinoic acid (RA) and arsenic (As) treatment are highly effective in combination in treating APL patients – exhibiting 100% complete remission in Phase III trials (Lo-Coco et al., 2013). RA binds to RARA and As to PML, and this binding results in transcriptional reactivation of target genes, reactivation of p53 signaling, and fusion protein degradation that restores cell differentiation and reduces the rate of senescence (Jeanne et al., 2010; Zhang et al., 2010). Further, these downstream anti-tumor effects of RA/As has been associated with reformation of wildtype (WT) PML nuclear condensates. Specifically, RA/As treatment results in degradation of PML/RARA fusion protein and As binding promotes WT PML condensate reformation (Jeanne et al., 2010; Niwa-Kawakita et al., 2017; Zhang et al., 2010). Overall, these insights into APL curative treatments (RA/As) associated with restoration of normal PML condensate formation is a case study of the potential for condensate therapeutic intervention in disease. To further strengthen the association between condensate restoration and clinical efficacy, mutations in the arsenic binding site of PML are present in As treatment resistant patients. In a cellular model, resistance mutations in PML induce dysregulated PML bodies that remain dysregulated upon arsenic treatment (Jeanne et al., 2010). This finding further strengthens the case for linking condensates as an avenue for therapeutic intervention and warrants studying whether restoration or modulation of dysregulated condensates might be a broader mechanism to be utilized for therapies across disease.

Modulating condensates with small molecules

Condensate liquid-to-solid transitions drive protein aggregate formation in various neurodegenerative diseases. A potential therapeutic strategy would be to prevent the liquid-to-solid transition using a therapeutic approach. Indeed, a small molecule called lipoamide can prevent the liquid-to-solid transition of stress granules in a *C. elegans* ALS model (Wheeler et al., 2019). In cultured motor neurons with a patient-derived mutation in the FUS protein, which is a stress granule component associated with ALS, lipoamide was able to prevent dieback of axons, and in a *Drosophila melanogaster* model of ALS, lipoamide was able to rescue motor defects (Wheeler et al., 2019).

Another study showed that a small molecule, 4,4′-dianilino-1,1′-binaphthyl-5,5′-disulfonic acid (bis-ANS), could modulate TDP43 condensates, another stress granule protein associated with ALS (Babinchak et al., 2020). Bis-ANS is a negatively charged small molecule that at low concentrations, promotes TDP43 droplet formation, and at high concentrations disrupts condensate formation through repulsive interactions between the negatively charged molecules – causing a re-entrant phase separation phenomenon (Babinchak et al., 2020). A similar behavior has been shown when using varying concentrations of negatively charged RNA molecules in protein droplets. Therefore, disrupting phase transitions of stress granules using various small molecules may be a viable therapeutic hypothesis for ALS.

Conclusion

In this introduction I have detailed how the study of membraneless organeless, called biomolecular condensates, has drastically changed the model for intracellular organization and

regulation of normal cellular functions. Quantitative biophysical frameworks, like phase separation, from the fields of soft matter physics and polymer chemistry are useful models to determine the formation, regulation, function, and physical properties of biomolecular condensates. In cells, the molecular interactions that drive the formation of condensates have been identified, and mutations within these features are associated with a number of human diseases. Further, condensate dysregulation phenotypes observed in disease may be reversible through interactions between drugs and condensates. In Chapters 2 and 3, I present the work that I have done to use insights from condensate biology to better understand human disease and to improve therapeutic development.

Figure 7

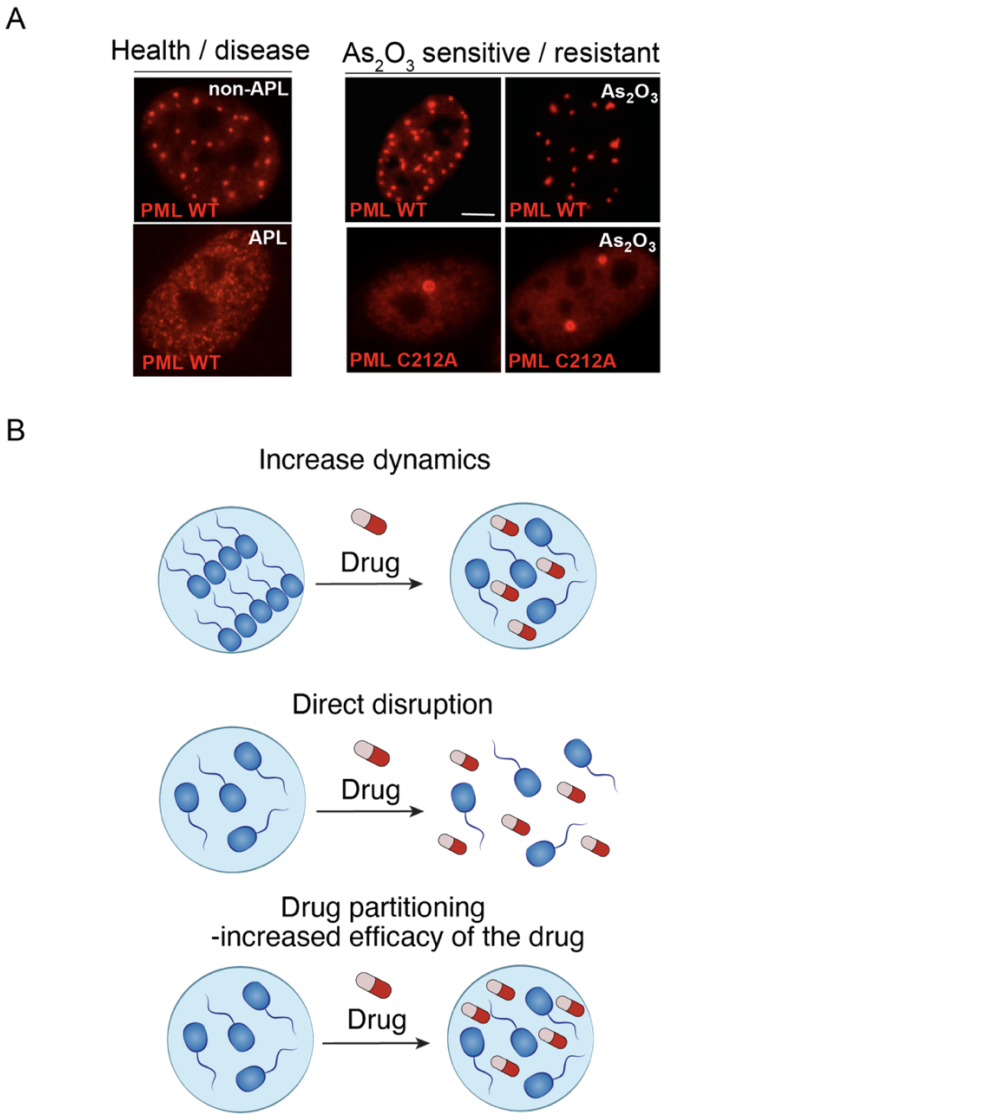


Figure 7. Drug behaviors and condensates. (A) PML body morphology in APL and non-APL cellular models (Left) and in As_2O_3 sensitive and resistant models. Figure 1A and Figure 3B, De Thé et al, *Cancer Cell* 2017. (B) Effects of drugs on condensates. Adapted from Boija et al, *Cancer Cell* 2020.

References

Alberti, S., Hyman, A.A., 2021. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. Nat Rev Mol Cell Biol 22, 196–213. https://doi.org/10.1038/s41580-020-00326-6

Alshareedah, I., Kaur, T., Banerjee, P.R., 2021. Chapter Six - Methods for characterizing the material properties of biomolecular condensates, in: Keating, C.D. (Ed.), Methods in Enzymology, Liquid-Liquid Phase Coexistence and Membraneless Organelles. Academic Press, pp. 143–183. https://doi.org/10.1016/bs.mie.2020.06.009

Babinchak, W.M., Dumm, B.K., Venus, S., Boyko, S., Putnam, A.A., Jankowsky, E., Surewicz, W.K., 2020. Small molecules as potent biphasic modulators of protein liquid-liquid phase separation. Nat Commun 11, 5574. https://doi.org/10.1038/s41467-020-19211-z

Banani, S.F., Afeyan, L.K., Hawken, S.W., Henninger, J.E., Dall'Agnese, A., Clark, V.E., Platt, J.M., Oksuz, O., Hannett, N.M., Sagi, I., Lee, T.I., Young, R.A., 2022. Genetic variation associated with condensate dysregulation in disease. Developmental Cell 57, 1776-1788.e8. https://doi.org/10.1016/j.devcel.2022.06.010

Banani, S.F., Lee, H.O., Hyman, A.A., Rosen, M.K., 2017. Biomolecular condensates: Organizers of cellular biochemistry. Nat Rev Mol Cell Biol 18, 285–298. https://doi.org/10.1038/nrm.2017.7

Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., Rosen, M.K., 2016. Compositional control of phase-separated cellular bodies. Cell 166, 651–663. https://doi.org/10.1016/j.cell.2016.06.010

Banjade, S., Rosen, M.K., 2014. Phase transitions of multivalent proteins can promote clustering of membrane receptors. eLife 3, e04123. https://doi.org/10.7554/eLife.04123

Basu, S., Mackowiak, S.D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., Geertsema, H., Ali, S., Jerković, I., Ewers, H., Mundlos, S., Meissner, A., Ibrahim, D.M., Hnisz, D., 2020. Unblending of transcriptional condensates in human repeat expansion disease. Cell 181, 1062-1079.e30. https://doi.org/10.1016/j.cell.2020.04.018

Bergeron-Sandoval, L.-P., Kumar, S., Heris, H.K., Chang, C.L.A., Cornell, C.E., Keller, S.L., François, P., Hendricks, A.G., Ehrlicher, A.J., Pappu, R.V., Michnick, S.W., 2021. Endocytic proteins with prion-like domains form viscoelastic condensates that enable membrane remodeling. Proceedings of the National Academy of Sciences 118, e2113789118. https://doi.org/10.1073/pnas.2113789118

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., Abraham, B.J., Afeyan, L.K., Guo, Y.E., Rimel, J.K., Fant, C.B., Schuijers, J., Lee, T.I., Taatjes, D.J., Young, R.A., 2018. Transcription factors activate genes through the phase-separation capacity of their activation domains. Cell 175, 1842–1855. https://doi.org/10.1016/j.cell.2018.10.042

Boija, A., Klein, I.A., Young, R.A., 2021. Biomolecular condensates and cancer. Cancer Cell 39, 174–192. https://doi.org/10.1016/j.ccell.2020.12.003

Boke, E., Mitchison, T.J., 2016. The balbiani body and the concept of physiological amyloids. Cell Cycle 16, 153–154. https://doi.org/10.1080/15384101.2016.1241605

Bolognesi, B., Lorenzo Gotor, N., Dhar, R., Cirillo, D., Baldrighi, M., Tartaglia, G.G., Lehner, B., 2016. A concentration-dependent liquid phase separation can cause toxicity upon Increased protein expression. Cell Rep 16, 222–231. https://doi.org/10.1016/j.celrep.2016.05.076

Bouchard, J.J., Otero, J.H., Scott, D.C., Szulc, E., Martin, E.W., Sabri, N., Granata, D., Marzahn, M.R., Lindorff-Larsen, K., Salvatella, X., Schulman, B.A., Mittag, T., 2018. Cancer mutations of the tumor suppressor SPOP disrupt the formation of active, phase-

separated compartments. Mol Cell 72, 19-36.e8. https://doi.org/10.1016/j.molcel.2018.08.027
Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M., E., Rengarajan, S., Volorio, A., McBride, M.J., Broye, L.C., Zou, L., Stamenkovic, I., Kadoch, C., Rivera, M.N., 2017. Cancer-specific retargeting of BAF complexes by a prion-like domain. Cell 171, 163-178.e19. https://doi.org/10.1016/j.cell.2017.07.036
Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., Hyman, A.A., 2009. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 324, 1729–1732. https://doi.org/10.1126/science.1172046
Brauch, K.M., Karst, M.L., Herron, K.J., de Andrade, M., Pellikka, P.A., Rodeheffer, R.J., Michels, V.V., Olson, T.M., 2009. Mutations in RNA binding protein gene cause familial dilated cardiomyopathy. J Am Coll Cardiol 54, 930–941. https://doi.org/10.1016/j.jacc.2009.05.038
Cascarina, S.M., Ross, E.D., 2022. Phase separation by the SARS-CoV-2 nucleocapsid protein: Consensus and open questions. J Biol Chem 298, 101677. https://doi.org/10.1016/j.jbc.2022.101677
Chandra, B., Michmerhuizen, N.L., Shirnekhi, H.K., Tripathi, S., Pioso, B.J., Baggett, D.W., Mitrea, D.M., Iacobucci, I., White, M.R., Chen, J., Park, C.-G., Wu, H., Pounds, S., Medyukhina, A., Khairy, K., Gao, Q., Qu, C., Abdelhamed, S., Gorman, S.D., Bawa, S., Maslanka, C., Kinger, S., Dogra, P., Ferrolino, M.C., Di Giacomo, D., Mecucci, C., Klco, J.M., Mullighan, C.G., Kriwacki, R.W., 2022. Phase separation mediates NUP98 fusion oncoprotein leukemic transformation. Cancer Discov 12, 1152–1169. https://doi.org/10.1158/2159-8290.CD-21-0674
Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., Cisse, I.I., 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science 361, 412-415. doi: 10.1126/science.aar4199.
Choi, J.-M., Holehouse, A.S., Pappu, R.V., 2020. Physical principles underlying the complex biology of intracellular phase transitions. Annual Review of Biophysics 49, 107–133. https://doi.org/10.1146/annurev-biophys-121219-081629
Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., Tjian, R., 2018. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science 361, eaar2555. https://doi.org/10.1126/science.aar2555
Chung, I., Leonhardt, H., Rippe, K., 2011. De novo assembly of a PML nuclear subcompartment occurs through multiple pathways and induces telomere elongation. Journal of Cell Science 124, 3603–3618. https://doi.org/10.1242/jcs.084681
Cioce, M., Lamond, A.I., 2005. CAJAL BODIES: A long history of discovery. Annual Review of Cell and Developmental Biology 21, 105–131. https://doi.org/10.1146/annurev.cellbio.20.010403.103738
Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., Lawrence, J.B., 2009. An architectural role for a nuclear non-coding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 33, 717–726. https://doi.org/10.1016/j.molcel.2009.01.026
Cohen, R.J., Benedek, G.B., 2002. Equilibrium and kinetic theory of polymerization and the sol-gel transition [WWW Document]. ACS Publications. https://doi.org/10.1021/j100216a005
Conicella, A.E., Dignon, G.L., Zerze, G.H., Schmidt, H.B., D'Ordine, A.M., Kim, Y.C., Rohatgi, R., Ayala, Y.M., Mittal, J., Fawzi, N.L., 2020. TDP-43 α-helical structure tunes liquid–liquid phase separation and function. Proceedings of the National Academy of Sciences 117, 5883–5894. https://doi.org/10.1073/pnas.1912055117

Dall'Agnese, A., Platt, J.M., Zheng, M.M., Friesen, M., Dall'Agnese, G., Blaise, A.M., Spinelli, J.B., Henninger, J.E., Tevonian, E.N., Hannett, N.M., Lazaris, C., Drescher, H.K., Bartsch, L.M., Kilgore, H.R., Jaenisch, R., Griffith, L.G., Cisse, I.I., Jeppesen, J.F., Lee, T.I., Young, R.A., 2022. The dynamic clustering of insulin receptor underlies its signaling and is disrupted in insulin resistance. Nat Commun 13, 7522. https://doi.org/10.1038/s41467-022-35176-7

de Thé, H., Chen, Z., 2010. Acute promyelocytic leukaemia: novel insights into the mechanisms of cure. Nat Rev Cancer 10, 775–783. https://doi.org/10.1038/nrc2943

de Thé, H., Chomienne, C., Lanotte, M., Degos, L., Dejean, A., 1990. The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor α gene to a novel transcribed locus. Nature 347, 558–561. https://doi.org/10.1038/347558a0

dos Santos, G.A., Kats, L., Pandolfi, P.P., 2013. Synergy against PML-RARa: targeting transcription, proteolysis, differentiation, and self-renewal in acute promyelocytic leukemia. Journal of Experimental Medicine 210, 2793–2802. https://doi.org/10.1084/jem.20131121

Du, M., Chen, Z.J., 2018. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. Science 361, 704–709. https://doi.org/10.1126/science.aat1022

Einstein, A., 1905. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. Annals of Physics, 322, 549560. http://dx.doi.org/10.1002/andp.19053220806

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C.-H., Eckmann, C.R., Myong, S., Brangwynne, C.P., 2015. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proceedings of the National Academy of Sciences 112, 7189–7194. https://doi.org/10.1073/pnas.1504822112

Erdős, G., Pajkos, M., Dosztányi, Z., 2021. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res 49, W297–W303. https://doi.org/10.1093/nar/gkab408

Fasciani, A., D'Annunzio, S., Poli, V., Fagnocchi, L., Beyes, S., Michelatti, D., Corazza, F., Antonelli, L., Gregoretti, F., Oliva, G., Belli, R., Peroni, D., Domenici, E., Zambrano, S., Intartaglia, D., Settembre, C., Conte, I., Testi, C., Vergyris, P., Ruocco, G., Zippo, A., 2020. MLL4-associated condensates counterbalance Polycomb-mediated nuclear mechanical stress in Kabuki Syndrome. Nat Genet 52, 1397–1411. https://doi.org/10.1038/s41588-020-00724-8

Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R.V., Brangwynne, C.P., 2016. Coexisting liquid phases underlie nucleolar subcompartments. Cell 165, 1686–1697. https://doi.org/10.1016/j.cell.2016.04.047

Flory, P.J., Krigbaum, W.R., 1951. Thermodynamics of High Polymer Solutions. Annual Review of Physical Chemistry 2, 383–402. https://doi.org/10.1146/annurev.pc.02.100151.002123

Gorter, E., Grendel, F., 1925. On bimolecular layers of lipids on the chromocytes of the blood. J Exp Med 41, 439–443.

Gouveia, B., Kim, Y., Shaevitz, J.W., Petry, S., Stone, H.A., Brangwynne, C.P., 2022. Capillary forces generated by biomolecular condensates. Nature 609, 255–264. https://doi.org/10.1038/s41586-022-05138-6

Guillén-Boixet, J., Kopach, A., Holehouse, A.S., Wittmann, S., Jahnel, M., Schlüßler, R., Kim, K., Trussina, I.R.E.A., Wang, J., Mateju, D., Poser, I., Maharana, S., Ruer-Gruß, M., Richter, D., Zhang, X., Chang, Y.-T., Guck, J., Honigmann, A., Mahamid, J., Hyman, A.A., Pappu, R.V., Alberti, S., Franzmann, T.M., 2020. RNA-Induced conformational switching and clustering of G3BP drive stress granule assembly by condensation. Cell 181, 346-361.e17. https://doi.org/10.1016/j.cell.2020.03.049

Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., Abraham, B.J., Boija, A., Decker, T.-M., Rimel, J.K., Fant, C.B., Lee, T.I., Cisse, I.I., Sharp, P.A., Taatjes, D.J., Young, R.A., 2019. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature 572, 543–548. https://doi.org/10.1038/s41586-019-1464-0

Hardy, W.B., 1915. Some problems of living matter. Proc. Phys. Soc. London 28, 99. https://doi.org/10.1088/1478-7814/28/1/312

Henninger, J.E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M.M., Andrews, J.O., Zamudio, A.V., Lazaris, C., Hannett, N.M., Lee, T.I., Sharp, P.A., Cissé, I.I., Chakraborty, A.K., Young, R.A., 2021. RNA-mediated feedback control of transcriptional condensates. Cell 184, 207-225.e24. https://doi.org/10.1016/j.cell.2020.11.030

Hofweber, M., Hutten, S., Bourgeois, B., Spreitzer, E., Niedner-Boblenz, A., Schifferer, M., Ruepp, M.-D., Simons, M., Niessing, D., Madl, T., Dormann, D., 2018. Phase separation of FUS is suppressed by its nuclear import receptor and arginine methylation. Cell 173, 706-719.e13. https://doi.org/10.1016/j.cell.2018.03.004

Huggins, M.L., 1942. Some properties of solutions of long-chain compounds. J. Phys. Chem. 46, 151–158. https://doi.org/10.1021/j150415a018

Jeanne, M., Lallemand-Breitenbach, V., Ferhi, O., Koken, M., Le Bras, M., Duffort, S., Peres, L., Berthier, C., Soilihi, H., Raught, B., de Thé, H., 2010. PML/RARA oxidation and arsenic binding initiate the antileukemia response of As2O3. Cancer Cell 18, 88–98. https://doi.org/10.1016/j.ccr.2010.06.003

Kilgore, H.R., Young, R.A., 2022. Learning the chemical grammar of biomolecular condensates. Nat Chem Biol 18, 1298–1306. https://doi.org/10.1038/s41589-022-01046-y

Kilgore, H.K., Mikhael, P.G., Overholt, K.G., Boija, A., Hannett, N., van Dongen, C., Lee, T.I., Chang, Y.-T. Barzilay, R., Young, R.A.Condensates create distinct chemical environments that influence intracellular distributions of molecules. Nature Chem Biol, in press.

Klein, I.A., Boija, A., Afeyan, L.K., Hawken, S.W., Fan, M., Dall'Agnese, A., Oksuz, O., Henninger, J.E., Shrinivas, K., Sabari, B.R., Sagi, I., Clark, V.E., Platt, J.M., Kar, M., McCall, P.M., Zamudio, A.V., Manteiga, J.C., Coffey, E.L., Li, C.H., Hannett, N.M., Guo, Y.E., Decker, T.-M., Lee, T.I., Zhang, T., Weng, J.-K., Taatjes, D.J., Chakraborty, A., Sharp, P.A., Chang, Y.T., Hyman, A.A., Gray, N.S., Young, R.A., 2020. Partitioning of cancer therapeutics in nuclear condensates. Science 368, 1386–1392. https://doi.org/10.1126/science.aaz4427

Klosin, A., Oltsch, F., Harmon, T., Honigmann, A., Jülicher, F., Hyman, A.A., Zechner, C., 2020. Phase separation provides a mechanism to reduce noise in cells. Science 367, 464–468. https://doi.org/10.1126/science.aav6691

Laflamme, G., Mekhail, K., 2020. Biomolecular condensates as arbiters of biochemical reactions inside the nucleus. Commun Biol 3, 1–8. https://doi.org/10.1038/s42003-020-01517-9

Lafontaine, D.L.J., Riback, J.A., Bascetin, R., Brangwynne, C.P., 2021. The nucleolus as a multiphase liquid condensate. Nat Rev Mol Cell Biol 22, 165–182. https://doi.org/10.1038/s41580-020-0272-6

Lancaster, A.K., Nutter-Upham, A., Lindquist, S., King, O.D., 2014. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics 30, 2501–2502. https://doi.org/10.1093/bioinformatics/btu310

Langdon, E.M., Qiu, Y., Ghanbari Niaki, A., McLaughlin, G.A., Weidmann, C.A., Gerbich, T.M., Smith, J.A., Crutchley, J.M., Termini, C.M., Weeks, K.M., Myong, S., Gladfelter, A.S., 2018. mRNA structure determines specificity of a polyQ-driven phase separation. Science 360, 922–927. https://doi.org/10.1126/science.aar7432

Li, C.H., Coffey, E.L., Dall'Agnese, A., Hannett, N.M., Tang, X., Henninger, J.E., Platt, J.M., Oksuz, O., Zamudio, A.V., Afeyan, L.K., Schuijers, J., Liu, X.S., Markoulaki, S.,

Lungjangwa, T., LeRoy, G., Svoboda, D.S., Wogram, E., Lee, T.I., Jaenisch, R., Young, R.A., 2020. MeCP2 links heterochromatin condensates and neurodevelopmental disease. Nature 586, 440–444. https://doi.org/10.1038/s41586-020-2574-4

Li, H., Ernst, C., Kolonko-Adamska, M., Greb-Markiewicz, B., Man, J., Parissi, V., Ng, B.W.-L., 2022. Phase separation in viral infections. Trends in Microbiology 30, 1217–1231. https://doi.org/10.1016/j.tim.2022.06.005

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., Russo, P.S., Jiang, Q.-X., Nixon, B.T., Rosen, M.K., 2012. Phase transitions in the assembly of multivalent signalling proteins. Nature 483, 336–340. https://doi.org/10.1038/nature10879

Lo-Coco, F., Avvisati, G., Vignetti, M., Thiede, C., Orlando, S.M., Iacobelli, S., Ferrara, F., Fazi, P., Cicconi, L., Di Bona, E., Specchia, G., Sica, S., Divona, M., Levis, A., Fiedler, W., Cerqui, E., Breccia, M., Fioritoni, G., Salih, H.R., Cazzola, M., Melillo, L., Carella, A.M., Brandts, C.H., Morra, E., von Lilienfeld-Toal, M., Hertenstein, B., Wattad, M., Lübbert, M., Hänel, M., Schmitz, N., Link, H., Kropp, M.G., Rambaldi, A., La Nasa, G., Luppi, M., Ciceri, F., Finizio, O., Venditti, A., Fabbiano, F., Döhner, K., Sauer, M., Ganser, A., Amadori, S., Mandelli, F., Döhner, H., Ehninger, G., Schlenk, R.F., Platzbecker, U., 2013. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. New England Journal of Medicine 369, 111–121. https://doi.org/10.1056/NEJMoa1300874

Lyon, A.S., Peeples, W.B., Rosen, M.K., 2021. A framework for understanding the functions of biomolecular condensates across scales. Nat Rev Mol Cell Biol 22, 215–235. https://doi.org/10.1038/s41580-020-00303-z

Lyons, H., Veettil, R.T., Pradhan, P., Fornero, C., De La Cruz, N., Ito, K., Eppert, M., Roeder, R.G., Sabari, B.R., 2023. Functional partitioning of transcriptional regulators by patterned charge blocks. Cell 186, 327-345.e28. https://doi.org/10.1016/j.cell.2022.12.013

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., Jahnel, M., Marrone, L., Chang, Y.T., Sterneckert, J., Tomancak, P., Hyman, A., Alberti, S., 2018. RNA buffers the phase separation behavior of prion-like RNA-binding proteins. Science 360, 918–921. https://doi.org/10.1126/science.aar7366

Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R.V., Mittag, T., 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science 367, 694–699. https://doi.org/10.1126/science.aaw8653

Mason, T.G., Ganesan, K., van Zanten, J.H., Wirtz, D., Kuo, S.C., 1997. Particle tracking microrheology of complex fluids. Phys. Rev. Lett. 79, 3282–3285. https://doi.org/10.1103/PhysRevLett.79.3282

McSwiggen, D.T., Mir, M., Darzacq, X., Tjian, R., 2019. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. Genes Dev 33, 1619–1634. https://doi.org/10.1101/gad.331520.119

Mensah, M.A., Niskanen, H., Magalhaes, A.P., Basu, S., Kircher, M., Sczakiel, H.L., Reiter, A.M., Elsner, J., Meinecke P., Biskup S., Chung B.H.Y., Dombrowsky, G., Eckmann-Sholz C., Hitz, M.P., Hoischen, A., Holterhus, P-M., Hülsemann, W., Kahrizi, K., Kalsheuer, V.M., Kan, A., Krumbiegel, M., Kurth, I., Leubner, J., Longardt, A.C., Moritz, J.D., Najmabadi, H., Skipalova, K., Blok, L.S., Tzschach, A., Wiedersberg, E., Zenker, M., Garcia-Cabau, C., Ruschow, R., Salvatella, X., Kraushar, M.L., Mundlos, S., Caliebe, A., Spielmann, M., Horn, D., Hnisz, D., 2023. Aberrant phase separation and nucleolar dysfunction in rare genetic diseases. Nature 614, 564–571. https://doi.org/10.1038/s41586-022-05682-1

Michmerhuizen, N.L., Klco, J.M., Mullighan, C.G., 2020. Mechanistic insights and potential therapeutic approaches for NUP98-rearranged hematologic malignancies. Blood 136, 2275–2289. https://doi.org/10.1182/blood.2020007093
Misteli, T., 2020. The self-organizing genome: Principles of genome architecture and function. Cell 183, 28–45. https://doi.org/10.1016/j.cell.2020.09.014
Mitrea, D.M., Cika, J.A., Stanley, C.B., Nourse, A., Onuchic, P.L., Banerjee, P.R., Phillips, A.H., Park, C.-G., Deniz, A.A., Kriwacki, R.W., 2018. Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. Nat Commun 9, 842. https://doi.org/10.1038/s41467-018-03255-3
Mitrea, D.M., Mittasch M., Gomes, B.F., Klein I.A., Murcko M.A., 2022. Modulating biomolecular condensates: A novel approach to drug discovery. Nat Rev Drug Disc 21, 841–862.
Miyake, N., Takahashi, H., Nakamura, K., Isidor, B., Hiraki, Y., Koshimizu, E., Shiina, M., Sasaki, K., Suzuki, H., Abe, R., Kimura, Y., Akiyama, T., Tomizawa, S., Hirose, T., Hamanaka, K., Miyatake, S., Mitsuhashi, S., Mizuguchi, T., Takata, A., Obo, K., Kato, M., Ogata, K., Matsumoto, N., 2020. Gain-of-function MN1 truncation variants cause a recognizable syndrome with craniofacial and brain abnormalities. Am J Hum Genet 106, 13–25. https://doi.org/10.1016/j.ajhg.2019.11.011
Montgomery Jr., T.S.H., 1898. Comparative cytological studies, with especial regard to the morphology of the nucleolus. Journal of Morphology 15, 265–582. https://doi.org/10.1002/jmor.1050150204
Mullock, B.M., Luzio, J.P., 2013. Theory of Organelle Biogenesis: A Historical Perspective, Madame Curie Bioscience Database [Internet]. Landes Bioscience.
Murray, D.T., Kato, M., Lin, Y., Thurber, K.R., Hung, I., McKnight, S.L., Tycko, R., 2017. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. Cell 171, 615-627.e16. https://doi.org/10.1016/j.cell.2017.08.048
Niwa-Kawakita, M., Ferhi, O., Soilihi, H., Le Bras, M., Lallemand-Breitenbach, V., de Thé, H., 2017. PML is a ROS sensor activating p53 upon oxidative stress. Journal of Experimental Medicine 214, 3197–3206. https://doi.org/10.1084/jem.20160301
Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., Baldwin, A.J., 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Molecular Cell 57, 936–947. https://doi.org/10.1016/j.molcel.2015.01.013
Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B., Dosztányi, Z., Uversky, V.N., Obradovic, Z., Kurgan, L., Dunker, A.K., Gough, J., 2013. D2P2: database of disordered protein predictions. Nucleic Acids Res 41, D508–D516. https://doi.org/10.1093/nar/gks1226
Observations on the organs and mode of fecundation in Orchideae and Asclepiadeae / by Robert Brown. [WWW Document], n.d. . Wellcome Collection. URL https://wellcomecollection.org/works/khgyheqh/items (accessed 3.20.23).
O'Flynn, B.G., Mittag, T., 2021. The role of liquid–liquid phase separation in regulating enzyme activity. Current Opinion in Cell Biology, Cell Signalling 69, 70–79. https://doi.org/10.1016/j.ceb.2020.12.012
Paiz, E.A., Allen, J.H., Correia, J.J., Fitzkee, N.C., Hough, L.E., Whitten, S.T., 2021. Beta turn propensity and a model polymer scaling exponent identify intrinsically disordered phase-separating proteins. J Biol Chem 297, 101343. https://doi.org/10.1016/j.jbc.2021.101343
Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L.A., Weigert, M., Myers, E.W., Grill, S., Drechsel, D., Hyman, A.A., Alberti, S., 2015. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell 162, 1066–1077. https://doi.org/10.1016/j.cell.2015.07.047

Patel, A., Malinovska, L., Saha, S., Wang, J., Alberti, S., Krishnan, Y., Hyman, A.A., 2017. ATP as a biological hydrotrope. Science 356, 753–756. https://doi.org/10.1126/science.aaf6846
Pederson, T., 2011. The Nucleolus. Cold Spring Harb Perspect Biol 3, a000638. https://doi.org/10.1101/cshperspect.a000638
Piovesan, D., Del Conte, A., Clementel, D., Monzon, A.M., Bevilacqua, M., Aspromonte, M.C., Iserte, J.A., Orti, F.E., Marino-Buslje, C., Tosatto, S.C.E., 2022. MobiDB: 10 years of intrinsically disordered proteins. Nucleic Acids Res 51, D438–D444. https://doi.org/10.1093/nar/gkac1065
Prouteau, M., Loewith, R., 2018. Regulation of cellular metabolism through phase separation of enzymes. Biomolecules 8, 160. https://doi.org/10.3390/biom8040160
Rai, A.K., Chen, J.-X., Selbach, M., Pelkmans, L., 2018. Kinase-controlled phase transition of membraneless organelles in mitosis. Nature 559, 211–216. https://doi.org/10.1038/s41586-018-0279-8
Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., Mahato, J., Panigrahi, R., Navalkar, A., Mehra, S., Gadhe, L., Chatterjee, D., Sawner, A.S., Maiti, S., Bhatia, S., Gerez, J.A., Chowdhury, A., Kumar, A., Padinhateeri, R., Riek, R., Krishnamoorthy, G., Maji, S.K., 2020. α-Synuclein aggregation nucleates through liquid–liquid phase separation. Nat. Chem. 12, 705–716. https://doi.org/10.1038/s41557-020-0465-9
Riback, J.A., Katanski, C.D., Kear-Scott, J.L., Pilipenko, E.V., Rojek, A.E., Sosnick, T.R., Drummond, D.A., 2017. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. Cell 168, 1028-1040.e19. https://doi.org/10.1016/j.cell.2017.02.027
Roden, C., Gladfelter, A.S., 2021. RNA contributions to the form and function of biomolecular condensates. Nat Rev Mol Cell Biol 22, 183–195. https://doi.org/10.1038/s41580-020-0264-6
Rodrigues e Silva, A.M., Geldsetzer, F., Holdorff, B., Kielhorn, F.W., Balzer-Geldsetzer, M., Oertel, W.H., Hurtig, H., Dodel, R., 2010. Who was the man who discovered the “Lewy bodies”? Movement Disorders 25, 1765–1773. https://doi.org/10.1002/mds.22956
Sabari, B.R., 2020. Biomolecular condensates and gene activation in development and disease. Developmental Cell 55, 84–96. https://doi.org/10.1016/j.devcel.2020.09.005
Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., Li, C.H., Guo, Y.E., Day, D.S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T.I., Cisse, I.I., Roeder, R.G., Sharp, P.A., Chakraborty, A.K., Young, R.A., 2018. Coactivator condensation at super-enhancers links phase separation and gene control. Science 361, 6400. doi: 10.1126/science.aar3958.
Schneider, J.W., Oommen, S., Qureshi, M.Y., Goetsch, S.C., Pease, D.R., Sundsbak, R.S., Guo, W., Sun, M., Sun, H., Kuroyanagi, H., Webster, D.A., Coutts, A.W., Holst, K.A., Edwards, B.S., Newville, N., Hathcock, M.A., Melkamu, T., Briganti, F., Wei, W., Romanelli, M.G., Fahrenkrug, S.C., Frantz, D.E., Olson, T.M., Steinmetz, L.M., Carlson, D.F., Nelson, T.J., 2020. Dysregulated ribonucleoprotein granules promote cardiomyopathy in RBM20 gene-edited pigs. Nat Med 26, 1788–1800. https://doi.org/10.1038/s41591-020-1087-x
Shin, Y., Chang, Y.-C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., Brangwynne, C.P., 2018. Liquid nuclear condensates mechanically sense and restructure the genome. Cell 175, 1481-1491.e13. https://doi.org/10.1016/j.cell.2018.10.057
Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A.V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., Chakraborty, A.K., 2019. Enhancer features that drive formation of transcriptional condensates. Mol Cell 75, 549-561.e7. https://doi.org/10.1016/j.molcel.2019.07.009

Smith, P.J., Sykes, H.R., Fox, M.E., Furlong, I.J., 1992. Subcellular distribution of the anticancer drug mitoxantrone in human and drug-resistant murine cells analyzed by flow cytometry and confocal microscopy and its relationship to the induction of DNA damage. Cancer research 52, 4000–8.
Snead, W.T., Gladfelter, A.S., 2019. The control centers of biomolecular phase separation: How membrane surfaces, post-translational modifications, and active processes regulate condensation. Mol Cell 76, 295–305. https://doi.org/10.1016/j.molcel.2019.09.016
Spannl, S., Tereshchenko, M., Mastromarco, G.J., Ihn, S.J., Lee, H.O., 2019. Biomolecular condensates in neurodegeneration and cancer. Traffic 20, 890–911. https://doi.org/10.1111/tra.12704
Stoeger, T., Battich, N., Pelkmans, L., 2016. Passive noise filtering by cellular compartmentalization. Cell 164, 1151–1161. https://doi.org/10.1016/j.cell.2016.02.005
Su, X., Ditlev, J.A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D.S., Taunton, J., Rosen, M.K., Vale, R.D., 2016. Phase separation of signaling molecules promotes T cell receptor signal transduction. Science 352, 595–599. https://doi.org/10.1126/science.aad9964
Taylor, N.O., Wei, M.-T., Stone, H.A., Brangwynne, C.P., 2019. Quantifying dynamics in phase-separated condensates using fluorescence recovery after photobleaching. Biophysical Journal 117, 1285–1300. https://doi.org/10.1016/j.bpj.2019.08.030
Trojanowski, J., Frank, L., Rademacher, A., Mücke, N., Grigaitis, P., Rippe, K., 2022. Transcription activation is enhanced by multivalent interactions independent of phase separation. Molecular Cell 82, 1878-1893.e10. https://doi.org/10.1016/j.molcel.2022.04.017
Tsang, B., Pritišanac, I., Scherer, S.W., Moses, A.M., Forman-Kay, J.D., 2020. Phase separation as a missing mechanism for interpretation of disease mutations. Cell 183, 1742–1756. https://doi.org/10.1016/j.cell.2020.11.050
van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E., Babu, M.M., 2014. Classification of intrinsically disordered regions and proteins. Chem. Rev. 114, 6589–6631. https://doi.org/10.1021/cr400525m
van Mierlo, G., Jansen, J.R.G., Wang, J., Poser, I., van Heeringen, S.J., Vermeulen, M., 2021. Predicting protein condensate formation using machine learning. Cell Reports 34, 108705. https://doi.org/10.1016/j.celrep.2021.108705
Vernon, R.M., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H., Forman-Kay, J.D., 2018. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. eLife 7, e31486. https://doi.org/10.7554/eLife.31486
Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L.M., Cortese, M.S., Lawson, J.D., Brown, C.J., Sikes, J.G., Newton, C.D., Dunker, A.K., 2005. DisProt: a database of protein disorder. Bioinformatics 21, 137–140. https://doi.org/10.1093/bioinformatics/bth476
Wagner, R., 1835. Einige Bemerkungen und Fragen über die Keimbläschen (vesicula germinativa). Archiv für Anatomie, Physiologie, 373-377.
Walsh, I., Martin, A.J.M., Di Domenico, T., Tosatto, S.C.E., 2012. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28, 503–509. https://doi.org/10.1093/bioinformatics/btr682
Wan, L., Chong, S., Xuan, F., Liang, A., Cui, X., Gates, L., Carroll, T.S., Li, Y., Feng, L., Chen, G., Wang, S.-P., Ortiz, M.V., Daley, S.K., Wang, X., Xuan, H., Kentsis, A., Muir, T.W., Roeder, R.G., Li, H., Li, W., Tjian, R., Wen, H., Allis, C.D., 2020. Impaired cell fate through gain-of-function mutations in a chromatin reader. Nature 577, 121–126. https://doi.org/10.1038/s41586-019-1842-7

Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R.V., Alberti, S., Hyman, A.A., 2018. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. Cell 174, 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006

Wang, S., Dai, T., Qin, Z., Pan, T., Chu, F., Lou, L., Zhang, L., Yang, B., Huang, H., Lu, H., Zhou, F., 2021. Targeting liquid–liquid phase separation of SARS-CoV-2 nucleocapsid protein promotes innate antiviral immunity by elevating MAVS activity. Nat Cell Biol 23, 718–732. https://doi.org/10.1038/s41556-021-00710-0

Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K.M., Bennett, R.E., Dujardin, S., Laskowski, P.R., MacKenzie, D., Kamath, T., Commins, C., Vanderburg, C., Roe, A.D., Fan, Z., Molliex, A.M., Hernandez-Vega, A., Muller, D., Hyman, A.A., Mandelkow, E., Taylor, J.P., Hyman, B.T., 2018. Tau protein liquid–liquid phase separation can initiate tau aggregation. The EMBO Journal 37, e98049. https://doi.org/10.15252/embj.201798049

Wheeler, R.J., Lee, H.O., Poser, I., Pal, A., Doeleman, T., Kishigami, S., Kour, S., Anderson, E.N., Marrone, L., Murthy, A.C., Jahnel, M., Zhang, X., Boczek, E., Fritsch, A., Fawzi, N.L., Sterneckert, J., Pandey, U., David, D.C., Davis, B.G., Baldwin, A.J., Hermann, A., Bickle, M., Alberti, S., Hyman, A.A., 2019. Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. https://doi.org/10.1101/721001

Wilson, E.B., 1899. The structure of the protoplasm. Science 10, 33–45. DOI:10.1126/science.10.237.33

Woodruff, J.B., Gomes, B.F., Widlund, P.O., Mahamid, J., Honigmann, A., Hyman, A.A., 2017. The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. Cell 169, 1066-1077.e10. https://doi.org/10.1016/j.cell.2017.05.028

Wright, P.E., Dyson, H.J., 2015. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16, 18–29. https://doi.org/10.1038/nrm3920

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., Hirose, T., 2018. Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. Molecular Cell 70, 1038-1053.e7. https://doi.org/10.1016/j.molcel.2018.05.019

Yoo, H., Triandafillou, C., Drummond, D.A., 2019. Cellular sensing by phase separation: Using the process, not just the products. J Biol Chem 294, 7151–7159. https://doi.org/10.1074/jbc.TM118.001191

Zhang, H., Elbaum-Garfinkle, S., Langdon, E., Taylor, N., Occhipinti, P., Bridges, A., Brangwynne, C.P., Gladfelter, A.S., 2015. RNA controls PolyQ protein phase transitions. Mol Cell 60, 220–230. https://doi.org/10.1016/j.molcel.2015.09.017

Zhang, X.-W., Yan, X.-J., Zhou, Z.-R., Yang, F.-F., Wu, Z.-Y., Sun, H.-B., Liang, W.-X., Song, A.-X., Lallemand-Breitenbach, V., Jeanne, M., Zhang, Q.-Y., Yang, H.-Y., Huang, Q.-H., Zhou, G.-B., Tong, J.-H., Zhang, Y., Wu, J.-H., Hu, H.-Y., de Thé, H., Chen, S.-J., Chen, Z., 2010. Arsenic trioxide controls the fate of the PML-RARα oncoprotein by directly binding PML. Science 328, 240–243. https://doi.org/10.1126/science.1183424

CHAPTER 2:

Genetic Variation Associated with Condensate Dysregulation in Disease

Originally published in *Developmental Cell* 57, 1776–1788, July 25, (2022)

Salman F. Banani[1,2,7], Lena K. Afeyan[1,3,7], Susana W. Hawken[1,4,7], Jonathan E. Henninger[1], Alessandra Dall'Agnese[1], Victoria E. Clark[1,5], Jesse M. Platt[1,6], Ozgur Oksuz[1], Nancy M. Hannett[1], Ido Sagi[1], Tong Ihn Lee[1], and Richard A. Young[1,3,8]*

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[4]Program of Computational & Systems Biology, Massachusetts Institute of Technology, MA 02139, USA
[5]Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[6]Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, Boston, MA, 02114, USA
[7]Equal contributions
[8]Lead contact
*Correspondence: Richard A. Young (young@wi.mit.edu)

Abstract

A multitude of cellular processes involve biomolecular condensates, which has led to the suggestion that diverse pathogenic mutations may dysregulate condensates. While proof-of-concept studies have identified specific mutations that cause condensate dysregulation, the full scope of pathological genetic variation that affects condensates is not yet known. Here we comprehensively map pathogenic mutations to condensate-promoting protein features in putative condensate-forming proteins and find over 36,000 pathogenic mutations that plausibly contribute to condensate dysregulation in over 1,200 Mendelian diseases and 550 cancers. This resource captures mutations presently known to dysregulate condensates and experimental tests confirm that additional pathological mutations do indeed affect condensate properties in cells. These findings suggest that condensate dysregulation may be a pervasive pathogenic mechanism underlying a broad spectrum of human diseases, provide a strategy to identify proteins and mutations involved in pathologically altered condensates, and serve as a foundation for mechanistic insights into disease and therapeutic hypotheses.

Introduction

How genetic variation gives rise to human disease is understood largely from the effects of mutations on the structure and function of individual protein molecules. Genetic and biochemical studies have revealed how mutations in protein coding sequences affect molecular-scale properties, such as conformation, stability, and catalytic activity, providing mechanistic hypotheses of disease causality that have led to valuable therapeutics (Stefl et al., 2013; Wan et al., 2004). However, underlying pathogenic mechanisms for many genetic diseases remain elusive, despite extensive cataloging of associated mutations. Recent studies have shown that disease-causing mutations may also affect properties related to mesoscale cellular organization (Kasza et al., 2019; Lewis et al., 2019). Many cellular proteins are compartmentalized within biomolecular condensates (Banani et al., 2017; Shin and Brangwynne, 2017), which are membraneless organelles that concentrate functionally related proteins and nucleic acids and organize many vital cellular processes, such as DNA replication, DNA repair, transcription, chromatin organization, RNA biosynthesis and homeostasis, ribosome biosynthesis, protein quality control, innate immunity, cell division, cell-cell adhesions, signaling, and synaptic transmission (Alberti, 2017; Beutel et al., 2019; Boija et al., 2018; Cai et al., 2019; Case et al., 2019; Cho et al., 2018; Du and Chen, 2018; Frottin et al., 2019; Gibson et al., 2019; Guo et al., 2019; Huang et al., 2019; Jiang et al., 2015; Kilic et al., 2019; King and Petry, 2020; Larson et al., 2017; Lu et al., 2020; Lyon et al., 2020; Milovanovic et al., 2018; Parker et al., 2019; Riback et al., 2020; Schwayer et al., 2019; Sheu-Gruttadauria and MacRae, 2018; Strom et al., 2017; Su et al., 2016; Woodruff et al., 2017; Zamudio et al., 2019; Zeng et al., 2016). A subset of condensate components directly governs the formation, maintenance, organization, composition, and physicochemical and material properties of the condensate (Banani et al., 2016; Feric et al., 2016; Jain et al., 2016; Li et al., 2012; Lin et al., 2015; Wang et al., 2018). Thus, a protein-coding mutation in a condensate-forming protein may affect not only the individual protein, but also the biomolecular condensate in which the protein is found. Specifically, mutations that affect regions of proteins that promote condensate formation can significantly alter the properties of condensates, including their formation (Ahn et al., 2021; Chandra et al., 2021; Li et al., 2020; Wan et al., 2019), material properties (Patel et al., 2015), localization (Boulay et al., 2017), or composition (Basu et al., 2020). These condensate-promoting features include modular interaction domains (MIDs) and stretches of low complexity sequences (LCSs) found within intrinsically disordered regions (IDRs) (Figure 1A). These observations have led us and others to

postulate that condensate dysregulation may play a role across a broad spectrum of diseases (Alberti and Dormann, 2019; Alberti and Hyman, 2021; Boija et al., 2021; Kim et al., 2013; Molliex et al., 2015; Tsang et al., 2020).

A resource that links data on pathogenic genetic variation to condensate-promoting protein features could promote further study of diseases likely to involve dysregulated condensates. To this end, we collected putative condensate-forming proteins, annotated condensate-promoting sequence features (MIDs and LCSs) onto these proteins, and mapped a broad spectrum of human disease variants associated with Mendelian diseases and cancers to these features. This approach produced a catalog of over 36,000 pathogenic mutations associated with 1,790 diseases that may involve condensate dysregulation as an underlying pathogenic mechanism. To demonstrate the utility of this approach and estimate its predictive accuracy, we performed experimental tests across 12 proteins from the catalog and found most tested mutations do indeed cause condensate dysregulation phenotypes in cells. This resource and its associated analyses provide a foundation for the study of condensate-associated disease mechanisms by facilitating the generation of novel mechanistic and therapeutic hypotheses.

Results

Generating a resource for the study of condensate dysregulation in disease

A set of putative condensate-forming proteins was defined by integrating existing databases of proteome-wide subcellular immunofluorescence (Yu et al., 2020), sequence-based predictions (Mierlo et al., 2021), and curation of phase-separating proteins from the literature (Li et al., 2019; Mészáros et al., 2019; You et al., 2019) (Figure 1B, Figure S1A, Table S4A-B, Methods). This approach defined 3,941 putative condensate-forming proteins.

Condensate-promoting features, consisting of MIDs and LCSs (Figure 1A), within these 3,941 putative condensate-forming proteins were then identified. MIDs, such as SH2, SH3, RRM, and Bromodomains, were defined by integrating annotations of the subset of conserved protein domains (Blum et al., 2020; Letunic et al., 2020; Lu et al., 2019; Mistry et al., 2020) known to participate in binding interactions (Bienz, 2020; Hentze et al., 2018; Lambert et al., 2018; Lunde et al., 2007; Pawson and Nash, 2003; Seet et al., 2006; Vaquerizas et al., 2009; Yun et al., 2011) (Table S4F). LCSs, such as prion-like domains (Alberti et al., 2009; Martin et al., 2020; Wang et al., 2018), low-complexity aromatic-rich kinked segments (LARKS) (Hughes et al., 2018), regions enriched with pi-interacting residues (Vernon et al., 2018), and acidic/basic charge blocks, were mapped using existing approaches where available or by scanning human protein sequences for statistically identified regions of low complexity (Figure S1A-C, Table S1). This analysis produced a map of condensate-promoting features across the set of putative condensate-forming proteins and recovered the MIDs and LCSs of known condensate-forming proteins with high fidelity (Figure S1D, Table S4B, Table S4F-G).

We then identified the pathogenic mutations that affect condensate-promoting features (Figure 1B), hypothesizing that such mutations would have a high likelihood of affecting condensate properties. We extracted pathological human disease variants from existing datasets of variants associated with Mendelian diseases and cancers (Methods). Variants were defined as pathogenic based on clinical assessments of pathogenicity provided in the source datasets for Mendelian variants, or integrated from independent knowledgebases for cancer variants (Chakravarty et al., 2017; Griffith et al., 2017; Landrum et al., 2017; Stenson et al., 2020; Tamborero et al., 2018) (Figure S1A, Table S4C, Methods). Such assessments of pathogenicity are largely based on established guidelines that integrate various sources of evidence, including

associations with clinical phenotypes, population or tumor frequencies, and computational predictions, as well as knowledge of functional or molecular properties of the mutation and the affected protein (Li et al., 2017; Richards et al., 2015). Within pathogenic variants, we focused on the types of variants where we could reasonably predict the effect of mutations on condensate-promoting features (Methods). These variant types consisted of missense variants, in-frame insertions and deletions (indels), as well as nonsense and frameshift variants (hereafter, referred to together as *truncating variants*). Together, these variant types comprised over 98% of the observed pathogenic mutations (Figure S2A). Truncating variants may lead to nonsense-mediated decay (NMD), confounding whether a truncating mutation imparts its effects primarily through the loss of a condensate-promoting feature versus the loss of the protein. To minimize this confounding effect, we chose to eliminate all truncations predicted to elicit NMD (Lindeboom et al., 2016) from the analyses (Table S4D, Methods). In total, we extracted 322,825 pathogenic variants associated with 5,342 Mendelian diseases and 659 cancer types for further study (Cerami et al., 2012; Consortium, 2017; Hoadley et al., 2018; Landrum et al., 2017; Stenson et al., 2020) (Figure 1B, Figure S1A).

We mapped these pathogenic variants to the condensate-promoting features annotated within condensate-forming proteins. Mutations were defined as affecting condensate-promoting features if they were missense mutations or in-frame insertions within the bounds of an MID or LCS, or if they were in-frame deletions and truncating mutations removing part of an MID or LCS (Methods). This resulted in a catalog of 36,777 pathogenic mutations found to affect condensate-promoting features in 1,745 of the putative condensate-forming proteins, spanning 1,233 distinct Mendelian phenotypes and 557 cancer types (Figure S1A, Figure S2B-D, Table S4H). This catalog recovered pathogenic mutations shown in the literature to cause condensate dysregulation with a sensitivity of 76%, including mutations in proteins such as UBQLN2, FUS, MECP2, TIA1, HNRNPA1, and SPOP (Bouchard et al., 2018; Conicella et al., 2016; Dao et al., 2019; Li et al., 2020; Mackenzie et al., 2017; Molliex et al., 2015; Patel et al., 2015; Quiroz et al., 2020) (Figure S2E, Figure S1D, Table S2). Thus, this catalog of pathogenic mutations, with annotations of associated diseases, disrupted condensate-promoting features, and affected condensate-forming proteins for each mutation (Table S4H), provides a foundation for further studies of condensate dysregulation in disease.

Figure 1

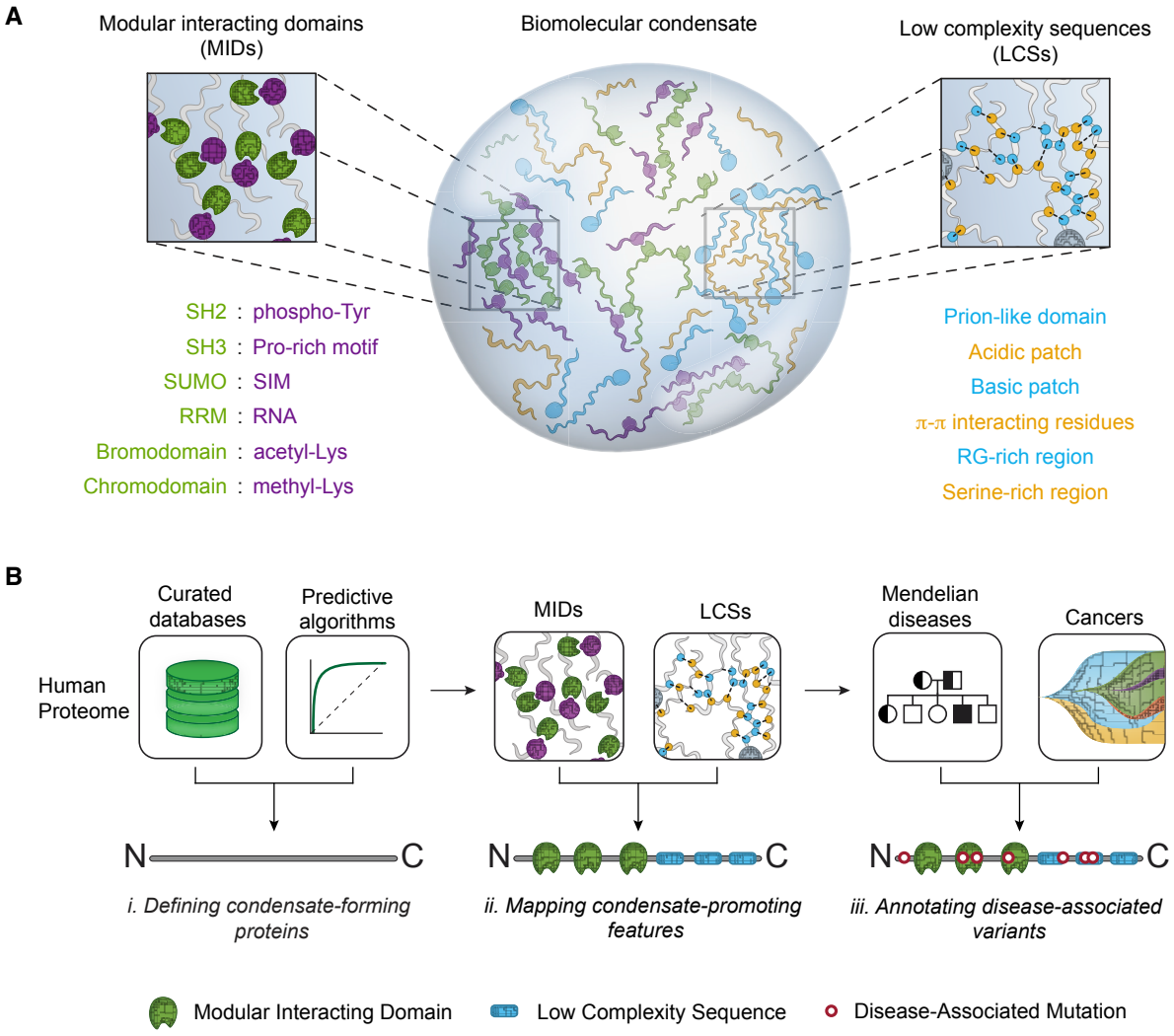


Figure 1. A proteome-wide map of pathogenic mutations in condensate-promoting features. See also Figures S1-S2 and Table S1.

A. Multivalent interacting features in proteins that promote biomolecular condensate formation, including modular interacting domains (MIDs, left, green and purple) and low complexity sequences (LCSs, right, blue and green).

B. Approach to generate a map of pathogenic mutations that affect condensate-promoting features across the proteome (see also Figure S1A). MIDs and LCSs were mapped across the proteome (left, top) and used to define multivalent proteins (middle). Mendelian and cancer variants were mapped across the proteome (left, bottom), in particular on across the set of multivalent proteins (middle), to identify pathogenic mutations that affect MIDs and LCSs (Methods). The approach allows analysis of diseases, condensates, and mutational signatures associated with pathogenic mutations that affect condensate-promoting features in multivalent proteins (right).

The spectrum of diseases predicted to involve dysregulated condensates

Thus far, a small fraction of known diseases has been shown to arise from condensate dysfunction, so most diseases have not been directly linked to pathogenic mechanisms involving condensates. Condensates that have been linked to specific diseases thus far have provided important new insights into the biological regulation of the condensate as well as the pathogenic mechanisms underlying the disease (Ahn et al., 2021; Boija et al., 2021; Cai et al., 2021; Chandra et al., 2021; Kim et al., 2013; Li et al., 2020; Min et al., 2019; Molliex et al., 2015; Nedelsky and Taylor, 2019; Patel et al., 2015; Quiroz et al., 2020; Ramaswami et al., 2013; Spannl et al., 2019; Zhang et al., 2020). Thus, we next asked what types of diseases were most associated with the mutations predicted to dysregulate condensates. We categorized Mendelian diseases and cancers by the organ systems or tissue types they involved (Methods). Mutations affecting condensate-promoting features were involved in nearly all types of Mendelian diseases and cancers (Figure 2A-B). The proportion of such mutations affecting a particular organ system was more or less comparable across all organ systems, and these mutations accounted for 5-10% of pathogenic mutations across Mendelian diseases and 15-25% of mutations across cancer types.

Specific mutations have been shown to cause dysregulation of a small subset of the biomolecular condensates described thus far (Basu et al., 2020; Kim et al., 2013; Li et al., 2020; Mackenzie et al., 2017; Molliex et al., 2015; Patel et al., 2015; Quiroz et al., 2020; Ramaswami et al., 2013), while the majority of known condensates have not directly been linked to human disease. To evaluate the breadth of known condensates that could be dysregulated in disease, we looked for associations with specific condensates among the set of disease-associated, condensate-forming proteins within our catalog (Methods). The mutations predicted to dysregulate condensates occurred in proteins associated with a broad range of functions and condensates, but were particularly evident among components of nuclear condensates, such as those involved in transcription, chromatin structure, RNA splicing and pre-ribosome biosynthesis (Figure 2C, Table S4I-K). Stratifying this analysis by disease type revealed known associations of condensates and diseases—including those of RNA granules with FTD, ALS, and other neurodegenerative phenotypes (Conicella et al., 2016; Kim et al., 2013; Mackenzie et al., 2017; Molliex et al., 2015; Ramaswami et al., 2013); of transcriptional condensates with polydactyly (Basu et al., 2020); of heterochromatin with Rett syndrome (Li et al., 2020); and of keratohyalin granules with atopic dermatitis (Quiroz et al., 2020)—and nominated numerous additional putative associations between known condensates and specific Mendelian diseases or cancers (Figure 2D, Table S4J-K). These results corroborate the hypothesis that condensate dysregulation may be an underlying pathogenic mechanism across a broad spectrum of human diseases.

It is important to note that some nominated mutations are also likely to contribute to pathogenesis via other known molecular-scale mechanisms, such as disruption of protein fold, catalysis, ligand binding, post-translational modifications, and subcellular localization signals (Figure S2G, Table S4B). Condensate dysregulation does not exclude these canonical models of protein dysfunction, but rather, provides an additional framework with which to better understand the pathogenic basis of disease and a foundation for mechanistic and therapeutic hypotheses.

Figure 2

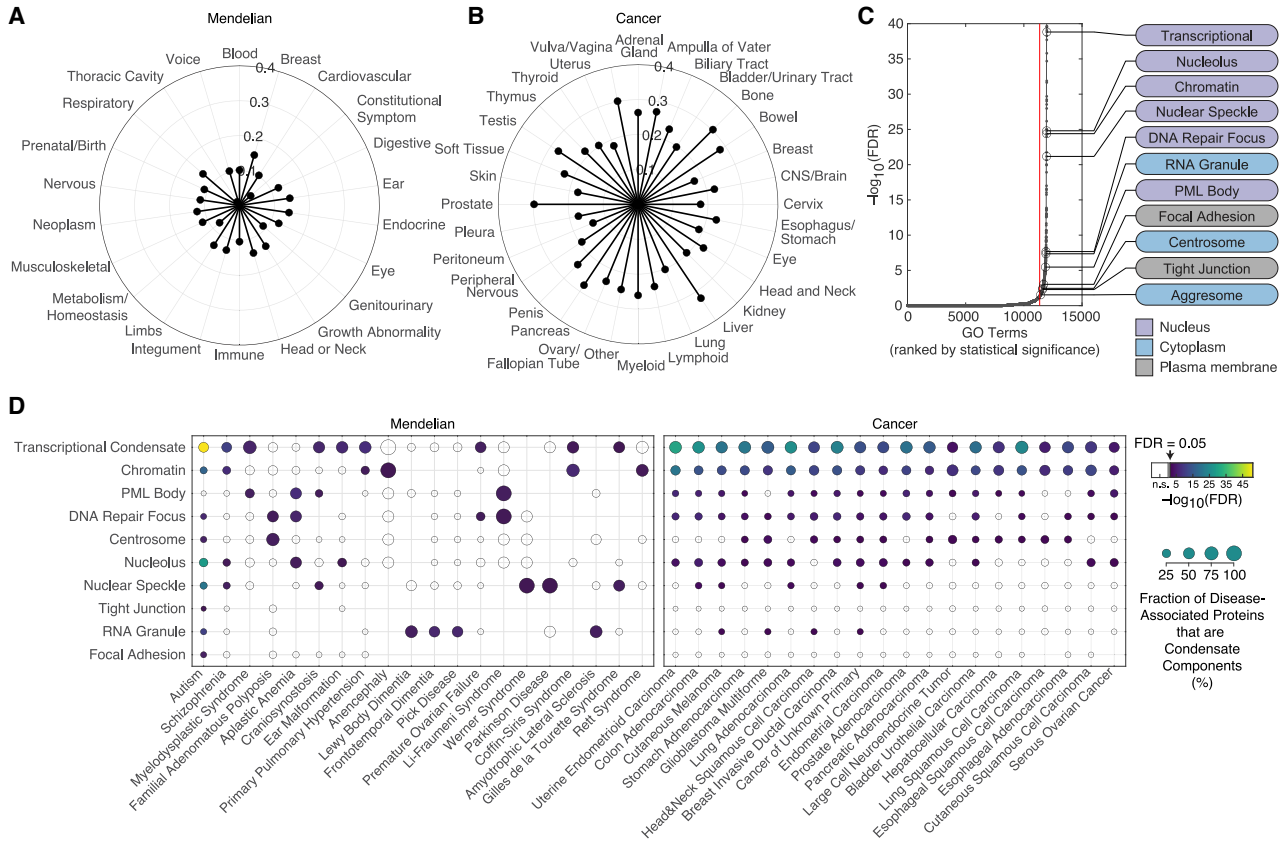


Figure 2. Condensate dysregulation across the spectrum of disease. See also Table S4I-K.

A. Proportion of pathogenic mutations (depicted as distance from center of radar plot) affecting condensate-promoting features in multivalent proteins across Mendelian diseases. Mendelian diseases are stratified by organ systems in which the diseases had a phenotypic effect (Methods).

B. Proportion of pathogenic mutations (depicted as distance from center of radar plot) affecting condensate-promoting features in multivalent proteins across cancers. Cancers are stratified by tissues of origin (Methods).

C. Enrichment of GO terms among the set of condensate-forming proteins that have pathogenic mutations that affect condensate-promoting features. GO terms (black dots) are ranked (*x*-axis) by statistical significance ($-\log_{10}$(FDR), *y*-axis). Red line denotes GO term rank corresponding to threshold for statistical significance (FDR < 0.05). The subset of significantly enriched GO terms that correspond to biomolecular condensates (Table S4I) are highlighted (black open circles and labels). Nuclear, cytoplasmic, and plasma membrane-associated condensates are indicated by purple, blue, or gray labels, respectively.

D. Significant associations between specific diseases and specific condensates. The set of condensate-forming proteins with pathogenic mutations affecting condensate-promoting features were mapped to specific condensates using Gene Ontology (see Methods) as well as associated with specific diseases. Overlaps between subsets of proteins associated with specific condensates (*y*-axis) and those associated with specific diseases (*x*-axis) were tested for statistical significance. Selected examples of Mendelian diseases (left) and cancer types (right) are shown (see also Table S4J-K). Filled data points correspond to a statistically significant association between the indicated disease with the indicated

condensate, with the data point color corresponding to the Benjamini-Hochberg adjusted *p*-value (FDR) for the enrichment of proteins defined as components of the indicated condensate based on GO (Methods) among the set of condensate-forming proteins that have pathogenic mutations involved in the indicated disease that affect condensate-promoting features. Unfilled datapoints correspond to a lack of a statistically significant enrichment. Size of data point is proportional to the fraction of the indicated disease-associated condensate-forming proteins that are components of the indicated condensates.

Pathogenic mutations affecting condensate-promoting features alter condensate properties in cells

To confirm that a subset of pathogenic mutations identified in this catalog can affect condensate properties in cells, we selected 25 putative condensate-forming proteins spanning a range of biological functions and diseases for study (Table S3). Murine embryonic stem cells (mESCs) were selected for use in this study because cell lines can be engineered to provide a consistent cellular environment for comparisons of multiple pairs of wild-type (WT) and mutant proteins and mESCs also have proven utility in the study of condensate properties (Cho et al., 2018; Guo et al., 2019; Henninger et al., 2021; Li et al., 2020; Sabari et al., 2018). Cell lines were generated in which genes encoding the 25 wild-type proteins were stably integrated and expressed with an mEGFP tag, and these were subjected to live-cell imaging with Airyscan confocal laser-scanning microscopy (Figure 3A). As controls, we used MECP2, a validated condensate-forming protein in mESCs (Li et al., 2020), and mEGFP, which exhibits a non-punctate distribution throughout the nucleus and cytoplasm (Figure 3B). The use of any one cell type for condensate studies is naturally limiting because condensate formation can depend on cell type, environmental stress and external signals. Nonetheless, approximately half of the proteins studied in mESCs (13/25) were found concentrated within punctate structures that exhibited dynamics typically observed in condensates, and two additional proteins formed puncta in another cell line or with exposure to oxidative stress (Figure S3A-C, Table S3).

We then asked whether pathological mutations that affect condensate-promoting features within the 13 proteins that formed condensates in mESCs affect measurable properties of the observed puncta. For each of the 13 proteins, we generated at least one analogous cell line expressing a representative missense or truncation mutation (Figure 3A, Table S3). Mutations were selected to represent the approximate proportion of mutation types and affected condensate-promoting features in the catalog (Figure S3D). For each WT and mutant pair of live cell lines, the area, number, and partitioning of the corresponding protein into condensates was measured (Figure 3A, Methods). We found that 87% of mutations tested (13/15) as well as a known condensate-disrupting mutation in MECP2 (Li et al., 2020) showed qualitative and quantitative differences in the properties of WT versus mutant puncta (Figure 3C, Figure S3E, Figure 4A-C). In 11 of 13 cases, the mutations caused significant reductions in partitioning of the protein into condensates. In the remaining two cases, one mutation enhanced the ability of the proteins to associate with condensates while the other caused the protein to form puncta in other cellular locations (Figure 3C, Figure S4B, Figure S4D).

Figure 3

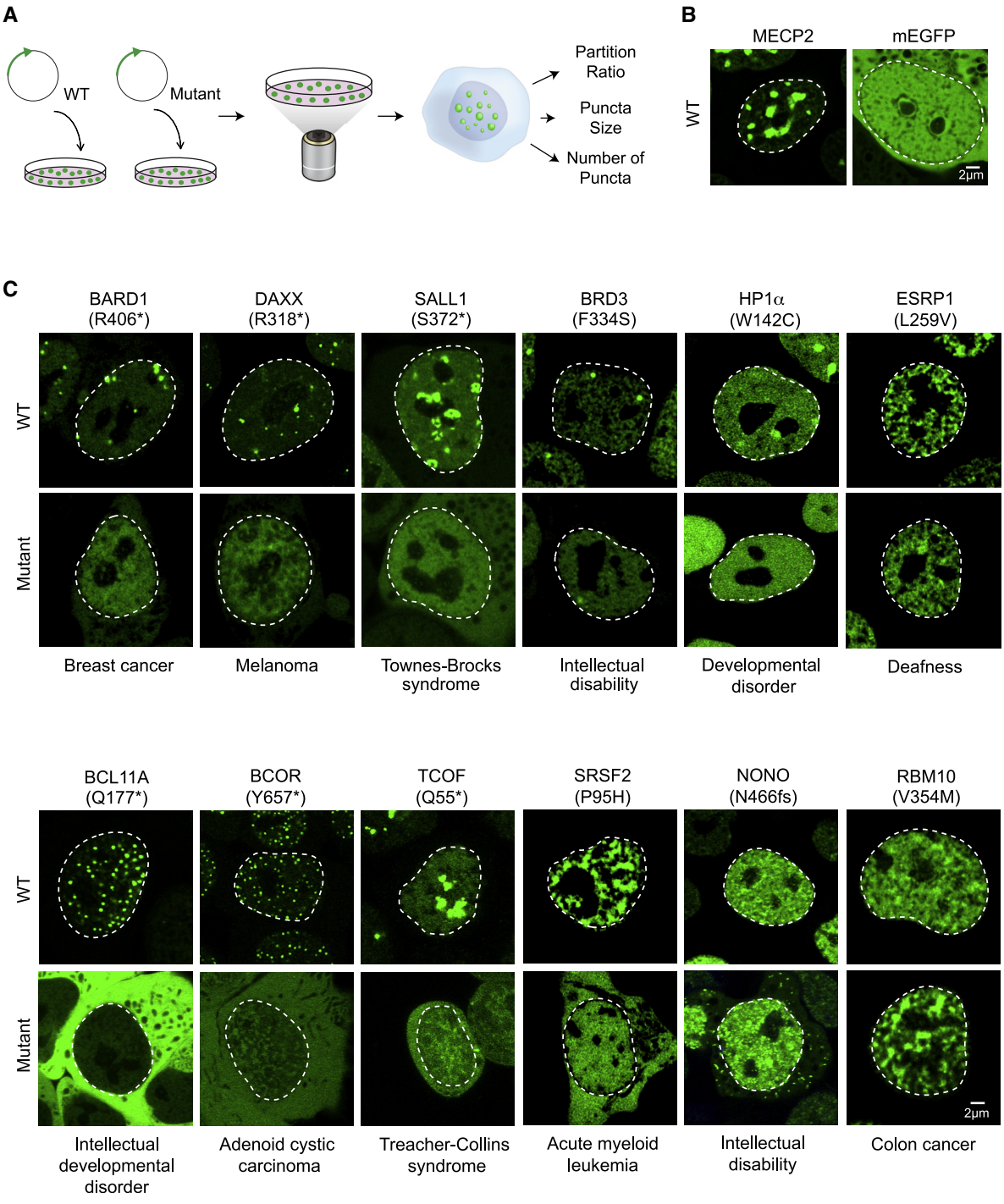


Figure 3. Pathogenic mutations in condensate-promoting features alter condensate properties in live cells. See also Figures S3-S4 and Table S3.

A. Experimental approach for testing the effect of a subset of identified mutations predicted to affect condensates. N-terminal mEGFP-tagged wild-type or mutant forms of candidate proteins were stably expressed in mESCs and condensate properties were assessed using live cell imaging and quantitative image analysis.
B. Representative images of wild-type MECP2 (positive control for condensate incorporation) mEGFP alone (negative control). Nuclei are outlined with white dashed lines.
C. Representative images of wild-type versus mutant mEGFP-tagged candidate proteins BARD1, DAXX, SALL1, BRD3, RBM10, BCL11A, NONO, BCOR, TCOF1, HP1A, SRSF2, ESRP1. Specific mutations that were tested along with their associated disease are indicated adjacent to the images.

Two lines of evidence indicate these observations in mESCs were relevant in humans: all thirteen candidates have previously been observed to occur in condensate-like puncta in at least one human cell line (Thul et al., 2017), and at least five of these occur in condensate-like punctate structures in disease-relevant human tissues or human cell lines (Figure S4F-G). While our experimental tests represent a relatively small sampling of mutations compared to the full catalog, our results suggest a predictive accuracy of the catalog to be between 60-98% (95% confidence interval) (Figure S4E). These results suggest that a substantial fraction of pathogenic mutations that were mapped to condensate-promoting features of condensate-forming proteins do produce condensate dysregulation phenotypes in cells, and that these phenotypes include reduced condensate incorporation, enhanced condensate incorporation, and altered condensate localization (Figure 4A-B).

Discussion

Much of human disease is understood through the lens of mutations affecting proteins at the molecular length scale, due in part to advances afforded by structural biology and hypotheses of protein function and disease causality that emerge from 3D structural models. This understanding plays a considerable role in therapeutic advances, as it enables medicinal chemistry employing structure-based drug design. In contrast, far less is known about how mutations affect properties that organize cellular processes at the mesoscale, such as the propensity to form biomolecular condensates, although this propensity has recently been linked to a variety of protein features. Thus, a map of known condensate-promoting protein features and the pathological mutations that affect these features could be a powerful tool in the investigation of disease mechanisms derived from disruption of this mesoscale organization.

The approach we present identified over 36,000 pathogenic mutations that plausibly contribute to condensate dysregulation across over 1,200 Mendelian diseases and 500 cancers. The premise of the approach is supported by many studies that have identified various types of MIDs and LCSs as predominant determinants of the formation and macroscopic properties of condensates as well as evidence that pathological mutations in these condensate-promoting features can lead to altered condensate properties (Banani et al., 2017; Bouchard et al., 2018; Choi et al., 2020; Molliex et al., 2015; Patel et al., 2015; Wan et al., 2019). Several observations suggest that this resource will prove to be a useful predictive guide to studying condensate-associated diseases. The analytical approach used here captured nearly 80% of known disease-causing mutations that affect condensates. Our experimental validation results show that nearly all of the thirteen tested mutations alter condensate properties in cells, and that these span condensate dysregulation phenotypes such as dissolution, enhancement, and mislocalization. Despite a small experimental sample size compared to the full catalog, our estimates suggest the predictive accuracy of the catalog to be between 60-98% (95% confidence interval). We thus expect this catalog of mutations to be substantially enriched for those that directly affect condensate properties.

This resource suggests that a substantial fraction of pathogenic mutations impart their phenotypic effects by altering the physicochemical properties of condensates that compartmentalize the diverse regulatory functions of cells. It predicts that mutations affecting condensate-promoting features of condensate-forming proteins contribute to diseases spanning all human organ systems, suggesting that the potential for novel disease mechanism discovery, therapeutic hypotheses, and consequent impact on medicine, is considerably vast. The mechanistic evaluation of these mutations will likely require evolving paradigms that address phase-separating systems across disciplines, including polymer physics, nonequilibrium thermodynamics, pharmacodynamics and pharmacokinetics, and medicinal chemistry. The therapeutic opportunities for diseases caused by condensate dysregulation have yet to be fully explored, but evidence that therapeutic small molecules can selectively interact with specific condensates (Babinchak et al., 2020; Howard and Roberts, 2020; Klein et al., 2020; Lemos et al., 2020; Viny and Levine, 2020; Wheeler et al., 2019) suggests that such therapies can be devised.

Limitations of the study

Validation of predicted condensate dysregulation is at present practically and technologically limited to relatively small experimental sample sizes. Our experimental studies do not establish a direct link between observed condensate dysregulation and the ultimate cellular and organismal defects that create the disease phenotype. Our analyses are primarily restricted to MIDs and LCSs, and while these are thought the be the predominant determinants of condensate properties, many additional protein features that we do not explicitly consider have been directly or indirectly associated with condensate properties (Bouchard et al., 2018; Gibson et al., 2019; Nott et al., 2015; Rai et al., 2018; Yoshizawa et al., 2018), suggesting that mutations not captured in our catalog may also affect condensates. We anticipate that technological and conceptual advances in condensate biology, as well as detailed molecular studies of specific proteins in disease-appropriate model systems may help to overcome these limitations in the future.

Figure 4

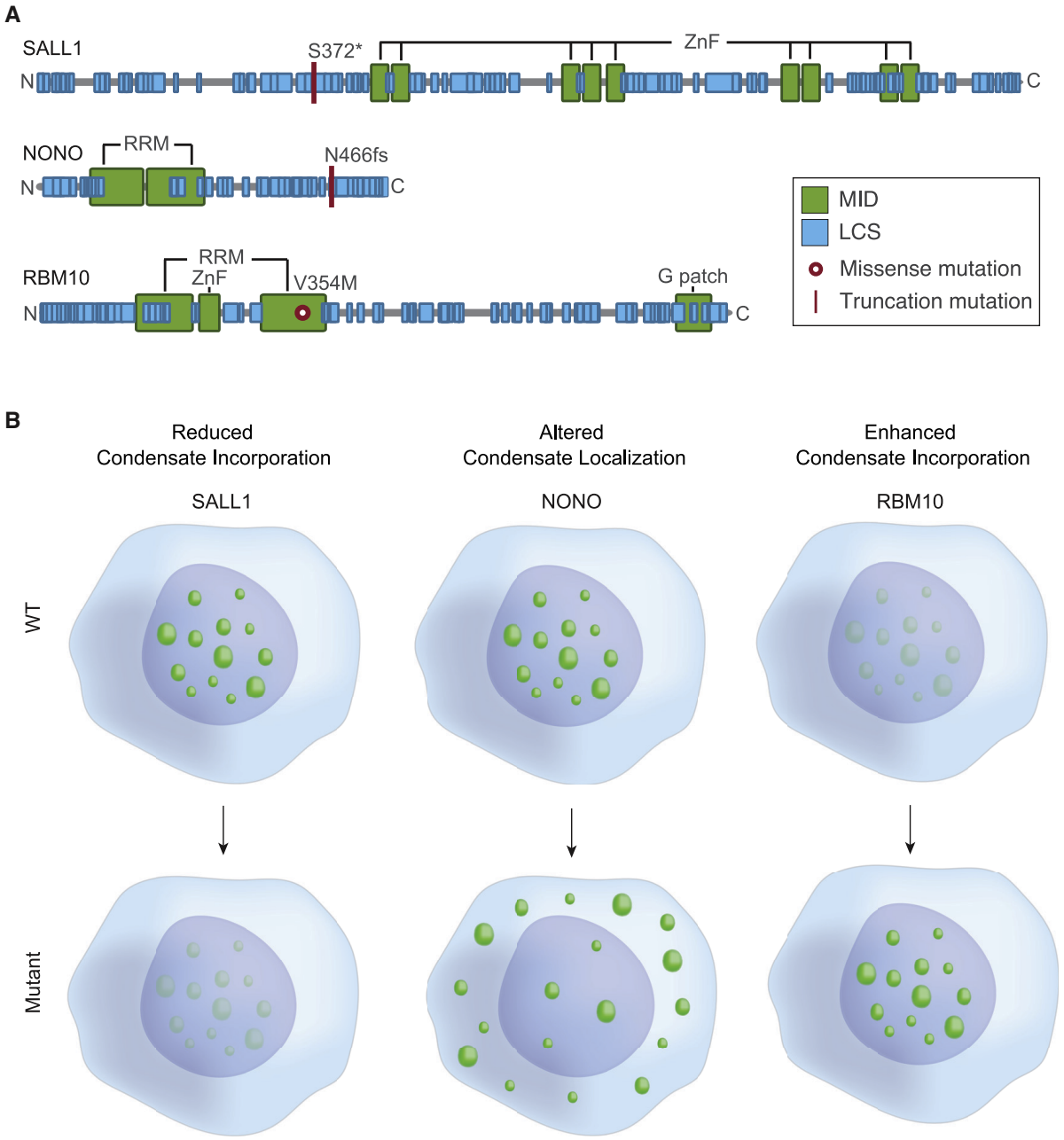


Figure 4. Mutations in condensate-promoting features cause diverse condensate dysregulation phenotypes.
Models for observed types of condensate dysregulation resulting from pathogenic mutations that affect condensate-promoting features of condensate-forming proteins, including altered condensate incorporation (left), enhanced condensate formation (middle), and altered condensate localization (right). Candidate where these phenotypes were observed (Figure 3C) are listed.

Acknowledgements

We thank Philip A. Sharp, Arup Chakraborty, Krishna Shrinivas, Brennan Decker, Fei Dong, Neal Lindeman, Isaac Klein, Charles Li, Eliot Coffey, Jie Wang, and members of the Young Lab for helpful discussions. We thank Caitlin Rausch for assistance with the graphical abstract. The work was supported by National Institutes of Health (NIH) grant R01 GM123511 (R.A.Y.), National Science Foundation (NSF) grant PHY1743900 (R.A.Y.), NIH grant 2 R01 MH104610-20 (R.A.Y.), the St. Jude Children's Research Hospital Collaborative on Cohesin, CTCF and the 3D Regulatory Nuclear Landscape of Pediatric Cancer Cells, funds from Novo Nordisk, Brigham and Women's Hospital Clinical Pathology Residency Program (S.F.B), NIH National Cancer Institute (NCI) T32 CA251062-02 (S.F.B.), NIH National Cancer Institute Award F31 CA250171 (L.K.A.), NCI F32 1F32CA254216 (J.E.H.), Hope Funds for Cancer Research (A.D.), DOD PRCRP Horizon Award W81XWH-20-10716 (V.E.C.), NREF Andrew T. Parsa Research Fellowship Grant (V.E.C.), NIH T32 5T32DK007191-45 (J.M.P.), and the Gruss-Lipper, Rothschild, and Zuckerman postdoctoral fellowships (I.S.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

Conceptualization, S.F.B, L.K.A., S.W.H., and R.A.Y.; Methodology, S.F.B, L.K.A., S.W.H. and R.A.Y.; Investigation, S.F.B, L.K.A., S.W.H., J.E.H., A.D., V.E.C., J.M.P., O.O., and N.M.H.; Formal Analysis, S.F.B., S.W.H., V.E.C., J.M.P.; Writing – Original Draft, S.F.B., L.K.A., S.W.H., and R.A.Y; Writing – Review and Editing, S.F.B., L.K.A., S.W.H., T.I.L., and R.A.Y; Visualization, S.F.B., L.K.A., S.W.H., and J.E.H.; Supervision, R.A.Y.; Funding Acquisition, R.A.Y.

Declaration of interests

R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics, and Dewpoint Therapeutics. S.F.B. and A.D. are consultants to Dewpoint Therapeutics. J.E.H. and O.O. are consultants to Camp4 Therapeutics. T.I.L. is a shareholder of Syros Pharmaceuticals and a consultant to Camp4 Therapeutics. All other authors declare no competing interests.

Figure S1

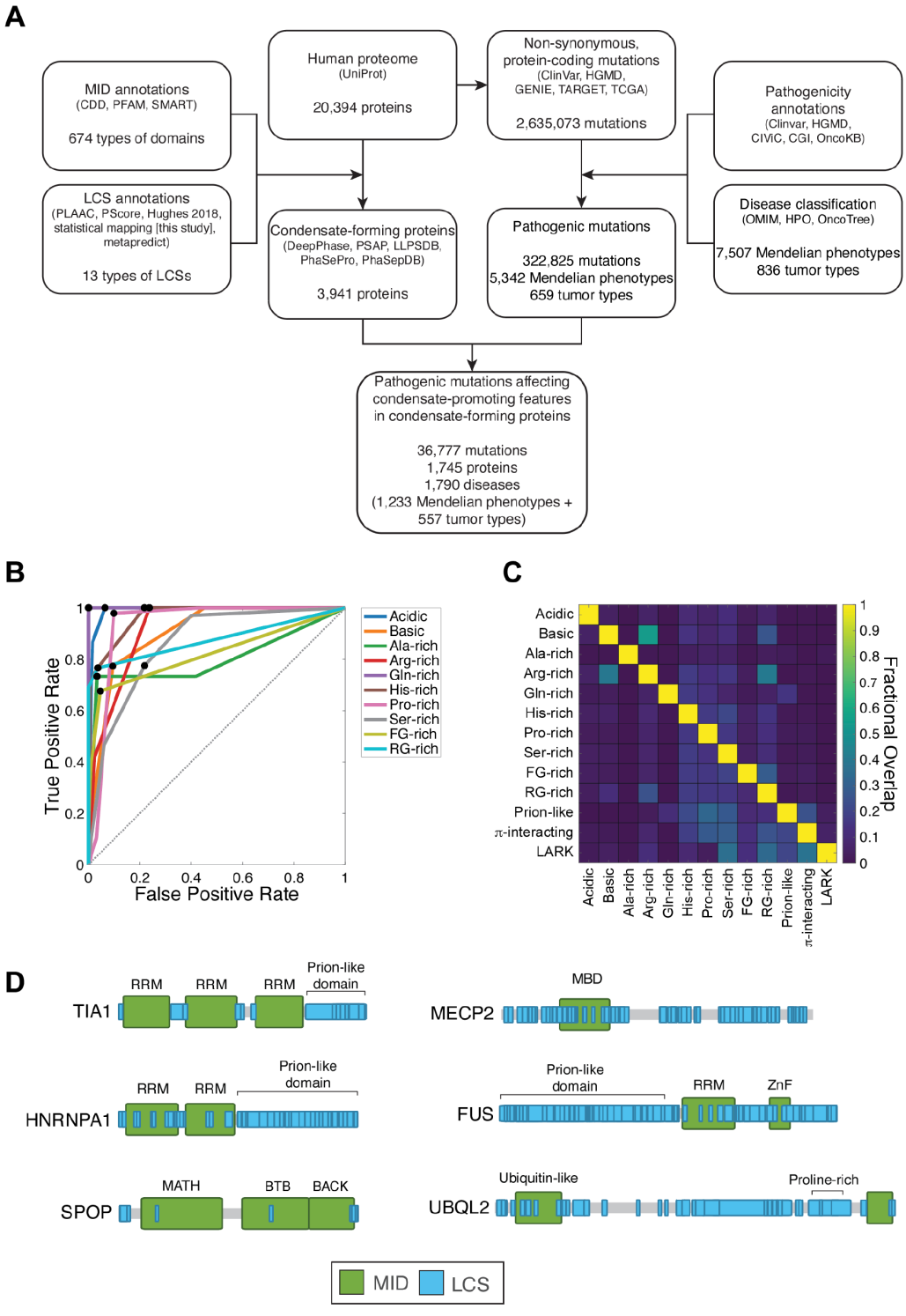


Figure S1. Mapping condensate-promoting features in condensate-forming proteins. Related to Figure 1.

A. Flow chart showing computational steps taken to define putative condensate-forming proteins, condensate-promoting features, pathogenic mutations, and disease annotations. Data sources are shown in parenthesis.
B. ROC curves for optimizing cutoffs for mapping indicated LCSs across the proteome, benchmarked against a set of validated LCSs (Supplementary Table 2).
C. Fractional overlap between two types of LCSs across the proteome. The fraction of residues mapped as a particular type of LCS (y-axis) that overlap with another type of LCS (x-axis) is indicated by the color scale.
D. Examples of known condensate-forming proteins and their condensate-promoting features, as identified by the analytical approach used in this study.

Figure S2

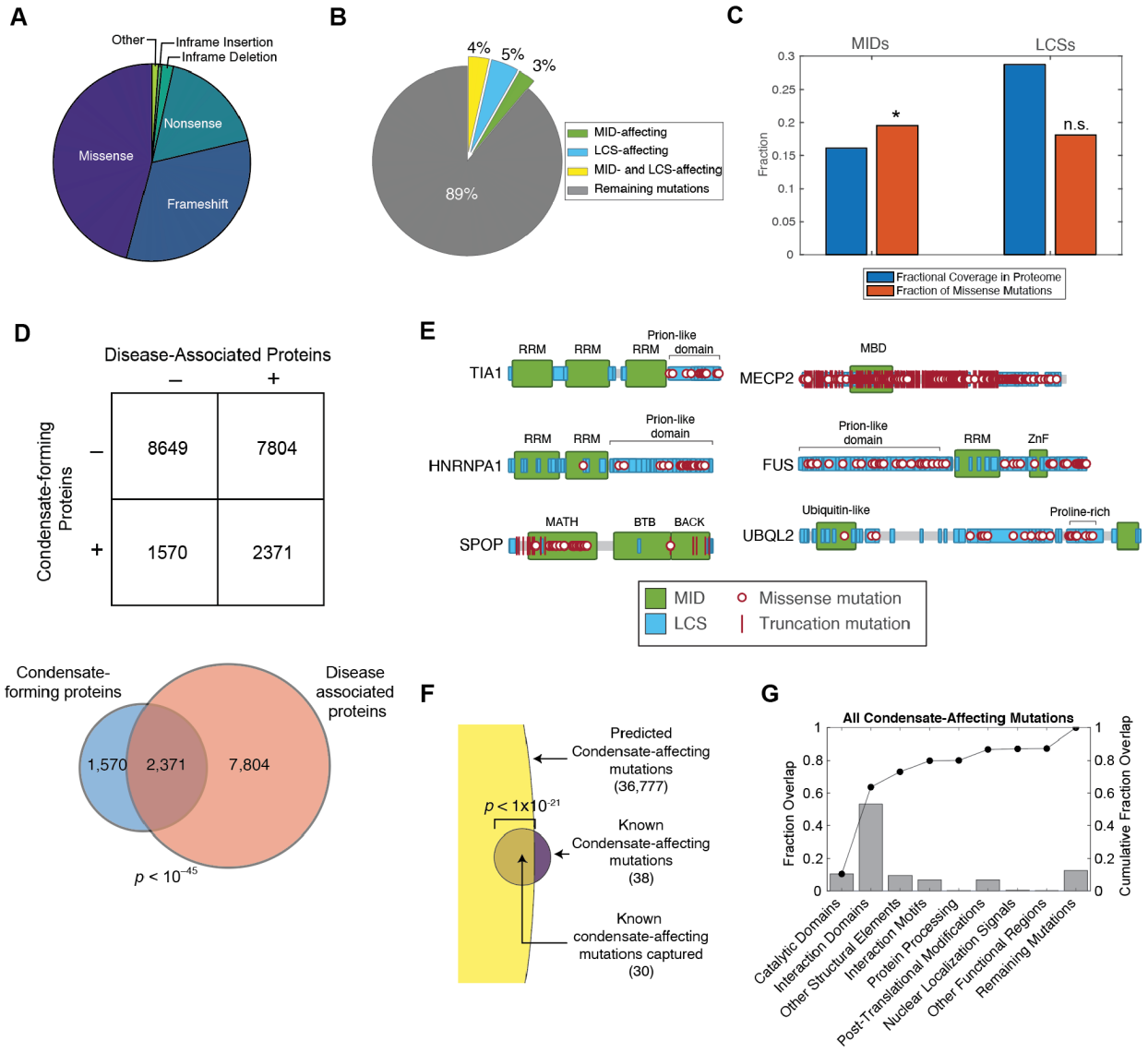


Figure S2. Mapping mutations to condensate-promoting features. Related to Figure 1.

A. Pie chart showing the types of pathogenic mutations examined in the study.
B. Pie chart showing the proportion of pathogenic mutations that affect MIDs, LCSs, or both MIDs and LCSs within putative condensate-forming proteins.
C. (Blue) Bar graph of fraction of proteome comprised of MIDs (left) or LCSs (right) along with (red) fraction of all pathogenic missense mutations that occur within MIDs or LCSs. Missense mutations were significantly enriched among MIDs (*), but not among LCSs (n.s., not significant). p-value < 10–250 and p-value ~ 1, respectively, binomial test.
D. Venn diagram showing a significant overlap between condensate-forming and proteins with pathogenic mutations (p < 10–45, Fisher exact test)
E. Examples of known condensate-forming proteins, their condensate-promoting features, and pathogenic mutations affecting these features as identified by the analytical approach used in this study.

F. Overlap of predicted condensate-affecting mutations with known condensate-affecting mutations curated from the literature (see also Table S4).
G. Bar graph of the cumulative proportion of pathogenic condensate-affecting mutations that overlap with canonical molecular-scale models. Moving from left to right, mutations in multiple categories were not double counted. Mutations with no overlap are denoted as “remaining mutations”.

Figure S3

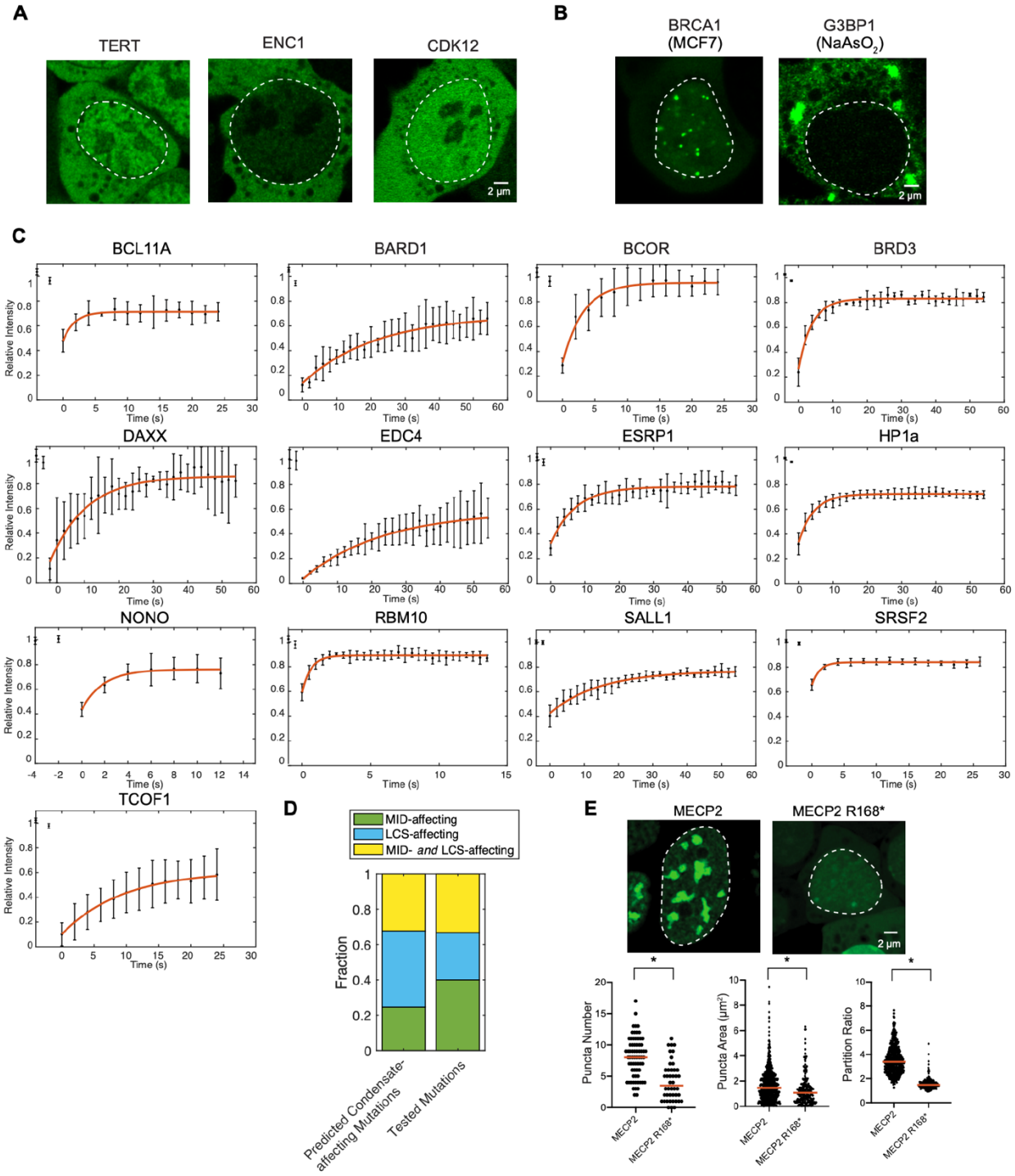


Figure S3. Pathogenic mutations in condensate-promoting features alter condensate properties in live cells. Related to Figure 3.

A. Representative images of 3 of the 12 mEGFP-tagged candidate proteins that did not incorporate into punctate structures indicative of condensates in mESCs. Nuclei indicated with dotted white line. See also Table S5.

B. Representative images of mGFP-tagged BRCA1 in MCF7 cells (left) and mEGFP-tagged G3BP1 in NaAsO2 treated mESCs (right), showing incorporation into punctate structures. Both candidates were found to be not punctate in unstressed mESCs. Nuclei are shown with white dotted lines. See also Supplementary Table 5.
C. Fluorescence recovery after photobleaching (FRAP) of 13 candidate proteins that incorporated into punctate structures in mESCs. Fluorescence intensities normalized to pre-bleach values are plotted over time. Data points represent the mean and error bars represent standard deviation of normalized fluorescent values over n = 3-4 puncta in independent cells that were analyzed for each candidate.
D. Fraction of predicted condensate-affecting mutations (left) or those selected for experimental evaluation (right) that were MID-affecting, LCS-affecting, or MID- and LCS-affecting.
E. Representative images (top) and quantification of puncta number (bottom, left), puncta area (bottom, middle), and partition ratio (bottom, right) for mEGFP-tagged MECP2 and mEGFP-tagged MECP2 R168* mutant in mESCs.

Figure S4

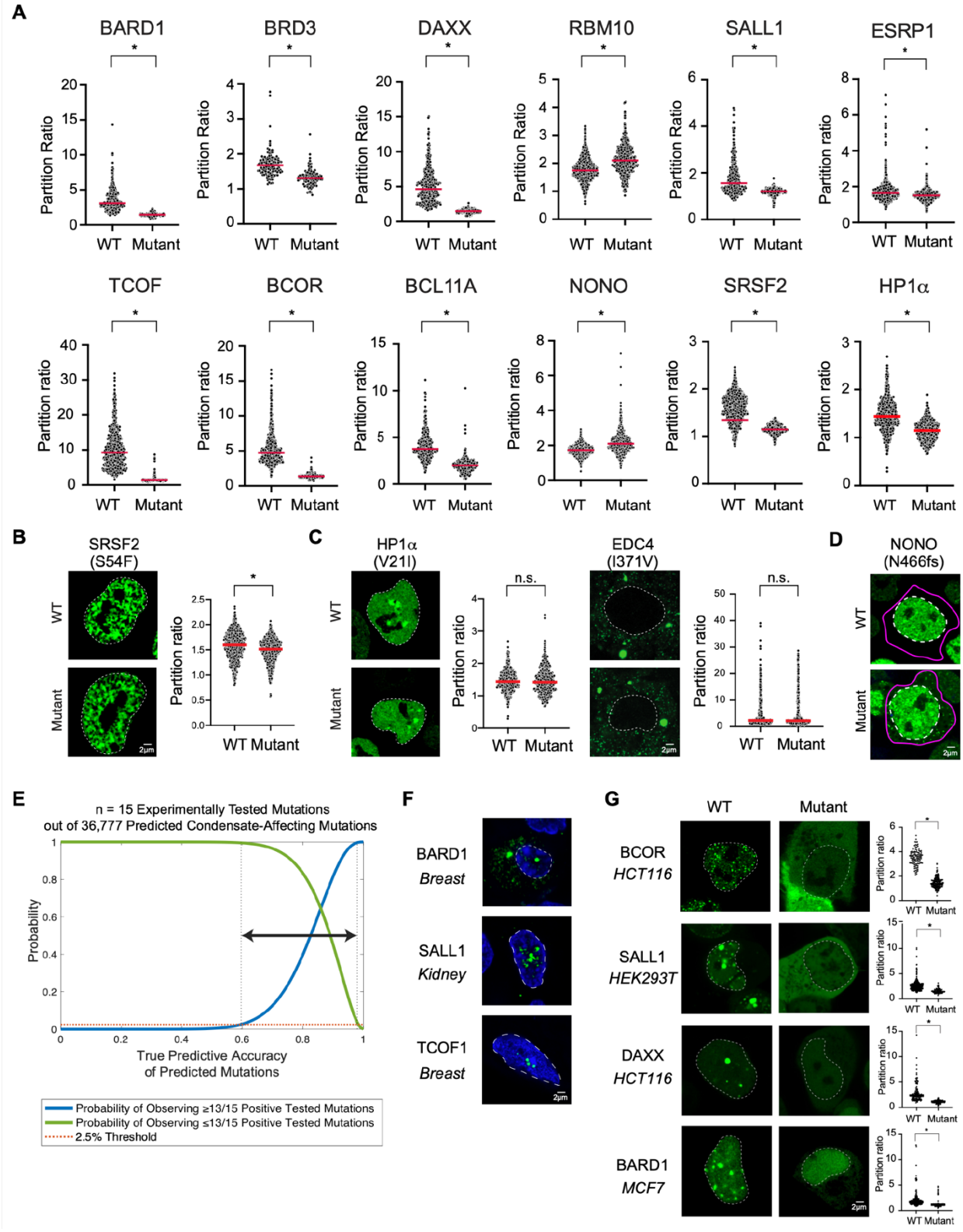

A
BARD1
BRD3
DAXX
RBM10
SALL1
ESRP1
TCOF
BCOR
BCL11A
NONO
SRSF2
HP1α
Partition Ratio
Partition ratio
WT
Mutant
B
SRSF2 (S54F)
C
HP1α (V21I)
EDC4 (I371V)
n.s.
D
NONO (N466fs)
E
n = 15 Experimentally Tested Mutations
out of 36,777 Predicted Condensate-Affecting Mutations
Probability
True Predictive Accuracy of Predicted Mutations
Probability of Observing ≥13/15 Positive Tested Mutations
Probability of Observing ≤13/15 Positive Tested Mutations
2.5% Threshold
F
BARD1 Breast
SALL1 Kidney
TCOF1 Breast
G
BCOR HCT116
SALL1 HEK293T
DAXX HCT116
BARD1 MCF7
2μm

Figure S4. Mutations in condensate-promoting features cause diverse condensate dysregulation phenotypes. Related to Figure 3.

A. Quantification of partition ratios of wild-type and mutant mEGFP-tagged cell lines for indicated candidates. *, p-values < 0.0001. See also Table S6.
B. Representative images (left) and quantification (right) for a second SRSF2 mutant tested in mEGFP-tagged mESC lines, showing a mild but significant effect on partitioning. *, p-value < 0.0001.
C. Representative images and quantification for a second HP1a mutant (left) and a WT/mutant pair for EDC4 (right) showing no significant difference in partitioning between the WT versus the mutant protein.
D. Representative images of mESC lines expressing WT or mutant NONO, reproduced from Fig. 3c, showing the cytoplasmic boundary of the cell of interest (magenta). Nuclear outline is shown as a white dashed line.
E. Estimation of the 95% confidence interval of the accuracy of the predicted catalog of condensate-affecting mutations. Under the assumption of randomly selected mutations, the probability of observing ≥ 13/15 (blue) or ≤ 13/15 (green) positive experimental outcomes from a catalog of n = 36,777 mutations as a function of all possible true catalog accuracies. Probabilities are computed using a hypergeometric distribution with the assumption of random sampling. Red dashed line indicates probability of 2.5% and its intersections with the red and blue curves denote an estimate for the 95% confidence interval for the true catalog accuracy to be between 60-98%.
F. Representative images of BARD1, SALL1, and TCOF1 immunofluorescence in human breast, kidney, and breast tissue, respectively. Nuclear outline is shown as a white dashed line.
G. Representative images of disease-relevant human cell lines expressing WT or mutant BCOR, SALL1, DAXX, and BARD1 (left) with quantification of partition ratios of wild-type and mutant cell lines for indicated candidates (right). *, p-values < 0.0001. Specific cell lines are shown adjacent to the image. Nuclear outline is shown as a white dashed line.

Table S1. LCSs used in this study and optimization of LCS mapping against a set of benchmark proteins. Related to Figure 1.

LCS	Amino acids	Test Protein	Curated coordinates	Refs.	Mapped coordinates	AUC	Optimal Frequency in 5-amino acid Window	TPR	FPR
Acidic patch	Asp, Glu	NPM1	34-39, 120-130, 161-188	(Mitrea et al., 2016)	118-134,159-189	0.99	0.5	1.0	0.06
Basic patch	Arg, Lys	MECP2	170-181, 184-194, 246-258, 263-274, 282-289, 301-310, 340-348	(Li et al., 2020)	26-39,105-115, 167-178, 249-258, 264-272,302-310	0.91	0.5	0.77	0.09
Alanine-rich region	Ala	HXD13	57-71	(Basu et al., 2020)	18-39,46-75,92-126	0.80	1.0	0.73	0.03
Argininerich region	Arg	SURF6	29-31, 56-58, 81-82, 118-120, 145-148, 152-159, 216-217, 221-223, 234-237, 246-249, 299-306, 321-326, 330331, 336-345	(Mitrea et al., 2016, 2018)	142-163, 213-225, 233-250, 295-310, 320-349	0.92	0.3	1.0	0.24
Glutaminerich region	Gln	HTT	17-40	(Peskett et al., 2018)	15-41, 49-68, 498-506, 593-601	1.00	0.5	1.0	$3x10^{-4}$
Histidinerich region	His	DYR1A	590-616	(Lu et al., 2018)	531-541, 596-622, 648-656	0.97	0.1	1.0	0.22
Proline-rich region	Pro	UBQL2	491-538	(Dao et al., 2018)	7-17, 313-327, 470-478, 490-537, 575-580	0.94	0.3	0.98	0.10
Serine-rich region	Ser	MED1	1078-1482	(Sabari et al., 2018)	808-816, 10211029, 1077-1150, 1163-1171, 12231285, 1366-1375, 1463-1471, 1532-1543	0.86	0.3	0.78	0.22
FG-rich region	Gly, Phe	NUP98	1-469	(Schmidt and Görlich, 2015)	3-68, 75-102, 113153, 224-278, 286-306, 315-395, 405-424, 433-453, 461-483, 871-884, 1052-1060	0.83	0.3	0.69	0.05
RG-rich region	Arg, Gly	FUS	211-285, 371-422, 453-526	(Wang et al., 2018)	211-222, 230-270, 375-411, 469-507	0.88	0.3	0.77	0.04

Prion-like domain	-	-	-	(Lancaster et al., 2014; Martin et al., 2020; Wang et al., 2018)	-	-	-	-	-
piinteracting residues	-	-	-	(Vernon et al., 2018)	-	-	-	-	-
LARKS	-	-	-	(Hughes et al., 2018)	-	-	-	-	-

List of LCSs within IDRs associated with condensate formation used in this study. Preexisting approaches were used when available to map certain LCSs across the proteome (prion-like domains, pi-interacting residues, and LARKS). The remainder of LCSs were mapped using a statistical approach (Methods) benchmarked against protein regions corresponding to previously characterized LCSs in test proteins identified in prior studies. These curated LCSs were used as 'gold standards' for benchmarking our LCS mapping parameter of frequency of corresponding amino acid types within 5-amino acid windows (see Methods). The optimal cutoff and its performance are indicated. AUC, area under the curve from ROC curves (Figure S1B). FG, phenylalanine-glycine; RG, argininegiycine; LARKS, low-complexity, aromatic-rich kinked segments; AUC, area under the curve; TPR, true positive rate; FPR, false positive rate; Refs., references.

Table S2. Pathological mutations known to affect condensates. Related to Figure 3.

Protein	Diseases	Mutations	References	Notes for uncaptured mutations
ANXA11	Amyotrophic lateral sclerosis	**Asp40Gly**, **Gly38Arg**, **His390Pro**, **Arg456His**	(Nahm et al., 2020)	
aSYNUCLEIN	Parkinson's disease	Ser129Glu, Ala53Thr, Glu46Lys	(Ray et al., 2020)	Mutation not in variant databases sourced in this study; Mutation does not affect defined MIDs or LCSs
DDX3X	Medulloblastoma, Intellectual disability	Ala222Pro, **Thr275Met**, Gly302Val, Leu353Phe, Met370Arg, Leu351Trp, Leu556Ser	(Fonseca et al., 2021; ValentinVega et al., 2016)	Mutation not in variant databases sourced in this study; Mutation does not affect defined MIDs or LCSs
FUS	Amyotrophic lateral sclerosis and frontotemporal lobar dementia	**Gly399Val**, **Gly187Ser**, **Gly156Glu**	(Burke et al., 2015; Patel et al., 2015)	
HNRNPA1	Amyotrophic lateral sclerosis	***Asp314Val**	(Molliex et al., 2015)	
HNRNPDL	Limb-girdle muscular dystrophy 1G	**Asp378His**, **Asp378Asn**	(Batlle et al., 2020)	
KEAP1	Lung squamous cell carcinoma	Arg320Glu, **Arg470Cys**	(Cloer et al., 2018)	Mutation not in variant databases sourced in this study
MECP2	Rett syndrome	**Arg168Ter**, **Arg255Ter**, **Arg270Ter**, **Arg294Ter**, Pro389Ter, **Arg306Cys**, Pro322Leu, Pro225Arg	(Li et al., 2020)	Mutation does not affect defined MIDs or LCSs
MLL4	Kabuki Syndrome	Gln4092Ter	(Fasciani et al., 2020)	Mutation not in variant databases sourced in this study
NF2	Cancers	Leu46Arg, Leu64Pro, Leu141Pro	(Meng et al., 2021)	Protein not captured in set of defined condensate-forming proteins
RBM20	Congenital dilated cardiomyopathy	**Arg636Ser**	(Schneider et al., 2020)	
SHP2	Noonan syndrome	**Glu76Lys**, **Arg498Leu**, Gln506Pro, Gly464Ala, **Tyr279Cys**, Tyr468Met	(Zhu et al., 2020)	Mutation not in variant databases sourced in this study; Mutation does not affect defined MIDs or LCSs
SPOP	Prostate cancer	**Phe133Val**, **Trp131Gly**	(Bouchard et al., 2018)	
TDP43	Amyotrophic lateral sclerosis	**Ala321Gly**, **Ala321Val**, Glu331Lys, **Met337Val**, Ala326Pro, Met337Pro	(Conicella et al., 2016)	Mutation not in variant databases sourced in this study;
TIA1	Amyotrophic lateral sclerosis and frontotemporal dementia	**Pro362Leu**, **Ala381Thr**, **Glu384Lys**	(Mackenzie et al., 2017)	
TAU	Alzheimer's disease	Pro301Leu	(Kanaan et al., 2020)	Mutations not in variant databases sourced in this study
UBQL2	Amyotrophic lateral sclerosis	**Pro506Ser**, **Pro506Thr**, **Pro506Ala**, Thr487Ile, **Pro497Leu**, **Pro497His**, **Pro497Ser**	(Dao et al., 2019; Sharkey et al., 2018)	Mutation does not affect defined MIDs or LCSs

Known pathological mutations that affect condensates curated from the literature. **Bolded** mutations, were captured among the set of mutations predicted to affect condensates in the catalog. Reasons for why certain mutations were not captured in our catalog are mentioned in the right-most column. *, this HNRNPA1 mutation is described in the corresponding study as Asp262Val, affecting a non-canonical isoform and maps to position 314 in the canonical isoform used in our analyses. See also Figure S2D.

Table S3. Selected protein candidates and mutations used in experimental tests. Related to Figure 3.

Protein	cDNA source	Disease(s)	Distribution in mESCs	Selected Mutation(s)	Mutation effect
MECP2	(Li et al., 2020)	Rett syndrome	Punctate	Arg186Ter	Reduced condensate incorporation
BARD1	MHS6278-211689242	Breast, ovarian, prostate, Pancreatic cancers	Punctate	Arg406Ter	Reduced condensate incorporation
BCL11A	Addgene 139809	Intellectual development disorders	Punctate	Gln177Ter	Reduced condensate incorporation
BCOR	MHS6278-202757783	Various cancers	Punctate	Tyr657Ter	Reduced condensate incorporation
BRD3	Unpublished	Intellectual disability	Punctate	Phe334Ser	Reduced condensate incorporation
HP1a	(Li et al., 2020)	Developmental disorder, Autism	Punctate	Val21Ile, Trp142Cys	Reduced condensate incorporation
DAXX	Addgene 52023	Various cancers	Punctate	Arg318Ter	Reduced condensate incorporation
EDC4	Addgene 66597	Congenital heart disease	Punctate	Ile371Val	Reduced condensate incorporation
ESRP1	MHS6278-202833454	Deafness, Ear malformation	Punctate	Leu259Val	Reduced condensate incorporation
NONO	Addgene 127655	Developmental delay	Punctate	Asn466fs	Enhanced condensate incorporation; Altered condensate localization
RBM10	Addgene 81958	Lung, bladder, colon, pancreatic cancers	Punctate	Val354Met	Enhanced condensate incorporation
SALL1	MMM1013-202859719	Townes-Brocks syndrome	Punctate	Ser372Ter	Reduced condensate incorporation
SRSF2	(Guo et al., 2019)	Acute myeloid leukemia, Myelodysplastic syndrome	Punctate	Ser54Phe, Pro95His	Reduced condensate incorporation
TCOF	MHS1010-202695722	Treacher-Collins syndrome	Punctate	Gln55Ter	Reduced condensate incorporation
ASXL1	MHS6278-213245938	Acute myeloid leukemia, Myelodysplastic syndrome	Not punctate	-	-
BCL6	Addgene 81869	Various cancers	Not punctate	-	-
BRCA1	Addgene 14999	Breast, ovarian cancers	Not punctate	-	-
DVL2	Addgene 24802	Neural tube defects	Not punctate	-	-
DYR1A	Addgene 101770	Autism, Intellectual disability	Not punctate	-	-
ENC1	MHS6278-202826591	Autism	Not punctate	-	-
G3BP1	Addgene 127104	Autism	Not punctate	-	-
HMGA2	Addgene 52727	Silver-Russel syndrome	Not punctate	-	-
NIPBL	Addgene 107716	Cornelia-deLange syndrome	Not punctate	-	-
NKX21	Addgene 119173	Choreoathetosis	Not punctate	-	-
SNCAP	MHS6278-202809062	Parkinson disease	Not punctate	-	-
TERT	Addgene 114315	Dyskeratosis congenita	Not punctate	-	-

25 protein candidates from catalog selected for experimental study, not including MECP2, which was used as a positive control (Li et al., 2020). cDNA source indicates Addgene catalog number, cDNA clone ID (Team et al., 2009), or a prior study. mESCs, mouse embryonic stem cells. See also Supplemental Discussion, *Selection of Candidates and Mutations for Experimental Validation*.

Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Richard A. Young (young@wi.mit.edu).

Materials availability

All unique/stable reagents generated in this study are available from the Lead Contact upon reasonable request with a completed Materials Transfer Agreement.

Data and code availability

Datasets used in this study are available from the sources cited in the Key Resources Table. Processed data used in the study is available in the Supplemental Tables and on the Young laboratory's website. Code generated during this study is available via https://github.com/bananisf/2022_Banani_Afeyan_Hawken_Condensate_Dysregulation.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

V6.5 murine embryonic stem cells (mESCs) were a gift from R. Jaenisch. Human cell lines (HCT116, MCF7, and HEK293T) were obtained from ATCC.

Cell culture conditions

mESCs were cultured in 2i/LIF media on tissue culture treated plates coated with 0.2% gelatin (Sigma G1890) in a humidified incubator with 5% CO2 at 37°C. Cells were passaged every 2-3 days using TrypLE Express (Gibco) for dissociation quenched with serum/LIF media. 2i/LIF media: DMEM/F12 (Gibco) supplemented with 1x N2 and 1x B27 (Gibco), 1x MEM Non-essential amino acids (Gibco), 100 U/mL Penicillin-Streptomycin (Gibco), 1mM L-glutamine (Gibco), 0.25% BSA Fraction V (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 3 μM CHIR99021 (Stemgent), 1 μM PD0325901 (Stemgent), and 1000 U/mL leukemia inhibitor factor (LIF) (ESGRO). Serum/LIF media: KnockOut DMEM (Gibco) supplemented with 15% fetal bovine serum (Sigma), 2 mM L-glutamine (Gibco), 1x MEM non-essential amino acids (Gibco), 100 U/mL penicillin-streptomycin (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), and 1000 u/mL leukemia inhibitor factor (LIF) (ESGRO).

Human cell lines HCT116 (ATCC), MCF7 (ATCC) and HEK293T (ATCC) were cultured in complete DMEM media (DMEM (Life Technologies 11995073), 10% Fetal Bovine Serum, FBS (Sigma Aldrich, F4135), 1% L-glutamine (GIBCO, 25030-081), 1% Penicillin Streptomycin (Life

Technologies, 15140163) at 37°C with 5% CO2 in a humidified incubator. For passaging, cells were washed in PBS (Life Technologies, AM9625) and TrypLE Express Enzyme (Life Technologies, 12604021) was used to detach cells from plates by incubating them in TrypLE at 37°C with 5% CO2 in a humidified incubator for 5 minutes. TrypLE was quenched with complete DMEM, described above, and cells were plated in new tissue culture-grade plates.

Human tissue samples

Human tissue samples were purchased from bioIVT. Healthy human breast tissue was from a 27-year-old female with infiltrating lobular carcinoma of the breast and stage group T4N2M0. Health human kidney tissue was from a diabetic 28-year-old male.

Human tissue sample storage conditions

Human tissue samples were kept frozen, embedded in OCT, and stored at –80°C until use.

METHOD DETAILS

Definition of Condensate-Forming Proteins

The set of 20,394 *Homo sapiens* proteins and their sequences in the UniProt/SwissProt database v2020_06 (Consortium et al., 2020) were defined as the human proteome in this study (Figure S1A, Table S4A). The list of known and predicted condensate-forming proteins was defined by integrating: (i) proteins with a DeepPhase (Yu et al., 2020) score (based on analysis of proteome-wide immunofluorescence) of ≥ 0.9, a validated threshold provided by the developers of the algorithm; (ii) proteins scoring within the top 10% of PSAP scores (Mierlo et al., 2021) (based on proteome-wide sequence-based analysis); and (iii) known phase-separating proteins curated from LLPSDB (Li et al., 2019) (accessed January, 2021), PhaSePro (Mészáros et al., 2019) (accessed January, 2021), and PhaSepDB v1.3 (You et al., 2019). The resulting list of proteins is shown in Table S4E.

Mapping of Protein Features

Multiple classes of canonical and condensate-promoting protein features were mapped onto the proteome as follows, with the mapping results provided in Table S4B.

Domains. Integrated domain annotations from CDD v3.18 (Lu et al., 2019), PFAM v33.1 (Mistry et al., 2020), SMART v7.1(Letunic et al., 2020) corresponding to the UniProt entries for the set of human proteins were obtained via InterPro v83.0 (Blum et al., 2020). We found that using integrated annotations provides a more comprehensive mapping of domains across the proteome. The integration in InterPro ensures that the same instance of a given domain within a protein from multiple domain databases is annotated as a single entry. Related domains (e.g. different subtypes of SH3 domains) were further grouped into a single parent entry using hierarchies provided in InterPro. *Catalytic domains* were defined as those having a molecular function annotation in InterPro including "*enzyme activity*" or ending with "*-ase activity*". For the analyses in Figure S2G, protein regions with UniProt annotations of *active site* were also included within the catalytic domains category. *Interaction Domains* (or *Modular Interaction Domains* [*MIDs*], as we refer to them in the context of condensate formation) were defined by combining the following subsets of domains: (i) domain entries that had any of the following molecular function annotations in InterPro: *protein complex*, *oligomerization*, *protein dimerization activity*, *protein tetramerization*, *protein homodimerization activity*, *protein heterodimerization activity*, *protein homooligomerization*, *DNA binding*, *RNA binding*, *protein binding*, *nucleic acid binding*,

actin binding, *microtubule binding*, *actin filament binding*, *histone binding*, *integrin binding*, *clathrin binding*, *cellulose binding*, *telomeric DNA binding*, *cadherin binding*, *starch binding*, *protein kinase binding*, *ubiquitin binding*, *tubulin binding*, *cytoskeletal protein binding*, *collagen binding*, *chitin binding*, *mismatched DNA binding*, *chromatin binding*, *double-stranded DNA binding*, or *phospholipid binding*; (ii) domain entries that matched (based on manual assessment) curated lists of domains from the literature corresponding to head-to-tail interacting domains (Bienz, 2020), RNA-binding domains (Hentze et al., 2018; Lunde et al., 2007), DNA-binding domains in transcription factors (Lambert et al., 2018; Vaquerizas et al., 2009), protein-protein interaction domains found in cell signaling (Pawson and Nash, 2003; Seet et al., 2006) (including those that recognize PTMs), or domains that recognize histone modifications (Yun et al., 2011); and (iii) manually curated list of domains not included in (i) or (ii) assembled from prior knowledge of or domain descriptions in InterPro documenting their known or suspected involvement in binding interactions. We note that the same list of MIDs were used for analyses involving canonical protein properties in Figure S2G as well as for the mapping of condensate-promoting features within condensate-forming proteins (Figure 1B, Figure S1A), given the known roles of MIDs in both canonical protein-protein interactions and in condensate formation. The final list of MIDs used in this study is provided in Table S4F.

It is important to note that the because the definition of a domain primarily relies on sequence conservation, it does not explicitly consider predictions of structure or disorder. While it turns out that most of the amino acid residues within most domains are often demonstrated or predicted to be folded/structured, this does not preclude occasional overlaps of domains with predicted intrinsically disordered regions and LCSs therein (see below).

Structural Elements. Structural elements were defined by integrating the following protein annotations: (i) all protein regions annotated in UniProt (Consortium et al., 2020) as having structural elements *helix*, *beta strand*, *turn*, *disulfide bond*, and *coiled coil*; and (ii) all conserved domains (see above), which are often structured, that did not meet the definition of catalytic domain or interaction domain, filtered to remove any regions predicted to be intrinsically disordered (see below).

Interaction Motifs. Interaction motifs were defined by integrating: (i) short linear interacting motifs (SLiM) annotations corresponding to UniProt entries from the ELM database (Kumar et al., 2019) (accessed November, 2020); (ii) all protein regions annotated in UniProt as *region* or *motif* (with the exception of those with a description including “Nuclear localization signal”); and (iii) molecular recognition features (MoRFs) that undergo coupled folding upon binding from MFIB database (Fichó et al., 2017) (version 26-06-2017). For (i), SLiMs were filtered to only include those annotated with the logic of *true positive*.

Protein Processing. Regions involved in protein processing were defined by using regions annotated in UniProt as *peptide, signal peptide*, *transit peptide*, *propeptide*, and *initiator methionine*.

Post-translational Modifications. Sites of PTMs were defined by integrating: (i) all protein regions annotated in UniProt as *modified residue*, *lipidation*, *glycosylation*, and *cross-link*; and (ii) all PTM sites corresponding to UniProt entries in PhosphoSitePlus (Hornbeck et al., 2015) (accessed November, 2020).

Nuclear Localization and Nuclear Export Signals. NLSs and NESs were defined by integrating: (i) all protein regions annotated in UniProt as *motif* and with a description containing “Nuclear localization signal”; and (ii) NLS and NES sites corresponding to UniProt in NLSdb (Bernhofer et al., 2017) (accessed November, 2020). For (ii), NLSs and NESs were filtered to only include those annotated as *Experimental* or *By Expert*.

Other Functional Regions. The miscellaneous category of *other functional regions* was defined by integrated all other protein regions annotated in UniProt presumed to be susceptible to disruption by mutation. The following annotations were integrated to define these regions: *site, binding site, metal binding, calcium binding, DNA binding, nucleotide binding,* and *mutagenesis*.

Intrinsically Disordered Regions. IDRs were mapped using metapredict (Emenecker et al., 2021) using a threshold of 0.2, which was within the recommended range of cutoffs suggested by the developers of the algorithm.

Low Complexity Sequences. The list of thirteen types of LCSs used in this study was assembled manually from literature evidence of their involvement in IDR-mediated phase separation (Table S1). Prion-like domains were mapped using PLAAC (Lancaster et al., 2014) with core length set to 60 and relative weighting of background probabilities set to 100, as done in prior work that globally mapped this LCS type across the proteome (Wang et al., 2018). pi-pi interacting residues were mapped using PSP (Vernon et al., 2018), and residues with a PScore of > 4, based on the confidence thresholds provided in the algorithm, were considered to be LCSs of this type. LARKS were obtained from a prior study (Hughes et al., 2018). For the remaining types of LCSs, established, validated approaches for mapping these LCSs to our knowledge do not exist to date, so an approach based on functions for analogous purposes in localCIDER (Holehouse et al., 2017) and on a previously applied procedure (Li et al., 2020) was used to map these LCS regions. Amino acid compositions in sliding 5-residue windows were computed for each protein. LCS regions were defined as stretches of ≥ 5 consecutive residues (at minimum 1 window length) that consisted of characteristic residues corresponding to the particular type of LCS occurring at a frequency above a predefined threshold, set as described below (Table S1). All identified regions were filtered for those that occurred within predicted IDRs, determined as described above. This approach performed well when benchmarked against a set of experimentally validated condensate-promoting LCSs, with a receiver operating characteristic (ROC) area under the curve (AUC) ranging from 0.8-1.0 across the mapped LCSs (Figure S1B, Table S1). Optimal cutoffs for frequencies of the characteristic amino acids within 5-residue window were determined for each LCS from the ROC curve as the point of minimum Euclidean distance from perfect performance (0% false positive rate, 100% true positive rate) (Figure S1B). LCS mapping results and the overlap between the different types of LCSs identified are shown in Figure S1C, Table S1, and Table S4G.

Annotation of Mendelian and Cancer Mutations

Variants associated with Mendelian diseases were obtained from ClinVar (Landrum et al., 2017) (accessed January 29, 2021) and HGMD v2020.4 (Stenson et al., 2020) in hg38. Cancer variants were obtained from AACR Project GENIE v8.1 (Consortium, 2017) and various TCGA and TARGET studies via cBioPortal (Cerami et al., 2012; Hoadley et al., 2018) (accessed January, 2021) (Figure S1A) (cBioPortal study identifiers: *ucec_tcga_pan_can_atlas_2018*,*skcm_tcga_pan_can_atlas_2018*,*coadread_tcga_pan_can_atlas_2018*,*luad_tcga_pan_can_atlas_2018*,*stad_tcga_pan_can_atlas_2018*,*lusc_tcga_pan_can_atlas_2018*,*blca_tcga_pan_can_atlas_2018*,*brca_tcga_pan_can_atlas_2018*,*hnsc_tcga_pan_can_atlas_2018*,*cesc_tcga_pan_can_atlas_2018*,*gbm_tcga_pan_can_atlas_2018*,*lihc_tcga_pan_can_atlas_2018*,*ov_tcga_pan_can_atlas_2018*,*lgg_tcga_pan_can_atlas_2018*,*esca_tcga_pan_can_atlas_2018*,*prad_tcga_pan_can_atlas_2018*,*paad_tcga_pan_can_atlas_2018*,*kirp_tcga_pan_can_atlas_2018*,*kirc_tcga_pan_can_atlas_2018*,*sarc_tcga_pan_can_atlas_2018*,*thca_tcga_pan_can_atlas_2018*,*acc_tcga_pan_can_atlas_2018*,*ucs_tcga_pan_can_atlas_2018*,*laml_tcga_pan_can_atlas_2018*,*dlbc_tcga_pan_can_atlas_2018*,*thym_tcga_pan_can_atlas_2018*,*meso_tcga_pan_can_atlas_2018*,*kich_tcga_pan_can_atlas_2018*,tgct_tcga_pan_can_atlas_2018,*chol_tcga_pan_can_atlas_2018*,*pcpg_tcga_pan_can_atlas_2018*,*uvm_tcga_pan_can_atlas_2018*,*wt_target_2018_pub*,*all_phase2_target_2018_pub*,*aml_target_2018_pub*,*nbl_target_2018_pub*, and *rt_target_2018_pub*). For cancer variants, genomic coordinates were converted from hg19 to hg38 using *liftover* (Kent et al., 2002). Deletions larger than 100kb were omitted from analysis. Variants were mapped to protein-coding sequence changes within our set of 20,394 human proteins in SwissProt/UniProt using Ensembl VEP v102 (Yates et al., 2019) and ID mappings

between Ensembl and UniProt. Given that the biological relevance of alternate isoforms is not comprehensively understood across the proteome, we chose to focus on protein isoforms considered to be the canonical isoforms (Consortium et al., 2020) which represent the best annotated and understood isoforms across all proteins (although we acknowledge that alternative splicing can affect IDRs (Buljan et al., 2012; Romero et al., 2006; Smith et al., 2020) and condensate-forming properties (Batlle et al., 2020; Gueroussov et al., 2017; Tsang et al., 2020), and therefore mutations in alternative isoforms can in principle affect condensate properties as well). Canonical isoforms are selected based on criteria such as prevalence, similarity to homologs, and in the absence of other information, sequence length (Consortium et al., 2020). A total of n = 2,644,688 DNA variants (62% of all variants in the source datasets) mapped to the 20,394 canonical protein isoforms in UniProt. Beyond this point, variants were counted as protein variants—i.e., DNA variants resulting in the same protein-coding alteration, regardless of their similarity or differences at the DNA level, were counted as the same variant. All synonymous variants were removed from further analysis. For non-synonymous variants, only the primary (often the most severe) protein-coding change of the variant was considered for classification of mutation types (e.g. missense, nonsense, frameshift, etc.) based on the established hierarchy of mutation effect severity within Ensembl variant annotations.

Pathogenicity of Mendelian variants was classified based on designations of clinical significance for ClinVar variants (*pathogenic* or *likely pathogenic*) or of variant class for HGMD variants (*DM* or *DM?*), and of cancer variants was determined by assessing the variants for their inclusion in CIViC (August 1, 2020 release) (Griffith et al., 2017), their inclusion the list of CGI's (Tamborero et al., 2018) *Validated Oncogenic Mutations*, or their designations of oncogenicity in OncoKB v2.10 (*oncogenic*, *likely oncogenic*, or *predicted oncogenic*) (Chakravarty et al., 2017) (Figure S1A). We note that these definitions of pathogenicity do rely on computational predictions of pathogenicity, but not to the same extent as clinical, biological/functional, or evolutionary evidence of pathogenicity (Li et al., 2017; Richards et al., 2015). The resulting 322,825 pathogenic mutations analyzed are shown in Table S4C.

Among pathogenic mutations, we chose to evaluate mutation types that were prevalent pathogenic mutations and where the effect of the mutation on the condensate-promoting feature at the protein level could reasonably be predicted. These mutation types included missense, frameshift, nonsense, in-frame deletion, and in-frame insertion mutations. Together these mutations accounted for 98.9% of pathogenic mutations (not shown), capturing the vast majority of pathogenic mutations. We did not evaluate mutations that affect splicing (0.5% of pathogenic mutations; splice region, splice donor, and splice acceptor mutations) or the start codon (0.3% of pathogenic mutations), or mutations that represented complex changes to the protein sequence (e.g. deletion-insertions, 0.2% of pathogenic mutations).

Nonsense and frameshift variants were considered together to be *truncating variants* and assessed for their predicted propensity to elicit NMD. Predictive rules for NMD were obtained from prior work (Lindeboom et al., 2016). A truncating variant was considered to elicit NMD if the corresponding premature stop codon it introduced occurred (i) >200 residues C-terminal to the start codon; (ii) >50 residues N-terminal to the final exon-exon junction; and (iii) in an exon ≤400 base pairs in length. Pathogenic variants classified as NMD-eliciting in this manner are shown in Table S4D and were omitted from further analyses of truncating variants.

Mutations were defined as affecting condensate-promoting features if they were missense mutations or in-frame insertions within the bounds of an MID or LCS, or if they were in-frame deletions and truncating mutations removing part of an MID or LCS. Truncation mutations can affect the valency of condensate-promoting features to varying degrees depending on the position of the truncation, and thus not all truncations are expected to lead to a substantial effect on condensates. We defined MID valency as total number of MIDs of any type and LCS content (a proxy for LCS valency) as the total number of LSC residues of any type and implemented a filter requiring that a truncation mutation remove at least 25% of the protein's total MID or LCS valency.

This cutoff was set based on studies of known condensate forming proteins that suggest that a fractional valency loss of 0.2-0.4 was necessary in these cases to observe substantial effects on condensate formation (Li et al., 2020; Quiroz et al., 2020). The catalog of 36,777 pathogenic mutations that affect condensate-promoting features within the set of putative condensate-forming proteins is shown in Table S4H.

Disease Terminology and Stratification of Diseases by Organ System

Different datasets use distinct terminologies for the same diseases, and we found that in some cases even within the same dataset, terminologies for the same diseases had semantic differences. For this reason, variants were mapped to a common disease nomenclature for analysis (Figure S1A). Mendelian variants were mapped to the list of 7,507 Mendelian phenotypes in OMIM (Amberger et al., 2015) (accessed January, 2021) (only phenotypes with the prefixes # or % were included) using links to OMIM provided in ClinVar or HGMD. 65% of pathogenic Mendelian mutations (n = 176,976 mutations) mapped to a Mendelian phenotype and were used for disease-related analyses. The Mendelian phenotypes were mapped to organ systems using HPO annotations (Köhler et al., 2018) (December 9, 2020 release) of OMIM phenotypes. Cancer variants were mapped to the list of 836 tumor types in OncoTree (Kundra et al., 2021) (accessed January, 2021; only terms at level 2 or greater were included, as level 1 indicated tissues of origin) using links to OncoTree provided in the cancer datasets. Nearly all of the pathogenic cancer variants (99.8%; n = 58,437 mutations) mapped to a tumor type and were used for disease-related analyses. Tumor types were mapped to tissues of origin using the hierarchy provided in OncoTree by mapping each tumor type to the corresponding term at level 1 of the hierarchy.

Gene Ontology Analysis

GO annotations associated with UniProt entries for human proteins were obtained from the Gene Ontology Resource (January 1, 2021 release) (Carbon et al., 2018). Annotations with the *NOT* qualifier were removed. Only annotations with the evidence codes *EXP*, *IDA*, *IPI*, *IMP*, *IGI*, *IEP*, *HTP*, *HAD*, *HMP*, *HGI*, or *HEP* were included in order restrict the analysis to annotations with supporting experimental evidence and to exclude computationally or phylogenetically derived annotations. A subset of GO terms that correspond to biomolecular condensates were manually curated from GO, and components of those condensates were defined as all proteins with GO annotations corresponding to those GO terms or any descendent terms thereof in the GO hierarchy. For known condensates without exact GO terms, a set of GO terms thought to best represent known properties of the condensate were used as the definitions for the condensate components. The correspondence between GO terms, known biomolecular condensates, and resultant proteins mapped to those condensates as used in this study is shown in Table S4I. For all analyses, the set of proteins associated with a particular GO term included all proteins annotated with the GO term or with all terms associated with the GO term at lower levels of the GO hierarchy.

For the analyses in Figure 2, all GO terms associated with the set of condensate-promoting proteins that had pathogenic mutations affecting condensate-promoting features were tested for statistical enrichment within the set. An analogous analysis performed by stratifying the set of condensate-promoting proteins that had pathogenic mutations affecting condensate-promoting features by the specific disease types associated with those mutations.

Selection of Candidates and Mutations for Experimental Validation

Candidate proteins for experimental validation (Table S3) were selected in a manner constrained only by the practical limitation of availability of DNA reagents by manually searching

Addgene or commercially available cDNA repositories from the MGC (Team et al., 2009) for availability of DNA material encoding full length proteins. This practical limitation is likely to bias the candidate selection toward proteins that are well-characterized (and therefore deposited by investigators in plasmid repositories), but we are not presently aware of any variables in this *ad hoc* selection process that would confound the selection of candidates toward those that that are more or less likely to harbor pathogenic mutations that impact condensates. 12 of these proteins (48%) did not show punctate localization in mESCs when ectopically expressed with a GFP (Figure S3A, Table S3). The false discovery rate for the set of condensate-forming proteins from which these candidates were selected is expected to be much less than 48%, based on the validation reported in the source databases or algorithms (Li et al., 2019; Mészáros et al., 2019; Mierlo et al., 2021; You et al., 2019; Yu et al., 2020). This suggests that the failure to detect punctate localization is unlikely to be due to a false designation of a protein as condensate-forming. Rather, as condensate formation can often to be cell-type and cell-state specific, this suggests that mESCs in resting state may not provide a conducive environment to observe condensation for all proteins (Figure S3B). These 12 proteins were not analyzed further. For the remaining 13 proteins that did show evidence of punctate localization in mESCs, pathogenic mutations in these protein candidates were selected *ad hoc* such that the relative distribution of selected mutations was similar to mutations in the full catalog with respect to the types of condensate-promoting features (MIDs or LCSs) they affected (Figure S3D). Given the relatively small sample size of mutations selected for experimental testing compared to the full catalog, a strictly random selection did not guarantee that the distribution of selected mutations would meet this criterion. While we do not expect our selection process for mutations to be biased toward mutations more or less likely to exhibit condensate effects, we cannot strictly rule this out this possibility. We initially selected 1-2 mutations per protein candidate, with the intention of testing additional mutations if the initial set of mutations did not show effects on condensate properties. However, we found that the majority of mutants (87%) selected in this initial cycle had effects measurable effects on condensate properties in cells (Figure 3C, Figure S4B-C)

mESC Cell Line Generation

Stable cell lines were generated by cloning WT and mutant gene sequences using NEBuilder HiFi DNA Assembly (NEB) into a doxycycline-inducible, N-terminal mEGFP-tagged expression construct with a hygromycin-resistance gene, which was integrated into mESCs using the PiggyBac transposon system (Systems Biosciences). 0.5 x 10^6 wildtype mESCs were plated in 6-well format and simultaneously transfected with 1 µg of the expression vector and 1 µg of the PiggyBac transposase using Lipofectamine 3000 (ThermoFisher), according to manufacturer instructions in serum/LIF media. The next day, media was changed to 2i, and cells were split into 100 mm gelatin-coated plates with 2i-media supplemented with 500 µg/mL hygromycin (ThermoFisher) for selection. Selection media was exchanged every day and un-transfected control cells were monitored to assess selection.

Imaging and Image Analysis in mESCs

Cells were grown on 35 mm glass plates (MatTek) coated with poly-L-ornithine (Sigma) for 30 minutes at 37°C followed by coating with laminin (Corning) for 2 hours at 37°C. Expression of the mEGFP-tagged protein was induced by adding doxycycline to the media at 1µg/mL for 24 hours. Cells were imaged live in a heated chamber at 37°C with humidified air and 5% CO_2 in 2i media supplemented with 5µM Draq5 dye (ThermoFisher) for nuclear staining. Images were acquired with a Zeiss LSM880 Confocal Microscope with Airyscan processing with a 63x Objective and 2x Zoom using ZenBlack acquisition software (W.M. Keck Microscopy Facility, MIT). Images were processed using Fiji is Just ImageJ (Fiji) (Schindelin et al., 2012). Image

analysis was conducted on *z*-stacks with 10-20 slices per cell at 0.2-0.36 μm per slice. Condensate partition ratio, cross-sectional area, number per cell were calculated using a custom script written in Python v3.6.9. A Python package, *cellpose* (Stringer et al., 2021), was used on the Draq5-DNA signal to segment nuclei in each cell. For each *z*-stack image, the maximum intensity projection was determined and puncta were identified as objects within the nuclear boundary (nucleoplasm) in which signal within the condensate was above a particular threshold cutoff. A threshold cutoff of 3 standard deviations above the mean of the image intensity was used for all candidates except for MECP2, in which a threshold cutoff of 2 standard deviations above the mean was used. Once identified, the area and number of each condensate was quantified. Partition ratios were calculated as the ratio of the mean pixel intensity within puncta relative to the mean pixel intensity of the nucleoplasmic signal, excluding signal within other segmented puncta.

FRAP analysis was performed on LSM880 Airyscan microscope with 488 nm laser. Photobleaching was performed by defining and exposing a region of interest in or around a punctum to 100% laser power. Images were collected every 0.5-2 seconds for up to a minute or until the majority of the fluorescence intensity was recovered. Fiji was used to calculate intensity values within the bleached region of interest during the timelapse before, during, and after bleaching. Fluorescence intensities in the region were normalized to pre-bleached values and the recovery profile was fit to a single exponential model.

Stress Condition Experiments

Cells were treated with 0.5mM $NaAsO_2$ (Sigma) solution in cell culture media for 60 minutes and imaged. Images were acquired using Zeiss LSM880 Confocal Microscope with Airyscan processing with a 63x Objective and 2x Zoom using ZenBlack acquisition software. Images were postprocessed using Fiji.

Immunofluorescence with Human Tissue Samples

Fresh frozen kidney and breast tissues were purchased from BioIVT. Tissue was embedded in OCT and frozen. Fresh frozen colon tissue was embedded in OCT and frozen at –80°C. Tissue was sectioned into 10 μm sections using the cryostat with temperature set at –15°C or –20°C, respectively. Sections were stored at –20°C until use.

For the immunofluorescence sections were brought to room temperature, they were fixed in 4% PFA in PBS for 10 minutes. Following three washes in PBS, tissues were permeabilized using 0.5% TX100 in PBS, washed three times in PBS and blocked with 4% BSA in PBS for 1 hours. Primary antibodies were diluted into 4% BSA in PBS and added to the tissue sample for overnight incubation at RT. Following three washes in PBS, samples were incubated with secondary antibodies diluted 1:500 in 4% BSA in PBS. Samples was washed in PBS, DNA was stained using 20μm/mL Hoechst 33258 (Life Technologies, H3569) for 5 minutes and mounted using Vectashield (VWR, 101098-042). Images were acquired using Zeiss LSM880 Confocal Microscope with Airyscan processing with a 63x Objective and 2x Zoom using ZenBlack acquisition software. Images were postprocessed using Fiji.

Primary antibodies used: TCOF1 (PA558309, Thermofisher); SALL1 (PA562057, Invitrogen); and BARD1 (ab226854, Abcam). Secondary antibody used: Goat anti-rabbit AlexaFluor-488 (A11008, Thermofisher).

Human Cell Line Experiments

Cells were transiently transfected with the 1μg of WT or mutant DNA constructs (same as those used for mESC experiments, see above) into 0.5 x 10^6 cells that were plated onto 35 mm glass plates (MatTek). Transfections were performed using Lipofectamine 3000 (ThermoFisher)

according to manufacturer instructions in complete DMEM media, described above. On the day following transfection, the media was changed and expression of the mEGFP-tagged protein was induced by adding doxycycline to the media at 1µg/mL for 24 hours, followed by imaging, processing, and analysis as described for mESCs above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise noted, all statistical analyses involving overlaps between sets of proteins and were done using one-tailed Fisher's exact tests, and *p*-values were adjusted for multiple comparisons using the Benjamini-Hoechberg procedure. Unless otherwise noted, all statistical analyses involving comparisons between distributions were done using two-sided Wilcoxon rank sum / Mann Whitney U tests, and *p*-values were adjusted for multiple comparisons using the Bonferroni correction.

Table S4. Results from analyses reported in this study. Related to Figures 1-3.

This file contains several tables containing additional information on the mapping of condensate-promoting features, canonical protein features, and pathogenic mutations across the proteome, as well as on the analyses reported in this study.

References

Ahn, J.H., Davis, E.S., Daugird, T.A., Zhao, S., Quiroga, I.Y., Uryu, H., Li, J., Storey, A.J., Tsai, Y.-H., Keeley, D.P., et al. (2021). Phase separation drives aberrant chromatin looping and cancer development. Nature 1–5. https://doi.org/10.1038/s41586-021-03662-5.

Alberti, S. (2017). The wisdom of crowds: regulating cell function through condensed states of living matter. J Cell Sci *130*, jcs.200295. https://doi.org/10.1242/jcs.200295.

Alberti, S., and Dormann, D. (2019). Liquid–Liquid Phase Separation in Disease. Annu Rev Genet *53*, 171–194. https://doi.org/10.1146/annurev-genet-112618-043527.

Alberti, S., and Hyman, A.A. (2021). Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. Nat Rev Mol Cell Bio *22*, 196–213. https://doi.org/10.1038/s41580-020-00326-6.

Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. Cell *137*, 146–158. https://doi.org/10.1016/j.cell.2009.02.044.

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res *43*, D789–D798. https://doi.org/10.1093/nar/gku1205.

Babinchak, W.M., Dumm, B.K., Venus, S., Boyko, S., Putnam, A.A., Jankowsky, E., and Surewicz, W.K. (2020). Small molecules as potent biphasic modulators of protein liquid-liquid phase separation. Nat Commun *11*, 5574. https://doi.org/10.1038/s41467-020-19211-z.

Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K. (2016). Compositional Control of Phase-Separated Cellular Bodies. Cell *166*, 651–663. https://doi.org/10.1016/j.cell.2016.06.010.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Bio *18*, 285–298. https://doi.org/10.1038/nrm.2017.7.

Basu, S., Mackowiak, S.D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., Geertsema, H., Ali, S., Jerković, I., Ewers, H., et al. (2020). Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. Cell *181*, 1062-1079.e30. https://doi.org/10.1016/j.cell.2020.04.018.

Batlle, C., Yang, P., Coughlin, M., Messing, J., Pesarrodona, M., Szulc, E., Salvatella, X., Kim, H.J., Taylor, J.P., and Ventura, S. (2020). hnRNPDL Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation. Cell Reports *30*, 1117-1128.e5. https://doi.org/10.1016/j.celrep.2019.12.080.

Bernhofer, M., Goldberg, T., Wolf, S., Ahmed, M., Zaugg, J., Boden, M., and Rost, B. (2017). NLSdb—major update for database of nuclear localization signals and nuclear export signals. Nucleic Acids Res *46*, gkx1021-. https://doi.org/10.1093/nar/gkx1021.

Beutel, O., Maraspini, R., Pombo-García, K., Martin-Lemaitre, C., and Honigmann, A. (2019). Phase Separation of Zonula Occludens Proteins Drives Formation of Tight Junctions. Cell *179*, 923-936.e11. https://doi.org/10.1016/j.cell.2019.10.011.

Bienz, M. (2020). Head-to-Tail Polymerization in the Assembly of Biomolecular Condensates. Cell *182*, 799–811. https://doi.org/10.1016/j.cell.2020.07.037.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2020). The InterPro protein families and domains database: 20 years on. Nucleic Acids Res *49*, gkaa977-. https://doi.org/10.1093/nar/gkaa977.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e16. https://doi.org/10.1016/j.cell.2018.10.042.

Boija, A., Klein, I.A., and Young, R.A. (2021). Biomolecular condensates and cancer. Cancer Cell *39*, 174–192. https://doi.org/10.1016/j.ccell.2020.12.003.

Bouchard, J.J., Otero, J.H., Scott, D.C., Szulc, E., Martin, E.W., Sabri, N., Granata, D., Marzahn, M.R., Lindorff-Larsen, K., Salvatella, X., et al. (2018). Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. Mol Cell *72*, 19-36.e8. https://doi.org/10.1016/j.molcel.2018.08.027.

Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M.E., Rengarajan, S., Volorio, A., McBride, M.J., et al. (2017). Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. Cell *171*, 163-178.e19. https://doi.org/10.1016/j.cell.2017.07.036.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., and Babu, M.M. (2012). Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. Mol Cell *46*, 871–883. https://doi.org/10.1016/j.molcel.2012.05.039.

Burke, K.A., Janke, A.M., Rhine, C.L., and Fawzi, N.L. (2015). Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. Mol Cell *60*, 231–241. https://doi.org/10.1016/j.molcel.2015.09.006.

Cai, D., Feliciano, D., Dong, P., Flores, E., Gruebele, M., Porat-Shliom, N., Sukenik, S., Liu, Z., and Lippincott-Schwartz, J. (2019). Phase separation of YAP reorganizes genome topology for long-term YAP target gene expression. Nat Cell Biol *21*, 1578–1589. https://doi.org/10.1038/s41556-019-0433-z.

Cai, D., Liu, Z., and Lippincott-Schwartz, J. (2021). Biomolecular Condensates and Their Links to Cancer Progression. Trends Biochem Sci https://doi.org/10.1016/j.tibs.2021.01.002.

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N.L., Lewis, S.E., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., et al. (2018). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res *47*, D330–D338. https://doi.org/10.1093/nar/gky1055.

Case, L.B., Ditlev, J.A., and Rosen, M.K. (2019). Regulation of Transmembrane Signaling by Phase Separation. Annu Rev Biophys *48*, 1–30. https://doi.org/10.1146/annurev-biophys-052118-115534.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov *2*, 401–404. https://doi.org/10.1158/2159-8290.cd-12-0095.

Chakravarty, D., Gao, J., Phillips, S., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. Jco Precis Oncol *2017*, 1–16. https://doi.org/10.1200/po.17.00011.

Chandra, B., Michmerhuizen, N.L., Shirnekhi, H.K., Tripathi, S., Pioso, B.J., Baggett, D.W., Mitrea, D.M., Iacobucci, I., White, M.R., Chen, J., et al. (2021). Phase Separation Mediates NUP98 Fusion Oncoprotein Leukemic TransformationPhase Separation Drives Oncogenesis by NUP98 Fusion Proteins. Cancer Discov *12*, 1152–1169. https://doi.org/10.1158/2159-8290.cd-21-0674.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, eaar4199. https://doi.org/10.1126/science.aar4199.

Choi, J.-M., Holehouse, A.S., and Pappu, R.V. (2020). Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. Annu Rev Biophys *49*, 1–27. https://doi.org/10.1146/annurev-biophys-121219-081629.

Cloer, E.W., Siesser, P.F., Cousins, E.M., Goldfarb, D., Mowrey, D.D., Harrison, J.S., Weir, S.J., Dokholyan, N.V., and Major, M.B. (2018). p62-Dependent Phase Separation of Patient-Derived KEAP1 Mutations and NRF2. Mol Cell Biol *38*, e00644-17. https://doi.org/10.1128/mcb.00644-17.

Conicella, A.E., Zerze, G.H., Mittal, J., and Fawzi, N.L. (2016). ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure *24*, 1537–1549. https://doi.org/10.1016/j.str.2016.07.007.

Consortium, T.A.P.G. (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov *7*, 818–831. https://doi.org/10.1158/2159-8290.cd-17-0151.

Consortium, T.U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., et al. (2020). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res *49*, D480–D489. https://doi.org/10.1093/nar/gkaa1100.

Dao, T.P., Kolaitis, R.-M., Kim, H.J., O'Donovan, K., Martyniak, B., Colicino, E., Hehnly, H., Taylor, J.P., and Castañeda, C.A. (2018). Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. Mol Cell *69*, 965-978.e6. https://doi.org/10.1016/j.molcel.2018.02.004.

Dao, T.P., Martyniak, B., Canning, A.J., Lei, Y., Colicino, E.G., Cosgrove, M.S., Hehnly, H., and Castañeda, C.A. (2019). ALS-Linked Mutations Affect UBQLN2 Oligomerization and Phase Separation in a Position- and Amino Acid-Dependent Manner. Structure *27*, 937-951.e5. https://doi.org/10.1016/j.str.2019.03.012.

Du, M., and Chen, Z. (2018). DNA-induced liquid phase condensation of cGAS activates innate immune signaling. Science *361*, eaat1022. https://doi.org/10.1126/science.aat1022.

Emenecker, R.J., Griffith, D., and Holehouse, A.S. (2021). metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. Biophys J *120*, 4312–4319. https://doi.org/10.1016/j.bpj.2021.08.039.

Fasciani, A., D'Annunzio, S., Poli, V., Fagnocchi, L., Beyes, S., Michelatti, D., Corazza, F., Antonelli, L., Gregoretti, F., Oliva, G., et al. (2020). MLL4-associated condensates counterbalance Polycomb-mediated nuclear mechanical stress in Kabuki syndrome. Nat Genet *52*, 1397–1411. https://doi.org/10.1038/s41588-020-00724-8.

Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R.V., and Brangwynne, C.P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell *165*, 1686–1697. https://doi.org/10.1016/j.cell.2016.04.047.

Fichó, E., Reményi, I., Simon, I., and Mészáros, B. (2017). MFIB: a repository of protein complexes with mutual folding induced by binding. Bioinformatics *33*, 3682–3684. https://doi.org/10.1093/bioinformatics/btx486.

Fonseca, M. de C., Oliveira, J.F. de, Araujo, B.H.S., Canateli, C., Prado, P.F.V. do, Neto, D.P.A., Bosque, B.P., Rodrigues, P.V., Godoy, J.V.P. de, Tostes, K., et al. (2021). Molecular and cellular

basis of hyperassembly and protein aggregation driven by a rare pathogenic mutation in DDX3X. Iscience *24*, 102841. https://doi.org/10.1016/j.isci.2021.102841.

Frottin, F., Schueder, F., Tiwary, S., Gupta, R., Körner, R., Schlichthaerle, T., Cox, J., Jungmann, R., Hartl, F.U., and Hipp, M.S. (2019). The nucleolus functions as a phase-separated protein quality control compartment. Sci New York N Y *365*, 342–347. https://doi.org/10.1126/science.aaw9157.

Gibson, B., Doolittle, L., Schneider, M., Jensen, L., Gamarra, N., Henry, L., Gerlich, D., Redding, S., and Rosen, M. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. Cell *179*, 470-484.e21. https://doi.org/10.1016/j.cell.2019.08.037.

Griffith, M., Spies, N.C., Krysiak, K., McMichael, J.F., Coffman, A.C., Danos, A.M., Ainscough, B.J., Ramirez, C.A., Rieke, D.T., Kujan, L., et al. (2017). CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet *49*, 170–174. https://doi.org/10.1038/ng.3774.

Gueroussov, S., Weatheritt, R.J., O'Hanlon, D., Lin, Z.-Y., Narula, A., Gingras, A.-C., and Blencowe, B.J. (2017). Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. Cell *170*, 324-339.e23. https://doi.org/10.1016/j.cell.2017.06.037.

Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature *572*, 543–548. https://doi.org/10.1038/s41586-019-1464-0.

Henninger, J.E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M.M., Andrews, J.O., Zamudio, A.V., Lazaris, C., Hannett, N.M., et al. (2021). RNA-Mediated Feedback Control of Transcriptional Condensates. Cell *184*, 207-225.e24. https://doi.org/10.1016/j.cell.2020.11.030.

Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. Nat Rev Mol Cell Bio *19*, 327–341. https://doi.org/10.1038/nrm.2017.130.

Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291-304.e6. https://doi.org/10.1016/j.cell.2018.03.022.

Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O.G., and Pappu, R.V. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J *112*, 16–21. https://doi.org/10.1016/j.bpj.2016.11.3200.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res *43*, D512–D520. https://doi.org/10.1093/nar/gku1267.

Howard, T.P., and Roberts, C.W.M. (2020). Partitioning of Chemotherapeutics into Nuclear Condensates—Opening the Door to New Approaches for Drug Development. Mol Cell *79*, 544–545. https://doi.org/10.1016/j.molcel.2020.07.029.

Huang, W.Y.C., Alvarez, S., Kondo, Y., Lee, Y.K., Chung, J.K., Lam, H.Y.M., Biswas, K.H., Kuriyan, J., and Groves, J.T. (2019). A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS. Science *363*, 1098–1103. https://doi.org/10.1126/science.aau5721.

Hughes, M.P., Sawaya, M.R., Boyer, D.R., Goldschmidt, L., Rodriguez, J.A., Cascio, D., Chong, L., Gonen, T., and Eisenberg, D.S. (2018). Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. Science *359*, 698–701. https://doi.org/10.1126/science.aan6398.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487–498. https://doi.org/10.1016/j.cell.2015.12.038.

Jiang, H., Wang, S., Huang, Y., He, X., Cui, H., Zhu, X., and Zheng, Y. (2015). Phase Transition of Spindle-Associated Protein Regulate Spindle Apparatus Assembly. Cell *163*, 108–122. https://doi.org/10.1016/j.cell.2015.08.010.

Kanaan, N.M., Hamel, C., Grabinski, T., and Combs, B. (2020). Liquid-liquid phase separation induces pathogenic tau conformations in vitro. Nat Commun *11*, 2809. https://doi.org/10.1038/s41467-020-16580-3.

Kasza, K.E., Supriyatno, S., and Zallen, J.A. (2019). Cellular defects resulting from disease-related myosin II mutations in Drosophila. Proc National Acad Sci *116*, 22205–22211. https://doi.org/10.1073/pnas.1909227116.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, and D. (2002). The Human Genome Browser at UCSC. Genome Res *12*, 996–1006. https://doi.org/10.1101/gr.229102.

Kilic, S., Lezaja, A., Gatti, M., Bianco, E., Michelena, J., Imhof, R., and Altmeyer, M. (2019). Phase separation of 53 BP 1 determines liquid-like behavior of DNA repair compartments. Embo J *38*, e101379. https://doi.org/10.15252/embj.2018101379.

Kim, H.J., Kim, N.C., Wang, Y.-D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*, 467–473. https://doi.org/10.1038/nature11922.

King, M.R., and Petry, S. (2020). Phase separation of TPX2 enhances and spatially coordinates microtubule nucleation. Nat Commun *11*, 270. https://doi.org/10.1038/s41467-019-14087-0.

Klein, I.A., Boija, A., Afeyan, L.K., Hawken, S.W., Fan, M., Dall'Agnese, A., Oksuz, O., Henninger, J.E., Shrinivas, K., Sabari, B.R., et al. (2020). Partitioning of cancer therapeutics in nuclear condensates. Science *368*, 1386–1392. https://doi.org/10.1126/science.aaz4427.

Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J.O.B., Danis, D., Gourdine, J.-P., Gargano, M., Harris, N.L., Matentzoglu, N., McMurry, J.A., et al. (2018). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res *47*, D1018–D1027. https://doi.org/10.1093/nar/gky1105.

Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J.A., Bukirova, D., Čalyševa, J., et al. (2019). ELM—the eukaryotic linear motif resource in 2020. Nucleic Acids Res *48*, D296–D306. https://doi.org/10.1093/nar/gkz1030.

Kundra, R., Zhang, H., Sheridan, R., Sirintrapun, S.J., Wang, A., Ochoa, A., Wilson, M., Gross, B., Sun, Y., Madupuri, R., et al. (2021). OncoTree: A Cancer Classification System for Precision Oncology. Jco Clin Cancer Informatics *5*, 221–230. https://doi.org/10.1200/cci.20.00108.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

Lancaster, A.K., Nutter-Upham, A., Lindquist, S., and King, O.D. (2014). PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics *30*, 2501–2502. https://doi.org/10.1093/bioinformatics/btu310.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2017). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res *46*, D1062–D1067. https://doi.org/10.1093/nar/gkx1153.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240. https://doi.org/10.1038/nature22822.

Lemos, C., Schulze, L., Weiske, J., Meyer, H., Braeuer, N., Barak, N., Eberspächer, U., Werbeck, N., Stresemann, C., Lange, M., et al. (2020). Identification of Small Molecules that Modulate Mutant p53 Condensation. Iscience *23*, 101517. https://doi.org/10.1016/j.isci.2020.101517.

Letunic, I., Khedkar, S., and Bork, P. (2020). SMART: recent updates, new developments and status in 2020. Nucleic Acids Res *49*, gkaa937-. https://doi.org/10.1093/nar/gkaa937.

Lewis, J.D., Caldara, A.L., Zimmer, S.E., Stahley, S.N., Seybold, A., Strong, N.L., Frangakis, A.S., Levental, I., Wahl, J.K., Mattheyses, A.L., et al. (2019). The desmosome is a mesoscale lipid raft-like membrane domain. Mol Biol Cell *30*, 1390–1405. https://doi.org/10.1091/mbc.e18-10-0649.

Li, C.H., Coffey, E.L., Dall'Agnese, A., Hannett, N.M., Tang, X., Henninger, J.E., Platt, J.M., Oksuz, O., Zamudio, A.V., Afeyan, L.K., et al. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. Nature 1–8. https://doi.org/10.1038/s41586-020-2574-4.

Li, M.M., Datto, M., Duncavage, E.J., Kulkarni, S., Lindeman, N.I., Roy, S., Tsimberidou, A.M., Vnencak-Jones, C.L., Wolff, D.J., Younes, A., et al. (2017). Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. J Mol Diagnostics *19*, 4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002.

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340. https://doi.org/10.1038/nature10879.

Li, Q., Peng, X., Li, Y., Tang, W., Zhu, J., Huang, J., Qi, Y., and Zhang, Z. (2019). LLPSDB: a database of proteins undergoing liquid–liquid phase separation in vitro. Nucleic Acids Res *48*, D320–D327. https://doi.org/10.1093/nar/gkz778.

Lin, Y., Protter, D.S.W., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Mol Cell *60*, 208–219. https://doi.org/10.1016/j.molcel.2015.08.018.

Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat Genet *48*, 1112–1118. https://doi.org/10.1038/ng.3664.

Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature *558*, 318–323. https://doi.org/10.1038/s41586-018-0174-3.

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., et al. (2019). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res *48*, D265–D268. https://doi.org/10.1093/nar/gkz991.

Lu, Y., Wu, T., Gutman, O., Lu, H., Zhou, Q., Henis, Y.I., and Luo, K. (2020). Phase separation of TAZ compartmentalizes the transcription machinery to promote gene expression. Nat Cell Biol *22*, 453–464. https://doi.org/10.1038/s41556-020-0485-0.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Bio *8*, 479–490. https://doi.org/10.1038/nrm2178.

Lyon, A.S., Peeples, W.B., and Rosen, M.K. (2020). A framework for understanding the functions of biomolecular condensates across scales. Nat Rev Mol Cell Bio 1–21. https://doi.org/10.1038/s41580-020-00303-z.

Mackenzie, I.R., Nicholson, A.M., Sarkar, M., Messing, J., Purice, M.D., Pottier, C., Annu, K., Baker, M., Perkerson, R.B., Kurti, A., et al. (2017). TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. Neuron *95*, 808-816.e9. https://doi.org/10.1016/j.neuron.2017.07.025.

Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R.V., and Mittag, T. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science *367*, 694–699. https://doi.org/10.1126/science.aaw8653.

Meng, F., Yu, Z., Zhang, D., Chen, S., Guan, H., Zhou, R., Wu, Q., Zhang, Q., Liu, S., Ramani, M.K.V., et al. (2021). Induced phase separation of mutant NF2 imprisons the cGAS-STING machinery to abrogate antitumor immunity. Mol Cell https://doi.org/10.1016/j.molcel.2021.07.040.

Mészáros, B., Erdős, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., Horváth, T., Murvai, N., Kovács, O.P., Kovács, M., et al. (2019). PhaSePro: the database of proteins driving liquid–liquid phase separation. Nucleic Acids Res *48*, D360–D367. https://doi.org/10.1093/nar/gkz848.

Mierlo, G. van, Jansen, J.R.G., Wang, J., Poser, I., Heeringen, S.J. van, and Vermeulen, M. (2021). Predicting protein condensate formation using machine learning. Cell Reports *34*, 108705. https://doi.org/10.1016/j.celrep.2021.108705.

Milovanovic, D., Wu, Y., Bian, X., and Camilli, P.D. (2018). A liquid phase of synapsin and lipid vesicles. Science *361*, eaat5671. https://doi.org/10.1126/science.aat5671.

Min, J., Wright, W.E., and Shay, J.W. (2019). Clustered telomeres in phase-separated nuclear condensates engage mitotic DNA synthesis through BLM and RAD52. Gene Dev *33*, 814–827. https://doi.org/10.1101/gad.324905.119.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2020). Pfam: The protein families database in 2021. Nucleic Acids Res *49*, gkaa913-. https://doi.org/10.1093/nar/gkaa913.

Mitrea, D.M., Cika, J.A., Guy, C.S., Ban, D., Banerjee, P.R., Stanley, C.B., Nourse, A., Deniz, A.A., and Kriwacki, R.W. (2016). Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. Elife *5*, e13571. https://doi.org/10.7554/elife.13571.

Mitrea, D.M., Cika, J.A., Stanley, C.B., Nourse, A., Onuchic, P.L., Banerjee, P.R., Phillips, A.H., Park, C.-G., Deniz, A.A., and Kriwacki, R.W. (2018). Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. Nat Commun *9*, 842. https://doi.org/10.1038/s41467-018-03255-3.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell *163*, 123–133. https://doi.org/10.1016/j.cell.2015.09.015.

Nahm, M., Lim, S.M., Kim, Y.-E., Park, J., Noh, M.-Y., Lee, S., Roh, J.E., Hwang, S.-M., Park, C.-K., Kim, Y.H., et al. (2020). ANXA11 mutations in ALS cause dysregulation of calcium homeostasis and stress granule dynamics. Sci Transl Med *12*, eaax3993. https://doi.org/10.1126/scitranslmed.aax3993.

Nedelsky, N.B., and Taylor, J.P. (2019). Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease. Nat Rev Neurol *15*, 272–286. https://doi.org/10.1038/s41582-019-0157-5.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., et al. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol Cell *57*, 936–947. https://doi.org/10.1016/j.molcel.2015.01.013.

Parker, M.W., Bell, M., Mir, M., Kao, J.A., Darzacq, X., Botchan, M.R., and Berger, J.M. (2019). A new class of disordered elements controls DNA replication through initiator self-assembly. Elife *8*, e48562. https://doi.org/10.7554/elife.48562.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein

FUS Accelerated by Disease Mutation. Cell *162*, 1066–1077. https://doi.org/10.1016/j.cell.2015.07.047.

Pawson, T., and Nash, P. (2003). Assembly of Cell Regulatory Systems Through Protein Interaction Domains. Science *300*, 445–452. https://doi.org/10.1126/science.1083653.

Peskett, T., Rau, F., O'Driscoll, J., Patani, R., Lowe, A., and Saibil, H. (2018). A Liquid to Solid Phase Transition Underlying Pathological Huntingtin Exon1 Aggregation. Mol Cell *70*, 588-601.e6. https://doi.org/10.1016/j.molcel.2018.04.007.

Quiroz, F.G., Fiore, V.F., Levorse, J., Polak, L., Wong, E., Pasolli, H.A., and Fuchs, E. (2020). Liquid-liquid phase separation drives skin barrier formation. Sci New York N Y *367*, eaax9554. https://doi.org/10.1126/science.aax9554.

Rai, A.K., Chen, J.-X., Selbach, M., and Pelkmans, L. (2018). Kinase-controlled phase transition of membraneless organelles in mitosis. Nature *559*, 211–216. https://doi.org/10.1038/s41586-018-0279-8.

Ramaswami, M., Taylor, J.P., and Parker, R. (2013). Altered Ribostasis: RNA-Protein Granules in Degenerative Disorders. Cell *154*, 727–736. https://doi.org/10.1016/j.cell.2013.07.038.

Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., Mahato, J., Panigrahi, R., Navalkar, A., Mehra, S., et al. (2020). α-Synuclein aggregation nucleates through liquid–liquid phase separation. Nat Chem 1–12. https://doi.org/10.1038/s41557-020-0465-9.

Riback, J.A., Zhu, L., Ferrolino, M.C., Tolbert, M., Mitrea, D.M., Sanders, D.W., Wei, M.-T., Kriwacki, R.W., and Brangwynne, C.P. (2020). Composition-dependent thermodynamics of intracellular phase separation. Nature *581*, 209–214. https://doi.org/10.1038/s41586-020-2256-2.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med *17*, 405–423. https://doi.org/10.1038/gim.2015.30.

Romero, P.R., Zaidi, S., Fang, Y.Y., Uversky, V.N., Radivojac, P., Oldfield, C.J., Cortese, M.S., Sickmeier, M., LeGall, T., Obradovic, Z., et al. (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc National Acad Sci *103*, 8390–8395. https://doi.org/10.1073/pnas.0507916103.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*, eaar3958. https://doi.org/10.1126/science.aar3958.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

Schmidt, H.B., and Görlich, D. (2015). Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. Elife *4*, e04251. https://doi.org/10.7554/elife.04251.

Schneider, J.W., Oommen, S., Qureshi, M.Y., Goetsch, S.C., Pease, D.R., Sundsbak, R.S., Guo, W., Sun, M., Sun, H., Kuroyanagi, H., et al. (2020). Dysregulated ribonucleoprotein granules promote cardiomyopathy in RBM20 gene-edited pigs. Nat Med *26*, 1788–1800. https://doi.org/10.1038/s41591-020-1087-x.

Schwayer, C., Shamipour, S., Pranjic-Ferscha, K., Schauer, A., Balda, M., Tada, M., Matter, K., and Heisenberg, C.-P. (2019). Mechanosensation of Tight Junctions Depends on ZO-1 Phase Separation and Flow. Cell *179*, 937-952.e18. https://doi.org/10.1016/j.cell.2019.10.006.

Seet, B.T., Dikic, I., Zhou, M.-M., and Pawson, T. (2006). Reading protein modifications with interaction domains. Nat Rev Mol Cell Bio *7*, 473–483. https://doi.org/10.1038/nrm1960.

Sharkey, L.M., Safren, N., Pithadia, A.S., Gerson, J.E., Dulchavsky, M., Fischer, S., Patel, R., Lantis, G., Ashraf, N., Kim, J.H., et al. (2018). Mutant UBQLN2 promotes toxicity by modulating intrinsic self-assembly. Proc National Acad Sci *115*, 201810522. https://doi.org/10.1073/pnas.1810522115.

Sheu-Gruttadauria, J., and MacRae, I.J. (2018). Phase Transitions in the Assembly and Function of Human miRISC. Cell *173*, 946-957.e16. https://doi.org/10.1016/j.cell.2018.02.051.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*, eaaf4382. https://doi.org/10.1126/science.aaf4382.

Smith, J.A., Curry, E.G., Blue, R.E., Roden, C., Dundon, S.E.R., Rodríguez-Vargas, A., Jordan, D.C., Chen, X., Lyons, S.M., Crutchley, J., et al. (2020). FXR1 splicing is important for muscle development and biomolecular condensates in muscle cells. J Cell Biol *219*. https://doi.org/10.1083/jcb.201911129.

Spannl, S., Tereshchenko, M., Mastromarco, G.J., Ihn, S.J., and Lee, H.O. (2019). Biomolecular condensates in neurodegeneration and cancer. Traffic *20*, 890–911. https://doi.org/10.1111/tra.12704.

Stefl, S., Nishi, H., Petukh, M., Panchenko, A.R., and Alexov, E. (2013). Molecular mechanisms of disease-causing missense mutations. J Mol Biol *425*, 3919–3936. https://doi.org/10.1016/j.jmb.2013.07.014.

Stenson, P.D., Mort, M., Ball, E.V., Chapman, M., Evans, K., Azevedo, L., Hayden, M., Heywood, S., Millar, D.S., Phillips, A.D., et al. (2020). The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Hum Genet *139*, 1197–1207. https://doi.org/10.1007/s00439-020-02199-3.

Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat Methods *18*, 100–106. https://doi.org/10.1038/s41592-020-01018-x.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245. https://doi.org/10.1038/nature22989.

Su, X., Ditlev, J.A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D.S., Taunton, J., Rosen, M.K., and Vale, R.D. (2016). Phase separation of signaling molecules promotes T cell receptor signal transduction. Science *352*, 595–599. https://doi.org/10.1126/science.aad9964.

Tamborero, D., Rubio-Perez, C., Deu-Pons, J., Schroeder, M.P., Vivancos, A., Rovira, A., Tusquets, I., Albanell, J., Rodon, J., Tabernero, J., et al. (2018). Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med *10*, 25. https://doi.org/10.1186/s13073-018-0531-8.

Team, T.M.P., Temple, G., Gerhard, D.S., Rasooly, R., Feingold, E.A., Good, P.J., Robinson, C., Mandich, A., Derge, J.G., Lewis, J., et al. (2009). The completion of the Mammalian Gene Collection (MGC). Genome Res *19*, 2324–2333. https://doi.org/10.1101/gr.095976.109.

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Blal, H.A., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*. https://doi.org/10.1126/science.aal3321.

Tsang, B., Pritišanac, I., Scherer, S.W., Moses, A.M., and Forman-Kay, J.D. (2020). Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. Cell *183*, 1742–1756. https://doi.org/10.1016/j.cell.2020.11.050.

Valentin-Vega, Y.A., Wang, Y.-D., Parker, M., Patmore, D.M., Kanagaraj, A., Moore, J., Rusch, M., Finkelstein, D., Ellison, D.W., Gilbertson, R.J., et al. (2016). Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. Sci Rep-Uk *6*, 25996. https://doi.org/10.1038/srep25996.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet *10*, 252–263. https://doi.org/10.1038/nrg2538.

Vernon, R.M., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H., and Forman-Kay, J.D. (2018). Pi-Pi contacts are an overlooked protein feature relevant to phase separation. Elife *7*, e31486. https://doi.org/10.7554/elife.31486.

Viny, A.D., and Levine, R.L. (2020). Drug modulation by nuclear condensates. Sci New York N Y *368*, 1314–1315. https://doi.org/10.1126/science.abc5318.

Wan, L., Chong, S., Xuan, F., Liang, A., Cui, X., Gates, L., Carroll, T.S., Li, Y., Feng, L., Chen, G., et al. (2019). Impaired cell fate through gain-of-function mutations in a chromatin reader. Nature *577*, 121–126. https://doi.org/10.1038/s41586-019-1842-7.

Wan, P.T.C., Garnett, M.J., Roe, S.M., Lee, S., Niculescu-Duvaz, D., Good, V.M., Jones, C.M., Marshall, C.J., Springer, C.J., Barford, D., et al. (2004). Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell *116*, 855–867. https://doi.org/10.1016/s0092-8674(04)00215-6.

Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., et al. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell *174*, 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006.

Wheeler, R.J., Lee, H.O., Poser, I., Pal, A., Doeleman, T., Kishigami, S., Kour, S., Anderson, E.N., Marrone, L., Murthy, A.C., et al. (2019). Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. Biorxiv 721001. https://doi.org/10.1101/721001.

Woodruff, J.B., Gomes, B.F., Widlund, P.O., Mahamid, J., Honigmann, A., and Hyman, A.A. (2017). The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. Cell *169*, 1066-1077.e10. https://doi.org/10.1016/j.cell.2017.05.028.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2019). Ensembl 2020. Nucleic Acids Res *48*, D682–D688. https://doi.org/10.1093/nar/gkz966.

Yoshizawa, T., Ali, R., Jiou, J., Fung, H.Y.J., Burke, K.A., Kim, S.J., Lin, Y., Peeples, W.B., Saltzberg, D., Soniat, M., et al. (2018). Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. Cell *173*, 693-705.e22. https://doi.org/10.1016/j.cell.2018.03.003.

You, K., Huang, Q., Yu, C., Shen, B., Sevilla, C., Shi, M., Hermjakob, H., Chen, Y., and Li, T. (2019). PhaSepDB: a database of liquid–liquid phase separation related proteins. Nucleic Acids Res *48*, D354–D359. https://doi.org/10.1093/nar/gkz847.

Yu, C., Shen, B., You, K., Huang, Q., Shi, M., Wu, C., Chen, Y., Zhang, C., and Li, T. (2020). Proteome-scale analysis of phase-separated proteins in immunofluorescence images. Brief Bioinform *22*. https://doi.org/10.1093/bib/bbaa187.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res *21*, 564–578. https://doi.org/10.1038/cr.2011.42.

Zamudio, A.V., Dall'Agnese, A., Henninger, J.E., Manteiga, J.C., Afeyan, L.K., Hannett, N.M., Coffey, E.L., Li, C.H., Oksuz, O., Sabari, B.R., et al. (2019). Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. Mol Cell *76*, 753-766.e6. https://doi.org/10.1016/j.molcel.2019.08.016.

Zeng, M., Shang, Y., Araki, Y., Guo, T., Huganir, R.L., and Zhang, M. (2016). Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity. Cell *166*, 1163-1175.e12. https://doi.org/10.1016/j.cell.2016.07.008.

Zhang, H., Zhao, R., Tones, J., Liu, M., Dilley, R.L., Chenoweth, D.M., Greenberg, R.A., and Lampson, M.A. (2020). Nuclear body phase separation drives telomere clustering in ALT cancer cells. Mol Biol Cell *31*, 2048–2056. https://doi.org/10.1091/mbc.e19-10-0589.

Zhu, G., Xie, J., Kong, W., Xie, J., Li, Y., Du, L., Zheng, Q., Sun, L., Guan, M., Li, H., et al. (2020). Phase Separation of Disease-Associated SHP2 Mutants Underlies MAPK Hyperactivation. Cell *183*, 490-502.e18. https://doi.org/10.1016/j.cell.2020.09.002.

CHAPTER 3:

Partitioning of Cancer Therapeutics in Nuclear Condensates

Originally published in *Science*, Volume 368, Issue 6497, 1386-1392 (2020).

Isaac A. Klein[1,2,#], Ann Boija[1,#], Lena K. Afeyan[1,3], Susana Wilson Hawken[1,3], Mengyang Fan[4,5], Alessandra Dall'Agnese[1], Ozgur Oksuz[1], Jonathan E. Henninger[1], Krishna Shrinivas[6,7], Benjamin R. Sabari[1], Ido Sagi[1], Victoria E. Clark[1,8], Jesse M. Platt[1,9], Mrityunjoy Kar[10], Patrick M. McCall[10,11,12], Alicia V. Zamudio[1,3], John C. Manteiga[1,3], Eliot L. Coffey[1,3], Charles H. Li[1,3], Nancy M. Hannett[1], Yang Eric Guo[1], Tim-Michael Decker[13], Tong Ihn Lee[1], Tinghu Zhang[4,5], Jing-Ke Weng[1,3], Dylan J. Taatjes[13], Arup Chakraborty[6,7,14-18], Phillip A. Sharp[3,18], Young Tae Chang[19], Anthony A. Hyman[11,20], Nathanael S. Gray[4,5], Richard A. Young[1,3*]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[4]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[5]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, USA
[6]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[9]Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, 55 Fruit Street, Blake Building, 4th Floor, GI Unit, Boston, MA, 02114, USA
[10]Max Planck Institute for the Physics of Complex Systems, Nöthnitzerstraße 38, 01187 Dresden, Germany
[11]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany
[12]Center for Systems Biology Dresden, Pfotenhauerstraße 108, 01307 Dresden, Germany
[13]Department of Biochemistry, University of Colorado, Boulder, CO 80303, USA
[14]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[15]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[16]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[17]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard, Cambridge, MA 02139, USA
[18]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[19]Department of Chemistry, Pohang University of Science and Technology, and Center for Self-assembly and Complexity, Institute for Basic Science (IBS), Pohang 37673, Republic of Korea
[20]Cluster of Excellence Physics of Life, Technical University of Dresden, 01062 Dresden, Germany
[#]Equal Contribution
*Correspondence to Richard A. Young at young@wi.mit.edu

Abstract

The nucleus contains diverse phase-separated condensates that compartmentalize and concentrate biomolecules with distinct physicochemical properties. Here we consider whether condensates concentrate small molecule cancer therapeutics such that their pharmacodynamic properties are altered. We found that antineoplastic drugs become concentrated in specific protein condensates in vitro and that this occurs through physicochemical properties independent of the drug target. This behavior was also observed in tumor cells, where drug partitioning influenced drug activity. Altering the properties of the condensate was found to impact the concentration and activity of drugs. These results suggest that selective partitioning and concentration of small molecules within condensates contributes to drug pharmacodynamics and that further understanding of this phenomenon may facilitate advances in disease therapy.

Main Text

The 5-10 billion protein molecules of cells are compartmentalized into both membrane- and non-membrane-bound organelles (*1–3*). Many non-membrane-bound organelles are phase-separated biomolecular condensates with distinct physicochemical properties that can absorb and concentrate specific proteins and nucleic acids (*4–17*). We reasoned that selective condensate partitioning might also occur with small molecule drugs whose targets occur within condensates (Figure 1A), and that the therapeutic index and efficacy of such compounds might therefore relate to their ability to partition into condensates that harbor their target. To test this idea, we focused our study on a collection of nuclear condensates previously reported in cell lines, demonstrated that they all occur in normal human cells and in tumor cells, and then developed in vitro condensate droplet assays with key components of each of the nuclear condensates to enable testing of small molecules.

Nuclear condensates have been described in diverse cultured cell lines and contain one or more proteins that can serve both as markers of the condensate and as a scaffold for condensate formation in droplet assays in vitro (*10–12*, *18–32*). Specifically, transcriptional condensates are marked by the condensate forming proteins MED1 and BRD4 (*10*, *12*, *19*), splicing speckles by SRSF2 (*11*, *20*), heterochromatin by HP1α (*21*, *22*) and nucleoli by FIB1 and NPM1 (*23–25*) (Figure S1A). To determine whether such condensates can also be observed in the cells of healthy and malignant human tissue, we obtained biopsies of breast ductal epithelium, invasive ductal carcinoma, normal colon, and colon cancer (Figures S1B, S1C). Immunofluorescence revealed nuclear bodies containing these marker proteins in both normal and transformed tissue (Figures 1B, 1C). There was a broad distribution of nuclear body sizes and numbers, as expected for dynamic biomolecular condensates, and no significant differences were observed between benign and malignant tissue (Figures S2A-C). However, tumor cells acquire large super-enhancers (SEs) at driver oncogenes (*33*) and these can form tumor-specific transcriptional condensates.

We developed an assay to model these nuclear condensates and study the behavior of small molecules within these droplets (Figure 1D). We produced and purified recombinant fluorescently-labeled versions of MED1, BRD4, SRSF2, HP1α, FIB1, and NPM1 (Figure S3), and confirmed the ability of these proteins to form droplets in an in vitro assay (Figures S4A, S4B). To investigate the partitioning behavior of small molecules, we added the dyes Fluorescein (332Da) and Hoechst (452Da), as well as fluorescently-labeled dextrans (4.4 kilodaltons (kDa), to solutions containing each of the six protein condensates. The dyes and dextrans appeared to diffuse through all the condensates without substantial partitioning (Figures 1E, S5, S6A-D). Small molecule drugs are generally smaller than 1 kDa, so these results suggested that small molecules can freely diffuse

through these nuclear condensates unless there are factors other than size that influence partitioning.

We next sought to determine whether diverse clinically important drugs with targets that reside in nuclear condensates also exhibit free diffusion across these condensates, or display a different behavior. Cisplatin and mitoxantrone, members of a class of antineoplastic compounds that modify DNA through platination or intercalation, can be modified to have fluorescent properties (cisplatin) (*34*) or are inherently fluorescent (mitoxantrone). When added to droplet formation buffer with purified MED1, BRD4, SRSF2, HP1α, FIB1, or NPM1, cisplatin was found to be selectively concentrated in MED1 droplets (Figures 2A, S7A), with a partition coefficient of up to

Figure 1

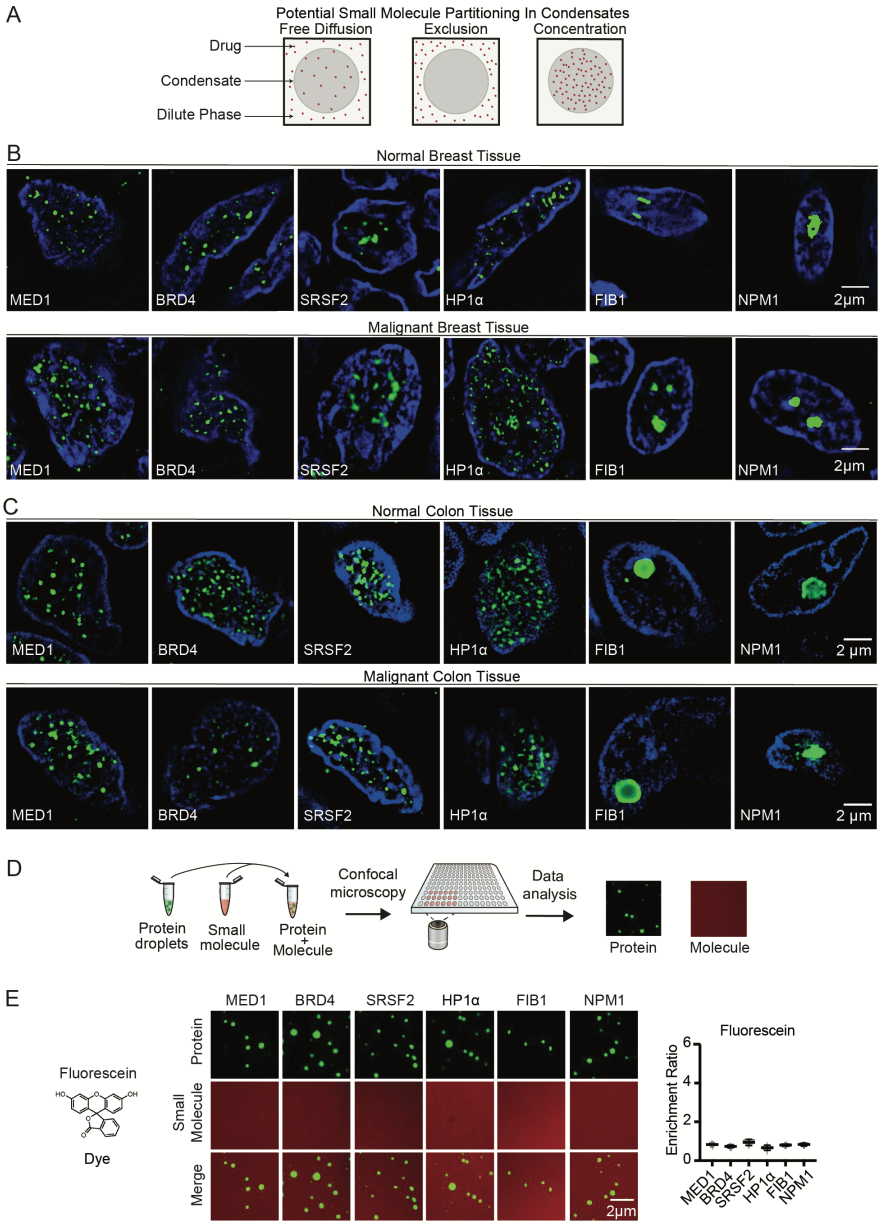


Fig. 1. Nuclear condensates in human tissue and in vitro.
A. Model illustrating potential behaviors of small molecules in nuclear condensates. (B-C) Immunofluorescence of scaffold proteins of various nuclear condensates in tissue biopsies from benign and malignant human breast
B. and benign and malignant colon tissue
C. in nuclei stained with Hoechst, imaged at 100x on a fluorescent confocal microscope (see also Figures S1, S2).
D. Schematic of in vitro droplet formation assay to measure small molecule partitioning into nuclear condensates.
E. In vitro droplet assay showing the behavior of fluorescein dye in the presence of six protein condensates formed in 125mM NaCl and 10% PEG, with 10μM protein and 5μM fluorescein, imaged at 150x on a confocal fluorescent microscope (see also Figures S3-S6). Quantification of enrichment of the drug is shown to the right, error bars represent SEM.

600 (Figures S8A-C). Fluorescent modification of cisplatin did not appear to contribute to this behavior in vitro, as the modified drug could be chased out of the condensate with unmodified cisplatin, and an isomer of cisplatin did not exhibit the same behavior (Figures S7B-D). Mitoxantrone was also concentrated in MED1 condensates, as well as in FIB1 and NPM1 condensates (Figures 2B, S7A, S8A-D). Consistent with these results, mitoxantrone is known to concentrate in the nucleolus where FIB1 and NPM1 reside (*35*, *36*). These results show that in contrast to the dyes tested above, small molecule drugs may concentrate in certain condensates even in the absence of the drug target.

We selected for further study antineoplastic drugs that target transcriptional regulators expected to be contained within transcriptional condensates in cells. These targets include: a) the estrogen receptor (ER), a transcription factor and nuclear hormone receptor, b) CDK7, a cyclin-dependent kinase that functions in transcription initiation and cell cycle control, and c) BRD4, a bromodomain protein and coactivator involved in oncogene regulation (Figures S9A, S9B). To monitor drug behavior with a confocal fluorescent microscope, we used a fluorescent tamoxifen analog (FLTX1), which targets ER, and modified fluorescent THZ1 and JQ1, which target CDK7 and BRD4, respectively (*37*, *38*). FLTX1 and THZ1 concentrated preferentially in MED1 droplets (Figures 2C, 2D, S7A), and this behavior was not attributable to the fluorescent moiety (Figures S7B, S7D). JQ1 concentration presented a different pattern, being concentrated in MED1, BRD4, and NPM1 droplets (Figures 2E, S7A, S7B). Reinforcing these results, we found that the small molecules that concentrate in MED1 condensates were also concentrated in condensates formed from purified whole Mediator complexes (Figure S10A) and in MED1 condensates formed in an alternative crowding agent (Figure S11A). The targets of these three compounds (ERα, CDK7, and the bromodomains of BRD4) are not present in these in vitro condensates, but are present in the SEs that form condensates with transcription factors and Mediator in vivo (*10*, *12*, *39*) (Figures S9A, S9B), suggesting that the ability of some small molecules to concentrate preferentially in the same condensate as their protein target may contribute to the pharmacological properties of these drugs.

To gain additional insight into the nature of interactions governing small molecule enrichment in condensates, we focused on the MED1-IDR condensate. Fluorescence recovery after photobleaching (FRAP) experiments showed that cisplatin molecules are highly mobile in this condensate (Figures S12A, S12B), suggesting that the condensate produces a physiochemical environment that facilitates drug concentration in a state of high dynamic mobility. To gain insights into the chemical features of small molecules that may contribute to selective association with MED1 in condensates, we used a fluorescent boron-dipyrromethene (BODIPY) library of 81

compounds with various combinations of chemical side groups (Figure S13A). Molecules that contained aromatic rings were found to preferentially concentrate in MED1 condensates (Figures S13A-D, S14A). These data suggest that pi-pi or pi-cation interactions are among the physicochemical properties that favor small molecule partitioning into MED1 condensates. Aromatic amino acids participate in pi-system interactions, and are overrepresented in the MED1 IDR relative to the other condensate forming proteins studied (Figure S3B). We generated a MED1 aromatic mutant protein (all 30 aromatic amino acids mutated to alanine) which retained the ability to form droplets in vitro, indicating that the aromatic amino acids are not required for droplet formation (Figure S14B, S14C), but small molecule probes containing aromatic rings and the polar molecule cisplatin no longer partitioned into condensates formed by the MED1 aromatic mutant protein (Figures S14D-F). These results suggest that the aromatic residues of MED1 condensates contribute to the physicochemical properties that selectively concentrate these small molecules.

Figure 2

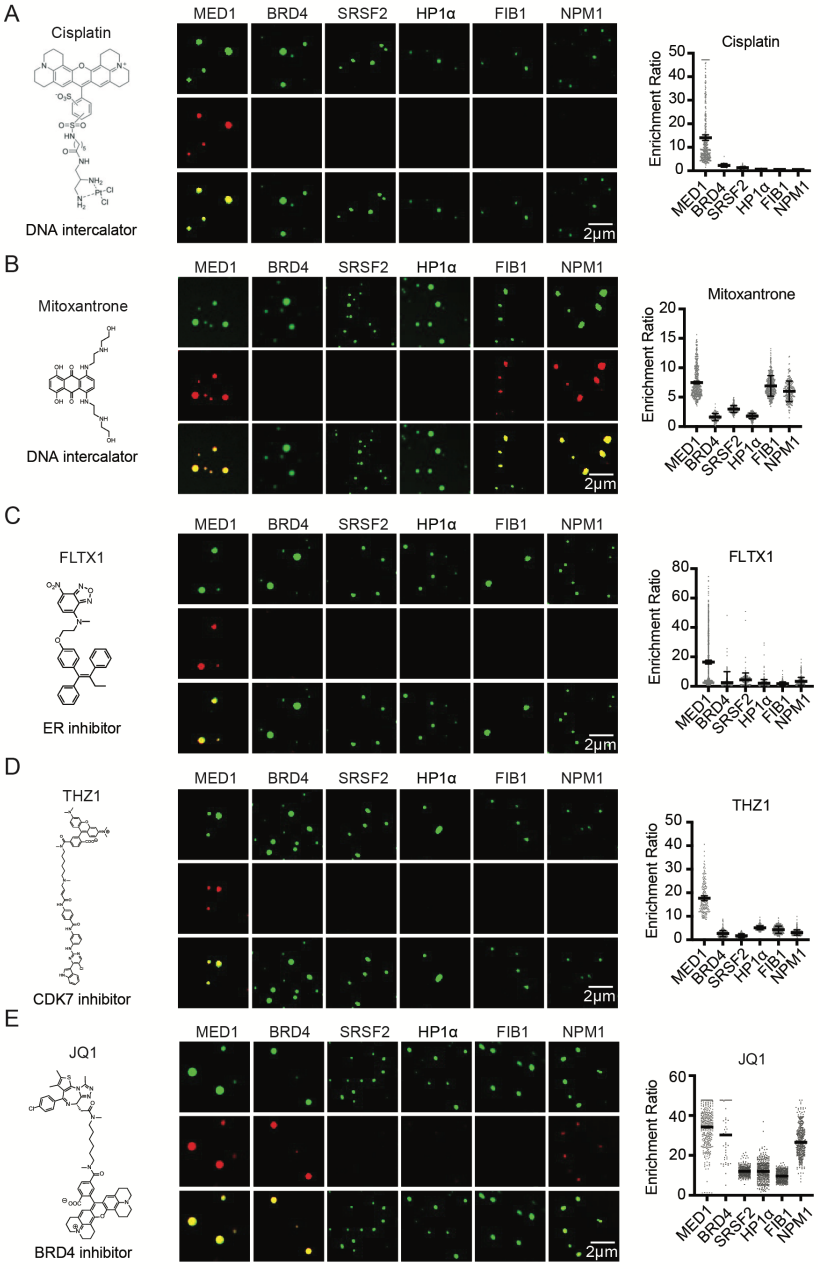


Fig. 2. The partitioning behavior of small molecule drugs in nuclear condensates in a droplet assay. Six nuclear condensates formed in 125mM NaCl and 10% PEG, with 10μM protein treated with either

A. 5μM Cisplatin-TMR
B. 50 μM Mitoxantrone
C. 100μM FLTX1
D. 5μM THZ1-TMR

E. 1μM JQ1-ROX imaged at 150x on a confocal fluorescent microscope (see also Figures S7-S11). Quantification of enrichment of the drug within droplets is shown to the right of each panel, error bars represent SEM (see also S12-S14).

We anticipated that the ability of small molecules to concentrate in specific condensates would influence target engagement and thus drug pharmacodynamics. To investigate this, we took advantage of the ability of condensates to incorporate DNA (Figure 3A, S15A), and measured the relative efficiency of DNA platination by cisplatin in MED1 condensates, where cisplatin is concentrated, versus HP1α condensates, where cisplatin freely diffuses (Figure 2A). DNA platination, visualized by size-shift on a bioanalyzer, was more prevalent in MED1 condensates than in HP1α condensates (Figure 3B), consistent with the expectation that elevated concentrations of cisplatin in the MED1 condensates yield enhanced target engagement. If cisplatin becomes concentrated in Mediator condensates in cells, we would expect that DNA colocalized within Mediator condensates would be preferentially platinated. To test this idea, we performed co-immunofluorescence in cisplatin-treated HCT116 colon cancer cells using an antibody that specifically recognizes platinated DNA (Figure S16A) (*40*) together with antibodies specific for MED1, HP1α, or FIB1. Consistent with cisplatin's preference for MED1 condensates in vitro, we found that platinated DNA frequently colocalized with MED1 condensates, but not with HP1α or FIB1 condensates (Figure 3C). To determine whether the ability of cisplatin to engage DNA is dependent on the presence of a MED1 condensate we treated cells with JQ1, which caused a loss of MED1 condensates (Figure S16B), and observed a concomitant reduction in platinated DNA at the *MYC* oncogene (Figures S16C, S16D). These results are consistent with the idea that concentration of small molecules in specific condensates can influence the efficiency of target engagement.

In cells, the preferential modification of DNA in MED1-containing condensates might be expected to selectively disrupt these condensates with prolonged treatment. To test this, HCT116 colon cancer cells were engineered to express GFP-tagged marker proteins for each of the 6 nuclear condensates (Figures S17A-G, S18A, S18B). When exposed to cisplatin, a selective and progressive reduction in MED1 condensates was observed (Figures 3D, S19A, S19B, S20A). Consistent with this, cisplatin treatment led to a preferential loss of MED1 ChIP-seq signal at SEs (Figures 3E S21A). Furthermore, high throughput sequencing data from platinated-DNA pull-down (*41*) revealed that cisplatin-modified DNA preferentially occurs at SEs, where MED1 is concentrated (*42*) (Figure 3F). These results are consistent with reports that cisplatin preferentially modifies transcribed genes (*41*, *43*), and argue that this effect is due to preferential condensate partitioning. Taken together, these results suggest a model where cisplatin preferentially modifies SE DNA, which in turn leads to dissolution of these condensates. Previous studies have shown that diverse tumor cells become highly dependent on SE driven oncogene expression (*44–48*), which might explain why platinum drugs, which are capable of general DNA modification, are effective therapeutics in diverse cancers (*49*).

Figure 3

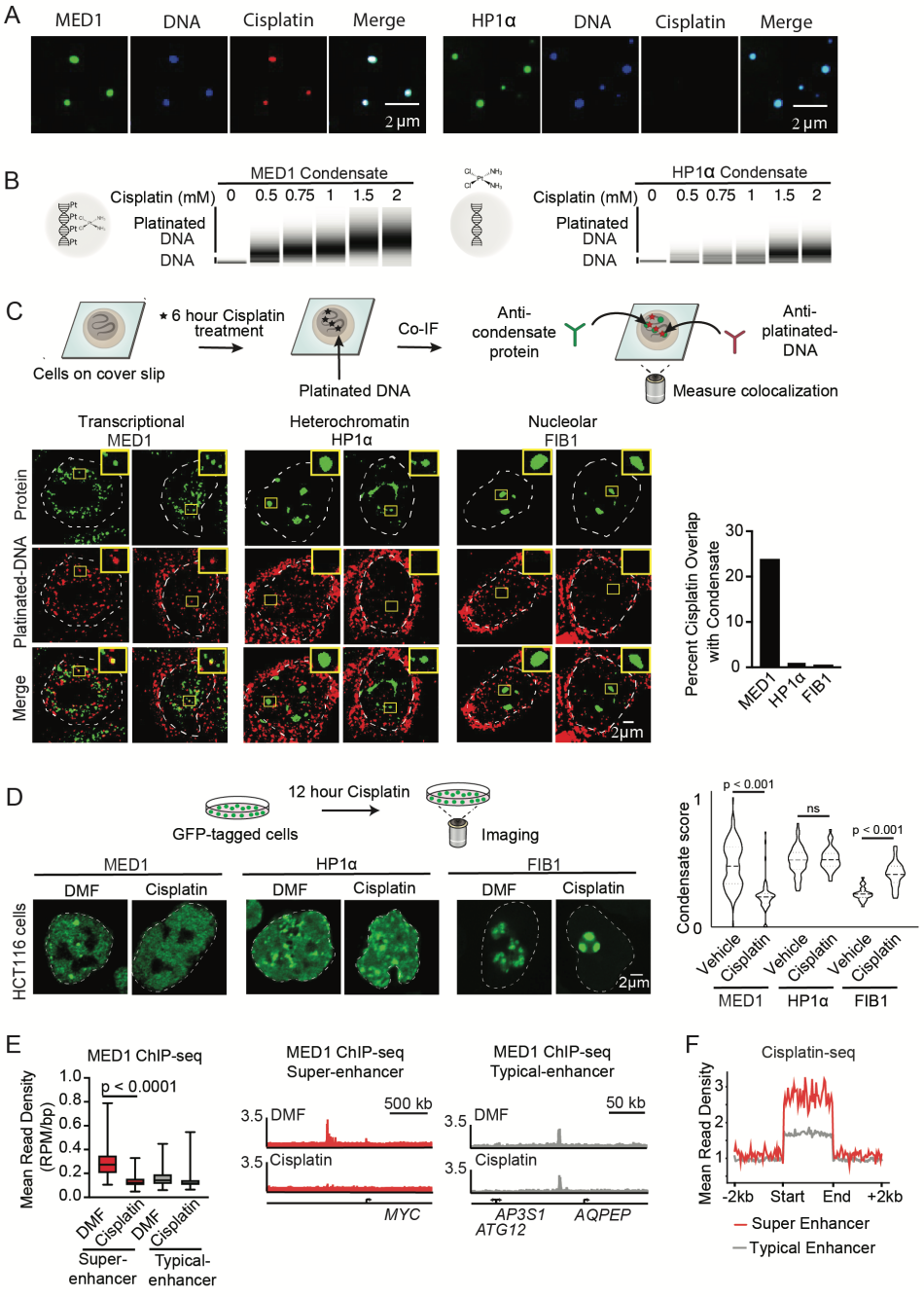


Fig. 3. Small molecule concentration within condensates influences drug activity.
A. In vitro droplet assay of MED1 and HP1α condensates formed in 125mM NaCl and 10% PEG, 5nM of 450bp DNA, 10μM MED1, and 5μM cisplatin-TR, imaged at 150x on a confocal fluorescent microscope (see also Figure S15).
B. Bioanalyzer tracings of DNA contained within either MED1 or HP1α droplets exposed to the indicated concentration of cisplatin.

We explored the behavior of another clinically important antineoplastic drug, tamoxifen, to assess whether drug response and resistance are associated with partitioning in condensates (Figure 4A). ERα incorporates into MED1 condensates in an estrogen-dependent manner in vitro (*12*); droplet assays confirmed this and revealed that the addition of tamoxifen leads to eviction of ERα from the MED1 condensates (Figure 4B). We further investigated the effects of estrogen and tamoxifen on MED1 condensates in breast cancer cells, focusing on the *MYC* oncogene due to its prominent oncogenic role and responsiveness to estrogen (*50*). MED1 condensates were observed on the *MYC* oncogene in the ER+ breast cancer cell line MCF7 (S9A, S22A-D). DNA FISH with MED1 IF revealed that estrogen enhances formation of MED1 condensates at the *MYC* oncogene and tamoxifen treatment reduces these (Figures S23A, S23B). Artificial MED1 condensates without ER concentrated FLTX1 at the site of the condensate (Figure S24A), indicating that ER is not required for the partitioning of FLTX1 into MED1 condensates in cells. These results are consistent with the model that ERα interacts with MED1 condensates in an estrogen-dependent, tamoxifen-sensitive manner to drive oncogene expression in breast cancer cells.

The mechanisms that produce drug resistance can provide clues to drug activity in the clinical setting. Endocrine therapy and tamoxifen resistance is an enduring clinical challenge and is associated with multiple mechanisms including ERα mutation and MED1 overexpression (Figure 4A, S25) (*51*–*55*). To investigate whether ERα mutations alter ERα behavior in condensates, we produced 4 patient-derived ERα mutant proteins and tested their partitioning in the presence of tamoxifen. In contrast to WT ERα, condensates composed of patient-derived ERα mutants and MED1 were not disrupted upon tamoxifen treatment (Figures 4B, S26A, S26B). The ERα point mutations reduce the affinity for tamoxifen approximately 10-fold (*52*), indicating that the drug concentration in the droplet is inadequate to evict these ER mutant proteins when this affinity is reduced.

Fig. 3. cont.

C. (Top) Schematic of an assay to determine the location of platinated DNA relative to various nuclear condensates. (Bottom) Co-immunofluorescence of platinated DNA and the indicated protein in HCT116 cells treated with 50μM cisplatin for 6 hours. Imaged at 100x on a confocal fluorescent microscope. Quantification of overlap shown to the right.

D. (Top) Schematic of a live cell condensate dissolution assay. (Bottom) HCT116 cells bearing endogenously mEGFP-tagged MED1, HP1α, or FIB1 treated with 50μM cisplatin for 12 hours. Quantification of MED1, HP1α, or FIB1 condensate score is shown to the right.

E. MED1 ChIP-seq in HCT116 cells treated with vehicle or 50μM cisplatin for 6 hours. (Left) Plotted are mean read density of MED1 at super-enhancers and typical-enhancers (error bars show min and max) and (Right) gene tracks of MED1 ChIP-Seq at the *MYC* super-enhancer and AQPEP typical-enhancer.

F. Metaplot of cisplatin-DNA-Seq in cisplatin treated Hela cells comparing super-enhancers and typical enhancers (*41*) (see also Figures S16-S21).

Figure 4

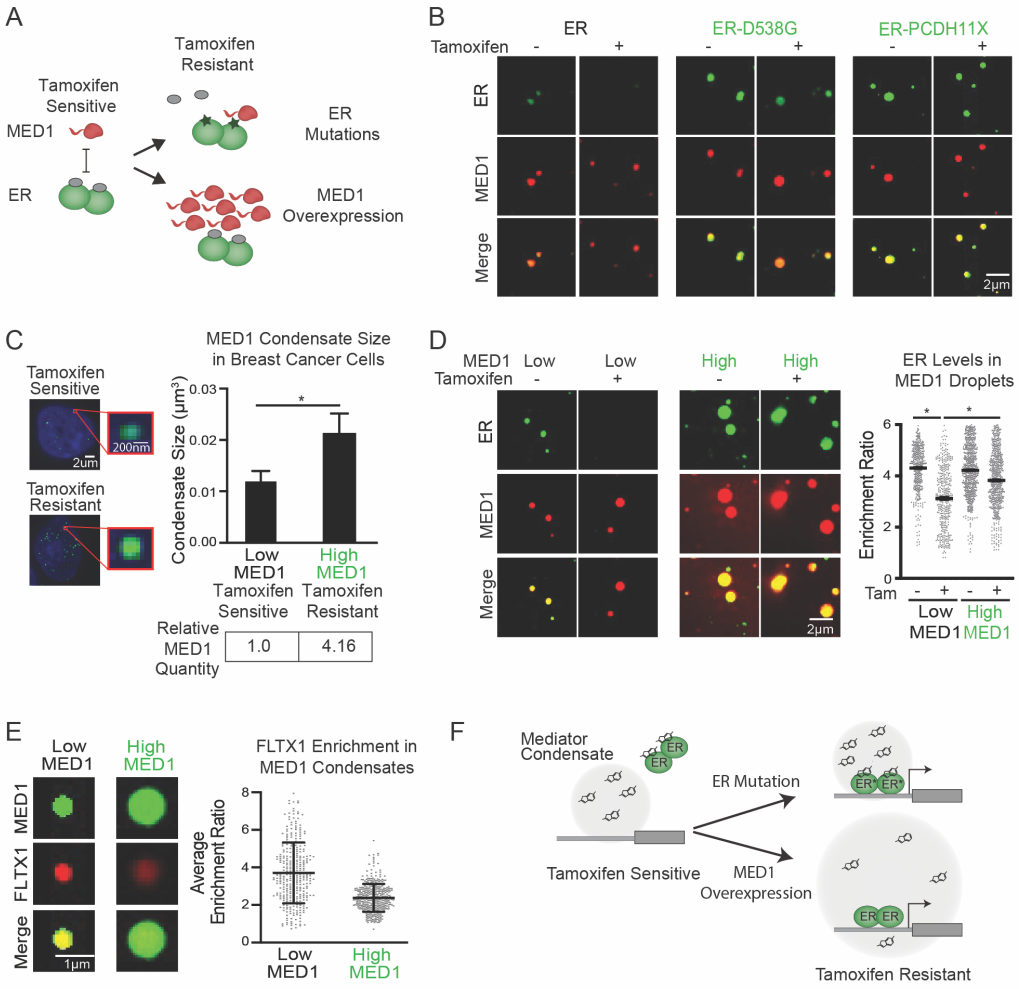


Fig. 4. Tamoxifen action and resistance in MED1 condensates.

A. Schematic showing tamoxifen resistance by ER mutation and MED1 overexpression in breast cancer.

B. In vitro droplets assay of the indicated form of GFP-tagged ER in the presence of estrogen, +/- 100μM tamoxifen. Droplets are formed in 125mM NaCl and 10% PEG with 10μM each protein and 100μM estrogen.

C. (Left) Immunofluorescence of MED1 in tamoxifen sensitive (MCF7) and resistant (TAMR7) ER+ breast cancer cell lines imaged at 100x on a confocal fluorescent microscope. (Top right) Quantification of MED1 condensate size in breast cancer cells. (Bottom right) Relative quantities of MED1 in the indicated breast cancer cell line by western blot, error bars show SEM.

D. In vitro droplets assays of ER in the presence of 100μM estrogen, +/- 100μM tamoxifen with either 5μM (Low) or 20μM (High) MED1. Droplets are formed with 5μM ER in 125mM NaCl and 10% PEG, imaged at 150x on a confocal fluorescent microscope, error bars are SEM.

E. In vitro droplet assay with either 5μM (Low) or 20μM (High) MED1 with 100μM FLTX1 in 125mM NaCl and 10% PEG, error bars are SD.

F. Models for tamoxifen resistance due to altered drug affinity (via ER mutation) or concentration (via MED1 overexpression) (see also Figures S22-S30).

MED1 overexpression is associated with tamoxifen resistance and poor prognosis in breast cancer (*51*), but it is not clear why overexpression of one subunit of the Mediator complex produces resistance. We considered the possibility that overexpressed MED1 is incorporated into transcriptional condensates, which contain clusters of Mediator molecules (*39*), thereby expanding their volumes and diluting the available tamoxifen (Figure S27A). We found that the tamoxifen-resistant breast cancer cell line TAMR7 (*56*), which was derived from the tamoxifen-sensitive cell line MCF7, produces 4-fold elevated levels of MED1 protein (Figure S27B). The volume of MED1-containing condensates is 2-fold larger in these cells (Figures 4C, S27C). When modeled in an in vitro droplet assay, we found that a 4-fold increase in MED1 levels led to a commensurate increase in droplet size (Figures S28A, S28B). Furthermore, we found that 100μM tamoxifen prevented ERα incorporation into MED1 condensates (Figures 4B, 4D), but was much less effective in preventing ERα incorporation into the larger MED1 condensates produced with higher MED1 levels (Figure 4D). To confirm that the levels of tamoxifen in the larger droplets are more dilute, we measured the enrichment of the fluorescent tamoxifen analog FLTX1 in MED1 droplets, and found that the larger condensates have lower concentrations of the drug (Figure 4E). These results were mirrored in cells, where a collection of tethered ERα molecules form a MED1 condensate that is eliminated by tamoxifen, but when MED1 is overexpressed tamoxifen is unable to dissociate the ERα-MED1 condensate (Figure S29A). Similarly, knockdown of MED1 in tamoxifen resistant breast cancer cells sensitizes cells to tamoxifen (*51*, *55*). These results support a model of tamoxifen resistance where MED1 overexpression causes the formation of larger transcriptional condensates, in which tamoxifen is diluted and thereby less effective in dissociating ER from the condensate (Figure 4F).

Our results show that drugs partition selectively into condensates, that this can occur through physicochemical properties that exist independent of their molecular targets, and that cells can develop resistance to drugs through condensate altering mechanisms. This may explain the surprising observation that inhibition of global gene regulators such as BRD4 or CDK7 can have selective effects on oncogenes that have acquired large SEs (*46*); selective partitioning of inhibitors like JQ1 and THZ1 into SE condensates will preferentially disrupt transcription at those loci. These results also have implications for future development of efficacious disease therapeutics; effective target-engagement will depend on measurable factors such as drug partitioning in condensates (Figures S30A-D). Condensate assays of the type described here may thus help optimize condensate partitioning, target engagement, and the therapeutic index of small molecule drugs.

Acknowledgments: We thank Chris Glinkerman for helpful comments, Wendy Salmon of the W.M Keck Microscopy Facility, and Tom Volkert, Jennifer Love, Stephen Mraz, and Sumeet Gupta of the Whitehead Genome Technologies Core for technical assistance. We thank the light microscopy facility at the MPI-CBG in Dresden for extensive help and support. **Funding:** This work was supported by NIH grants GM123511, CA213333, CA155258 (RAY), NSF grant PHY1743900 (RAY), NIH grant GM117370 (DJT), Max Planck Society (AAH), American Society of Clinical Oncology Young Investigator Award (IAK), American Cancer Society Postdoctoral Fellowship (IAK), Ovarian Cancer Research Alliance Mentored Investigator Award (IAK), Swedish Research Council Postdoctoral Fellowship (VR 2017-00372) (AB), Hope Funds for Cancer Research (AD), Gruss-Lipper Postdoctoral Fellowship and by the Rothschild Postdoctoral Fellowship (IS), NIH grant T32:5T32DK007191-45 (JMP), German Research Foundation DFG postdoctoral fellowship DE 3069/1-1 (T-MD), Cancer Research Institute Irvington Fellowship (YEG), Damon Runyon Cancer Research Foundation Fellowship (2309-17) (BRS), ELBE postdoctoral fellowship (PM). **Author contributions:** Conceptualization (IAK, AB, RAY), Data curation (SWH, JEH), Formal analysis (IAK, AB, LKH, SWH, JH, PMM, MK, CHL), Funding acquisition (DJT, AC, PAS, YTC, AAH, NSG, RAY), Investigation (IAK, AB, LKA, SWH, MF, AD, OO, KS, BRS, IS, VEC, JMP, MK, PMM, AVZ, YEG), Methodology (IAK, AB, LKA, AD, OO), Project administration (IAK, AB, TIL, RAY), Resources (MF, JCM, ELC, CHL, NMH, TMD, TZ, DJT, AC, YTC, AAH, NSG, RAY), Software (JEH, KS), Supervision (RAY), Validation (IAK, AB, LKA, SWH), Visualization (IAK, AB, LKA, SWH, AD, JEH, KS, BRS, MK, PMM, AZ), Writing - original draft (IAK, AB, RAY), Writing – reviewing and editing (all authors). **Competing interests:** RAY is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics and Dewpoint Therapeutics. AAH is a founder and shareholder in Dewpoint Therapeutics, and a shareholder in Careway Therapeutics. NSG is a founder and shareholder of Syros Pharmaceuticals. IAK is a shareholder and member of the Scientific Advisory Board of Dewpoint Therapeutics. JKW is a co-founder, member of the Scientific Advisory Board, and shareholder of DoubleRainbow Biosciences. TIL is a shareholder of Syros Pharmaceuticals and a consultant to Camp4 Therapeutics. AC is on the Scientific Advisory Board of Dewpoint Therapeutics and Omega Therapeutics. PAS is a is a shareholder and consultant to Dewpoint Therapeutics. All other authors declare no competing interests. YTC is an inventor on patents US9513294 B2, China Patent ZL 201380067802.8, EP 2,938,619 B1, Japan 6300380 held by POSTECH University that cover Diversity-oriented Fluorescent Library Approaches. IAK, AB, and RAY are inventors on patent application submitted by The Whitehead Institute that covers small molecule drug partitioning in and acting upon biomolecular condensates.

Data and materials availability: All data is available in the main text or the supplementary materials. High-throughput sequencing data sets are available in GEO (GSE149085).

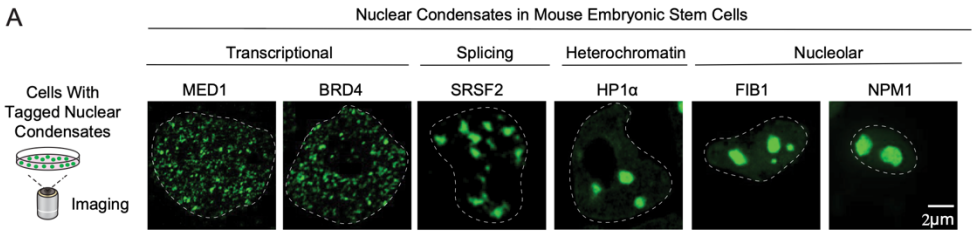
Supplementary Materials:

Materials and Methods

Figures S1-S30

Supplementary Figures

Figure S1



B

Tumor Tissues

Type	Tumor Grade	Age	Sex	Biosample Diagnosis	AJCC/UICC Stage Group	Other
Breast	Grade II	84	Female	Infiltrating ductal carcinoma	T2N1M0	ER+ PR+ Hercep test-0 Post Menopausal
Colon	Grade II	63	Female	Adenocarcinoma	T2NXM0	

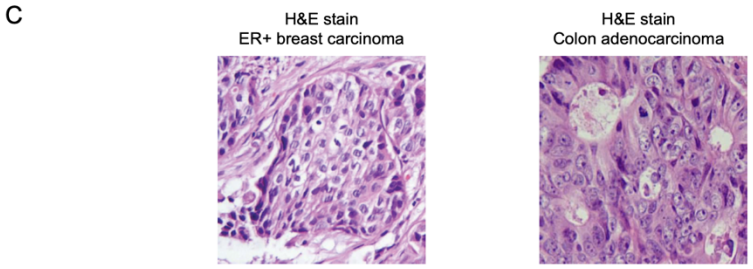


Figure S1. Nuclear condensates in cell lines and human tumor tissue.
A. Mouse embryonic stem cells expressing either endogenously mEGFP-tagged proteins (MED1, BRD4, SRSF2), mCherry-tagged proteins (HP1α) or transfected with constructs expressing GFP-tagged proteins (NPM1, FIB1) were imaged by confocal fluorescent microscopy.
B. Clinical data from biopsied breast and colon cancer specimen.
C. H&E staining of ER positive breast carcoinoma and colon adenocarcinoma.

Figure S2

A

Percent Nuclear Volume of Nuclear condensates

	Transcriptional		Splicing	Heterochromatin	Nucleolar		Total
	MED1	BRD4	SRSF2	HP1α	FIB1	NPM1	
Normal Breast Tissue	2.1 +/- 6.9%	1.8 +/- 2.0%	7.3 +/- 0.2%	2.8 +/- 1.2%	11 +/- 1.4%	8.3 +/- 1.2%	33.3%
Malignant Breast Tissue	2.8 +/- 2.0%	1.9 +/- 3.3%	6.9 +/- 4.3%	2.9 +/- 3.3%	7.6 +/- 2.3%	5.8 +/- 3.3%	27.9%
Normal Colon Tissue	4.1 +/- 1.0%	10 +/- 3.7%	4.9 +/- 1.6%	7.9 +/- 3.7%	11 +/- 9.9%	6.6 +/- 1.5%	44.5%
Malignant Colon Tissue	2.6 +/- 1.5%	2.0 +/- 0.8%	6.1 +/- 2.1%	8.8 +/- 7.2%	13 +/- 7.3%	10 +/- 11%	42.5%

B

Average Volume of Nuclear condensates

	Transcriptional		Splicing	Heterochromatin	Nucleolar	
	MED1	BRD4	SRSF2	HP1α	FIB1	NPM1
Normal Breast Tissue	0.021 +/- 0.018	0.026 +/- 0.013	0.018 +/- 0.005	0.052 +/- 0.028	0.059 +/- NA	2.653 +/- 2.209
Malignant Breast Tissue	0.015 +/- 0.003	0.028 +/- 0.005	0.037 +/- 0.026	0.018 +/- 0.009	0.143 +/- 0.119	0.816 +/- 0.692
Normal Colon Tissue	0.034 +/- 0.008	0.084 +/- 0.055	0.039 +/- 0.009	0.048 +/- 0.015	0.762 +/- 0.957	0.950 +/- 0.861
Malignant Colon Tissue	0.040 +/- 0.017	0.023 +/- 0.004	0.046 +/- 0.007	0.038 +/- 0.023	0.927 +/- 1.521	0.431 +/- 0.270

C

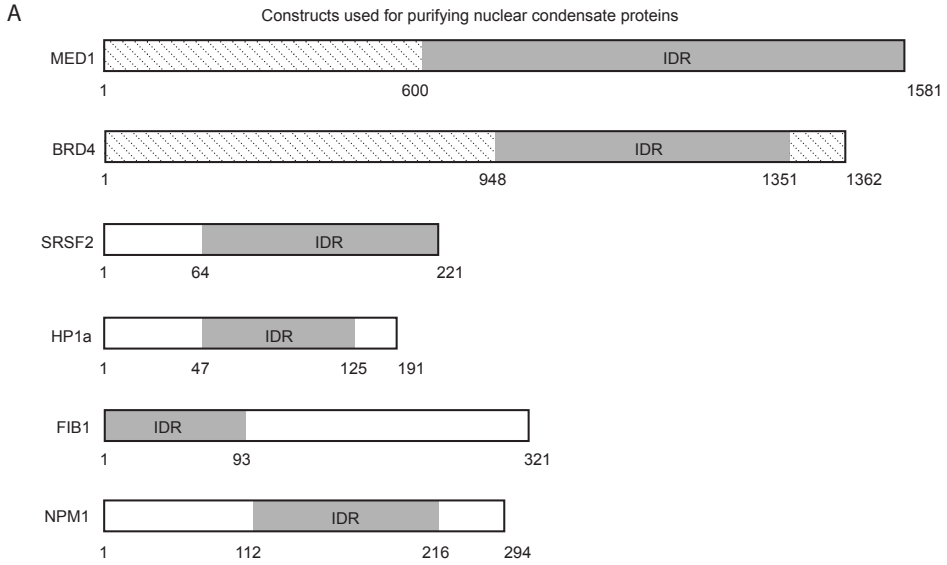
Average Number of Nuclear condensates

	Transcriptional		Splicing	Heterochromatin	Nucleolar	
	MED1	BRD4	SRSF2	HP1α	FIB1	NPM1
Normal Breast Tissue	93 +/- 54.9	53 +/- 20.9	37 +/- 20.5	55 +/- 57.3	18 +/- NA	4 +/- 5.2
Malignant Breast Tissue	78 +/- 41.1	72 +/- 28.0	32 +/- 16.1	133 +/- 81.0	21 +/- 6.4	8 +/- 6.0
Normal Colon Tissue	60 +/- 7.9	15 +/- 8.1	47 +/- 20.1	74 +/- 49.2	10 +/- 10.4	5 +/- 2.9
Malignant Colon Tissue	24 +/- 12.9	66 +/- 19.3	47 +/- 28.9	42 +/- 18.9	13 +/- 14.6	11 +/- 7.4

Figure S2. Volume and number of nuclear condensates in normal and tumor tissue.
A. Volume of nuclear condensates in normal and malignant breast tissue (upper) and in normal and malignant colon tissue (lower). Values indicate percent nuclear volume and standard deviation. There were no significant differences between the individual nuclear condensates in normal and malignant states.
B. Table showing average volume of nuclear condensates in normal and malignant tissue.
C. Table showing average number of nuclear condensates in normal and malignant tissue.

Figure S3

A



B

	Aromatic Amino Acid Analysis							
	Full-length Protein				IDR			
	F	W	Y	Total	F	W	Y	Total
MED1	40	4	26	70	19	1	10	30
BRD4	25	5	18	48	10	1	2	13
FIB1	13	1	7	21	6	0	0	6
HP1α	6	4	5	15	3	1	1	5
SRSF2	4	0	9	13	0	0	4	4
NPM1	7	2	4	13	2	0	0	2

	Basic/Acidic Patch Analysis			
	Total		IDR	
	CIE+	CIE-	CIE+	CIE-
MED1	11	10	7	3
BRD4	11	9	3	3
FIB1	3	0	2	0
HP1α	4	4	3	1
SRSF2	4	2	4	2
NPM1	3	4	3	2

Figure S3. Nuclear condensate forming proteins.

A. Schematic representation of constructs used for purifying nuclear condensate proteins. The IDR (intrinsically disordered region) alone was used for MED1 and BRD4 proteins and the full length was used for HP1α, SRSF2, NPM1, and FIB1 proteins.

B. (Upper) Number of hydrophobic amino acids Phenylalanine (F), Tryptophan (W), and Tyrosine (Y) in the IDR and full-length protein. MED1 IDR has the highest number of hydrophobic residues. (Lower) Table of Positive Charged Interaction Elements (CIE+) and Negative Charged Interaction Elements (CIE-) of the IDR or full length nuclear condensate protein (*63*). These results indicate that MED1 protein might participate in interactions governed by the pi-system.

Figure S4

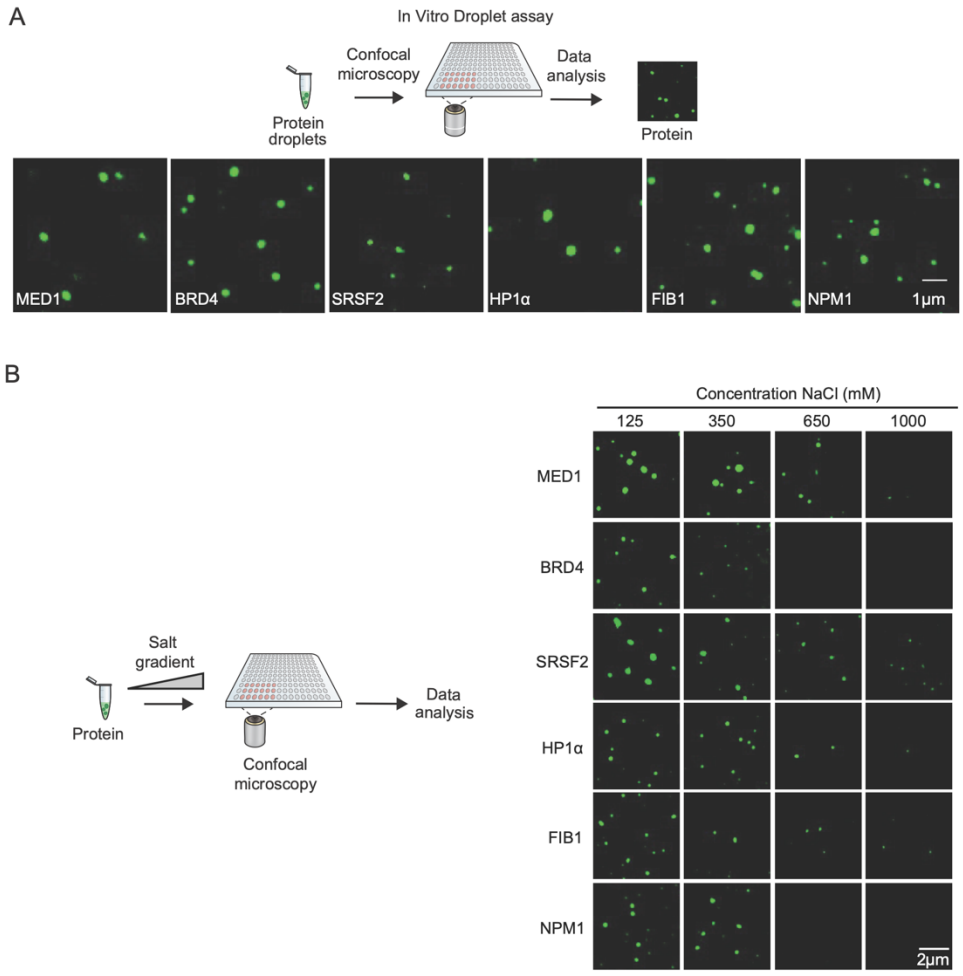
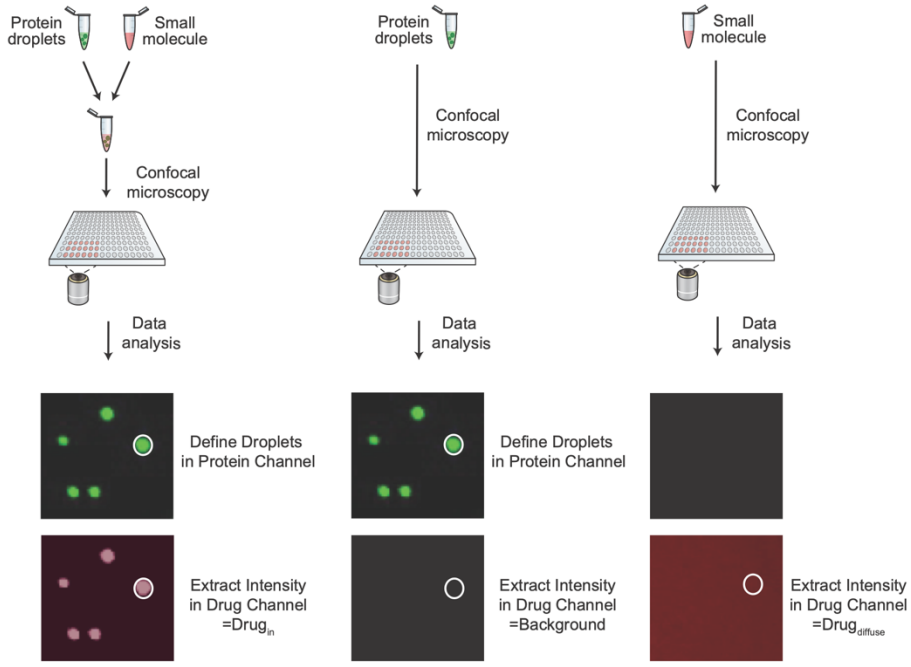


Figure S4. In vitro droplets of condensate forming proteins.
A. Confocal microscopy of in vitro droplet formation assays of the indicated GFP-tagged protein in 125mM NaCl and 10% PEG. MED1 and BRD4 proteins are the IDR portion only.
B. Confocal microscopy images of MED1, BRD4, SRSF2, HP1α, FIB1, and NPM1 nuclear condensates at the indicated concentration of salt (125mM, 350mM, 650mM, 1000mM NaCl), experiments were performed with 10μM protein in 10% PEG.

Figure S5



$$\text{Enrichment Ratio} = \frac{\text{Drug}_{in} - \text{Background}}{\text{Drug}_{diffuse}}$$

Figure S5. Schematic representation of enrichment ratio calculations. Droplets are defined in the protein channel and maximum intensity of drug is measured in that area to obtain $drug_{in}$ (left panel), background is measured in the drug channel in areas defined by the protein channel in an in vitro droplet reaction containing protein but no drug (middle panel), and $drug_{diffuse}$ intensity is measured in a droplet reaction without the protein (right panel).

Figure S6

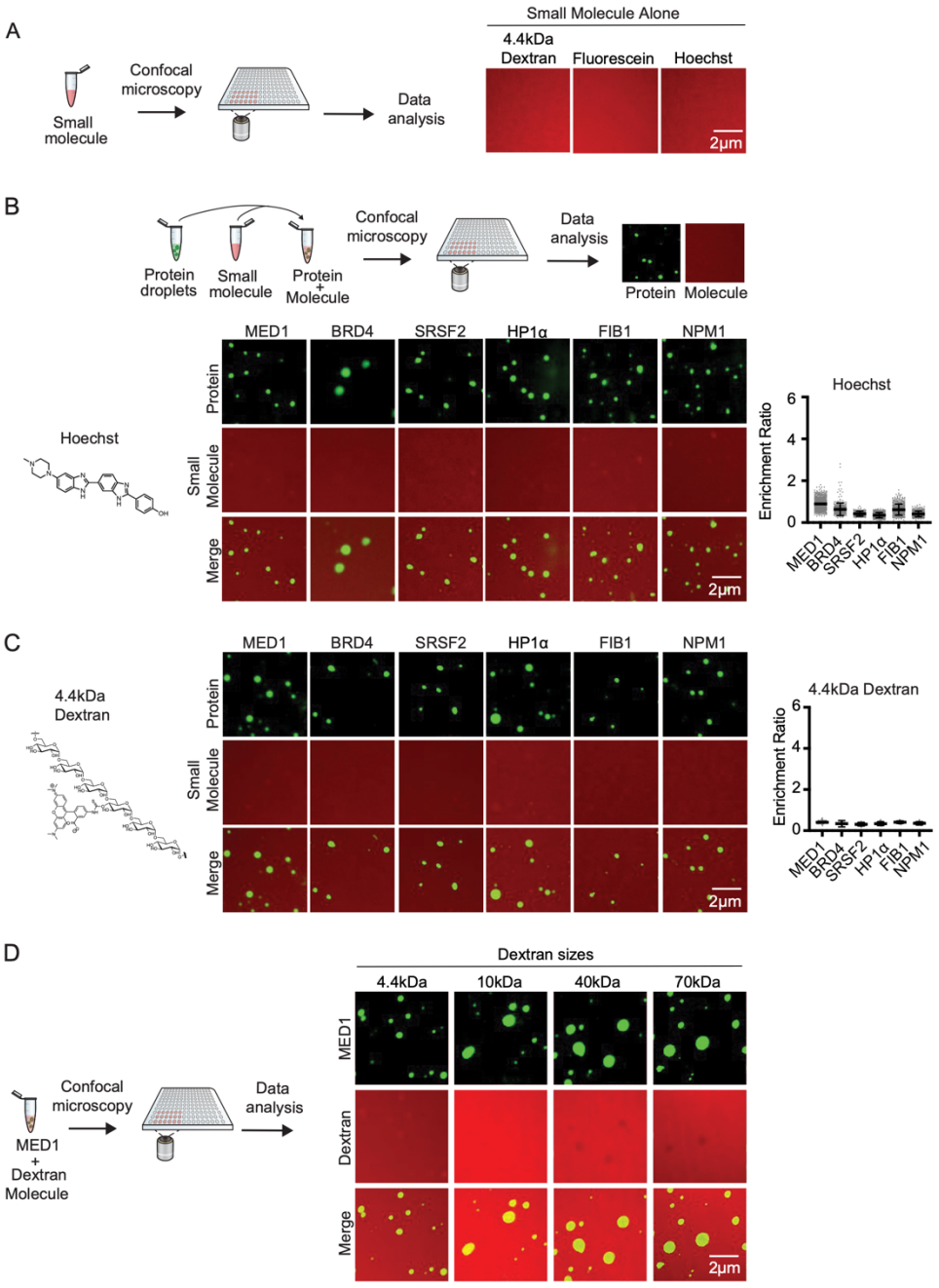


Figure S6. Small molecule partitioning in nuclear condensates.
A. Confocal microscopy of in vitro droplet formation assays of the indicated small molecule alone (4.4kDa dextran, fluorescein, and hoechst) without any protein added to the reaction. All small molecules alone show a diffuse fluorescent signal indicating that the molecule alone does not form droplets. (B-C) Confocal microscopy images showing the behavior of hoechst
B. and 4.4kDa dextran

C. relative to six nuclear condensates formed in vitro, in 125mM NaCl and 10% PEG. Quantification shown to the right, error bars represent SEM. Both hoechst and dextran diffuse freely through the condensates tested without being excluded or concentrated. Schematic of the assay shown at top.
D. Confocal microscopy images of fluorescently-labeled 4.4kDa, 10kDa, 40kDa, and 70kDa dextran in MED1 condensates. Experiments were performed with 10μM protein and 0.1mg/ml TRITC-labeled dextran, in 125mM salt and 16% ficoll. Dextran of smaller sizes (4.4kDa and 10kDa) are able to freely diffuse through the condensates while larger sizes of dextran (40kDa and 70kDa) are partially excluded from MED1 condensates. This indicates that the effective pore sizes of the condensates studied is at least 10kDa (*65*).

Figure S7

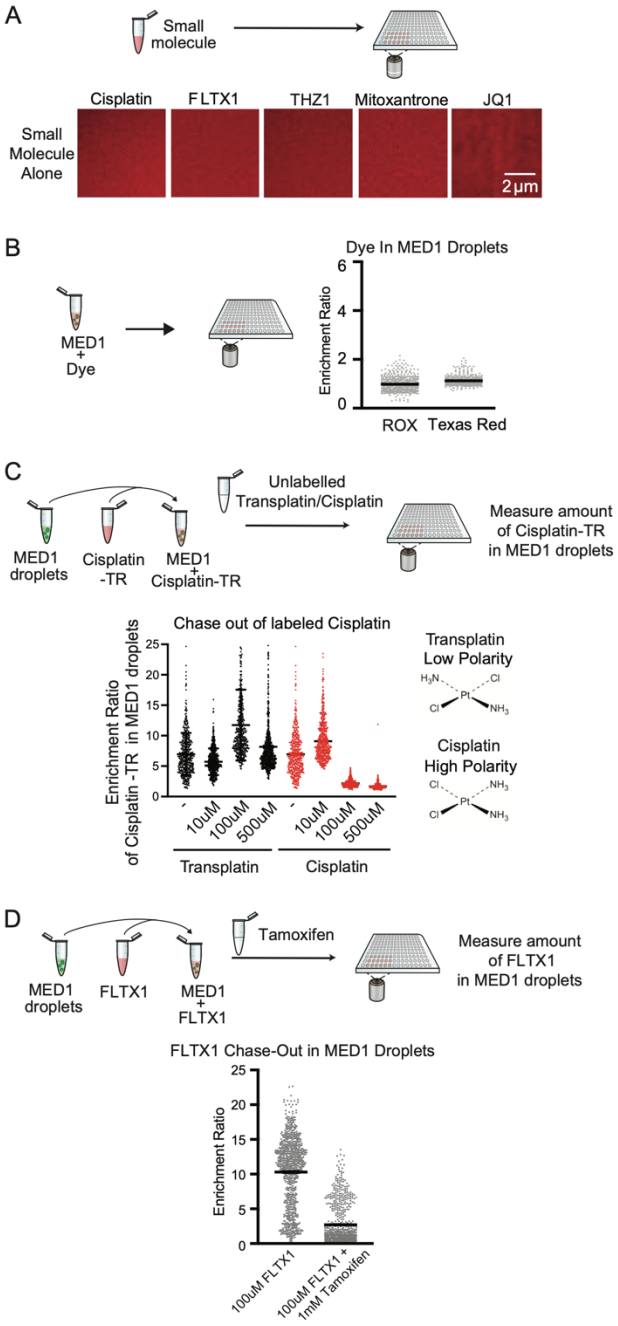
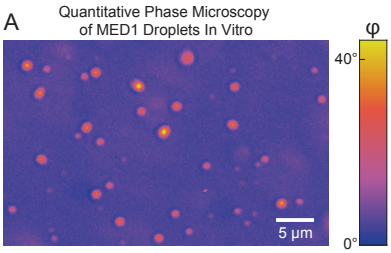


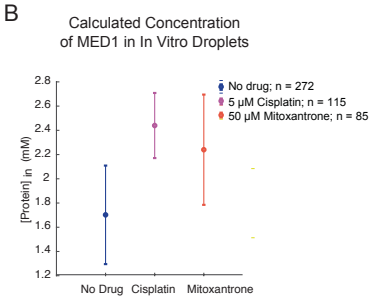
Figure S7. Properties of small molecule drugs, not their fluorescent moiety, govern partitioning into condensates.

A. Confocal microscopy of in vitro droplet formation assays of the indicated small molecule drug alone (cisplatin, FLTX1, THZ1, mitoxantrone, and JQ1) without any protein added to the reaction. All small molecule drugs alone show a diffuse fluorescent signal indicating that the molecules alone do not form droplets.

B. ROX and Texas Red enrichment in MED1 droplets formed in 125mM NaCl and 10% PEG measured by confocal microscopy. Neither of the two dyes used to visualized drugs were enriched in MED1 condensates.
C. Schematic of in vitro droplet drug chase out experiment. Labeled cisplatin is added to MED1 droplets to form MED1 droplets concentrated with cisplatin-TR. Unlabeled transplatin or unlabeled cisplatin is added to the droplet mixture and the amount of labeled cisplatin-TR remaining in the droplet is measured after chase out. Transplatin, a clinically ineffective trans-isomer of cisplatin, is not able to chase out cisplatin-TR, while high concentrations of unlabeled cisplatin is able to chase out cisplatin-TR.
D. Schematic of in vitro droplet drug chase out experiment. Graph showing FLTX1 enrichment in MED1 droplets upon tamoxifen addition measured by confocal microscopy. Tamoxifen was able to chase-out FLTX1 from MED1 droplets. All error bars shown represent SEM.

Figure S8





C

Calculated Partition Coefficient
Based on Dilute Phase Spectroscopy

Cisplatin-TR (µM)	2	4	6	8
% in Supernatent	24	39	37	41
% in Droplet	76	61	63	59
Partitioning Coefficient	630	310	340	290

Mitoxantrone (µM)	2	4	6
% in Supernatent	65	60	59
% in Droplet	35	40	41
Partitioning Coefficient	110	130	140

Figure S8. Small molecule drugs can be concentrated into MED1 condensates by 100-folds.

A. Quantitative phase microscopy of MED1 droplets formed in 125 mM NaCl and 10% PEG. Colorbar indicates optical phase delay, φ, in degrees. From phase images, we calculate the average MED1 concentration in individual condensates.

B. Graph showing MED1 concentration in in vitro droplets upon the addition of no drug, 5 μM cisplatin or 50 μM mitoxantrone. Datapoints are population averages (n = 272, 115 and 85 individual condensates for each condition). Error bars denote standard deviation.

C. Varying concentration of cisplatin or Mitoxantrone was added to MED1 droplets and the concentration of drug remaining in solution was measured by uv-spectroscopy. Combining the spectroscopy measurements with an estimate of the total volume of the MED1 condensate phase obtained from the measurements in (B), we estimate the partition ratio of cisplatin to be up to 600-fold and the partitioning ratio of Mitoxantrone to be approximately 100-fold.

Figure S9

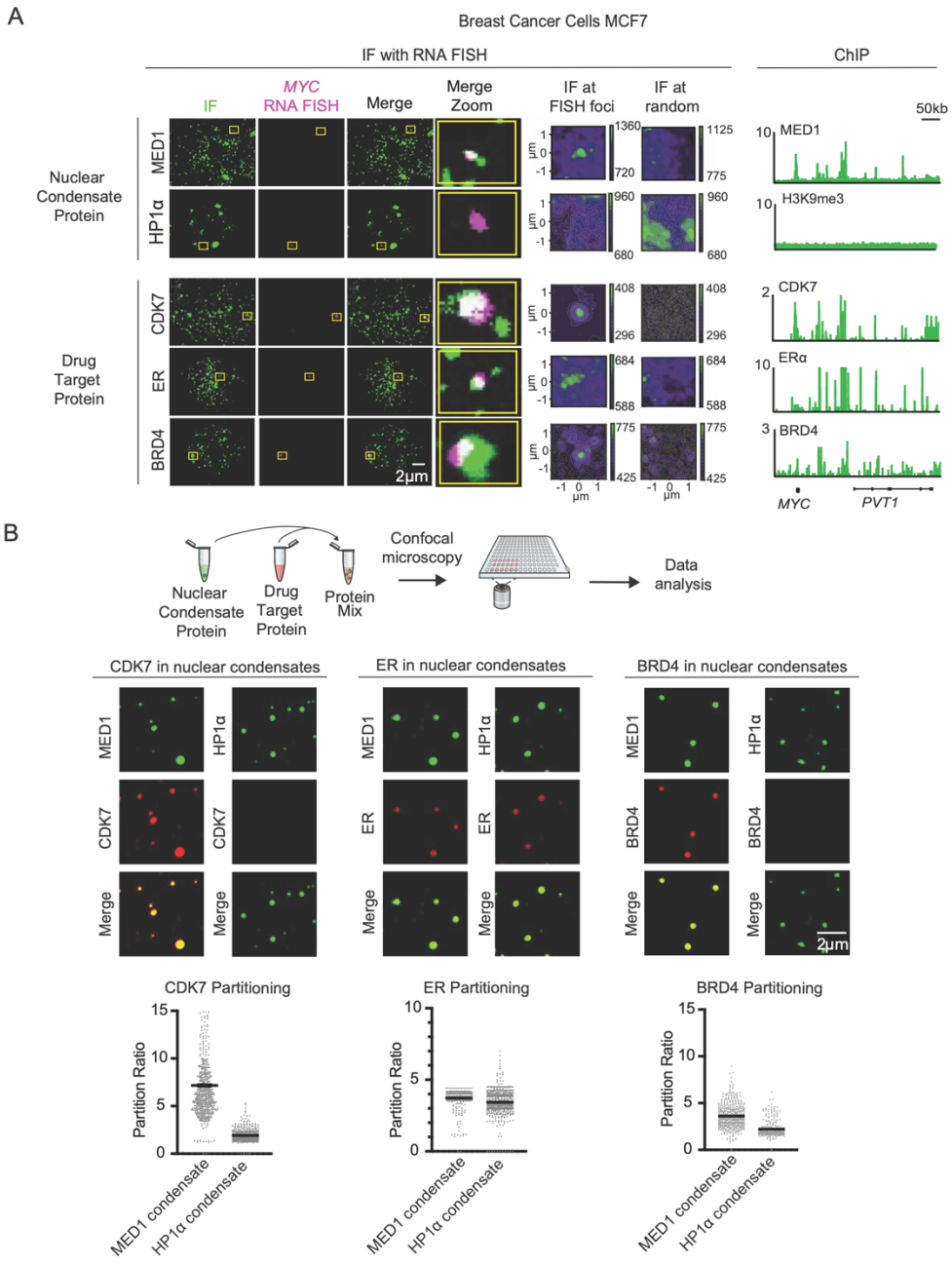


Figure S9. Association of drug targets with transcriptional condensates.
A. Immunofluorescence of MED1, HP1α, CDK7, ER, and BRD4 together with MYC *RNA* FISH. Consistent with the finding that MED1, a marker of transcriptional condensates, is present in puncta at the MYC oncogene, CDK7, ER, and BRD4 are also found in puncta at MYC. These results mirror those obtained by ChIP-Seq at this locus. In contrast, signal for HP1α, a marker of heterochromatin condensates, is not found at *MYC*. Average and random image analysis shown to the right.

B. (Top) Schematic of in vitro droplet assay showing mixing of nuclear condensate protein (MED1 or HP1α) with various drug target proteins (CDK7, ER, or BRD4), with partitioning into the nuclear condensate measured by confocal microscopy. (Middle) In vitro droplet assays with MED1, ER, HP1α and BRD4 at 10μM, CDK7 at 200nM. Droplets are formed in 125mM NaCl, 10% PEG and droplet formation buffer. All drug targets tested were concentrated in MED1 condensates. ER was found to be concentrated both in MED1 and HP1α condensates, consistent with previous reports and its ability to associate with both co-activators and co-repressors (*12*, *66*). (Bottom) Quantification of target protein enrichment in the indicated condensates, error bars represent SEM.

Figure S10

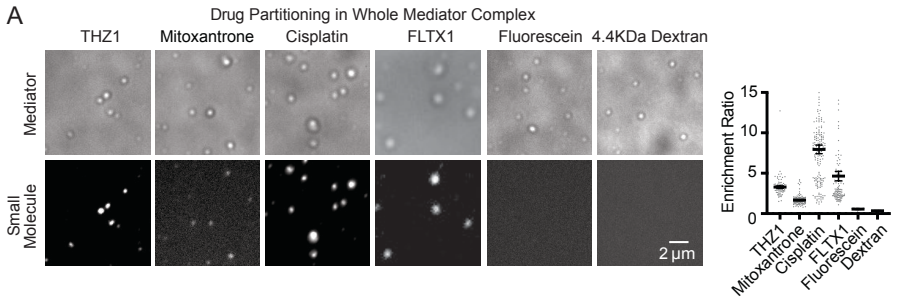


Figure S10. Partitioning behavior of various small molecule drugs in whole Mediator complex. Confocal microscopy images of drugs (THZ1, mitoxantrone, cisplatin, FLTX1, fluorescein, and 4.4kDa dextran) in whole mediator complex condensates. Mediator was imaged in brightfield while the small molecule was imaged by the channel in which it fluoresces. Experiments were performed in 10% PEG and 125mM NaCl. The partitioning behavior of various small molecule drugs into whole Mediator complex recapitulate the partitioning behavior of drugs into MED1 condensates. Quantification of enrichment shown to the right, error bars represent SEM.

Figure S11

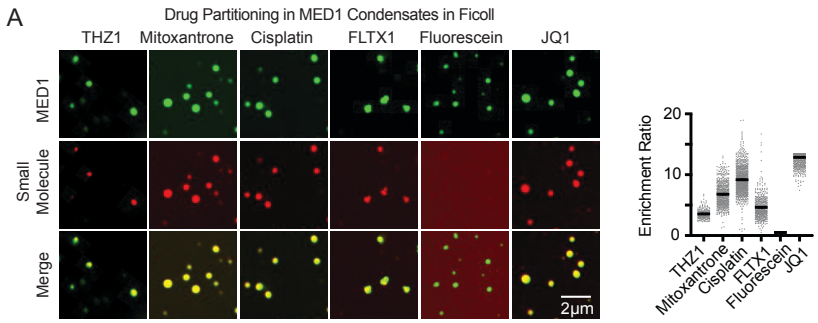


Figure S11. Partitioning behavior of various small molecule drugs into MED1 condensates formed in ficoll. Confocal microscopy images of small molecule drugs (THZ1, mitoxantrone, cisplatin, FLTX1, fluorescein, and JQ1) concentration behavior in MED1 condensates in the presence of 125mM NaCl and 20% ficoll. The partitioning behavior of small molecules are similar regardless of crowder used to form MED1 droplets. Quantification of enrichment shown to the right, error bars represent SEM.

Figure S12

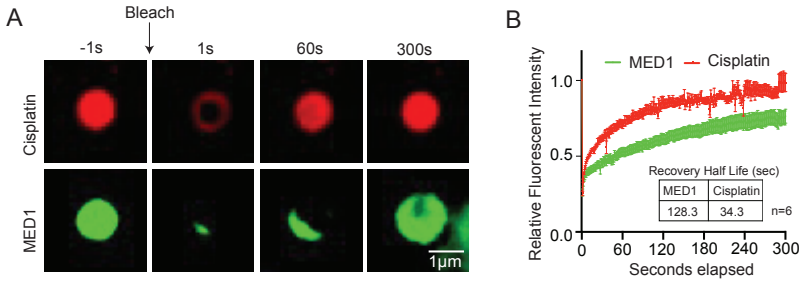


Figure S12. Cisplatin molecules are highly mobile in MED1 droplets.
A. Confocal microscopy images showing fluorescence recovery after photobleaching (FRAP) of TR-cisplatin and MED1 in condensates formed in the presence of 125mM NaCl and 10% PEG with 5μM TR-cisplatin and 10μM protein.
B. Quantification of FRAP (error bars represent SEM).

Figure S13

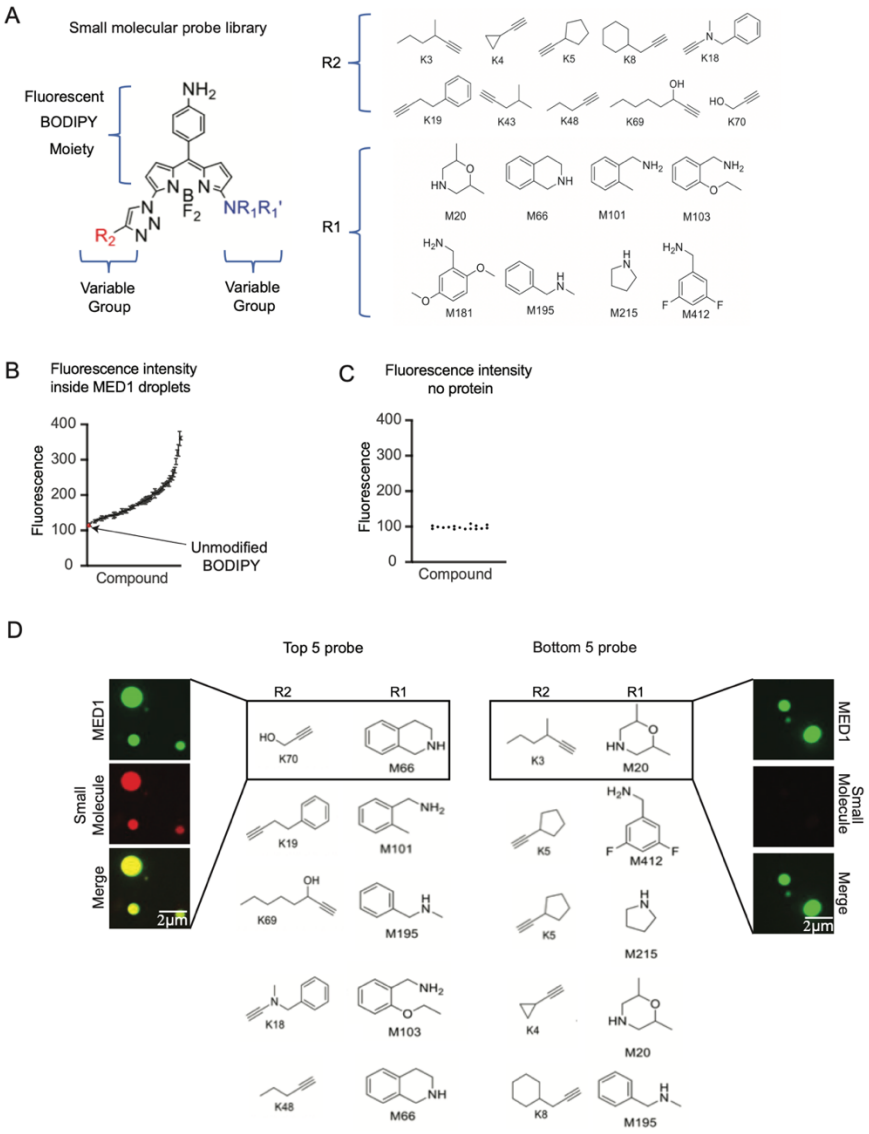


Figure S13. Specific chemical moieties govern concentration in MED1 condensates.
A. Depiction of small molecule boron-dipyrromethene (BODIPY) library.
B. Fluorescence intensity of probe library in MED1 droplets measured by confocal microscopy. Experiments were performed in 125mM NaCl and 10% PEG, with 10μM MED1 and 1μM small molecule. The fluorescence of the BODIPY molecule alone is highlighted in red.
C. Fluorescent intensity of a random selection of 18 probes from the library without MED1 protein demonstrating they have similar fluorescent intensity.
D. Top 5 (left) and bottom 5 (right), R2 and R1 sidechains, ranked by fluorescent intensity. This screen of 81 compounds suggests that pi-system interactions mediate compound accumulation

in condensates, a larger screen will further define the chemical features that mediate this phenomenon.

Figure S14

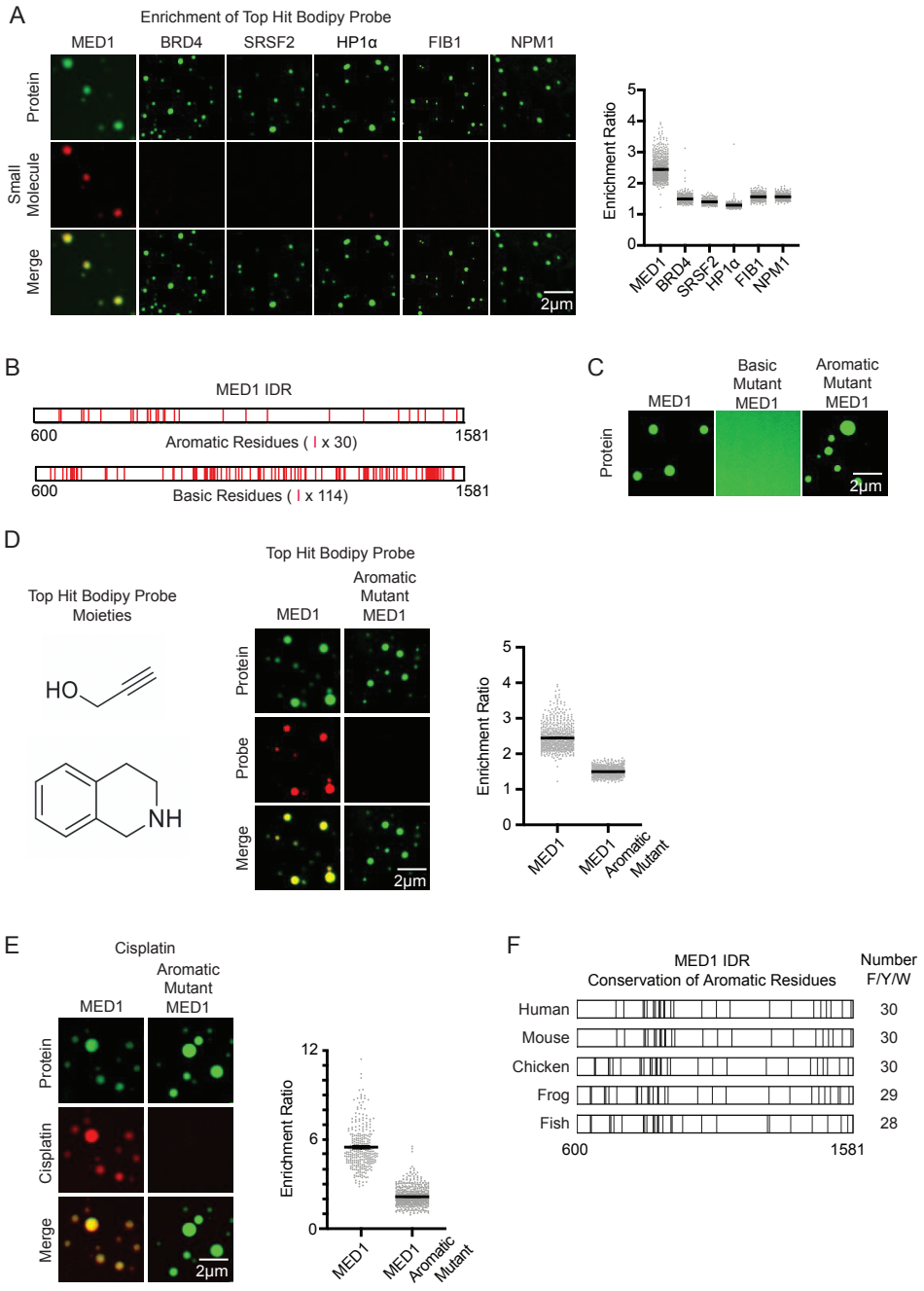


Figure S14. Aromatic residues of MED1 contribute to small molecule partitioning into MED1 condensates but are dispensable for condensate formation.

A. Confocal microscopy images of MED1, BRD4, SRSF2, HP1α, FIB1, and NPM1 nuclear condensates formed in 125mM NaCl and 10% PEG together with 5μM of the small molecule probe that ranked the

highest in fluorescent intensity within MED1 condensates. The probe was specifically concentrated into MED1 condensates, indicating that chemical features of the probe selectively interact with those of MED1 condensates. The top-ranking probes that concentrated in MED1 condensates showed a preference for BODIPY molecules that are modified with an aromatic ring. This suggests that the pi-system might be contributing to the interaction between small molecules and MED1.

B. Schematic of the MED1 IDR mutant proteins. The pi-system governs the interactions of supramolecular assemblies, where pi-pi or pi-cation interactions play prominent roles. To test if these interactions govern small molecule partitioning into MED1 condensates, and encouraged by the observation that the MED1 IDR is enriched for both aromatic and basic amino acids residues relative to other proteins studied here (Figure S3B), we generated an aromatic MED1 IDR mutant (all 30 aromatic residues changed to alanine) and a basic MED1 IDR mutant (all 114 basic residues changed to alanine).

C. We tested the ability of MED1 mutants to form droplets by confocal microscopy using MED1 wildtype, MED1 basic mutant (all basic amino acids replaced with alanine), and MED1 aromatic mutant (all aromatic amino acids replaced with alanine) in the presence of 125mM NaCl and 10% PEG. The MED1 basic mutant showed an impaired ability to form droplets in vitro, indicating that the basic residues of MED1 are required for the homotypic interactions that govern droplet formation. The MED1 aromatic mutant formed droplets similar to those of MED1 wildtype protein. D. Role of MED1 aromatic residues in incorporation of aromatic small molecule probes. Confocal microscopy images and their quantification for the top hit BODIPY probe together with MED1 or MED1 aromatic mutant, which show that the partitioning behavior of the aromatic probe into MED1 aromatic mutant droplets is substantially reduced. Experiments were performed in 10% PEG and 125mM NaCl with 10μM protein and 5μM small molecule.

E. Confocal microscopy images and their quantification for cisplatin together with MED1 or MED1 aromatic mutant, which show that the partitioning behavior of cisplatin into MED1 aromatic mutant droplets is substantially reduced. Experiments were performed in 10% PEG and 125mM NaCl with 10μM protein and 5μM cisplatin-TR. Taken together, these results suggest that the pi-system contributes to small molecule partitioning into MED1 condensates.

F. Conservation of aromatic amino acids in the MED1 IDR across species, with the total number of aromatic residues for each species. All error bars represent SEM.

Figure S15

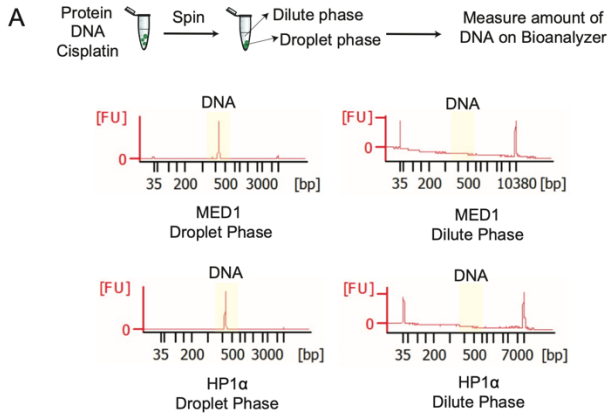


Figure S15. DNA can be compartmentalized and concentrated in nuclear condensates. (Top) Schematic of droplet assay showing protein, DNA, and cisplatin mixed in droplet forming conditions, then spun down to separate the droplet phase from the dilute phase. The amount of DNA in the two phases is subsequently measured using a Bioanalyzer. DNA is enriched in MED1 and HP1α droplet phase (left) compared to MED1 and HP1α dilute phase (right).

Figure S16

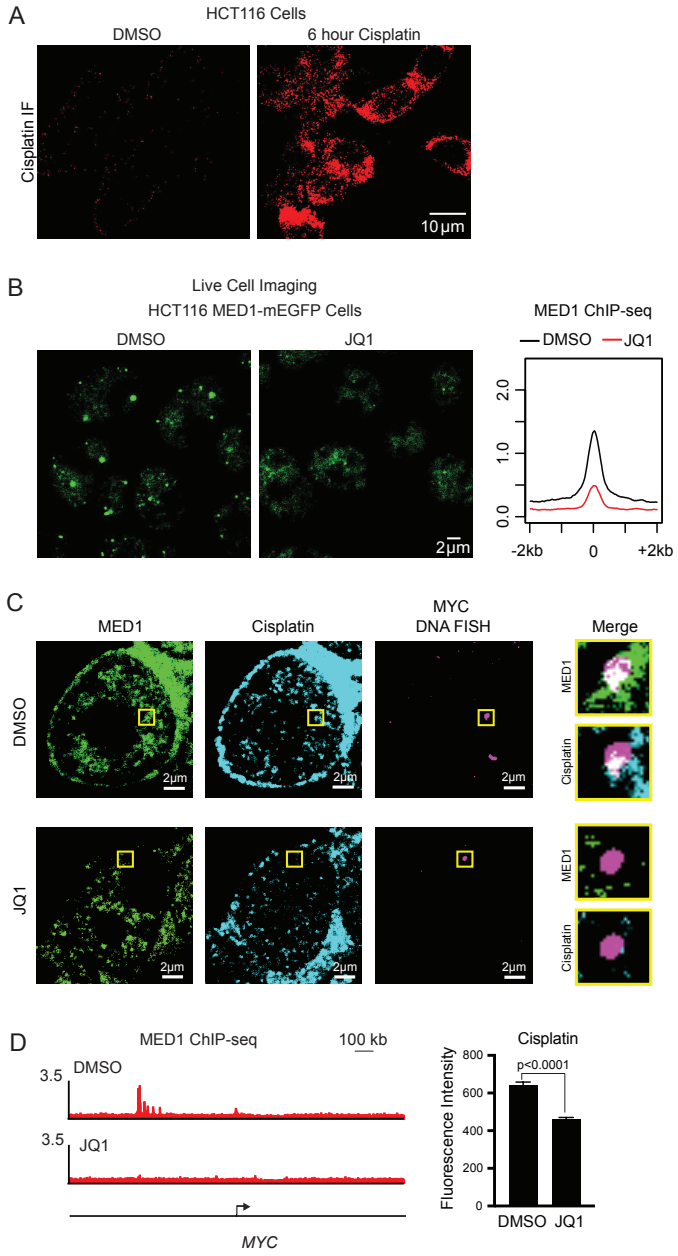


Figure S16. Concentration of small molecules in specific condensates can influence target engagement.

A. HCT116 cells were treated with DMSO or 50µM cisplatin for 6 hours followed by cisplatin immunofluorescence. The antibody only recognizes platinated DNA in cells treated with cisplatin, supporting antibody specificity.

B. (Left) mEGFP-MED1 tagged HCT116 cells treated with JQ1 for 24 hours result in diminution of MED1 condensates. (Right) Metaplot of MED1 ChIP-Seq in DMSO vs JQ1 treated HCT116 cells.

C. Cells were treated with JQ1 and then cisplatin to determine whether diminution of MED1 condensates leads to reduced DNA platination at *MYC* locus. *MYC* DNA FISH and MED1 immunofluorescence showed

a loss of signal for platinated DNA after JQ1 treatment, indicating that the presence of a MED1 condensate contributes to DNA platination at this locus.
D. (Left) MED1 ChIP-Seq track at *MYC* in DMSO or JQ1 treated HCT116 cell showing loss of MED1 loading after JQ1 treatment. (Right) Quantification of cisplatin IF signal at *MYC* DNA FISH foci in HCT116 cells with DMSO or JQ1 treatment, error bars represent SEM.

Figure S17

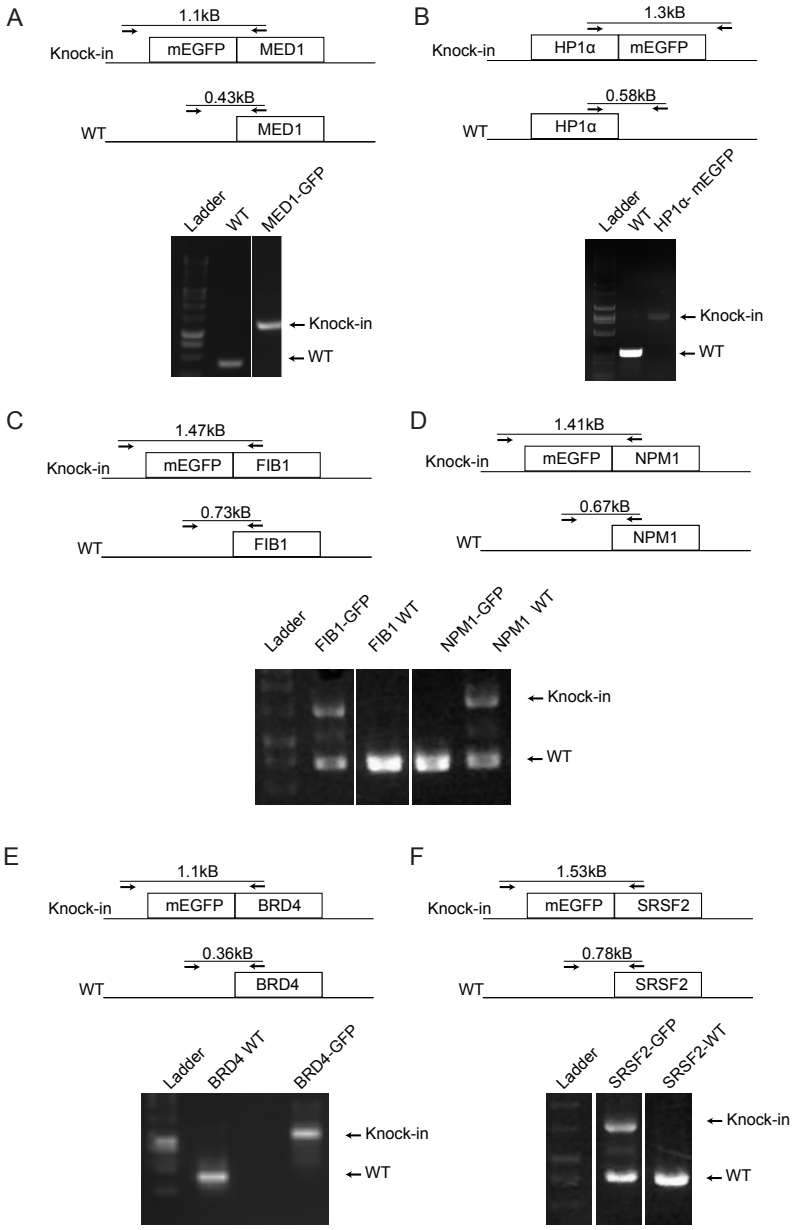


Figure S17. Genotyping of endogenously tagged cell lines. Schematic image and genotyping agarose gel showing mEGFP tagged

A. MED1
B. HP1α
C. FIB1
D. NPM1
E. BRD4
F. SRSF2 in HCT116 colon cancer cells.

Figure S18

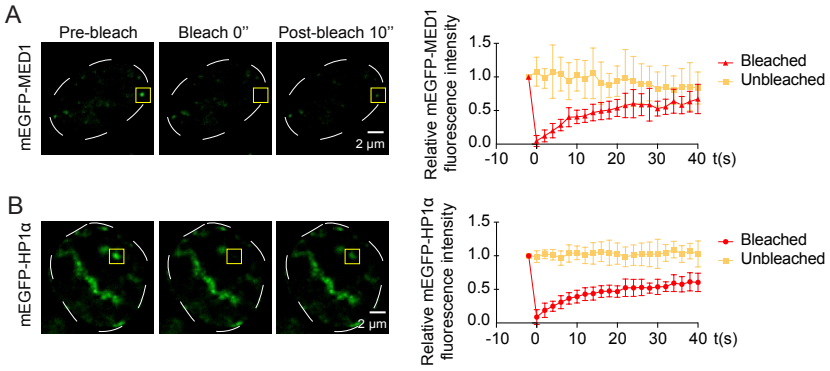


Figure S18. Nuclear condensates in cells are highly dynamic. FRAP of mEGFP-tagged
A. MED1
B. HP1α in HCT116 cell lines (error bars represent SEM) (n=7).

Figure S19

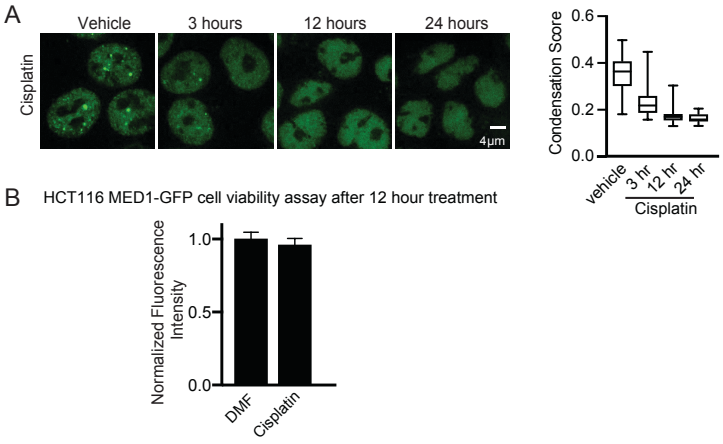


Figure S19. Dissolution of MED1 condensates in cells upon prolonged cisplatin treatment.
A. HCT116 cells endogenously GFP-tagged MED1 treated with DMF or 50μM cisplatin for 3, 6, or 12 hours. Quantification shown to the right, error bars are SD.
B. Cell viability assay of HCT116 cells expressing GFP-MED1 treated for 12 hours with DMF or 50μM Cisplatin.

Figure S20

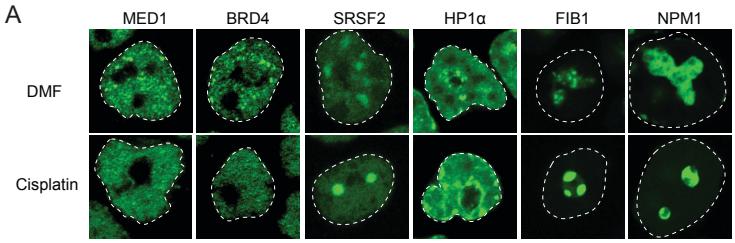


Figure S20. Effect of cisplatin on various nuclear condensates.
A. HCT116 cells bearing either endogenously GFP-tagged MED1, BRD4, HP1α, FIB1, NPM1, or SRSF2 treated with 50μM cisplatin for 12 hours. Cisplatin specifically disrupts MED1 and BRD4 condensates, consistent with cisplatin and BRD4 being selectively concentrated in MED1 condensates.

Figure S21

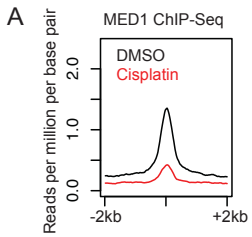


Figure S21. Decreased MED1 genomic occupancy upon cisplatin treatment. Graph shows MED1 ChIP-seq after 6 hours of DMSO or 50μM cisplatin treatment, MED1 genomic levels are reduced after cisplatin treatment.

Figure S22

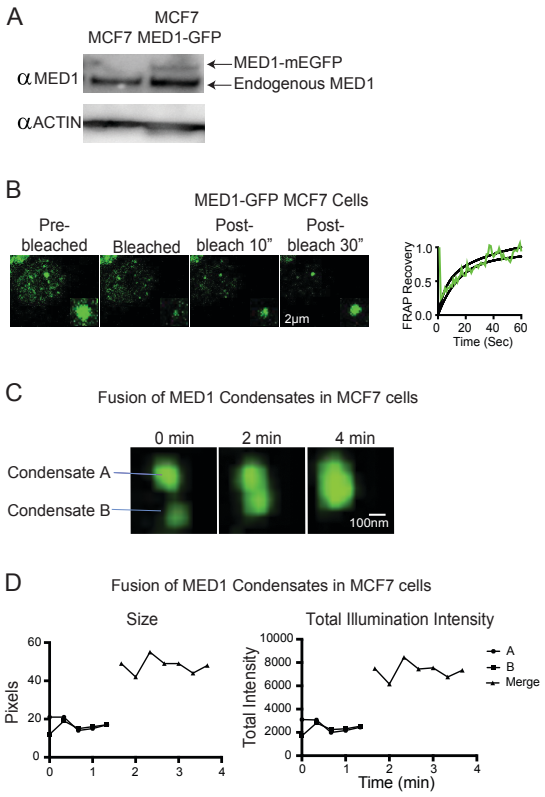


Figure S22. Characterization of MED1 condensates in MCF7 cells.
A. Western blot of MED1 in MCF7 cells and MCF cells infected with MED1-mEGFP lentiviral vector.
B. FRAP of MED1-mEGFP in MCF7 cells expressing this fusion protein by virtue of a lentiviral vector. Quantification shown to the right, black bars represent 95% confidence interval of the best fit line.
C. MCF7 cells expressing MED1-mEGFP were grown in estrogen-free conditions then stimulated with 100nM estrogen for 15 minutes and imaged for 4 minutes on a confocal fluorescent microscope.
D. Quantification of size and intensity of fusing MED1 condensates shown in (C).

Figure S23

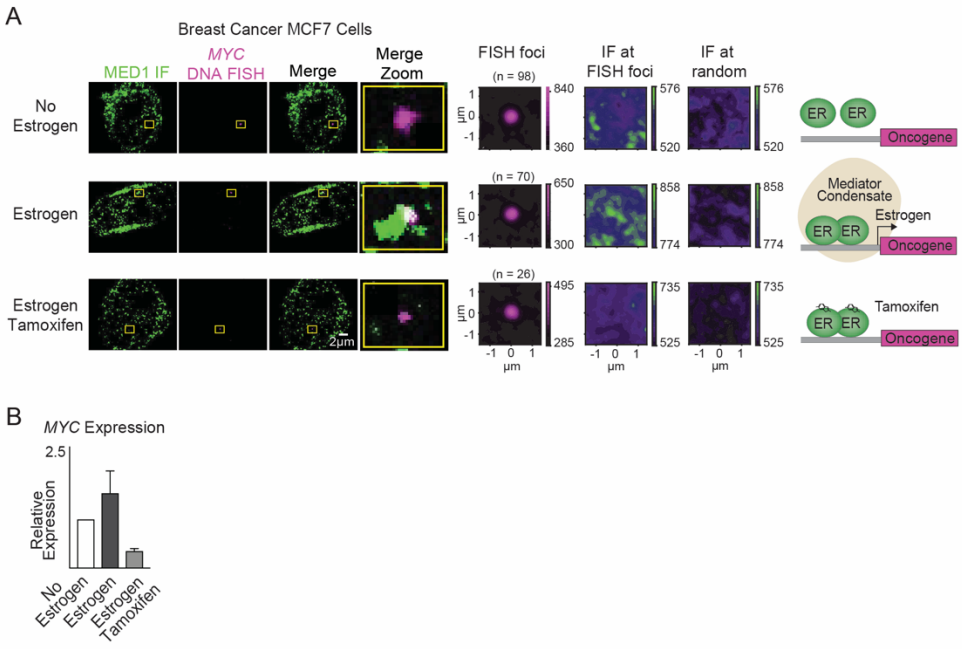


Figure S23. Estrogen and tamoxifen dependent MED1 condensate formation at the *MYC* oncogene.

A. DNA FISH and immunofluorescence in estrogen-starved MCF7 cells treated with 100nM estrogen or 100nM estrogen and 5μM tamoxifen for 24 hours. Average image analysis and random image analysis shown to the right.

B. RT-qPCR showing relative *MYC* RNA expression in estrogen-starved, estrogen stimulated, or estrogen and tamoxifen treated MCF7 cells, error bars represent SEM.

Figure S24

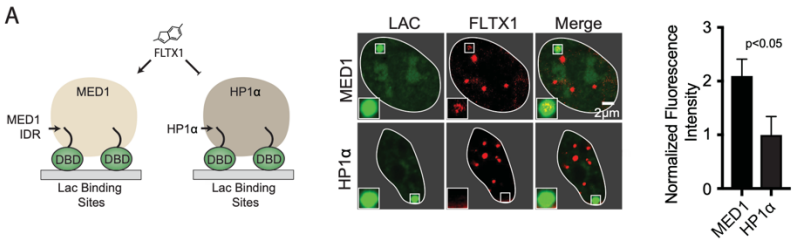


Figure S24. FLTX1 concentrates in MED1 condensates in cells. (Left) Schematic of MED1 or HP1α tethered to the LAC array in U2OS cells generating a MED1 or HP1α condensate. (Middle) Representative images of isolated U2OS cell nuclei with either MED1 or HP1α tethered to the LAC array exposed to FLTX1. Zoomed image of the Lac array shown inset, merged images shown on the right. (Right) Quantification of FLTX1 enrichment at the LAC array with either MED1 or HP1α tethered, error bars represent SEM. ESR1 is not expressed in this osteosarcoma cell line (*67*).

Figure S25

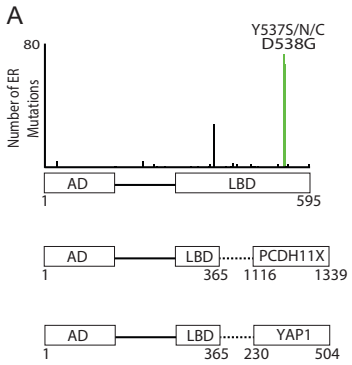


Figure S25. Patient derived hormonal therapy resistant mutations of ESR1. Plot of ER mutation frequency derived from a 220 patient set from the cBioPortal database showing locations of ER point mutations with hotspots at 537 and 538 (*68*).

Figure S26

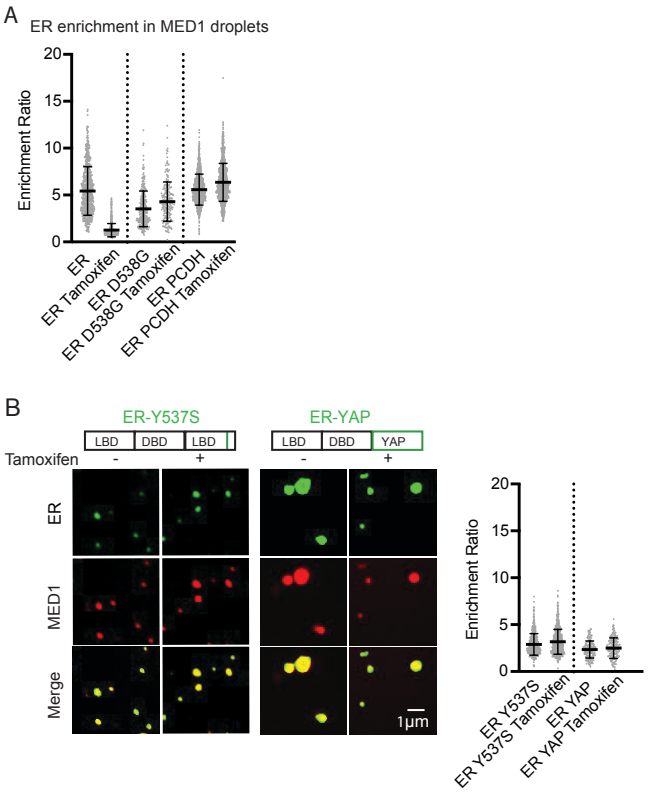
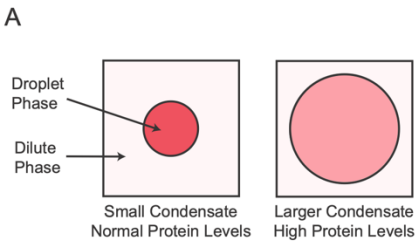
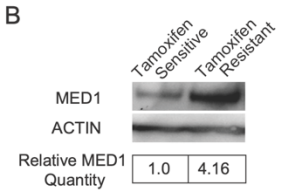


Figure S26. Enrichment ratios of ER and ER mutants in MED1 droplets.
A. Quantification of ER or ER mutant enrichment ratios in MED1 droplets in the presence of either estrogen or estrogen and tamoxifen.
B. (Left) Representative images of ER mutants partitioning in MED1 droplets, enrichment ratios shown to the right. Experiments for both (A) and (B) are performed in 125mM NaCl, 10% PEG, 10μM of each protein, 100μM estrogen with or without 100μM of the indicated ligand. All error bars represent SD.

Figure S27





C

Quantification of MED1 condensates in breast cancer cells		
	Tamoxifen Sensitive	Tamoxifen Resistant
Sample Size	4994	4923
Mean Volume (uM^3)	0.012	0.021
Median Volume (uM^3)	0.003	0.004
Mean Condensates Per Nucleus	463	482

Figure S27. MED1 overexpression in tamoxifen resistant breast cancer cells.
A. Schematic demonstrating drug concentration in a condensate upon increase in condensate volume by scaffold protein overexpression. Assuming limited drug in a system (see Figure 4E), the concentration of drug in a MED1 droplet is expected to decrease upon condensate volume expansion
B. Western blot of MED1 and Actin in MCF7 cells (tamoxifen sensitive) and TAMR7 cells (tamoxifen resistant derivative of MCF7) showing that MED1 levels are higher TAMR7 cells. Quantification from the western blot is shown below, which is an average of 3 experiments.
C. Quantification of MED1 condensates in tamoxifen sensitive and resistant cell lines showing the volume of the MED1 condensates and the number of condensates per nucleus.

Figure S28

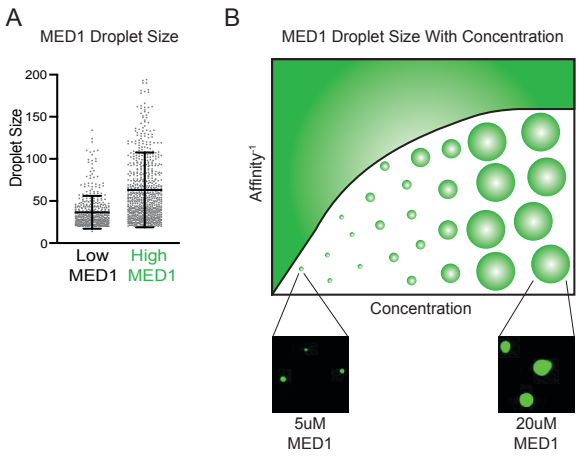


Figure S28. MED1 condensates increase in size with increasing MED1 concentration.
A. Droplet size in pixels from in vitro droplet assays performed with either 5μM (Low) or 20μM (High) MED1-GFP in 125mM NaCl and 10% PEG. Quantification shown to the right, error bars represent SD.
B. Schematic phase diagram of MED1, demonstrating that when the total concentration of MED1 increases, the size of droplet increases while maintaining the concentration of protein within the droplet phase.

Figure S29

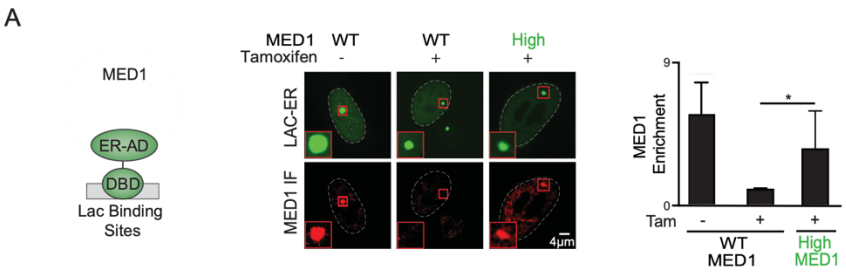
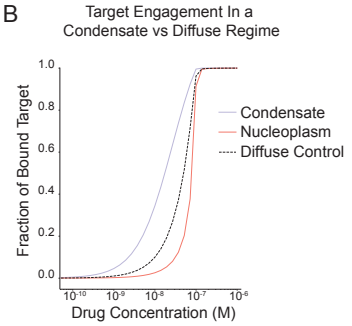


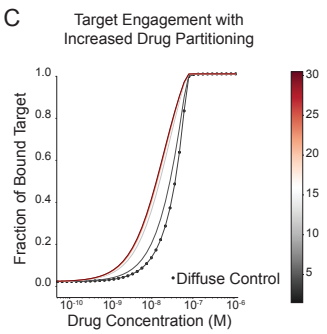
Figure S29. MED1 Condensation at the Lac Array.
A. (Left) Schematic of the Lac array assay. U2OS cells bearing 50,000 copies of the Lac binding site are transfected with a construct expressing the Lac DNA binding domain (DBD) to the estrogen receptor ligand binding domain (LBD). When the transcriptional apparatus is recruited to that site a mediator condensate is detectable by immunofluorescence (*12*) (Middle) U2OS-Lac cells were transfected with a construct expressing the Lac DBD fused to the ER LBD and GFP +/- a construct overexpressing MED1. Cells were grown in estrogen deprived media, and treated with 10nM estrogen +/- 10nM tamoxifen then fixed and subjected to MED1 IF. Top panel shows the location of ER-LBD at the Lac array, bottom panel shows MED1 IF. Inset image shows zoom. (Right) Quantification of MED1 enrichment relative at the Lac array, error bars represent SD.

Figure S30

A

Model Parameters	Value
Target concentration	100nM
Kd drug	0.15 nM
Condensate volume fraction	0.02
Drug partitioning	20
Target partitioning	10





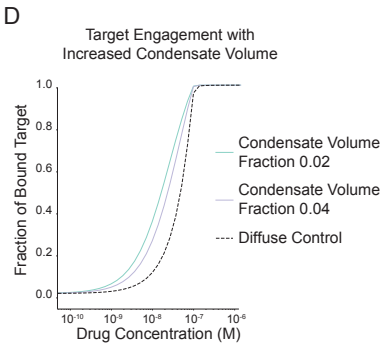


Figure S30. In silico model of small molecule partitioning in condensates. To demonstrate the behavior of a small molecule drug engaging a target contained within a condensate, we developed a simple model in which a drug and target are both contained within a condensate with percent target engagement as the readout. In this model, target partitioning is not affected by drug binding

A. Table of the values used to build a model of drug engagement within condensates, derived from known values of ER and tamoxifen. Condensate volume fraction value derived from analysis of MED1 IF on human ER+ breast carcinoma biopsies (Figure S1A).

B. Target binding as a function of drug concentration in simulations. The dashed line represents a system in which target and drug are freely diffusing through the cells. Red and blue lines represent a system in which target and drug are concentrated into a condensate. The blue line represents target engagement in the condensate where the drug and target are concentrated, the red line represents target engagement in the dilute phase of the nucleoplasm. Overall, these data show that drug engages a higher percent of target molecules inside a condensate that outside, at a given concentration.

C. Fraction of bound target at a given concentration of drug at various partitioning coefficients of drug. Dotted line represents the target engagement in a diffuse regime. Overall, this simulation shows that as the partitioning coefficient of drug in a condensate increases the percent of target bound at a given concentration. (D) Target engagement by drug in the setting of larger condensates. Simulation of target binding as a function of drug concentration in the setting of normal condensate volume (2% of the volume of the nucleus) versus larger condensate volume (4% of the volume of the nucleus). Diffuse control shown by the dashed line. Overall, these data show that a drug may be less effective in binding its target in larger condensates.

Materials and Methods

Cell lines

Cell lines were obtained as indicated, TamR7 (ECACC 16022509). V6.5 murine embryonic stem cells were a gift from R. Jaenisch of the Whitehead Institute. V6.5 are male cells derived from a C57BL/6(F) x 129/sv(M) cross. MCF7 cells were a gift from the R. Weinberg of the Whitehead Institute and HCT116 cells were from ATCC (CCL-247) were used. V6.5 murine embryonic stem endogenously tagged with MED1-mEGFP (*10*), BRD4-mEGFP (*10*), SRSF2-mEGFP (*11*), or HP1α-mEGFP were used. Cells were tested negative for mycoplasma. The CRISPR/Cas9 system was used to generate genetically modified endogenously tagged ESCs and HCT116 cells. Target-specific sequences were cloned into a plasmid containing sgRNA backbone, a codon-optimized version of Cas9, and BFP or mCherry. A homology directed repair template was cloned into pUC19 using NEBuilder HiFi DNA Master Mix (NEB E2621S). The homology repair template consisted of mCherry or mEGFP cDNA sequence flanked on either side by 800 bp homology arms amplified from genomic DNA using PCR. To generate genetically modified cell lines, 750,000 cells were transfected with 833 ng Cas9 plasmid and 1,666 ng non-linearized homology repair template using Lipofectamine 3000 (Invitrogen L3000). Cells were sorted 48 hours after transfection for the presence of BFP or mCherry fluorescence proteins encoded on the Cas9 plasmid to enrich for transfected cells. This population was allowed to expand for 1 week before sorting a second time for the presence of mCherry or mEGFP. For mES cells, 40,000 mCherry positive cells were plated in serial dilution in a 6-well plate and allowed to expand for a week before individual colonies were manually picked into a 96-well plate. 24 colonies were screened for successful targeting using PCR genotyping to confirm insertion. For HCT116, single cells were plated in a 96 well plate and allowed to grow until confluence, then screened for successful targeting using PCR genotyping to confirm insertion.

PCR genotyping was performed using Phusion polymerase (Thermo Scientific F531S). Products were amplified according to kit recommendations and visualized on a 1% agarose gel. The following primers were used for PCR genotyping:

HP1α-mCherry_fwd (mES): AACGTGAAGTGTCCACAGATTG
HP1α-mCherry_rev (mES): TTATGGATGCGTTTAGGATGG
HP1α-GFP_fwd (HCT116): CCAAGGTGAGGAGGAAATCA
HP1α-GFP_rev (HCT116): CACAGGGAAGCAGAAGGAAG
MED1α-GFP_fwd (HCT116): GAAGTTGAGAGTCCCCATCG
MED1-GFP_rev (HCT116): CGAGCACCCTTCTCTTCTTG
BRD4-GFP_fwd (HCT116): CTGCCTCTTGGGCTTGTTAG
BRD4-GFP_rev (HCT116): TTTGGGGAGAGGAGACATTG
SRSF2-GFP_fwd (HCT116): CAAGTCTCCTGAAGAGGAAGGA
SRSF2-GFP_rev (HCT116): AAGGGCTGTATCCAAACAAAAAC
FIB1-GFP_fwd (HCT116): CCTTTTAATCAGCAACCCACTC
FIB1-GFP_rev (HCT116): GTGACCGAGTGAGAATTTACCC
NPM1-GFP_fwd (HCT116): TCAAATTCCTGAGCTGAAGTGA
NPM1-GFP_rev (HCT116): AACACGGTAGGGAAAGTTCTCA

Cell culture

V6.5 murine embryonic stem (mES) cells were grown in 2i + LIF conditions. mES cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates. The media used for 2i + LIF media conditions is as follows: 967.5 mL DMEM/F12 (GIBCO 11320), 5 mL N2 supplement (GIBCO 17502048), 10 mL B27 supplement (GIBCO 17504044), 0.5mML-glutamine (GIBCO 25030), 0.5X

non-essential amino acids (GIBCO 11140), 100 U/mL Penicillin-Streptomycin (GIBCO 15140), 0.1 mM b-mercaptoethanol (Sigma), 1 uM PD0325901 (Stemgent 04- 0006), 3 uM CHIR99021 (Stemgent 04-0004), and 1000 U/mL recombinant LIF (ESGRO ESG1107). TrypLE Express Enzyme (Life Technologies, 12604021) was used to detach cells from plates. TrypLE was quenched with FBS/LIF-media ((DMEM K/O (GIBCO, 10829-018), 1X nonessential amino acids, 1% Penicillin Streptomycin, 2mM L-Glutamine, 0.1mM b-mercaptoethanol and 15% Fetal Bovine Serum, FBS, (Sigma Aldrich, F4135)). Cells were spun at 1000rpm for 3 minutes at RT, resuspended in 2i media and $5x10^6$ cells were plated in a 15 cm dish.

MCF7 cells and HCT116 cells were grown in complete DMEM media (DMEM (Life Technologies 11995073), 10% Fetal Bovine Serum, FBS, (Sigma Aldrich, F4135), 1% L-glutamine (GIBCO, 25030-081), 1% Penicillin Streptomycin (Life Technologies, 15140163)). For growth in estrogen-free conditions MCF7 cells in regular media were washed 3x with PBS then the media was changed to estrogen free media containing phenol red-free DMEM (Life Technologies 21063029), 10% charcoal stripped FBS (Life Technologies A3382101), 1% L-glutamine (GIBCO, 25030-081) and 1% Penicillin Streptomycin (Life Technologies, 15140163) for 48 hours prior to use. TamR7 cells were grown in TAMR7 media (Phenol red-free DMEM/F12 (Life Technologies 21041025, 1% L-glutamine (GIBCO, 25030-081)1% Penicillin Streptomycin (Life Technologies, 15140163), 1% Fetal Bovine Serum, FBS, (Sigma Aldrich, F4135), 6ng/mL insulin (Santa Cruz Biotechnology, sc-360248)). For passaging, cells were washed in PBS (Life Technologies, AM9625). TrypLE Express Enzyme (Life Technologies, 12604021) was used to detach cells from plates. TrypLE was quenched with indicated media.

Live cell imaging

Cells were grown on glass dishes (Mattek P35G-1.5-20-C). Before imaging the cells, culture medium was replaced with phenol red-free media, and imaged using the Andor Revolution Spinning Disk Confocal microscope. Raw Andor images were processed using FIJI. For imaging mESC, coated glass dishes were used (5 µg/ml of poly-L-ornithine (Sigma-Aldrich, P4957) for 30 minutes at 37 °C, and with 5 µg/ml of laminin (Corning, 354232) for 2–16 hours at 37 °C). For imaging FIB1 and NPM1 in mES cells, vectors encoding GFP-tagged NPM1 or FIB1 were transfected as described above with Lipofectamine 3000 per package instructions.

Immunofluorescence of tissue samples

Fresh frozen breast and colon tissues were purchased from BioIVT. Frozen breast tissue was fixed in 2% PFA in PBS for 30minutes-1hour. Fixed tissue was incubated in 30% sucrose in PBS at 4°C for 4 days. Tissue was embedded in OCT and frozen. Fresh frozen colon tissue was embedded in OCT and frozen. Tissue was sectioned into 10µm sections using the cryostat with temperature set at -25°C or -30°C. Sections were stored at -20°C. For IF, sections were brought to room temperature, they were fixed in 4% PFA in PBS for 10 minutes. Following three washes in PBS, tissues were permeabilized using 0.5%TX100 in PBS, washed three times in PBS and blocked with 4% BSA in PBS for 30 minutes. Primary antibodies were diluted into 4% BSA in PBS and added to the tissue sample for O/N incubation at RT. Following three washes in PBS, samples were incubated with secondary antibodies diluted 1:500 in 4% BSA in PBS. Samples was washed in PBS, DNA was stained using 20µm/mL Hoechst 33258 (Life Technologies, H3569) for 5 minutes and mounted using Vectashield (VWR, 101098-042). Images were acquired using the Elyra Super-Resolution Microscope at Harvard Center for Biological Imaging. Images were post-processed using Fiji Is Just ImageJ (https://fiji.sc/).

Nuclear volume quantification of condensates

For image acquisition: 10 z-slices were imaged. The outline of the nuclei were defined manually in Fiji Is Just ImageJ (https://fiji.sc/) and the volume of each nucleus was calculated as nuclear

area (μm) * number of z-slices imaged (10) * voxel depth (0.1μm). The volume of condensates in the nucleus was measured using a custom Python script and the scikit-image package. Condensates were segmented from 3D images of the protein channel on two criteria: (1) an intensity threshold that was three s.d. above the mean of the image; (2) size thresholds (10 pixel minimum condensate size). The estimated volume of the segmented objects was then calculated by multiplying the width (μm) * height (μm) * voxel depth (0.1μm). For each protein factor, the average and s.d. volume of condensates in the healthy and malignant tissue was reported. The number of condensates per nucleus was defined as the number of segmented objects contained within the perimeter of the defined nucleus. For each protein factor, the average and s.d. number of condensates per nucleus in the healthy and malignant tissue was reported. Percentage of nuclear volume occupied by the condensates was calculated as follows: (Σ volume of all detected condensates in the nucleus)/(estimated nuclear volume).

Antibodies

The following antibodies were used for Immunofluorescence: NPM1 (Abcam ab10530), BRD4 (ab128874), MED1 (ab64965), HP1α (ab109028), FIB1 (ab5821), SRSF2 (ab11826), ER (ab32063), CDK7 (Santa Cruz sc-7344), Cisplatin modified DNA (ab103261), 568 goat anti rat (Life Technologies A11077), Goat anti-Rabbit IgG Alexa Fluor 488 (Life Technologies A11008).

Protein purification

Human cDNA was cloned into a modified version of a T7 pET expression vector. The base vector was engineered to include a 5' 6xHIS followed by either BFP, mEGFP or mCherry and a 14 amino acid linker sequence "GAPGSAGSAAGGSG." NEBuilder® HiFi DNA Assembly Master Mix (NEB E2621S) was used to insert these sequences (generated by PCR) in-frame with the linker amino acids. All expression constructs were sequenced to ensure sequence identity.

For protein expression plasmids were transformed into LOBSTR cells (gift of Chessman Lab) and grown as follows. A fresh bacterial colony containing the tagged MED1 constructs were inoculated into LB media containing kanamycin and chloramphenicol and grown overnight at 37°C. Cells were diluted 1:30 in 500ml room temperature LB with freshly added kanamycin and chloramphenicol and grown 1.5 hours at 16°C. IPTG was added to 1mM and growth continued for 20 hours. Cells were collected and stored frozen. Cells containing all other expression plasmids were treated in a similar manner except they were grown for 5 hours at 37°C after IPTG induction.

Cell pellets of SRSF2 were resuspended in 15ml of denaturing buffer (50mM Tris 7.5, 300mM NaCl, 10mM imidazole, 8M Urea) with cOmplete protease inhibitors (Roche,11873580001) and sonicated (ten cycles of 15 seconds on, 60 sec off). The lysates were cleared by centrifugation at 12,000g for 30 minutes and added to 1ml of Ni-NTA agarose (Invitrogen, R901-15) that had been pre-equilibrated with 10 volumes of the same buffer. Tubes containing this agarose lysate slurry were rotated for 1.5 hours at room temperature, then centrifuged for 10 minutes at 3,000 rpm, washed with 2 X 5ml of lysis buffer and eluted with 3 X 2ml lysis buffer with 250mM imidazole. Elutions were incubated for at least 10 minutes rotating at room temperature and centrifuged for 10 minutes at 3,000 rpm to collect protein. Fractions were run on a 12% acrylamide gel and proteins of the correct size were dialyzed first against buffer containing 50mM Tris pH 7.5, 500mM NaCl, 1Mm DTT and 4M Urea, followed by the same buffer containing 2M Urea and lastly 2 changes of buffer with 10% Glycerol, no Urea. Any precipitate after dialysis was removed by centrifugation at 3,000rpm for 10 minutes. All other proteins were purified in a similar manner by resuspending cell pellets in 15ml of buffer containing 50mM Tris pH7.5, 500 mM NaCl, cOmplete protease inhibitors, sonicating, and centrifuging at 12,000xg for 30 minutes at 4°C. The lysate was added to 1ml of pre-equilibrated Ni-NTA agarose, and rotated at 4°C for 1.5 hours.

The resin slurry was centrifuged at 3,000 rpm for 10 minutes, washed with 2 X 5ml lysis buffer with 50mM imidazole and eluted by incubation for 10 or more minutes rotating 3 X with 2ml lysis buffer containing 250mM imidazole followed by centrifugation and gel analysis. Fractions containing protein of the correct size were dialyzed against two changes of buffer containing 50mM Tris 7.5, 125mM NaCl, 10% glycerol and 1mM DTT at 4°C or the same buffer with 500mM NaCl for the HP1α construct.

The following human proteins or protein fragments were used for production:
NPM1 - full length, amino acids 1-294.
SRSF2 - full length, amino acids 1-221.
HP1α - full length, amino acids 1-191.
MED1 - amino acids 600-1581.
MED1 - aromatic mutant amino acids 600-1581, all aromatic residues changed to alanine.
MED1 - basic mutant amino acids 600-1581, all basic residues changed to alanine.
BRD4 - amino acids 674-1351.
FIB1- full length, amino acids 1-321.
ER and ER mutants - full length, amino acids 1-595 (WT).

Cbioportal data acquisition
For frequency of patient mutations, cbioportal (http://www.cbioportal.org/) was queried for mutations in ESR1 that are present in any breast cancer sequencing data set.

Drugs and small molecules
Drugs and small molecules were obtained and processed as follows. Hoescht 33258 (Life Technologies H3569) was obtained and utilized in liquid form, Fluorescein (Sigma F2456) was dissolved in DMSO at 10mM then diluted further in droplet formation buffer for use. Dextrans measuring 4.4kDa (Sigma T1037), 10kDa (Invitrogen D1816), 40kDa (Invitrogen D1842), or 70kDa (Invitrogen D1864) conjugated to either TRITC or FITC, ROX (Life technologes 12223012), and Texas Red (Sigma Aldrich 60311-02-6), were diluted in droplet formation buffer. FLTX1 (AOBIO 4054) was dissolved in DMSO then diluted further in droplet formation buffer. THZ1-TMR and JQ1-ROX was synthesized as below to achieve the molecular structure displayed in Figure 2D-E. Cisplatin conjugated to texas red (Ursa Bioscience) was dissolved in DMF to 2mM and diluted for further use in droplet formation buffer. Mitoxantrone (Sigma F6545) was dissolved in DMSO and diluted for further use in droplet formation buffer. Chemical structures were made using ChemDraw software.

Unlabeled molecules were used for live cell and chase out experiments as below: JQ1 (Cayman Chemical 11187), cisplatin (Selleck S1166), transplatin (Toku-E T108), tamoxifen (Sigma Aldrich T5648), 4-hydroxytamoxifen (Sigma H7904).

In vitro droplet assay
Recombinant BFP, GFP, or mCherry fusion proteins were concentrated and desalted to an appropriate protein concentration and 125mM NaCl using Amicon Ultra centrifugal filters (30K MWCO, Millipore). Recombinant protein was added to droplet formation buffer (50mM Tris-HCl pH 7.5, 10% glycerol, 1mM DTT) with the indicated amount of salt and the indicated crowding agent (Ficoll or PEG). The protein solution was immediately loaded onto glass bottom 384 well plate (Cellvis P384-1.5H-N) and imaged with an Andor confocal microscope with a 150x objective. Unless indicated, images presented are of droplets settled on the glass coverslip. For each experiment at least 10 images were taken. Images were post-processed using Fiji Is Just ImageJ (https://fiji.sc/).

Drug and small molecule concentrations used in the droplet experiments are as follows:
Texas red-cisplatin - 5μM
FLTX1 - 100μM
Mitoxantrone - 50μM
Fluorescein - 1μM
Hoechst - 1mg/mL
Labeled dextrans - 0.05mg/mL
THZ1-TMR - 5μM
JQ1-ROX - 1μM
ROX - 1μM
TR - 5μM

For chase-out experiments 5μM labeled cisplatin-TR was added to a MED1 droplet reaction (10 μM MED1, 50mM Tris-HCl pH 7.5, 10% glycerol, 1mM DTT, 10% PEG) in order to form MED1 droplets concentrated with Cisplatin-TR. Unlabeled transplatin or unlabeled cisplatin (vehicle, 10μM, 100μM, or 500μM) were added to the droplet mixture and the amount of labeled cisplatin-TR remaining in the droplet is measured after chase out. 100μM fluorescent FLTX1 was added to a MED1 droplet reaction (10 μM MED1, 50mM Tris-HCl pH 7.5, 10% glycerol, 1mM DTT, 10% PEG) in order to form MED1 droplets concentrated with FLTX1. 1mM of the non-fluorescent version of the drug, tamoxifen, was added to the droplet mixture and the amount of fluorescent FLTX1 remaining in the droplet is measured after chase out. For assaying eviction of ER from MED1 condensates, fluorescently labeled ER and MED1 were mixed in droplet formation buffer at the indicated concentrations with the indicated components in the presence of 100μM estrogen (Sigma E8875). For conditions with tamoxifen treatment, 4-hydroxytamoxifen (Sigma H7904) was then added to a final concentration of 100μM and imaged as above on a confocal fluorescent microscope. Images were post-processed using Fiji Is Just ImageJ (FIJI).

For droplet assay with fluorescent DNA a 451 basepair DNA fragment was commercially synthesized in a vector with flanking M13F and M13R primer binding sites. Primers M13F and M13R were commercially synthesized covalently bound to a Cy5 fluorophore and this fragment was amplified using these primers. The DNA fragment was then purified from PCR reactions and diluted in droplet formation buffer for use in the droplet assay as described. For testing the ability of recombinant CDK7 to partition in MED1 or HP1α droplets, recombinant CDK activating complex (Millipore 14-476) was supplied at 0.4mg/mL in 150mM NaCl at pH 7.5. One vial of Cy5 monoreactive dye (Amersham PA23001) was resuspended in 30uL of 0.2M Sodium Bicarbonate at pH 9.3 in 150mM NaCl. 5uL of this reaction was added to 5uL of protein and incubated at RT for 1 hour. Free dye was removed by passing through a Zeba Spin Desalting Columns, 40MWCO (87764, Thermo Scientific) as described in the package insert into droplet formation buffer with 1mM DTT in 125mM NaCl at a final concentration of 1uM. This protein was used in the droplet assay as needed.

For screening of a modified BODIPY library, 81 modified BODIPY molecules were selected from a larger library collection as previous described (*56*). These molecules were diluted to 1mM in DMSO then to 10μM in droplet formation buffer. Droplets of MED1-IDR-BFP were formed in Droplet formation buffer with 125mM NaCl and 10% PEG with 5μM protein, probe was added to this reaction to a final concentration of 1μM, the mixture was added to one well of a 384-well plate and imaged on an Andor confocal fluorescent microscope at 150x in the 488 (BODIPY) and 405 (protein) channels. Images were post-processed using Fiji Is Just ImageJ (FIJI). Images were quantified by the aforementioned pipelines to quantify the maximum 488 signal intensity in

droplets defined by the 405 channel. These values were then ranked to quantify the top and bottom “hits”. To ensure that the fluorescent intensity of the probes were equivalent, 1μM of 18 random probes in droplet formation buffer was imaged as above and the average fluorescent intensity in the field determined. The same approach was taken to measure the fluorescent intensity of BODIPY alone (Sigma 795526), both in MED1 droplet and in the diffuse state.

FRAP of in vitro droplets with drug

For FRAP of in vitro droplets, 5 pulses of laser at a 50μs dwell time was applied to the MED1 channel and 20 pulses of laser at a 100μs dwell time was applied to the Cisplatin channel. Recovery was imaged on an Andor microscope every 1s for the indicated time periods. Fluorescence intensity was measured using FIJI. Post bleach FRAP recovery data was averaged over 6 replicates for each channel.

Calculating drug enrichment ratios

To analyze in vitro droplet experiments, custom Python scripts using the scikit-image package were written to identify droplets and characterize their size, shape and intensity. Droplets were segmented from average images of captured channels on various criteria: (1) an intensity threshold that was three s.d. above the mean of the image; (2) size thresholds (20 pixel minimum droplet size); and (3) a minimum circularity (circularity=4π·areaperimeter2) (circularity=4π·areaperimeter2) of 0.8 (1 being a perfect circle). After segmentation, mean intensity for each droplet was calculated while excluding pixels near the phase interface, and background-corrected by subtracting intensity of dark images of droplet formation buffer only. Droplets identified in the channel of the fluorescent protein from ten independent fields of view were quantified for each experiment. The maximum intensity of signal within the droplets was calculated for each channel, the maximum intensity in the drug channel was termed “maximum drug intensity”. To obtain the intensity of drug or dye alone in the diffuse state (termed “diffuse drug intensity”), the compound was added to droplet formation buffer at same concentration used in the droplet assay. This was then imaged on a confocal fluorescent microscope, the resulting image was processed in FIJI to obtain the fluorescent intensity of the field. To obtain the fluorescent intensity of protein droplets that bleed through in the drug channel (termed “background intensity”) protein droplets were imaged in the fluorescent channel in which the drug fluoresces and processed as above to obtain the average maximum intensity within the droplet across 10 images. The enrichment ratio was obtained by the following formula [(maximum drug intensity)-(background intensity)]/(diffuse drug intensity). The box plots show the distributions of all droplets. Each dot represents an individual droplet.

Chromatin immunoprecipitation (ChIP) and sequencing

MCF7 cells were grown in complete DMEM media to 80% confluence. 1% formaldehyde in PBS was used for crosslinking of cells for 15 minutes, followed by quenching with Glycine at a final concentration of 125mM on ice. Cells were washed with cold PBS and harvested by scraping cells in cold PBS. Collected cells were pelleted at 1000 g for 3 minutes at 4°C, flash frozen in liquid nitrogen and stored at 80°C. All buffers contained freshly prepared cOmplete protease inhibitors (Roche, 11873580001). Frozen crosslinked cells were thawed on ice and then resuspended in lysis buffer I (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, protease inhibitors) and rotated for 10 minutes at 4°C, then spun at 1350 rcf., for 5 minutes at 4°C. The pellet was resuspended in lysis buffer II (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, protease inhibitors) and rotated for 10 minutes at 4°C and spun at 1350 rcf. for 5 minutes at 4C. The pellet was resuspended in sonication buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 0.1% SDS, and 1% Triton X-100, protease inhibitors) and then sonicated on a Misonix 3000 sonicator for 10 cycles at 30 s each on ice (18-21 W) with 60 s on ice between cycles. Sonicated lysates were

cleared once by centrifugation at 16,000 rcf. for 10 minutes at 4° C. Input material was reserved and the remainder was incubated overnight at 4°C with magnetic beads bound with CDK7 Bethyl A300-405A antibody to enrich for DNA fragments bound by CDK7. Beads were washed twice with each of the following buffers: wash buffer A (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 0.1% Na-Deoxycholate, 1% Triton X-100, 0.1% SDS), wash buffer B (50 mM HEPES-KOH pH 7.9, 500 mM NaCl, 1 mM EDTA pH 8.0, 0.1% Na-Deoxycholate, 1% Triton X-100, 0.1% SDS), wash buffer C (20 mM Tris-HCl pH8.0, 250 mM LiCl, 1 mM EDTA pH 8.0, 0.5% Na-Deoxycholate, 0.5% IGEPAL C-630, 0.1% SDS), wash buffer D (TE with 0.2% Triton X-100), and TE buffer. DNA was eluted off the beads by incubation at 65°C for 1 hour with intermittent vortexing in elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS). Cross-links were reversed overnight at 65°C. To purify eluted DNA, 200 mL TE was added and then RNA was degraded by the addition of 2.5 mL of 33 mg/mL RNase A (Sigma, R4642) and incubation at 37°C for 2 hours. Protein was degraded by the addition of 10 mL of 20 mg/mL proteinase K (Invitrogen, 25530049) and incubation at 55°C for 2 hours. A phenol:chloroform:isoamyl alcohol extraction was performed followed by an ethanol precipitation. The DNA was then resuspended in 50 mL TE and used for sequencing. ChIP libraries were prepared with the Swift Biosciences Accel-NGS 2S Plus DNA Library Kit, according to the kit instructions. Following library preparation, ChIP libraries were run on a 2% gel on the PippinHT with a size-collection window of 200–600 bases. Final libraries were quantified by qPCR with the KAPA Library Quantification kit from Roche, and sequenced in single-read mode for 40 bases on an Illumina HiSeq 2500.

HCT116 cells were grown in complete DMEM media to 80% confluence followed by treatment with JQ1 or DMSO for 24 hours, followed by cell permeabilization (10min at 37°C with the solution of tx100 in PBS at 1:1000 in media) and subsequently treated with DMF or Cisplatin for 6 hours. 1% formaldehyde in PBS was used for crosslinking of cells for 15 minutes, followed by quenching with Glycine at a final concentration of 125mM on ice. Cells were washed with cold PBS and harvested by scraping cells in cold PBS. Collected cells were pelleted at 1000 g for 3 minutes at 4°C, flash frozen in liquid nitrogen and stored at 80°C. All buffers contained freshly prepared cOmplete protease inhibitors (Roche, 11873580001). Frozen crosslinked cells were thawed on ice and then resuspended in lysis buffer I (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, protease inhibitors) and rotated for 10 minutes at 4°C, then spun at 1350 rcf., for 5 minutes at 4°C. The pellet was resuspended in lysis buffer II (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, protease inhibitors) and rotated for 10 minutes at 4°C and spun at 1350 rcf. for 5 minutes at 4°C. The pellet was resuspended in sonication buffer (20 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, protease inhibitors) and then sonicated on a Misonix 3000 sonicator for 10 cycles at 30 s each on ice (18-21 W) with 60 s on ice between cycles. Sonicated lysates were cleared once by centrifugation at 16,000 rcf. for 10 minutes at 4°C. Input material was reserved and the remainder was incubated overnight at 4°C with magnetic beads bound with MED1 antibody (Bethyl A300-793A) to enrich for DNA fragments bound by MED1. Beads were washed with each of the following buffers: washed twice with sonication buffer (20 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS), once with sonication buffer with high salt (20 mM Hepes pH 7.5, 500 mM NaCl, 1 mM EDTA 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS), once with LiCl wash buffer (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), and once with TE buffer. DNA was eluted off the beads by incubation with agitation at 65°C for 15 minutes in elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS). Cross-links were reversed for 12 hours at 65°C. To purify eluted DNA, 200 mL TE was added and then RNA was degraded by the addition of 2.5 mL of 33 mg/mL RNase A (Sigma, R4642) and incubation at 37°C for 2 hours. Protein was degraded by the addition of 4 ul of 20 mg/mL proteinase K (Invitrogen, 25530049) and incubated at 55°C for 30 minutes. DNA was

purified using Qiagen PCR purification kit, eluted in 30 µl Buffer EB, and used for sequencing. ChIP libraries were prepared with the Swift Biosciences Accel-NGS 2S Plus DNA Library Kit, according to the kit instructions. Following library preparation, ChIP libraries were run on a 2% gel on the PippinHT with a size-collection window of 200–400 bases. Final libraries were quantified by qPCR with the KAPA Library Quantification kit from Roche, and sequenced in single-read mode for 50 bases on an Illumina HiSeq 2500.

ChIP-Seq data were aligned to the mm9 version of the mouse reference genome using bowtie with parameters –k 1 –m 1 –best and –l set to read length. Wiggle files for display of read coverage in bins were created using MACS with parameters –w –S –space = 50 –nomodel –shiftsize = 200, and read counts per bin were normalized to the millions of mapped reads used to make the wiggle file. Reads-per-million-normalized wiggle files were displayed in the UCSC genome browser. For ER, MED1, BRD4, and H3K9me3 ChIP-Seq in MCF7 cells, published datasets were used (GEO GSE60270, GSM1348516, and GSM945857, respectively).

Purification of CDK8-Mediator

The CDK8-Mediator samples were purified as described (*57*) with modifications. Prior to affinity purification, the P0.5M/QFT fraction was concentrated, to 12 mg/mL, by ammonium sulfate precipitation (35%). The pellet was resuspended in pH 7.9 buffer containing 20mM KCl, 20mM HEPES, 0.1mM EDTA, 2mM MgCl2, 20% glycerol and then dialyzed against pH 7.9 buffer containing 0.15M KCl, 20mM HEPES, 0.1mM EDTA, 20% glycerol and 0.02% NP-40 prior to the affinity purification step. Affinity purification was carried out as described, eluted material was loaded onto a 2.2mL centrifuge tube containing 2mL 0.15M KCl HEMG (20mM HEPES, 0.1mM EDTA, 2mM MgCl2, 10% glycerol) and centrifuged at 50K RPM for 4h at 4°C. This served to remove excess free GST-SREBP and to concentrate the CDK8-Mediator in the final fraction. Prior to droplet assays, purified CDK8-Mediator was concentrated using Microcon-30kDa Centrifugal Filter Unit with Ultracel-30 membrane (Millipore MRCF0R030) to reach 300nM of Mediator complex. Concentrated CDK8-Mediator was added to the droplet assay to a final concentration of 200nM. Droplet reactions contained 10% PEG-8000 and 125mM salt.

Immunofluorescence with RNA FISH

Cells were plated on coverslips and grown for 24 hours followed by fixation using 4% paraformaldehyde, PFA, (VWR, BT140770) in PBS for 10 minutes. After washing cells three times in PBS, the coverslips were put into a humidifying chamber or stored at 4°C in PBS. Permeabilization of cells were performed using 0.5% Triton X-100 (Sigma Aldrich, X100) in PBS for 10 minutes followed by three PBS washes. Cells were blocked with 4% IgG-free Bovine Serum Albumin, BSA, (VWR, 102643-516) for 30 minutes and the indicated primary antibody (see table S2) was added at a concentration of 1:500 in PBS for 4-16 hours. Cells were washed with PBS three times followed by incubation with secondary antibody at a concentration of 1:5000 in PBS for 1 hour. After washing twice with PBS, cells were fixed using 4% paraformaldehyde, PFA, (VWR, BT140770) in PBS for 10 minutes. After two washes of PBS, Wash buffer A (20% Stellaris RNA FISH Wash Buffer A (Biosearch Technologies, Inc., SMF-WA1-60), 10% Deionized Formamide (EMD Millipore, S4117) in RNase-free water (Life Technologies, AM9932) was added to cells and incubated for 5 minutes. 12.5 mM RNA probe (Stellaris) in Hybridization buffer (90% Stellaris RNA FISH Hybridization Buffer (Biosearch Technologies, SMF-HB1-10) and 10% Deionized Formamide) was added to cells and incubated overnight at 37°C. After washing with Wash buffer A for 30 minutes at 37°C, the nuclei were stained with 20 mg/mL Hoechst 33258 (Life Technologies, H3569) for 5 minutes, followed by a 5 minute wash in Wash buffer B (Biosearch Technologies, SMF-WB1-20). Cells were washed once in water followed by mounting the coverslip onto glass slides with Vectashield (VWR, 101098-042) and finally sealing the coverslip with nail polish (Electron Microscopy Science Nm, 72180). Images were acquired at an

RPI Spinning Disk confocal microscope with a 100x objective using MetaMorph acquisition software and a Hammamatsu ORCA-ER CCD camera (W.M. Keck Microscopy Facility, MIT). Images were post-processed using Fiji Is Just ImageJ (FIJI).

RNA FISH image analysis

For analysis of RNA FISH with immunofluorescence, custom Python scripts were written to process and analyze 3D image data gathered in FISH and immunofluorescence channels. FISH foci were automatically called using the scipy ndimage package. The ndimage find_objects function was then used to call contiguous FISH foci in 3D. These FISH foci were then filtered by various criteria, including size, circularity of a maximum *z*-projection (circularity=4π·areaperimeter2;0.7)(circularity=4π·areaperimeter2;0.7), and being present in a nucleus (determined by nuclear mask). The FISH foci were then centered in a 3D box (length size (*l*) = 3.0 μm). The immunofluorescence signals centered at FISH foci for each FISH and immunofluorescence pair were then combined, and an average intensity projection was calculated, providing averaged data for immunofluorescence signal intensity within a *l*×*l* square centered at FISH foci. As a control, this same process was carried out for immunofluorescence signals centered at an equal number of randomly selected nuclear positions. These average-intensity projections were then used to generate 2D contour maps of the signal intensity. Contour plots were generated using the matplotlib Python package. For the contour plots, the intensity-color ranges presented were customized across a linear range of colors (*n* = 15). For the FISH channel, black to magenta was used. For the immunofluorescence channel, we used chroma.js (an online color generator) to generate colors across 15 bins, with the key transition colors chosen as black, blue–violet, medium blue and lime. This was done to ensure that the reader's eye could more-readily detect the contrast in signal. The generated color map was used in 15 evenly spaced intensity bins for all immunofluorescence plots. The averaged immunofluorescence, centered at FISH or at randomly selected nuclear locations, is plotted using the same color scale, set to include the minimum and maximum signal from each plot.

Cisplatin treatments followed by immunofluorescence

HCT116 cells were plated in 24-well plate at 50k cells per well to yield 100k cells after 21 hours (doubling time of HCTs). Cells were permeabilized using a solution of Tx100 in media at 0.55 pmol/cell for 12 minutes at 37°C. Cells were then washed with 500 µl media and treated with 500 µl of 50 µM cisplatin in media for 6 hours. After 6 hours, the cells were washed once with room temperature PBS and then fixed with 500 µL 4% formaldehyde in PBS for 12 min at room temperature. The cells were then washed 3 more times with PBS. Coverslips were put into a humidifying chamber or stored at 4°C in PBS. Permeabilization of cells were performed using 0.5% Triton X-100 (Sigma Aldrich, X100) in PBS for 10 minutes followed by three PBS washes. Cells were blocked with 4% IgG-free Bovine Serum Albumin, BSA, (VWR, 102643-516) for 30 minutes and the indicated primary antibody was added at a concentration of 1:500 in PBS for 4-16 hours. Cells were washed with PBS three times followed by incubation with secondary antibody at a concentration of 1:5000 in PBS for 1 hour. Samples was washed in PBS, DNA was stained using 20µm/mL Hoechst 33258 (Life Technologies, H3569) for 5 minutes and mounted using Vectashield (VWR, 101098-042). Images were acquired at an RPI Spinning Disk confocal microscope with a 100x objective using MetaMorph acquisition software and a Hammamatsu ORCA-ER CCD camera (W.M. Keck Microscopy Facility, MIT). Images were post-processed using Fiji Is Just ImageJ (FIJI).

Cisplatin/condensate co-IF

For the analysis of co-immunofluorescence data, custom python scripts were written to both process and analyze the 3D image data from IF and DAPI channels. Nuclei were detected using the Triangle thresholding method and a nuclear mask was applied the IF channels. Manual

minimal thresholds were applied to the 488 channel to determine nuclear puncta for protein of interest (MED1, HP1a, or FIB1). The triangle thresholding method was applied to the 561 channel to determine nuclear puncta for cisplatin. Percentage of cisplatin overlap was calculated by the number of defined nuclear cisplatin puncta that overlapped with the protein of interest puncta divided by the total number of nuclear cisplatin puncta.

Cisplatin-seq analysis

Cisplatin-seq fastq files for rep1 24-hour treated cells were downloaded from https://www.ncbi.nlm.nih.gov/sra/SRX1962532[accn] (sequencing run ID SRR3933212) (*40*). Reads were aligned to the human genome build hg19 (GRCh37) using Bowtie2 to get aligned .bam files (*58*). H3K27Ac ChIP-seq reads in HELA cells were used to call super-enhancers using the ROSE algorithm (*47*, *59*). Super-enhancers were separated from typical enhancers using the super-enhancer table output by ROSE algorithm. The typical enhancers were broken down further by their H3K27Ac signal. The last decile of enhancers was extracted based on H3K27Ac signal to get the low H3kK7Ac category of enhancers. Each category of enhancer *(*super-enhancers, typical enhancers, and low H3K27ac signal enhancers) was broken down into their constituents, and constituents that overlapped with blacklist regions were excluded. Black list regions were downloaded from ENCODE file https://www.encodeproject.org/files/ENCFF001TDO/. Each enhancer constituent was then extended by 2kb at either end. The 24-hour treated cisplatin-seq reads were mapped to each of the three categories of 2kb-extended enhancers using the bamToGFF.py script. For each category of enhancer, the constituent region and flanking regions were separately split into 50 equally-sized bins and the reads in each bin were counted. The average read count per bin across all enhancer constituents and flanking regions was used to create the meta-plot.

Cisplatin Treatments followed by live cell imaging

HCT116 cells with the indicated GFP knock-in were plated at 35k per well of a glass bottom 8-well chamber slide. Following incubation at 37°C overnight, cells were treated with 50μM cisplatin in DMEM or a 1:1000 dilution of DMSO for 12 hours. Prior to imaging, cells were additionally treated with a 1:5000 dilution of Hoechst 33342 to stain DNA and 2μM propidium iodide to stain dead cells. For the quantified dataset of GFP-tagged MED1, HP1 or FIB1 in HCT116 cells, cells were imaged using an Andor confocal microscope at 100X magnification. For representative images of each of the six tagged lines treated with vehicle or 50μM cisplatin, cells were imaged on the Zeiss LSM 880 confocal microscope with Airyscan detector with 63x objective at 37°C. Images were post-processed using Fiji Is Just ImageJ (FIJI).

Condensate score analysis

Nuclei were segmented from images of treated cells by custom Python scripts using the *scikit-image*, *open-cv*, and *scipy-ndimage* Python packages. Nuclei were segmented by median filter, thresholding, separated by the watershed algorithm, and labeled by the *scikit-image label* function. For each nuclei, the fluorescence signal in the GFP channel (corresponding to either MED1, HP1α or FIB1) was maximally-projected if z-stacks were acquired. A grey-level co-occurrence matrix (GLCM) was then generated from the projected signal, and the 'correlation' texture property from the GLCM was calculated per nucleus. One-way ANOVA followed by Sidak's multiple comparisons test was performed on the correlation values across conditions using GraphPad Prism version 8.2.0 for Mac (www.graphpad.com). Finally, to derive the condensation score, these values were subtracted from 1.

FRAP of HCT116 mEGFP tagged cell lines

FRAP was performed on Andor confocal microscope with 488nm laser. Bleaching was performed over a $r_{bleach} \approx 1\,um$ using 100% laser power and images were collected every two seconds.

Fluorescence intensity was measured using FIJI. Background intensity was subtracted, and values are reported relative to pre-bleaching time points. Post bleach FRAP recovery data was averaged over 7 replicates for each cell-line and condition.

Determination of partitioning by spectrophotometry and quantitative phase microscopy

Derivation of expression for drug partition coefficient in condensates

Here we derive briefly an expression for the partition coefficient of a client molecule into a condensed phase in terms of quantities that are readily measurable experimentally. We consider a sample composed of two coexisting phases, named dilute and condensed, with volume fractions ϕ_{dilute} and ϕ_{cond} such that $\phi_{dilute} + \phi_{cond} = 1$. If a client molecule (e.g. a drug) is also present in the sample at an average concentration of c_{total}, then mass conservation requires that

$$c_{total} = c_{dilute}\phi_{dilute} + c_{cond}\phi_{cond}, \quad (1)$$

where c_{dilute} and c_{cond} are the concentrations of the client in the dilute and condensed phases, respectively. Finally, we define the partition coefficient of the client into the condensed phase as $P = c_{cond}/c_{dilute}$. With this definition and the requirement that the phase volume fractions sum to 1, Eq 1 can be written as

$$c_{total} = c_{dilute}(1 - \phi_{cond}) + c_{dilute}P\phi_{cond}, \quad (2)$$

which can be simplified and rearranged to yield

$$P = 1 + \left(\frac{c_{total}}{c_{dilute}} - 1\right)(\phi_{cond})^{-1}. \quad (3)$$

We estimate the ratio c_{total}/c_{dilute} from fluorescence spectroscopy measurements, as described in a subsequent section, while ϕ_{cond} we infer from the lever rule (*60*) as follows: denoting the concentration of scaffold protein (e.g. MED1) by *s*, mass conservation gives $s_{total} = s_{dilute}\phi_{dilute} + s_{cond}\phi_{cond}$, in analogy with Eq. 1. Again using the requirement that the volume fractions of coexisting phases sum to 1, this can be rearranged to yield

$$\phi_{cond} = \frac{s_{total} - s_{dilute}}{s_{cond} - s_{dilute}}, \quad (4)$$

where s_{total} and s_{dilute} are measured spectrophotometrically from optical absorbance at 280 nm, and s_{cond} is measured from quantitative phase microscopy, using a coherence-controlled holographic microscope (Q-Phase, Telight (formerly TESCAN), Brno, CZ) equipped with 40x dry objectives (NA = 0.90).

UV-Vis fluorescence spectroscopy measurements and analysis

Uv-vis spectroscopy (TECAN Spark20M) was used to estimate the absolute concentration of drug in solution using Beer-Lambert law with Eq 5,

$$A = \text{Log10}(I0/I) = \varepsilon cL \quad (5)$$

where A is the measured absorbance (in Absorbance Units (AU)), I0 is the intensity of the incident light at a given wavelength, I is the transmitted intensity, L the path length through the sample, and c the concentration of the absorbing species. For each species and wavelength, ε is a constant known as the molar absorptivity or extinction coefficient. This constant is a fundamental molecular property in a given solvent, at a particular temperature and pressure, and has units of 1/M*cm.

The partitioned drug was measured by using spin down assay. Known concentration of drug was added with the protein and kept for the droplet formation. After 30 minutes, the mixture was centrifuged at 15,000 rpm for 10 minutes. The supernatant was collected and measured the concentrations of the drug. The partitioned drug was calculated by subsracting from the total known concentration of drug added.

Quantitative phase microscopy measurements and analysis

Quantitative phase measurements were performed using a coherence-controlled holographic microscope (Q-Phase, Telight (formerly TESCAN), Brno, CZ) equipped with 40x dry objectives

(NA = 0.90) as follows. Immediately following phase separation, samples were loaded into a custom temperature-controlled flowcell, sealed and allowed to settle under gravity prior to imaging. Flowcells were constructed with a PEGylated coverslip and a sapphire slide as bottom and top surfaces, respectively, using parafilm strips as spacers. Peltier elements affixed to the sapphire slide enabled regulation of flowcell temperature, as previously described (*61*). Temperature was maintained at 21.00 ± 0.02 °C during measurements.

Q-PHASE software was used to construct compensated phase images from acquired holograms, which were subsequently analyzed in MATLAB using custom code. As details regarding the calculation of protein concentration from quantitative phase images will be discussed extensively elsewhere (McCall et al, forthcoming), only a conceptual overview will be given here. Briefly, each phase image is spatially segmented based on intensity, and a window containing each segmented object is fit to a spatial function of the form

$$\varphi(x,y) = \frac{2\pi}{\lambda}\Delta n H(x,y|R), \quad (6)$$

where $\varphi(x,y)$ is the phase intensity at pixel location (x,y), λ is the illumination wavelength, Δn is the refractive index difference between MED1 condensates and the surrounding dilute phase, and $H(x,y|R)$ is the projected height of a sphere of radius R. The fitting parameters in Eq. 6 are Δn and R. We assume that no PEG partitions into the condensates and calculate the average scaffold concentration in each filtered condensate as

$$s_{cond} = \frac{\Delta n + (n_{dilute} - n_0)}{dn/ds}. \quad (7)$$

Here n_0 is the refractive index of buffer in the absence of scaffold and PEG, n_{dilute} is the refractive index of the dilute phase, and both are measured at 21.00 ± 0.01 °C using a J457 digital refractometer (Rudolph Research Analytic, Hackettstown, NJ). The refractive index increment of the scaffold protein, dn/ds, is estimated from amino acid composition (*62*).

Cisplatin-DNA engagement assay

MED1-IDR-BFP and HP1a-BFP droplets were formed by mixing 10 μM protein with the droplet formation buffer containing 50mM Tris-HCl pH 7.5, 100mM NaCl, 10% PEG 8000, 10% glycerol, 1mM DTT and 5ng/μl DNA in a 10 μl reaction volume. The droplet reactions were incubated for 30 min at RT. Next, increasing concentrations of activated Cisplatin (0, 0.5, 0.75, 1, 1.5, and 2 mM) were added to the droplet reactions and incubated for another 30 min at RT. The reactions are then treated with 1 μl of Proteinase K (Invitrogen, 20 mg/ml) for 4 hr at 55 °C. Platination of DNA was visualized by size-shift on a bioanalyzer. Measurements from two chip runs were compiled into a single electropherogram.

Amino acid and basic/acidic patch analysis

Basic and acidic patches were determined by identifying charged interaction elements (CIEs) as previously described by (*63*). For each protein, the net charge per residue (NCPR) along the protein sequence was calculated using a sliding window of 5 amino acids with a step size of 1 amino acid using the localCIDER software (*64*). Stretches of 4 or more amino acids with NCPR < -0.35 were identified as acidic patches (CIE-), while stretches of 4 or more amino acids with NCPR > +0.35 ewere identified as basic patches (CIE+). The number of acidic and basic patches within the total protein and the IDR specifically was counted. Separately, the number of aromatic residues within the whole protein and the IDR was also counted.

Cell survival assay

HCT116 cells were plated in 24-well plate at 50k cells per well to yield 100k cells after 21 hours (doubling time of HCTs). Cells were then treated with either 50μM cisplatin or DMF in DMEM media for 12 hours. At 12 hours, CellTiter-Glo Reagent was added to each well, following the

CellTiter-Glo Luminescent Viability Assay. Luminescence was then measured, averaging 5 wells for each condition.

In silico modeling

We developed a simplified model of drug-target interactions in the presence of a condensate. The relevant species are the drug (D), target (T), and the drug-target complex (D-T). We assume there are only 2-types of phases, the bulk/dilute nuclear phase (n) and the condensate phase (c), which is present with volume fraction $f = V_{condensate}/V_{nucleus}$. At equilibrium, the following partitioning conditions are obeyed:

$$\frac{[D]_c}{[D]_n} = p_D; \frac{[T]_c}{[T]_n} = p_T;$$

where p_D, p_T are the partition coefficients of the drug and target. $[D]_c$ represents the concentration of species D in condensate phase (and similarly for other components/phases). In this model, the drug and target complex with phase-independent disassociation constant of K_D.

$$[D] + [T] \leftrightarrow^{K_D} [D-T]$$
$$K_D = \frac{[D][T]}{[D-T]}$$

To solve for equilibrium concentrations of various species, which are present at overall levels $[D]_0, [T]_0$, we write down the species balance as:

$$f([D]_c + [D-T]_c) + (1-f)([D]_n + [D-T]_n) = [D]_0$$
$$f([T]_c + [D-T]_c) + (1-f)([T]_n + [D-T]_n) = [DT]_0$$

We solve these 6 concentrations with 2-equations and 4 constraints (2 from partitioning and 2 from reaction equilibria). In Fig S31, the fraction of bound target is defined as:

$$Fraction_{bound}, c = \frac{[D-T]_c}{[D]_c + [T]_c}$$

A similar expression is used for the fraction of bound target in the nuclear (bulk or dilute) phase. In case of controls plotted, we plot fraction when there is only 1 phase (f=0).

Generation and analysis of MCF7 mEGFP-MED1 cells

To generate MCF7 mEGFP-MED1 cells, a lentiviral construct containing the full length MED1 with a N-terminal mEGFP fusion connected by a 10 amino acid GS linker was cloned, containing a puromycin selection marker. Lentiviral particles were generated in HEK293T cells. 250,000 MCF7 cells were plated in one well of a 6 well plate and viral supernatant was added. 48 hours later puromycin was added at 1µg/mL for 5 days for selection.

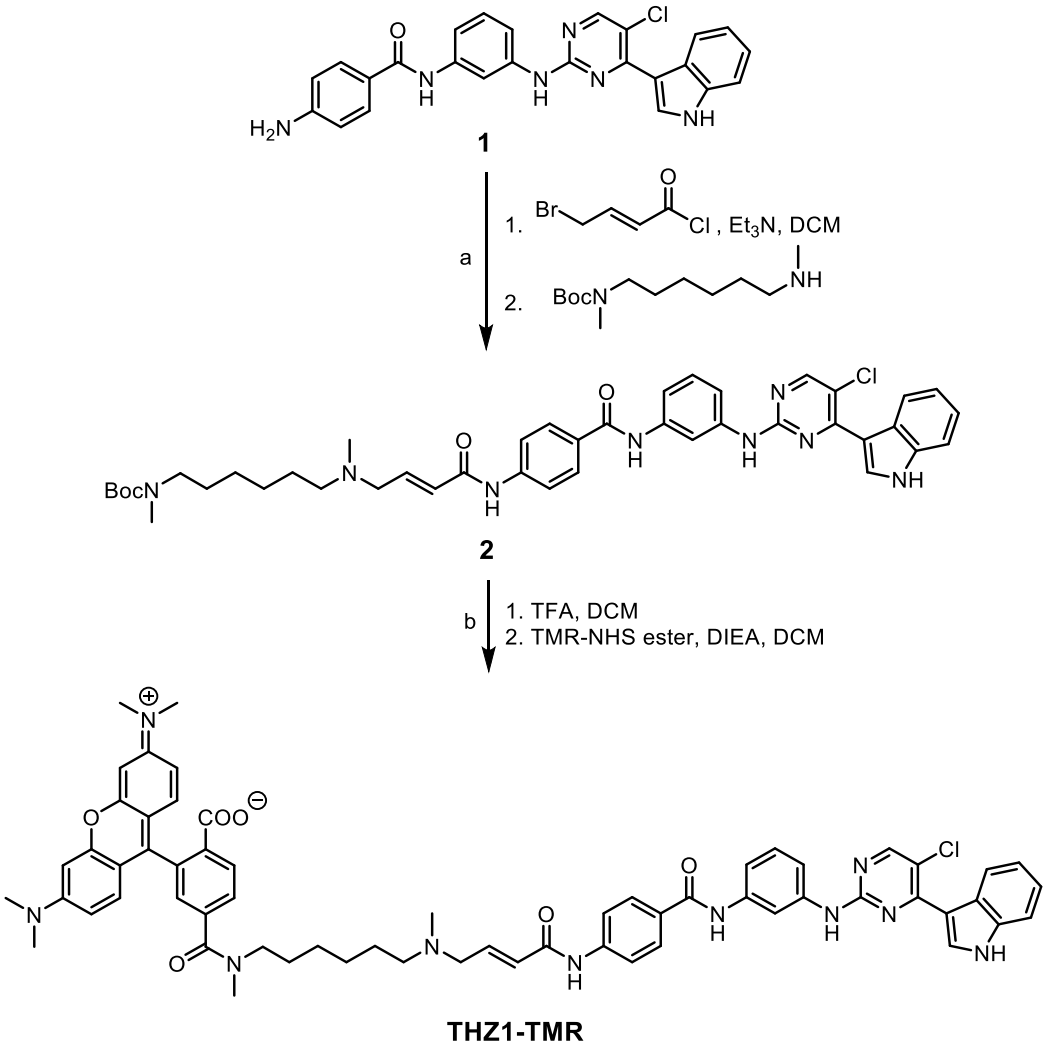
For live-cell FRAP experiments, the tagged MED1-mEGFP MCF7 cells were plated on Poly-L-Ornithine coated glass-bottom tissue culture plate. 20 pulses of laser at a 50µs dwell time were applied to the array, and recovery was imaged on an Andor microscope every 1 s for the indicated time periods. Quantification was performed in FIJI. The instrument background was subtracted from the average signal intensity in the bleached puncta then divided by the instrument background subtracted from a control puncta. These values were plotted every second, and a best fit line with 95% confidence intervals was calculated. For observing fusions of MED1-GFP foci, MED1-mEGFP MCF7 cells were grown for 3 days in estrogen-free conditions then plated on glass-bottomed plates. 15 minutes prior to imaging, cells were treated with 100nM estrogen and

placed on the Andor confocal microscope and imaged at 150x for 4 minutes. Images were post-processed in FIJI. Fluorescent intensity calculations were made in FIJI.

Chemistry

Unless otherwise noted, reagents and solvents were obtained from commercial suppliers and were used without further purification. Mass spectra were obtained on a Waters Micromass ZQ instrument. Preparative HPLC was performed on a Waters Sunfire C18 column (19 mm × 50 mm, 5 μM) using a gradient of 15−95% methanol in water containing 0.05% trifluoroacetic acid (TFA) over 22 min (28 min run time) at a flow rate of 20 mL/min.

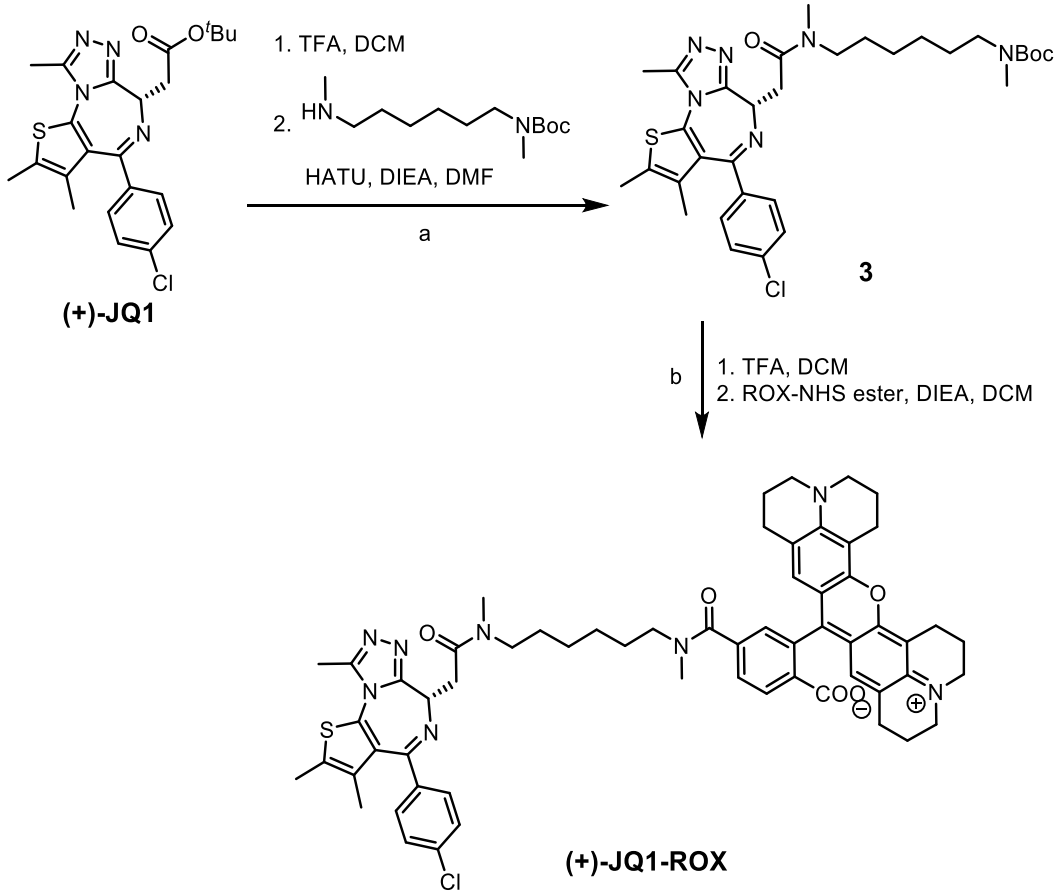
Scheme 1: Synthesis of THZ1-TMR



Reagents and conditions: (a) (*E*)-4-bromobut-2-enoyl chloride, triethyl amine, DCM, 0 °C~r.t., then *tert*-butyl methyl(6-(methylamino)hexyl)carbamate, r.t.~50 °C; (b) trifluoroacetic acid, DCM, r.t., then TMR-NHS ester, diisopropylethyl amine, DCM, r.t.~40 °C *tert*-butyl *E*)-(6-((4-((4-((3-((5-chloro-4-(1H-indol-3-yl)pyrimidin-2-yl)amino)phenyl) carbamoyl)phenyl)amino)-4-oxobut-2-en-1-yl)(methyl)amino)hexyl)(methyl)carbamate (2). To a solution of 1 (20 mg, 0.044 mmol, prepared

according to patent WO2014/63068) and triethyl amine (29 mg, 0.27 mmol) in 0.8 mL DCM was added (*E*)-4-bromobut-2-enoyl chloride (0.24 mL, 0.2 M in DCM). The solution was stirred for 6 hours. Then *tert*-butyl methyl(6-(methylamino)hexyl)carbamate (13 mg, 0.052 mmol) in 0.4 mL DCM was added. The mixture was warmed to 50 °C and kept overnight. The mixture was concentrated in vacuo, then purified by preparative HPLC to provide intermediate 2 (6 mg, 19%). LC/MS (ESI) m/z = 765 (M + H)+. (*E*)-4-((6-((4-((4-((3-((5-chloro-4-(1H-indol-3-yl)pyrimidin-2-yl)amino)phenyl)carbamoyl)phenyl)amino)-4-oxobut-2-en-1-yl)(methyl)amino)hexyl)(methyl)carbamoyl)-2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)benzoate (THZ1-TMR). To a solution of 2 (6 mg, 0.0078 mmol) in 0.5 mL DCM was added 0.1 mL TFA. The resultant solution was stirred at room temperature for 1 h, and then concentrated in vacuo to obtain free amine as TFA salt, which was dissolved in 0.5 mL DCM again. To this solution DIEA (5mg, 0.039 mmol) and TMR-NHS ester (5 mg, 0.0094 mmol) were added in sequence. The mixture was warmed to 40 °C and kept overnight. The mixture was concentrated in vacuo, then purified by preparative HPLC to provide THZ1-TMR (2 mg, 23%). LC/MS (ESI) m/z = 1077 (M + H)+.

Scheme 2: Synthesis of (+)-JQ1-ROX



Reagents and conditions: (a) trifluoroacetic acid, DCM, r.t., then *tert*-butyl methyl(6-(methylamino)hexyl)carbamate, 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate, diisopropylethyl amine, DMF, r.t.; (b) trifluoroacetic acid, DCM, r.t., then ROX-NHS eater, diisopropylethyl amine, DCM, r.t.~40 °C *tert*-butyl (*S*)-(6-(2-(4-(4-chlorophenyl)-2,3,9-trimethyl-6H-thieno[3,2-f][1,2,4]triazolo[4,3-a][1,4]diazepin-6-yl)-N-methylacetamido)hexyl)(methyl)carbamate (3) To a solution of (+)-JQ1 (25 mg, 0.055 mmol) in 2

mL DCM was added 0.4 mL TFA. The resultant solution was stirred at room temperature for 1 h, and then concentrated in vacuo to obtain free amine as TFA salt, which was dissolved in 0.8 mL DMF. To this solution was added *tert*-butyl methyl(6-(methylamino)hexyl)carbamate (16 mg, 0.065 mmol) in 0.5 mL DMF, DIEA (35mg, 0.28 mmol) and HATU (24 mg, 0.064 mmol) in sequence. The mixture was stirred at r.t. for 6 hours. Then purified by preparative HPLC to provide intermediate 3 (15 mg, 43%). LC/MS (ESI) m/z = 627 (M + H)+.
(+)JQ1-ROX. To a solution of 3 (15 mg, 0.024 mmol) in 2 mL DCM was added 0.4 mL TFA. The resultant solution was stirred at room temperature for 1 h, and then concentrated in vacuo to obtain free amine as TFA salt, which was dissolved in 1 mL DCM again. To this solution DIEA (16mg, 0.12 mmol) and ROX-NHS ester (13mg, 0.021 mmol) were added in sequence. The mixture was warmed to 40°C and kept overnight. The mixture was concentrated in vacuo, then purified by preparative HPLC to provide (+)JQ1-ROX (6 mg, 28), LC/MS (ESI) m/z = 1043 (M + H)+.

Immunofluorescence with DNA FISH

MCF7 cells were grown in estrogen-free DMEM for 3 days on Poly-L-ornithine coated coverslips in 24 well plates at an initial seeding density of 50,000 cells per well. Cells were then treated with vehicle, 10nM estradiol, or 10nM estradiol and 5uM 4-hydroxytamoxifen for 45 minutes. HCT116 cells were treated with 1μM JQ1 for 24 hours, followed by cell permeabilization (10min at 37°C with the solution of tx100 in PBS at 1:1000 in media) and subsequently DMF or 50μM Cisplatin for 6 hours.

Cells on cover slips were then fixed in 4% paraformaldehyde. Immunofluorescence was performed as described above. After incubating the cells with the secondary antibodies, cells were washed three times in PBS for 5min at RT, fixed with 4% PFA in PBS for 10min and washed three times in PBS. Cells were incubated in 70% ethanol, 85% ethanol and then 100% ethanol for 1 minute at RT. Probe hybridization mixture was made mixing 7μL of FISH Hybridization Buffer (Agilent G9400A), 1μl of FISH probes (SureFISH 8q24.21 MYC 294kb G101211R-8) and 2μL of water. 5μL of mixture was added on a slide and coverslip was placed on top (cell-side toward the hybridization mixture). Coverslip was sealed using rubber cement. Once rubber cement solidified, genomic DNA and probes were denatured at 78°C for 5 minutes and slides were incubated at 16°C in the dark O/N. The coverslip was removed from slide and incubated in pre-warmed Wash Buffer 1 (Agilent, G9401A) at 73°C for 2 minutes and in Wash Buffer 2 (Agilent, G9402A) for 1 minute at RT. Slides were air dried and nuclei were stained in 20μm/mL Hoechst 33258 (Life Technologies, H3569) in PBS for 5 minutes at RT. Coverslips were washed three times in PBS, followed by mounting the coverslip onto glass slides, sealing, imaging, and post-processing as described above.

RT-qPCR

MCF7 cells were estrogen deprived for 3 days then stimulated with either 10nM estrogen or 10nM estrogen and 5μM 4-hydroxytamoxifen for 24 hours. RNA was isolated by AllPrep Kit (Qiagen 80204) followed by cDNA synthesis using High-Capacity cDNA Reverse Transcription Kit (Applies Biosystems 4368814). qPCR was performed in biological and technical triplicate using Power SYBR Green mix (Life Technologies #4367659) on a QuantStudio 6 System (Life Technologies). The following oligos was used in the qPCR; Myc fwd AACCTCACAACCTTGGCTGA, MYC rev TTCTTTTATGCCCAAAGTCCAA, GAPDH fwd TGCACCACCAACTGCTTAGC, GAPDH rev GGCATGGACTGTGGTCATGAG. Fold change was calculated and *MYC* expression values were normalized to *GAPDH* expression.

LAC binding assay

Constructs were assembled by NEB HIFI cloning in pSV2 mammalian expression vector containing an SV40 promoter driving expression of a mCherry-LacI fusion protein. The intrinsically disordered region of MED1, HP1α, or the activation domain of *ESR1* was fused by the c-terminus to this recombinant protein, joined by the linker sequence GAPGSAGSAAGGSG. For experiments comparing FLTX1 enrichment at the array, U2OS-Lac cells were plated onto chambered coverglass (1.5 Borosilicate Glass, Nunc Lab-Tek, 155409) and transfected with either MED1 IDR or HP1α constructs with lipofectamine 3000 (Thermofisher L3000015). After 24 hours, cells were treated with either 1uM FLTX1 or vehicle (DMF). After 30 minutes, cells were imaged on the Zeiss LSM 880 confocal microscope with Airyscan detector with 63x objective at 37°C. For experiments with high MED1, cells grown in DMEM were plated on glass coverslips and transfected using lipofectamine 3000 (Thermofisher L3000015). A construct with a mammalian expression vector containing a PGK promoter driving the expression of MED1 fused to GFP was co-transfected in high MED1 conditions. 24 hours after transfection, cells were treated for 45 minutes with 4-Hydroxytamoxifen (Sigma-Aldrich H7904) reconstituted in DMSO. Following treatment, cells were fixed and immunofluorescence was performed with a MED1 antibody as described above. Cells were then imaged using the RPI Spinning Disk confocal microscope with a 100x objective. Images were post-processed in FIJI.

For analysis of Lac array data comparing MED1 or HP1α tethered, a region of interest was called using the signal in the Lac array (561 channel). The average fluorescent signal for FLTX1 (488 channel) was then measured in the region of interest and divided by the average fluorescence in the region of interest at the Lac array. This value was then divided in the drug treated condition by the vehicle treated condition and all values were normalized to the HP1α condition. For analysis of Lac array data for MED1 overexpression, enrichment was calculated by dividing the average fluorescent signal for MED1 immunofluorescence at the region of interest, defined by the ER tethered at the lac array, by MED1 immunofluorescence signal at a random nuclear region. Enrichment of MED1 was plotted over each concentration of tamoxifen in wildtype or high MED1 conditions.

Western blot

Cells were lysed in Cell Lytic M (Sigma-Aldrich C2978) with protease inhibitors (Roche, 11697498001). Lysate was run on a 3%–8% Tris-acetate gel or 10% Bis-Tris gel or 3-8% Bis-Tris gels at 80 V for ~2 hrs, followed by 120 V until dye front reached the end of the gel. Protein was then wet transferred to a 0.45 µm PVDF membrane (Millipore, IPVH00010) in ice-cold transfer buffer (25 mM Tris, 192 mM glycine, 10% methanol) at 300 mA for 2 hours at 4°C. After transfer the membrane was blocked with 5% non-fat milk in TBS for 1 hour at room temperature, shaking. Membrane was then incubated with 1:1,000 of the indicated antibody (ER ab32063, MED1 ab64965) diluted in 5% non-fat milk in TBST and incubated overnight at 4°C, with shaking. In the morning, the membrane was washed three times with TBST for 5 minutes at room temperature shaking for each wash. Membrane was incubated with 1:5,000 secondary antibodies for 1 hr at RT and washed three times in TBST for 5 minutes. Membranes were developed with ECL substrate (Thermo Scientific, 34080) and imaged using a CCD camera or exposed using film or with high sensitivity ECL. Quantification of western blot was performed using BioRad image lab.

References

1. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science (80-.).* **357**, eaaf4382 (2017).
2. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
3. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
4. R. J. Ries, S. Zaccara, P. Klein, A. Olarerin-George, S. Namkoong, B. F. Pickering, D. P. Patil, H. Kwak, J. H. Lee, S. R. Jaffrey, m6A enhances the phase separation potential of mRNA. *Nature*. **571**, 424–428 (2019).
5. E. M. Langdon, A. S. Gladfelter, A New Lens for RNA Localization: Liquid-Liquid Phase Separation. *Annu. Rev. Microbiol.* **72**, 255–271 (2018).
6. E. M. Langdon, Y. Qiu, A. Ghanbari Niaki, G. A. McLaughlin, C. A. Weidmann, T. M. Gerbich, J. A. Smith, J. M. Crutchley, C. M. Termini, K. M. Weeks, S. Myong, A. S. Gladfelter, mRNA structure determines specificity of a polyQ-driven phase separation. *Science (80-.).* **360**, 922–927 (2018).
7. M.-T. Wei, Y.-C. Chang, S. F. Shimobayashi, Y. Shin, C. P. Brangwynne, Nucleated transcriptional condensates amplify gene expression. *bioRxiv*, 737387 (2019).
8. T. J. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. P. Bazett-Jones, T. Pawson, J. D. D. Forman-Kay, A. J. J. Baldwin, Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell*. **57**, 936–947 (2015).
9. T. J. Nott, T. D. Craggs, A. J. Baldwin, Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nat. Chem.* **8**, 569–575 (2016).
10. B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, I. L. Tong, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, Coactivator condensation at super-enhancers links phase separation and gene control. *Science (80-.).* **361**, eaar3958 (2018).
11. Y. E. Guo, J. C. Manteiga, J. E. Henninger, B. R. Sabari, A. Dall'Agnese, N. M. Hannett, J.-H. Spille, L. K. Afeyan, A. V. Zamudio, K. Shrinivas, B. J. Abraham, A. Boija, T.-M. Decker, J. K. Rimel, C. B. Fant, T. I. Lee, I. I. Cisse, P. A. Sharp, D. J. Taatjes, R. A. Young, Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*. **572**, 543–548 (2019).
12. A. Boija, I. A. Klein, B. R. Sabari, A. Dall'Agnese, E. L. Coffey, A. V Zamudio, C. H. Li, K. Shrinivas, J. C. Manteiga, N. M. Hannett, B. J. Abraham, L. K. Afeyan, Y. E. Guo, J. K. Rimel, C. B. Fant, J. Schuijers, T. I. Lee, D. J. Taatjes, R. A. Young, Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*. **175**, 1842-1855.e16 (2018).
13. J. J. Bouchard, J. H. Otero, D. C. Scott, E. Szulc, E. W. Martin, N. Sabri, D. Granata, M. R. Marzahn, K. Lindorff-Larsen, X. Salvatella, B. A. Schulman, T. Mittag, Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. *Mol. Cell*. **72**, 19-36.e8 (2018).
14. K. Shrinivas, B. R. Sabari, E. L. Coffey, I. A. Klein, A. Boija, A. V Zamudio, J. Schuijers, N. M. Hannett, P. A. Sharp, R. A. Young, A. K. Chakraborty, Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell*. **75**, 549-561.e7 (2019).

15. E. P. Bentley, B. B. Frey, A. A. Deniz, Physical Chemistry of Cellular Liquid-Phase Separation. *Chemistry*. **25**, 5600–5610 (2019).
16. M. Boehning, C. Dugast-Darzacq, M. Rankovic, A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer, X. Darzacq, M. Zweckstetter, RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
17. P. Cramer, Organization and regulation of gene transcription. *Nature*. **573**, 45–54 (2019).
18. S. Alberti, S. Saha, J. B. Woodruff, T. M. Franzmann, J. Wang, A. A. Hyman, A User's Guide for Phase Separation Assays with Purified Proteins. *J. Mol. Biol.* **430**, 4806–4820 (2018).
19. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, Perspective A Phase Separation Model for Transcriptional Control. *Cell*. **169**, 13–23 (2017).
20. Y. Chen, A. S. Belmont, Genome organization around nuclear speckles. *Curr. Opin. Genet. Dev.* **55**, 91–99 (2019).
21. A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, G. J. Narlikar, Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*. **547**, 236–240 (2017).
22. A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, G. H. Karpen, Phase separation drives heterochromatin domain formation. *Nat. Publ. Gr.* **547**, 241–245 (2017).
23. M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, C. P. Brangwynne, Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*. **165**, 1686–1697 (2016).
24. D. M. Mitrea, J. A. Cika, C. S. Guy, D. Ban, P. R. Banerjee, C. B. Stanley, A. Nourse, A. A. Deniz, R. W. Kriwacki, Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *Elife*. **5** (2016), doi:10.7554/eLife.13571.
25. D. M. Mitrea, J. A. Cika, C. B. Stanley, A. Nourse, P. L. Onuchic, P. R. Banerjee, A. H. Phillips, C.-G. Park, A. A. Deniz, R. W. Kriwacki, Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation. *Nat. Commun.* **9**, 842 (2018).
26. S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, M. K. Rosen, Compositional Control of Phase-Separated Cellular Bodies. *Cell*. **166**, 651–663 (2016).
27. A. Patel, H. O. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. A. Royer, M. Weigert, E. W. W. Myers, S. Grill, D. Drechsel, A. A. A. Hyman, S. Alberti, A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*. **162**, 1066–1077 (2015).
28. S. Alberti, The wisdom of crowds: regulating cell function through condensed states of living matter. *J. Cell Sci.* **130**, 2789–2796 (2017).
29. A. E. Posey, A. S. Holehouse, R. V. Pappu, in *Methods in enzymology* (2018; https://linkinghub.elsevier.com/retrieve/pii/S0076687918304099), vol. 611, pp. 1–30.
30. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science (80-.).* **357**, eaaf4382 (2017).
31. V. N. Uversky, Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* **44**, 18–30 (2017).
32. Y. S. Mao, B. Zhang, D. L. Spector, Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).
33. D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, R. A. Young, XSuper-enhancers in the control of cell identity and disease. *Cell*. **155**, 934–47 (2013).
34. Y. H. Chu, M. Sibrian-Vazquez, J. O. Escobedo, A. R. Phillips, D. T. Dickey, Q. Wang, M.

Ralle, P. S. Steyger, R. M. Strongin, Systemic Delivery and Biodistribution of Cisplatin in Vivo. *Mol. Pharm.* **13**, 2677–2682 (2016).
35. S. Vibet, K. Mahéo, J. Goré, P. Dubois, P. Bougnoux, I. Chourpa, Differential subcellular distribution of mitoxantrone in relation to chemosensitization in two human breast cancer cell lines. *Drug Metab. Dispos.* **35**, 822–8 (2007).
36. P. J. Smith, H. R. Sykes, M. E. Fox, I. J. Furlong, Subcellular distribution of the anticancer drug mitoxantrone in human and drug-resistant murine cells analyzed by flow cytometry and confocal microscopy and its relationship to the induction of DNA damage. *Cancer Res.* **52**, 4000–8 (1992).
37. N. Kwiatkowski, T. Zhang, P. B. Rahl, B. J. Abraham, J. Reddy, S. B. Ficarro, A. Dastur, A. Amzallag, S. Ramaswamy, B. Tesar, C. E. Jenkins, N. M. Hannett, D. McMillin, T. Sanda, T. Sim, N. D. Kim, T. Look, C. S. Mitsiades, A. P. Weng, J. R. Brown, C. H. Benes, J. A. Marto, R. A. Young, N. S. Gray, Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature*. **511**, 616–20 (2014).
38. J. Marrero-Alonso, A. Morales, B. García Marrero, A. Boto, R. Marín, D. Cury, T. Gómez, L. Fernández-Pérez, F. Lahoz, M. Díaz, Unique SERM-like properties of the novel fluorescent tamoxifen derivative FLTX1. *Eur. J. Pharm. Biopharm.* **85**, 898–910 (2013).
39. W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, I. I. Cisse, Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*. **361**, 412–415 (2018).
40. M. J. Tilby, C. Johnson, R. J. Knox, J. Cordell, J. J. Roberts, C. J. Dean, Sensitive detection of DNA modifications induced by cisplatin and carboplatin in vitro and in vivo using a monoclonal antibody. *Cancer Res.* **51**, 123–9 (1991).
41. X. Shu, X. Xiong, J. Song, C. He, C. Yi, Base-Resolution Analysis of Cisplatin-DNA Adducts at the Genome Scale. *Angew. Chem. Int. Ed. Engl.* **55**, 14246–14249 (2016).
42. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. **153**, 307–319 (2013).
43. X. Rovira-Clave, S. Jiang, Y. Bai, G. L. Barlow, S. Bhate, A. Coskun, G. Han, B. Zhu, C.-M. Ho, C. Hitzman, S.-Y. Chen, F.-A. Bava, G. Nolan, Subcellular localization of drug distribution by super-resolution ion beam imaging. *bioRxiv*, 557603 (2019).
44. Y. Wang, T. Zhang, N. Kwiatkowski, B. J. Abraham, T. I. Lee, S. Xie, H. Yuzugullu, T. Von, H. Li, Z. Lin, D. G. Stover, E. Lim, Z. C. Wang, J. D. Iglehart, R. A. Young, N. S. Gray, J. J. Zhao, CDK7-dependent transcriptional addiction in triple-negative breast cancer. *Cell*. **163**, 174–86 (2015).
45. M. R. Mansour, J. M. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A. D. Durbin, J. Etchin, L. Lawton, S. E. Sallan, L. B. Silverman, M. L. Loh, S. P. Hunger, T. Sanda, R. A. Young, A. T. Look, B. J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A. D. Durbin, J. Etchin, L. Lawton, S. E. Sallan, L. B. Silverman, M. L. Loh, S. P. Hunger, T. Sanda, R. A. Young, A. T. Look, An Oncogenic Super-Enhancer Formed Through Somatic Mutation of a Noncoding Intergenic Element. *Science (80-.).* **346**, 1373–1377 (2014).
46. J. E. Bradner, D. Hnisz, R. A. Young, Transcriptional Addiction in Cancer. *Cell*. **168**, 629–643 (2017).
47. J. Lovén, H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, R. A. Young, Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. **153**, 320–334 (2013).
48. Z. Nie, G. Hu, G. Wei, K. Cui, A. Yamane, W. Resch, R. Wang, D. R. Green, L. Tessarollo, R. Casellas, K. Zhao, D. Levens, c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell*. **151**, 68–79 (2012).
49. S. Dasari, P. Bernard Tchounwou, Cisplatin in cancer therapy: Molecular mechanisms of action. *Eur. J. Pharmacol.* **740** (2014), pp. 364–378.

50. D. Dubik, T. C. Dembinski, R. P. C. Shiu4, “Stimulation of c-myc Oncogene Expression Associated with Estrogen-induced Proliferation of Human Breast Cancer Cells1” (1987).
51. A. Nagalingam, M. Tighiouart, L. Ryden, L. Joseph, G. Landberg, N. K. Saxena, D. Sharma, Med1 plays a critical role in the development of tamoxifen resistance. *Carcinogenesis*. **33**, 918–30 (2012).
52. S. W. Fanning, C. G. Mayne, V. Dharmarajan, K. E. Carlson, T. A. Martin, S. J. Novick, W. Toy, B. Green, S. Panchamukhi, B. S. Katzenellenbogen, E. Tajkhorshid, P. R. Griffin, Y. Shen, S. Chandarlapaty, J. A. Katzenellenbogen, G. L. Greene, Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *Elife*. **5** (2016), doi:10.7554/eLife.12792.
53. C. S. Ross-Innes, R. Stark, A. E. Teschendorff, K. A. Holmes, H. R. Ali, M. J. Dunning, G. D. Brown, O. Gojis, I. O. Ellis, A. R. Green, S. Ali, S. F. Chin, C. Palmieri, C. Caldas, J. S. Carroll, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. **481**, 389–393 (2012).
54. M. Murtaza, S. J. Dawson, D. W. Y. Tsui, D. Gale, T. Forshew, A. M. Piskorz, C. Parkinson, S. F. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*. **497**, 108–112 (2013).
55. J. Cui, K. Germer, T. Wu, J. Wang, J. Luo, S. C. Wang, Q. Wang, X. Zhang, Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer Res.* **72**, 5625–5634 (2012).
56. S. Hole, A. M. Pedersen, S. K. Hansen, J. Lundqvist, C. W. Yde, A. E. Lykkesfeldt, New cell culture model for aromatase inhibitor-resistant breast cancer shows sensitivity to fulvestrant treatment and cross-resistance between letrozole and exemestane. *Int. J. Oncol.* **46**, 1481–1490 (2015).
57. L. Zhang, J. C. Er, H. Jiang, X. Li, Z. Luo, T. Ramezani, Y. Feng, M. K. Tang, Y. T. Chang, M. Vendrell, A highly selective fluorogenic probe for the detection and: In vivo imaging of Cu/Zn superoxide dismutase. *Chem. Commun.* **52**, 9093–9096 (2016).
58. K. D. Meyer, A. J. Donner, M. T. Knuesel, A. G. York, J. M. Espinosa, and D. J. Taatjes, Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *EMBO J.* **27**, 1447–57 (2008).
59. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).
60. W. A. Whyte, S. Bilodeau, D. A. Orlando, H. A. Hoke, G. M. Frampton, C. T. Foster, S. M. Cowley, R. A. Young, Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*. **482**, 221–225 (2012).
61. M. Rubinstein, R. H. Colby, *Polymer Phyics* (Oxford University Press, 2003).
62. M. Mittasch, P. Gross, M. Nestler, A. W. Fritsch, C. Iserman, M. Kar, M. Munder, A. Voigt, S. Alberti, S. W. Grill, M. Kreysing, Non-invasive perturbations of intracellular flow reveal physical principles of cell organization. *Nat. Cell Biol.* **20**, 344–351 (2018).
63. H. Zhao, P. H. Brown, P. Schuck, On the distribution of protein refractive index increments. *Biophys. J.* **100**, 2309–2317 (2011).
64. C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V Pappu, M. K. Rosen, Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell*. **63**, 72–85 (2016).
65. A. S. Holehouse, R. K. Das, J. N. Ahad, M. O. G. Richardson, R. V Pappu, CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **112**, 16–21 (2017).
66. M.-T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. Chih-Hsiung Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V Pappu, C. P. Brangwynne, Phase behaviour of disordered

proteins underlying low density and high permeability of liquid organelles. *nature.com* (2017), doi:10.1038/NCHEM.2803.
67. R. Delage-Mourroux, P. G. V. Martini, I. Choi, D. M. Kraichely, J. Hoeksema, B. S. Katzenellenbogen, Analysis of estrogen receptor interaction with a repressor of estrogen receptor activity (REA) and the regulation of estrogen receptor transcriptional activity by REA. *J. Biol. Chem.* **275**, 35848–35856 (2000).
68. Broad Institute Cancer Cell Line Encyclopedia (CCLE), (available at https://portals.broadinstitute.org/ccle).
69. cBioPortal for Cancer Genomics, (available at https://www.cbioportal.org/).

CHAPTER 4:

Discussion and Future Directions

Condensate dysregulation in disease

The compartmentalization of biomolecules – protein, RNA, and DNA – into biomolecular condensates involved in diverse cellular processes has completely transformed how we conceptualize cellular organization, function, pathology, and therapeutic strategies. Condensates provide a new layer of regulation with which to interrogate a host of longstanding mysteries across various fields of biology. Despite these exciting developments, condensate biology is still a burgeoning field, and many questions remain with respect to condensates in disease and therapeutic development. I have outlined the main questions in each area below, and I will elaborate further upon each topic throughout the discussion.

1) In Chapter 2, I show that amino acid sequence features of proteins can be utilized to link human genetics to condensate dysregulation. Genetic variation affecting the other biomolecules found within condensates, DNA and RNA, and other features of proteins, such as protein concentration, were not examined. What are the models with which genetic variation impacting RNA and non-coding DNA molecules or protein concentration could contribute to condensate dysregulation?

2) Several studies have now identified *associations* between dysregulated condensates and disease. Is there a causal relationship between altered condensate properties and disease? How can we develop assays to test this?

3) Looking beyond genetics, how do environmental factors, like diet or chemical exposures, contribute to condensate dysregulation in disease?

4) In Chapter 3, I show that small molecule cancer therapeutics selectively partition in specific condensates in the absence of their target based on their physicochemical properties. What are the implications of small molecule selective partitioning for pharmacokinetic profile more broadly?

5) How can we determine the chemical rules that govern small molecule selective partitioning?

6) Similar to small molecules, other therapeutic modalities, like RNA-based drugs, might selectively partition into specific condensates, how might this behavior play a role in their efficacy, toxicity, and design?

Amino acid sequence features and beyond

Protein amino acid sequence features: The focus of Chapter 2 was primarily on amino acid sequence features of proteins that have been shown to contribute to condensate formation, which we term condensate-promoting features. These features include structured modular interaction domains (MIDs) and disordered low complexity sequences (LCSs) in putative condensate-forming proteins. We find that pathogenic mutations spanning diverse disease classes affect these proteins features and therefore link condensate dysregulation as a potential pathogenic mechanism in a broad spectrum of human disease. Beyond condensate-promoting features, there are other crucial components involved in condensate formation, including non-coding DNA, RNA, and protein concentration. Changes to DNA molecules, RNA molecules, and protein expression

levels have all been linked to disease, and therefore it is worth discussing the ways in which genetic variation affecting each could contribute to condensate dysregulation.

RNA: Cases of disease mutations affecting RNA have already been linked to condensates dysregulation. In neurodegenerative diseases like ALS, repeat expansion mutations produce repeat-containing RNA transcripts which in turn form aberrant condensates in the nucleus, sequestering critical RNA-binding proteins away from their normal location and function (Jain and Vale, 2017). In many repeat expansion disorders, disease phenotypes manifest above a critical threshold number of nucleotide repeats and this threshold number correlates with the RNA transcripts produced forming condensates (Jain and Vale, 2017; Lee and McMurray, 2014). Increasing the number of repeats alters the condensate physical properties to more solid-like, implicating loss of condensate dynamics as a potential pathogenic mechanism (Jain and Vale, 2017). Only a handful of repeat expansion disorders have been linked to condensate dysregulation thus far through ectopic RNA condensates, but there are several more nucleotide repeat expansions that have yet to be characterized, causing ~20 different neurological diseases (La Spada and Taylor, 2010). Aberrant condensate dysregulation arising from RNA mutations may therefore be a broader mechanism for this class of diseases.

Beyond repeat expansion disorders, RNA in general is a central component of condensates and its sequence, length, structure, and modification can influence physical properties of condensates. Mutations altering any of these features could in theory modulate the condensate in which the RNA molecule resides. Predicting effects of mutations on RNA sequence and length is straight forward, but RNA structure prediction and identifying RNA modification sites has proven to be more challenging (Chen and Lin, 2016; D'Esposito et al., 2022; Schlick and Pyle, 2017; Schroeder, 2018). Better computational and experimental tools for RNA structure prediction and detecting sites of RNA modification will increase our ability to map, predict, and test the effects of RNA mutations on condensates.

Non-coding DNA: DNA is a component of several condensates, especially those which reside on the genome. Mutations in non-coding DNA regions of the genome, called enhancers, have been shown to drive tumorigenesis and may be associated with aberrant acquisition of condensates on chromatin. In tumor cells, oncogenes acquire clusters of enhancers called super-enhancers, containing myriad DNA binding sites for protein factors of the transcriptional apparatus. At super-enhancer driven genes, the increase in transcriptional apparatus local concentration causes the assembly of transcriptional condensates, which can drive increased oncogene expression (Cho et al., 2018; Sabari et al., 2018). In B cell lymphomas, the immunoglobin H (IgH) super-enhancer is translocated upstream of the MYC oncogene and thus recruits high levels of transcriptional machinery such as MED1, transcription factors, coactivators, and histone modifications, potentially forming a transcriptional condensate that is absent in a normal B cells. (Klein et al., 2011; Lovén et al., 2013). Acquisition of super-enhancers via DNA mutations affecting enhancers has been shown for other cancers as well (Mansour et al., 2014). Although there is no direct evidence for this model in B cell lymphomas, we do have evidence, that addition of transcription factor DNA binding sites upstream of genes above a threshold number can generate a switch-like activation of gene expression that correlates with transcriptional condensate formation (Shrinivas et al., 2019). It is possible that the acquisition of transcriptional condensates on the genome is a general feature of cancers driven by these types of chromosomal translocations. If so, then determining the threshold number of DNA binding sites needed in cells to drive transcriptional condensate formation and gene activation will allow for identification of the translocations that act via this condensate mechanism.

Protein concentration: Condensate formation is a concentration-dependent process and increases or decreases in normal expression levels of proteins have effects on condensate assembly. Overexpression of proteins is a key feature of many cancers and could affect condensates in a number of ways. For example, overexpression of a particular protein could 1) promote condensate assembly of a typically soluble proteins 2) alter the physical properties of liquid-like condensates to a more gel-like or solid-like state 3) cause ectopic condensate formation in inappropriate cellular locations and 4) allow for abnormal partitioning of the protein into condensates in which it does not normally reside. Several oncogenes, such as MYC, are upregulated in cancer types and associated with altered chromatin structure, transcription rate, metabolic pathways, and many other functions (Kuzyk and Mai, 2014; Lin et al., 2012; Peterson and Ayer, 2011). Does MYC form condensates and if so, how might MYC condensates be dysregulated in cancers with MYC overexpression?

In contrast to overexpression, some mutations cause a premature stop codon (a nonsense mutation) in protein-coding genes generating a truncated transcript that is degraded in a process called nonsense-mediated decay (NMD). Thus, NMD reduces the levels of protein expression. If the transcript encodes a condensate-forming protein, then this decrease in concentration of the protein may result in increased protein solubility and reduced condensate formation. Diseases associated with altered protein expression levels through overexpression or NMD should be broadly mapped and examined for their effects on condensates.

Overall, there is still a wealth of genetic variation that likely implicates condensates as a pathogenic mechanism in disease and the large-scale mapping and experimental approaches presented in Chapter 2 provide tools to do explore and test broad mutation types and condensate alterations.

Beyond disease associations

Thus far, the link between condensate dysregulation and disease has been principally associative. For diseases that are shown to be associated with condensates, it remains unknown whether the dysregulation of condensate properties or another effect on the protein (ablation of an interaction with a binding partner, reduced catalytic activity, etc.) is the key driver of the disease. Deconvolving these two mechanisms is difficult as they likely that have strong influences on one another. However, we can get a better measure of the contribution of condensate dysregulation to the disease pathophysiology through experimental studies. For example, in healthy cells, the protein UTX IDR is critical for UTX condensation and downstream tumor suppressor activity (Shi et al., 2021). A nonsense mutation in UTX (UTX Q555*) causes loss of UTX IDR and a decrease in UTX condensation *in vitro* and in cells. If condensate formation via the IDR is critical for tumor suppressor activity and loss of condensation is what drives the loss of the activity, as opposed to loss of a crucial protein binding partner or another protein function, then we might expect that replacing the UTX IDR with one from a condensate-forming protein with an entirely different function might rescue condensation and tumor suppressor activity. Replacing UTX IDR with an IDR from eIF4GII of similar length rescued UTX condensation and tumor suppressor activity (Shi et al., 2021). Several mutations tested in Chapter 2 are truncation mutations that primary affected LCSs within IDRs, such as BARD1 R406*, DAXX R318*, SALL1 S372*, BCL11A Q177*, BCOR Y657*, and TCOF Q55* (See Chapter 2, Figure 3). Rescue experiments, like those used in the study of UTX mutations, will be critical to disentangle the molecular scale and mesoscale properties that contribute to disease pathogenesis.

Beyond human genetics

Human genetics is a powerful link between condensate-promoting proteins and human disease, but it is not the only driver of disease—environmental factors also contribute to a wide variety of diseases. For example, poor diet and chemical exposures contribute to cellular oxidative stress – the accumulation of reactive oxygen species (ROS). Over time, prolonged high ROS levels can damage nearly every biomolecule and function in the cell. How might oxidative stress be impacting condensates to contribute to disease? As discussed in the introduction, it has already been shown that increased ROS levels have effects on insulin receptor (IR) condensates by decreasing IR mobility and thus decreasing IR signaling (Dall'Agnese et al., 2022). In plants, normal increases in ROS levels during development induce phase separation of a transcription factor via cysteine oxidation and the formation of disulfide bonds to promote homotypic TF interactions, altering the condensate physical properties. Although this example is under physiological conditions in plants, the principles of amino acid oxidation causing changes to condensate physical properties, potentially promoting phase transitions, can be applied to cases of oxidative stress in human disease. Outside of oxidative stress, environmental factors that damage biomolecules (UV exposure, ionizing radiation, increased cholesterol levels, etc.) may also have detrimental effects on condensate properties over time.

Condensates and therapeutic development

Selective partitioning of drugs, as shown in Chapter 3, is a powerful model to analyze and hypothesize how drugs interact with condensates and whether these interactions may provide insights for better drug design. In the drug design process, evaluating the drug pharmacokinetic profile is necessary to understand drug efficacy and safety. In this next section, I discuss the models by which condensate partitioning might affect the pharmacokinetic profile of small molecule drugs.

Small molecule condensate partitioning and pharmacokinetic profile: Efficacy and toxicity

A key insight from Chapter 3 is that the co-partitioning of a drug and its target in a condensate can increase drug potency and efficacy. Given this insight, it will be important to generate small molecules that specifically target a condensate of interest, doing so through the addition of condensate-specific chemical moieties. One way to discover these moieties would be through small molecule-based screens in cells in which the condensate and small molecules are fluorescently labeled. Utilizing condensates as a platform to increase potency and by extension efficacy of small molecules will be an important extension of this work.

Just as co-partitioning of drugs and their targets in the same condensate can increase potency, partitioning of drugs in an incorrect condensate in which the target does not reside could decrease drug potency and increase toxicity. Small molecule partitioning in the incorrect condensate could sequester the drug away from its desired target, or it could modulate the condensate physical properties, doing so by specifically binding an off-target protein or altering the condensate microenvironment to cause aberrant phase transitions. Small molecules could also *de novo* promote ectopic condensate formation if they had off-target interactions with condensate-forming proteins. All of these scenarios are contributors to potential drug toxicity—limiting these off-target effects would be crucial in the drug design process.

To optimize the potency, efficacy and toxicity of drugs that interact with condensates, utilizing rational design would be highly beneficial, however there is currently not enough known about

physicochemical properties governing drug partitioning. Further drug-condensate screens will be important to determine these chemical rules.

Determining chemical grammar of condensates

Determining the chemical rules or "grammar" with which small molecules selectively partition into some condensates could aid in rational drug design. As discussed, condensates regulating different cellular processes are comprised of different protein and nucleic acid molecules, and it has been proposed that condensate composition can dictate its internal chemical microenvironment (Kilgore and Young, 2022). As shown in Chapter 3, the small molecule cisplatin selectively partitions in MED1 condensates and this partitioning behavior is dependent on aromatic amino acids present in the MED1 IDR, implicating π-π or cation-π interactions in MED1-cisplatin interactions. For these types *in vitro* experiments with fluorescently labeled condensate wildtype scaffold proteins or proteins with particular amino acid subtypes are highly useful as one can assess the partitioning behavior of small molecules in the presence or absence of certain interactions. In theory, large screens can be done with this type of model system across different types of condensates and large small molecule libraries. One limitation in this approach is the simplicity of the *in vitro* condensate, a single purified protein, as compared to a more complex intracellular condensates, comprised of 100s to 1000s of different biomolecules. The microenvironment of the *in vitro* condensate is almost certainly different than that in the cell. In this regard, computational tools to determine, in a natural context, what biomolecules partition into condensates and if there are any characteristic amino acid patterns or interactions present on the biomolecules that may dictate the condensate chemical microenvironment will be very important. These chemical patterns could then be applied to rational drug design. For example, if charge blocks are characteristic of IDRs that are specific to active transcriptional condensates, then generating charge block peptides and conjugating them to drugs that should be localized to transcriptional condensates may aid in drug potency.

Beyond small molecules

Condensates can provide insights for other drug modalities such as nucleic-acid based drugs. Specifically, RNA-based drugs (ASOs, siRNAs, mRNAs, etc.) are an emerging class of therapeutics that have proven to be effective across diverse disease types.

RNA is a central feature of condensates, and its sequence, length, structure, and modification can have impacts of condensate physical properties and function (Roden and Gladfelter, 2021). RNA can also interact with other RNAs or proteins to aid in condensate assembly and regulation. Utilizing condensates in RNA-based drug design is similar in principle to how condensates can inform small molecule drug design. RNA-based drugs could aberrantly concentrate in condensates without their RNA target or modulate condensate physical properties. We already have evidence that build-up of low concentration of RNA can promote protein condensate formation, and at high concentrations disrupt those same condensates – a phenomenon known as re-entrant phase separation. If large amounts of RNA-based drugs build up in condensates, could it be causing a re-entrant phase transition? What effect would this have on its efficacy if the condensate is necessary for finding its target? In the cell, RNA-based drugs may also aberrantly interact with RNA-binding proteins to promote ectopic condensate assembly. In fact, it has already been shown that antisense oligonucleotides (ASOs) with a phosphorothioate backbone (PS-ASOs) modification promote ectopic formation of condensates through binding to paraspeckle proteins, resulting in higher toxicity and lower efficacy (limited therapeutic window) (Shen et al., 2014). 2'-O-methyl (MOE) chemical modification to the PS-ASO reduces paraspeckle protein binding and aberrant condensate formation, thus lowering toxicity (Shen et al., 2019). Insights

into the chemical modifications on RNA that promote or disfavor aberrant condensate formation could aid in rational design of ASOs, a class of drugs that has proven to be efficacious but often suffers from limited therapeutic window (Gagliardi and Ashizawa, 2021; Roberts et al., 2020).

Concluding thoughts

The emerging field of biomolecular condensates has yielded numerous insights for human disease and drug design. When assessing pathogenic mechanisms of disease mutations, researchers will now need to expand their models from lock-and-key, stoichiometric interactions to models that include effects on non-stoichiometric, large assemblies of biomolecules. When designing new targeted therapies, drug developers will not only need to take into account the physicochemical microenvironment of condensates, but also to consider condensates themselves as viable drug targets. In order to best apply condensate principles to disease and drug design, more still needs to be known about the rules governing biomolecule partitioning into condensates, condensate chemical microenvironment, and predicting effects of mutations or environmental factors on condensate physical properties, such as viscosity, viscoelasticity, and surface tension. Pushing forward the field of condensate biology over the next decade will continue to require the work of physicists, chemists, biologists, and computational scientists alike to develop novel biological insights, technological innovations, and life-saving treatments for patients.

References

Chen, W., Lin, H., 2016. Recent Advances in Identification of RNA Modifications. Noncoding RNA 3, 1. https://doi.org/10.3390/ncrna3010001

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., Cisse, I.I., 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science 361, 412–415. https://doi.org/10.1126/science.aar4199

Dall'Agnese, A., Platt, J.M., Zheng, M.M., Friesen, M., Dall'Agnese, G., Blaise, A.M., Spinelli, J.B., Henninger, J.E., Tevonian, E.N., Hannett, N.M., Lazaris, C., Drescher, H.K., Bartsch, L.M., Kilgore, H.R., Jaenisch, R., Griffith, L.G., Cisse, I.I., Jeppesen, J.F., Lee, T.I., Young, R.A., 2022. The dynamic clustering of insulin receptor underlies its signaling and is disrupted in insulin resistance. Nat Commun 13, 7522. https://doi.org/10.1038/s41467-022-35176-7

D'Esposito, R.J., Myers, C.A., Chen, A.A., Vangaveti, S., 2022. Challenges with simulating modified RNA: Insights into role and reciprocity of experimental and computational approaches. Genes (Basel) 13, 540. https://doi.org/10.3390/genes13030540

Gagliardi, M., Ashizawa, A.T., 2021. The challenges and strategies of antisense oligonucleotide drug delivery. Biomedicines 9, 433. https://doi.org/10.3390/biomedicines9040433

Jain, A., Vale, R.D., 2017. RNA phase transitions in repeat expansion disorders. Nature 546, 243–247. https://doi.org/10.1038/nature22386

Kilgore, H.R., Young, R.A., 2022. Learning the chemical grammar of biomolecular condensates. Nat Chem Biol 18, 1298–1306. https://doi.org/10.1038/s41589-022-01046-y

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., Casellas, R., Nussenzweig, M.C., 2011. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell 147, 95–106. https://doi.org/10.1016/j.cell.2011.07.048

Kuzyk, A., Mai, S., 2014. c-MYC-induced genomic instability. Cold Spring Harb Perspect Med 4, a014373. https://doi.org/10.1101/cshperspect.a014373

La Spada, A.R., Taylor, J.P., 2010. Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat Rev Genet 11, 247–258. https://doi.org/10.1038/nrg2748

Lee, D.-Y., McMurray, C.T., 2014. Trinucleotide expansion in disease: why is there a length threshold? Curr Opin Genet Dev 0, 131–140. https://doi.org/10.1016/j.gde.2014.07.003

Lin, C.Y., Lovén, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., Young, R.A., 2012. Transcriptional amplification in tumor cells with elevated c-Myc. Cell 151, 56–67. https://doi.org/10.1016/j.cell.2012.08.026

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., Young, R.A., 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 153, 320–334. https://doi.org/10.1016/j.cell.2013.03.036

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., Loh, M.L., Hunger, S.P., Sanda, T., Young, R.A., Look, A.T., 2014. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science 346, 1373–1377. https://doi.org/10.1126/science.1259037

Peterson, C.W., Ayer, D.E., 2011. An extended Myc network contributes to glucose homeostasis in cancer and diabetes. FBL 16, 2206–2223. https://doi.org/10.2741/3848

Roberts, T.C., Langer, R., Wood, M.J.A., 2020. Advances in oligonucleotide drug delivery. Nat Rev Drug Discov 19, 673–694. https://doi.org/10.1038/s41573-020-0075-7

Roden, C., Gladfelter, A.S., 2021. RNA contributions to the form and function of biomolecular condensates. Nat Rev Mol Cell Biol 22, 183–195. https://doi.org/10.1038/s41580-020-0264-6

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., Li, C.H., Guo, Y.E., Day, D.S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T.I., Cisse, I.I., Roeder, R.G., Sharp, P.A., Chakraborty, A.K., Young, R.A., 2018. Coactivator condensation at super-enhancers links phase separation and gene control. Science 361, eaar3958. https://doi.org/10.1126/science.aar3958

Schlick, T., Pyle, A.M., 2017. Opportunities and challenges in RNA structural modeling and design. Biophys J 113, 225–234. https://doi.org/10.1016/j.bpj.2016.12.037

Schroeder, S.J., 2018. Challenges and approaches to predicting RNA with multiple functional structures. RNA 24, 1615–1624. https://doi.org/10.1261/rna.067827.118

Shen, W., De Hoyos, C.L., Migawa, M.T., Vickers, T.A., Sun, H., Low, A., Bell, T.A., Rahdar, M., Mukhopadhyay, S., Hart, C.E., Bell, M., Riney, S., Murray, S.F., Greenlee, S., Crooke, R.M., Liang, X., Seth, P.P., Crooke, S.T., 2019. Chemical modification of PS-ASO therapeutics reduces cellular protein-binding and improves the therapeutic index. Nat Biotechnol 37, 640–650. https://doi.org/10.1038/s41587-019-0106-2

Shen, W., Liang, X., Crooke, S.T., 2014. Phosphorothioate oligonucleotides can displace NEAT1 RNA and form nuclear paraspeckle-like structures. Nucleic Acids Res 42, 8648–8662. https://doi.org/10.1093/nar/gku579

Shi, B., Li, W., Song, Y., Wang, Z., Ju, R., Ulman, A., Hu, J., Palomba, F., Zhao, Y., Le, J.P., Jarrard, W., Dimoff, D., Digman, M.A., Gratton, E., Zang, C., Jiang, H., 2021. UTX condensation underlies its tumor-suppressive activity. Nature 597, 726–731. https://doi.org/10.1038/s41586-021-03903-7

Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A.V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., Chakraborty, A.K., 2019. Enhancer features that drive formation of transcriptional condensates. Mol Cell 75, 549-561.e7. https://doi.org/10.1016/j.molcel.2019.07.009