Insights from biomolecular condensates into disease and drug development

By

Lena K. Afeyan

B.A., Molecular and Cellular Biology
Harvard University, 2017

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Signature of Author: ____

Department of Biology
March 18, 2022

Certified by: ____

Richard A. Young
Member, Whitehead Institute
Professor of Biology
Thesis Supervisor

Accepted by: ____

Amy Keating
Professor of Biology
Co-Chair, Biology Graduate Committee

Insights from biomolecular condensates into disease and drug development

By

Lena K. Afeyan

Submitted to the Department of Biology on May 18
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the
Massachusetts Institute of Technology

ABSTRACT

The cell concentrates and compartmentalizes proteins and nucleic acids into diverse phase-separated biomolecular condensates. The study of condensates has yielded myriad fundamental insights into cell biology, ranging from a better understanding of cellular organization, to the exploration of novel mesoscale functions resulting from the emergent properties of these liquid-like compartments. A consideration of the role of condensates in disease and drug development could facilitate similar paradigm shifts, with early evidence suggesting that condensate dysregulation may be a feature of many diseases, and that therapeutics can modulate condensates. In the studies presented in this thesis, we developed and validated a strategy for nominating patient mutations across the spectrum of disease that may cause condensate dysregulation, providing nominated mutations as a resource to the biomedical community for the acceleration of the study of condensates in disease. Further, we tested the hypothesis that small molecule therapeutics can concentrate into condensates, showing that clinically important cancer therapeutics display differential partitioning and that condensate partitioning can affect therapeutic activity (Klein et al., 2020). Lastly, we have expanded upon this condensate partitioning work to show that antisense oligonucleotides (ASOs), nucleic acid-based therapeutics targeting RNA, partition into and modulate certain condensates, and that specific chemical modifications can alter this partitioning behavior. Ultimately, by considering the implications of a condensate model in disease and drug development, this thesis aims to leverage recent insights into biomolecular condensates to facilitate the development of novel disease mechanistic and therapeutic hypotheses.

Thesis Supervisor: Richard A. Young
Title: Member, Whitehead Institute; Professor of Biology

DEDICATION

To my parents, the most brilliant people I know, thank you for everything.

ACKNOWLEDGEMENTS

There are many people I would like to thank for their support and guidance during my PhD. First, I would like to thank Rick Young, my PI, for allowing me to join his lab and to work on science I was inspired and motivated by throughout my PhD. Rick is a visionary scientist, and I have been incredibly lucky to learn from him about how to find and go after big, paradigm-shifting questions in science, how to communicate my science to the world, and how to lead scientific teams with rigor and optimism. I would also like to thank Tony Lee, who has been a fantastic mentor and friend in the Young lab and on the volleyball court. Tony is a brilliant scientific thinker and has gone out of his way many times to create opportunities for me to learn and to advocate for me. I am also grateful to many members of the Young lab for helping me become a better scientist and for bringing your curiosity and intellect to all of our conversations. Specifically, I am extremely thankful to Isaac Klein, who has been among the most influential mentors in my life and is one of the most impressive scientists I have had the honor to learn from. Thank you to Susana and Ale, whose friendship in lab and support at and beyond the bench has been so important throughout my time in the Young lab. Thank you, as well, to Jon and Victoria for great discussions and fun outings, and to Nancy and Mary Anne for their support. I am also thankful to my friends in the MIT Biotech Group, which has been a big part of my graduate experience at MIT, for all the great times and for all we built together.

I am also grateful to my thesis committee. Thank you to Phil Sharp, Angela Koehler, Eliezer Calo and Richard Kriwacki for dedicating your time to serve on my committee. Presenting to, answering questions from, and discussing science with Phil Sharp, whose deep knowledge of scientific discoveries over several decades is remarkable, has been one of the highlights of my scientific career. Angela has also offered her unique perspective and encouragement in and beyond the committee, including her support of the biotech group. Eliezer has also been a very generous and thoughtful collaborator.

To my best friends- for the countless calls, dinners, drinks, talks, hugs, and festivities throughout my PhD to pick me up in the tough times and to celebrate the victories- thank you and I am so lucky to get to experience all the big life moments to come with my team. A special thanks to Lilit- for a lifetime of unconditional love, loyalty, and fierce friendship, to Lori- for providing my life with so many adventures, laugher-filled memories and a space to talk about anything and everything, to Rose- for rushing to my side when I needed you most and listening with your heart, to Chelsea- for your outstanding patience, scheming, and the many ways you've always had my back, and to Monica- for your brilliant insights, intensity, and vulnerability as we problem solve life together.

Lastly and most importantly, a huge thank you to my family, whose strength, passion, and spirit inspire me every day. Dad- you are the most innovative and inspirational person in my life- thank you for believing in me to an almost unreasonable degree. Mom- you are the wisest person I know- thank you for helping me keep perspective and balance, and for making me feel seen and cared for every day. Taleen- you showed up for me through this PhD with your unparalleled empathy in a way that I'll forever be grateful for- I owe so much of my happiness and pride in writing this to you. Armine- thank you for creating so many wonderful moments for me and all of us with your stunning vision for memorable activities, and for your advice along the way. Alex- thanks for the check-ins, harmonies, hours listening during car rides, scientific chats, sets, and spikes. To Mormor (and Morfar, whose presence and humor are sorely missed), for your inspirational optimism and the bright light you shine on my life and the whole family, tack. Also thank you to my amazing cousins, aunts, and uncles, in and beyond Boston, for cheering me on and making me feel so loved- especially to my cousins Sandra and Raffi for their sage advice and for bringing their kids- Zaven, Alek, and Anna- into my life.

STATEMENT ON WORK PRESENTED

Chapter 1

I wrote Chapter 1 with feedback from Rick Young and Tony Lee. Several figures were modified from figures published in referenced papers. Figure 4 was developed in collaboration with a graphic designer working with the Whitehead Institute, Caitlin Rausch.

Chapter 2

I contributed to the conceptualization, methodology, and investigation in analyses presented in all figures. I performed imaging analyses presented in Figure 3 and developed graphics presented in Figures 1 and 4. I wrote the manuscript, in collaboration with Salman Banani and Susana Wilson Hawken, with edits from Rick Young, Tony Lee, and other authors.

Chapter 3

I performed experiments in Figures 3 and 4, and Supplementary Figures S16, S17, S18, S19, S20, S24, and S29. The manuscript was written primarily by Isaac Klein and Ann Boija, with input, including edits and descriptions of experimental results and methods relevant to experiments I performed, from me and other authors.

Chapter 4

I performed all experiments presented in the main figures and supplemental figures of this Chapter, except the experiment presented in Figure S1C. Quantitative and computational analyses of the results of these experiments were done by me and Susana Wilson Hawken. I generated all graphics in BioRender, except the graphic in Figure 1A, which was adapted from a figure generated by Jon Henninger. I wrote Chapter 4, with edits on methods from Susana Wilson Hawken and feedback from Rick Young.

Chapter 5

I wrote Chapter 5.

TABLE OF CONTENTS

CHAPTER 1:
INTRODUCTION

Overview

The billions of protein molecules in cells, as well as nucleic acids deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), are organized into both membrane-bound and non-membrane-bound organelles (Shin and Brangwynne, 2017). These organelles facilitate diverse cellular functions in the cytoplasm and nucleus of the cell. We have recently come to appreciate that many non-membrane-bound, or “membraneless” organelles, are phase-separated biomolecular condensates with distinct physicochemical properties that facilitate the concentration and compartmentalization of specific proteins and nucleic acids, and display liquid-like behaviors such as dynamic internal rearrangement and coalescence (Banani et al., 2017; Hyman et al., 2014).

Studies of phase-separated condensate formation, integrating theories of polymer physics, nonequilibrium thermodynamics, advanced microscopy techniques, and molecular biology, have yielded fundamental insights into the organization and dynamic regulation of cellular processes (Hyman et al., 2011; Brangwynne et al., 2015; Holehouse and Pappu 2018, Choi et al., 2020). In 2009, a pioneering study on P granules in *Caenorhabditis elegans* from the lab of Anthony Hyman demonstrated that these assemblies display liquid properties and proposed that phase transitions may provide a fundamental mechanism for organizing cellular processes (Brangwynne et al., 2009). The study of condensates has, since then, revolutionized cell biology, with nearly every known cellular process now having been shown to involve condensate formation in some way (Banani et al., 2017). Such processes include 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 (Lyon et al., 2021).

Condensates endow mesoscale properties that have provided novel models and explanations for longstanding mysteries in fields such as transcription, germ-cell specification, and ribosome biogenesis (Lyon et al., 2021). In other words, the study of the properties of these biomolecules when they come together on the scale of thousands of molecules has yielded novel insights not captured by the study of individual protein molecules and stoichiometric binding events (Shin and Brangwynne 2017).

Analogously, these new principles have profound potential implications for disease and drug development (Alberti and Hyman, 2021; Tsang et al., 2020; Boija et al., 2021). In this thesis, I will describe our efforts to begin exploring these implications and how they can be exploited to improve therapies for patients. Studies of how condensates are altered in disease have already begun to offer new models for disease mechanisms emerging from changes of mesoscale properties in the cell. Similarly, understanding how therapeutics such as small molecules and nucleic acid-based therapies interact with condensates- including whether they are able to enter specific condensates over others and whether they can be used to modulate condensate properties or disrupt condensates- could be critical for designing more efficacious drugs and understanding why certain drugs may fail unexpectedly.

In Chapter 2 of this thesis, I describe how we computationally created and experimentally validated a resource for predicting how protein-coding patient mutations may affect specific condensates in disease. In this work, I and my co-authors defined a set of condensate-forming proteins, mapped sequence features identified from many studies that promote condensate formation, and overlayed pathogenic patient mutations onto this map. We used this map to predict which mutations are likely to cause condensate dysregulation, tested these predictions in live cells, and present a catalog of nominated mutations as a resource to the biomedical community.

Shifting focus to drug development, in Chapter 3, I focus on our efforts to study the partitioning behavior of small molecule therapeutics into biomolecular condensates (Klein et al., 2020). Specifically, we hypothesized that, much like proteins and nucleic acids concentrate preferentially into specific condensates based upon the physicochemical environment of these compartments, small molecule therapeutics may also display differential partitioning. Such partitioning behavior could drastically alter the therapeutic activity of these small molecules and may thus be a key consideration in the development and improvement of therapies.

Building upon this work, Chapter 4 describes our study of the potential effect of condensate partitioning behavior of nucleic acid-based therapies such as antisense oligonucleotides (ASOs) on the therapeutic index of these drugs. As nucleic acids are key components of biomolecular condensates, we hypothesized that these drugs were likely to partition into condensates and that this partitioning may affect both the efficacy and toxicity of this class of drugs. Our work has suggested that ASOs do indeed partition into diverse biomolecular condensates and that modifications that affect their therapeutic index also alter their condensate partitioning behavior.

As we now know that many components of condensates are mutated in disease and that many drug targets are found within biomolecular condensates, the study of condensates is likely to facilitate myriad novel disease mechanistic and therapeutic hypotheses. To conclude this thesis, in Chapter 5, I discuss these and other ways in which an understanding of biomolecular condensates may shift paradigms in the study of disease and in the development of therapeutics.

Organization of biomolecules in the cell into organelles

Cells must organize the billions of proteins and nucleic acid molecules they contain such that specific proteins and nucleic acids can carry out complex, often multi-step, processes with precise spatiotemporal control (Hyman and Brangwynne, 2011).

Membrane-bound organelles

The eukaryotic cell contains several lipid membrane-bound organelles that compartmentalize key functionalities. For example, the nucleus of the cell is enveloped by a double membrane in eukaryotic cells called the nuclear membrane, which separates the chromosomes of the cell from the cytoplasm. Other organelles include the mitochondria, which contain outer and inner membranes and facilitate cellular respiration, the membranous endoplasmic reticulum (ER), which plays roles in the production of proteins (rough ER) and lipids (smooth ER), and the Golgi apparatus, which is composed of folded membranes that facilitate protein modifications (Alberts et al., 2014).

In each of these organelles, membranes play a variety of roles in organizing the molecules that carry out these functions. The nuclear membrane, for example, is selectively permeable to certain molecules, and requires active transport of others into and out of the nucleus through nuclear pores (DeMagistris and Antonin, 2018; Alberts et al., 2014). This barrier is thought to function in part to protect the genetic information of the cell from molecules that could damage it. Also, by separating processes such as transcription and translation by keeping many of their components in distinct compartments, namely the nucleus and the cytoplasm, the nuclear membrane provides additional spatiotemporal organization in these key cellular processes (DeMagistris and Antonin, 2018).

Membraneless organelles

The cell also contains organelles that are not encapsulated by a membrane. These “membraneless organelles” can contain thousands of molecules and range in size (Banani et al., 2017). The nucleolus is the longest-studied and most prominent membraneless organelle, first visualized with light microscopy over 200 years ago by Fontana, Valentin, and Wagner, taking up as much as 25% of the nucleus and playing key roles in ribosome biogenesis (Pederson, 1998; Pederson, 2011). Other examples of membraneless organelles include nuclear speckles, which are involved in RNA splicing, and stress granules, which store a variety of RNAs and proteins in response to cell stress. Although these compartments in the cell could be visualized by microscopy, until recently, how they were formed, held together, and kept separate from the rest of the nucleus and cytoplasm in the absence of a membrane was not well understood.

In 2009, a study from the lab of Anthony Hyman proposed that phase transitions may represent a fundamental physicochemical mechanism for organizing cellular processes into membraneless organelles (Brangwynne et al., 2009). Phase transitions in the cell can include liquid-liquid phase separation (LLPS), a process by which molecules condense out of the cellular milieu to form a distinct dense phase and a dilute phase, as well as liquid-to-solid transitions, resulting in the formation of gels, sols, or suspensions (Hyman and Brangwynne, 2011). Phase separation has provided a predictive biophysical framework for understanding the formation of membraneless organelles in which millions of biomolecules are organized and diverse cellular processes are carried out (Choi et al., 2020; Lyon et al., 2021).

Biomolecular condensates across the cell

Phase transitions in cell biology: early studies and re-emergence

The study of phase separation as a driving force for organization of the cell, while the topic of intense study in the last decade, builds upon many theoretical and experimental works in the late 19th and early 20th centuries (Graham, 1861; Wilson, 1899; Hardy, 1899; Leduc, 1911; Gibbs, 1961). Studies utilizing polymer physics to examine the behavior of biological macromolecules such as albumin and globulin began to model the cytoplasm of the cell as a colloid, namely particles suspended in a liquid, in which colloidal phase separation provided molecular organization. Similarly, scientists such as E.B. Wilson and others viewed the cell as a densely packed emulsion of liquid protein phases or “coascervates” (Wilson, 1899)

In the mid 20th century, advances in structural biology shifted the focus of cell biologists toward lock and key or stoichiometric protein interactions and the propensity of proteins to form solid crystals as a means to extract atomic level information on these macromolecules. This shift reducd the pace of progress in the study of biopolymers as colloids capable of phase separation (Haas, 1999; Piazza, 2000).

Advances in confocal microscopy in the late 20th century led to the visualization of several non-membrane-bound organelles in the cell, referred to as "puncta" or "bodies" and questions about the principles that governed their formation inspired the re-emergence of the study of phase transitions in the cell. These membraneless organelles or assemblies have been termed biomolecular condensates (Figure 1) (Banani et al., 2017).

P granules and a phase separation mechanism for the formation of membraneless organelles

The first of many studies to revisit principles of phase transitions in the cell focused on the formation of germ granules in *Caenorhabditis elegans* (*C. elegans*), called P granules. In 2009, P granules were known to mediate germ-cell specification in *C. elegans* embryos, but the mechanism by which they dissolved and formed in the posterior of the embryo was not well understood. Brangwynne et al. visualized these granules, formed by RNA and RNA-binding proteins, by tagging two constitutive P granule components in one-cell embryos with green fluorescent protein (GFP). Three-dimensional (3D) particle tracking by fluorescence microscopy indicated that P granules appeared to transition between a soluble and a condensed phase and, when they reattached to the nucleus in the four-cell stage, appeared similar to liquid drops wetting a surface, occasionally fusing (Brangwynne et al., 2009).

These observations inspired the hypothesis that rapid molecular rearrangements of the RNA and the RNA-binding proteins in these P granules could give rise to the formation of liquid-like compartments. To test this hypothesis, Brangwynne and colleagues performed experiments in which they applied shear stresses across large nuclear-associated P granules and showed that P granules exhibit classic liquid behaviors, including coalescence. Further, photobleaching part of these P granules and observing the recovery time of GFP-tagged P granule proteins indicated that the fluorescence recovery occurred on a rapid time scale consistent with the rapid rearrangements of molecules in liquids such as glycerol and colloidal liquids. These and other experiments led to the proposal that P granule formation and localization can be regulated by altering the saturation concentration, namely the concentration above which condensation occurs, in the posterior of the embryo via specific polarity proteins, including MEX-5 and PAR-1 (Brangwynne et al., 2009).

Biomolecular condensates in the organization of diverse cellular processes

More broadly, these pioneering experiments suggested that phase transitions in the cell may provide a fundamental mechanism for spatiotemporally organizing molecules into membraneless organelles. Above the saturation concentration (Csat), proteins and nucleic acids could form two separate phases, with dynamic rearrangement of biomolecules within and across these phases, without needing a membrane to separate the dense and soluble phases (Figure 2A) (Hyman et al., 2014).

Biomolecular condensates, previously termed "assemblies", formed by LLPS are dynamic, liquid-like organelles made up of proteins and nucleic acids. Condensates range in size from <100nM (transcriptional condensates, human colon cancer cells) to >10uM (nucleolus, *Xenopus laevis*) and can contain 100s-1000s of molecules (Forman-Kay et al., 2022; Cho et al., 2018; Brangwynne, 2013). Condensate formation via LLPS occurs in conditions of molecular supersaturation facilitated by changes in protein/ RNA concentrations, charge state, salt/ proton concentration, or temperature (Choi et al., 2020). Several previously visualized condensates have been shown to form via liquid-liquid phase separation, including the nucleolus, stress granules and DNA damage repair sites (Brangwynne et al., 2011; Patel et al., 2015; Altmeyer et al., 2015).

The study of condensate formation in the cell has also begun to provide models and explanations for longstanding mysteries in processes not previously thought to occur in distinct organelles. Transcription is an example of such a process. This fundamental cellular process had previously been studied in great detail to identify individual components that came together stoichiometrically to transcribe a gene into mRNA. However, at key cell identity genes and oncogenes, several distal regulatory elements and hundreds of molecules were seen to come together into super-enhancers, which was not seemingly consistent with previous stoichiometric studies of transcription (Hnisz et al., 2017). These observations are consistent with a condensate model by which transcription factors bind to regulatory elements and concentrate and compartmentalize the many proteins required to initiate and carry out transcription via liquid-liquid phase separation into a condensate (Sabari et al., 2018; Cho et al., 2018; Boija et al., 2018; Guo et al., 2019).

The list of condensate-associated processes is continually growing, and includes transcription, DNA replication, DNA repair, heterochromatic and euchromatic organization, RNA biosynthesis and homeostasis, and ribosome biosynthesis, protein quality control, innate immunity, cell division, cell-cell adhesions, signaling, and synaptic transmission (Figure 1) (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).

Principles of condensate formation by phase separation

Forces driving phase separation

An LLPS model for condensate formation provides a framework allowing for formalisms and models in soft matter physics and polymer chemistry to be applied to yield insights in cell biology (Choi et al., 2020). At a simple thermodynamic level, LLPS in the cell occurs when the enthalpic forces of intra- and inter-molecular interactions of macromolecules drive the formation of two distinct phases, a dense phase and a dilute phase, overcoming the enthalpy of interactions between macromolecules and water, as well as entropic forces favoring one higher entropy phase (Figure 2B) (Brangwynne et al., 2015). This balance can be summarized by the Chi parameter, a molecular interaction parameter helpful for modeling the variables upon which LLPS depends using the Flory Huggins and Cahn-Hilliard equations (Flory, 1941; Huggins, 1941; Cahn, 1958). The saturation concentration of a phase-separating system can depend upon a variety of variables, including interaction strength, or pH, salt concentration, and temperature, which can affect the interactions of biomolecules (Choi et al., 2020).

Figure 1

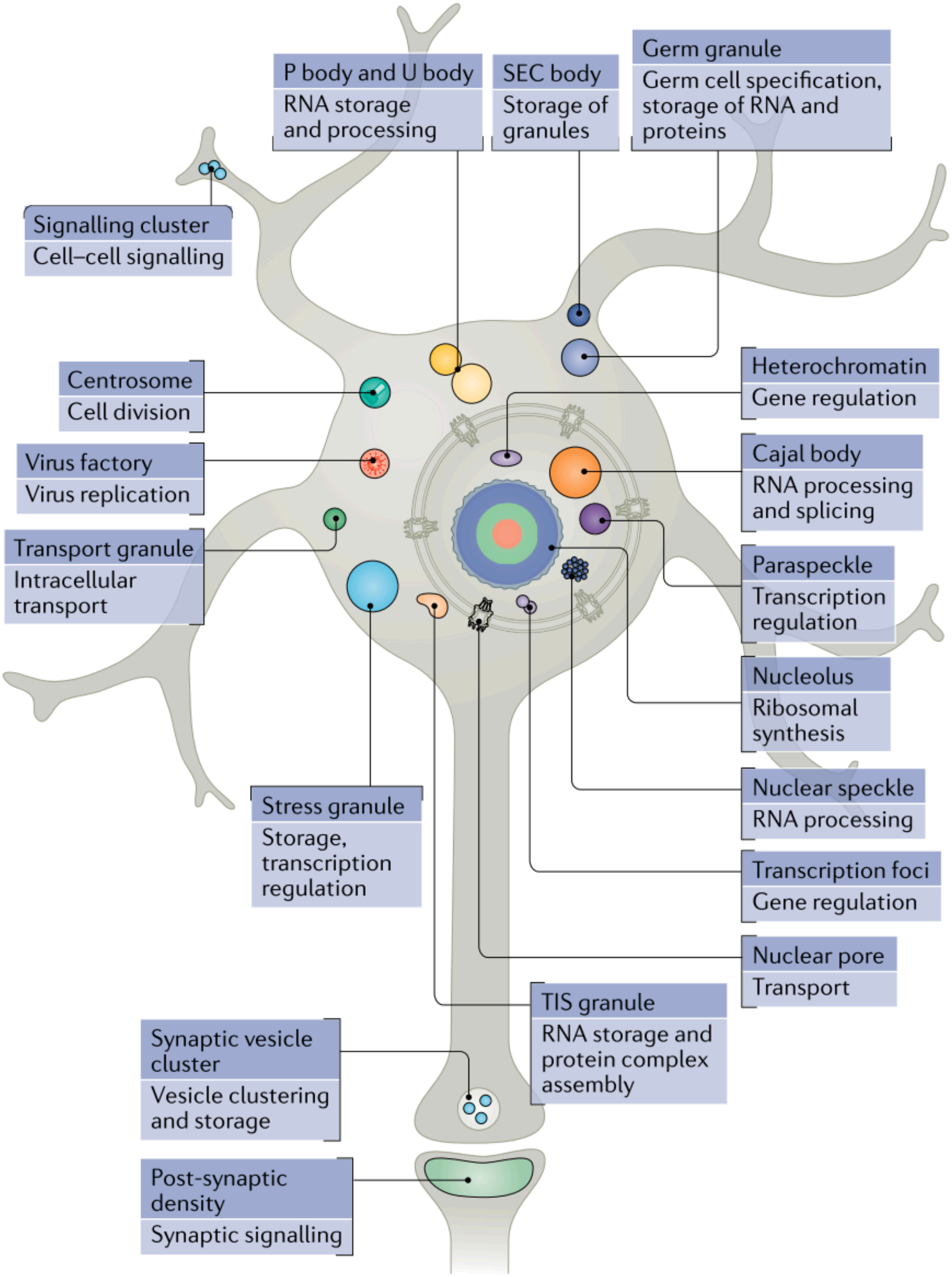


Figure 1. Biomolecular condensates organize diverse cellular processes. (A) Depiction of biomolecular condensates across the cell and the processes they facilitate. Adapted from Alberti and Hyman, 2021.

LLPS depends on weak, multivalent interactions between biomolecules that favor the formation of higher order assemblies (Banani et al., 2017). These interactions can include hydrophobic, pi-pi, pi-cation, and electrostatic interactions, especially in condensates containing RNA (Figure 3A) (Pak et al., 2016). Multivalent biomolecules are those that contain multiple sites of interaction, such as multiple binding motifs on an RNA or multiple RNA-binding domains in an RNA-binding protein (RBP) (Hyman et al., 2014). A study of the actin-regulatory signaling pathway demonstrated the importance of multivalency in the condensate formation of proteins. In this study, Li et al. showed that multivalency- specifically that of Src homology 2 (SH2) domains in Nck interacting with phospho-tyrosine residues in neprhin, as well as SH3 domains in Nck and proline-rich motifs in N-WASP- can shift Csat (Li et al., 2012). Oligomers resulting from highly multivalent interactions have markedly decreased solubility, which promotes LLPS and condensate formation (Hyman et al., 2014).

Multivalent interacting elements are often interspersed in the protein sequence or 3D fold of a protein between elements or regions of the protein that do not significantly drive attractive interactions. A sticker-and-spacer model, derived from principles of polymer physics used to describe associative polymers, has provided a framework to describe the contribution of various sequence elements, terming interacting elements as stickers, and non-interacting elements as spacers (Choi et al., 2020). Conformational flexibility of spacers, as is found in intrinsically disordered regions (IDR) of many proteins, combined with the multivalency of stickers in these regions, have been found to promote phase separation. Consistent with this observation, intrinsically disordered regions, namely those that do not adopt an ordered stable three-dimensional structure, are enriched within phase-separating proteins (Nott et al., 2015; Choi et al., 2020).

The protein components of condensates have traditionally been classified as either scaffold proteins or client proteins (Banani et al., 2016). Scaffold proteins are those that drive condensate formation and without which the condensate would not form. Client proteins are those that preferentially partition into the condensate based upon the physicochemical environment of the condensate (Ditlev et al., 2018).

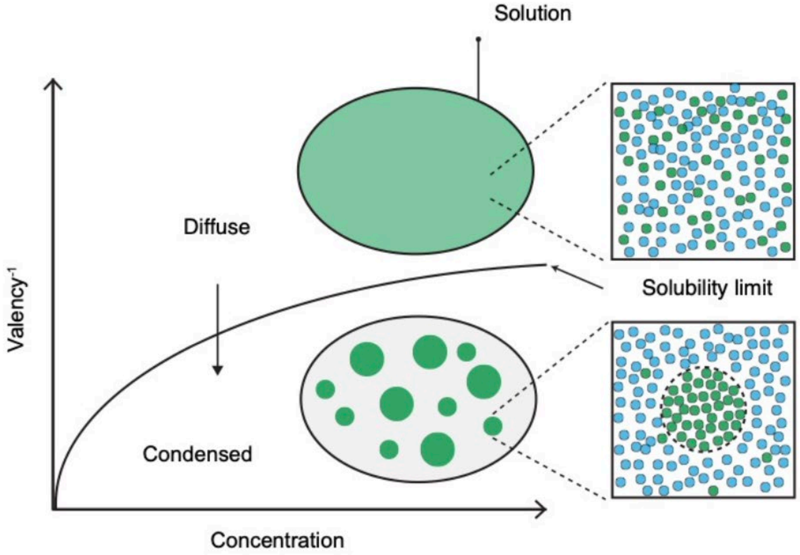
While modeling condensates as liquid compartments can provide helpful formalisms, it is important to consider the assumptions underlying existing equations, many of which are meant to describe systems in equilibrium. The cell, however, is not in equilibrium, containing dynamic structures in a steady-state flux of both substrates and products, with energy consumption in processes such as enzyme-catalyzed hydrolysis of ATP. It will be important to continue to evolve these models to consider the non-equilibrium behavior of the living cell (Milin and Deniz, 2018; Choi et al., 2020).

Providing experimental evidence for biomolecular condensate formation via LLPS

Several experiments are typically used to ask whether a biomolecular condensate of interest forms via liquid-liquid phase separation. In cells, immunofluorescence and live-cell imaging are often used to visualize the morphology and dynamics of the condensate. In live-cell imaging experiments, proteins that comprise the condensate are typically fused to a fluorescent tag, such as green fluorescent protein (GFP) and visualized by confocal microscopy. In vitro, phase separation can be observed using droplet assays, in which purified proteins are combined with physiological levels of salt and molecular crowders such as polyethylene glycol (PEG) or ficoll to simulate the crowded cellular environment, and are subsequently visualized by microscopy (Alberti et al., 2017).

Figure 2

A.



B.

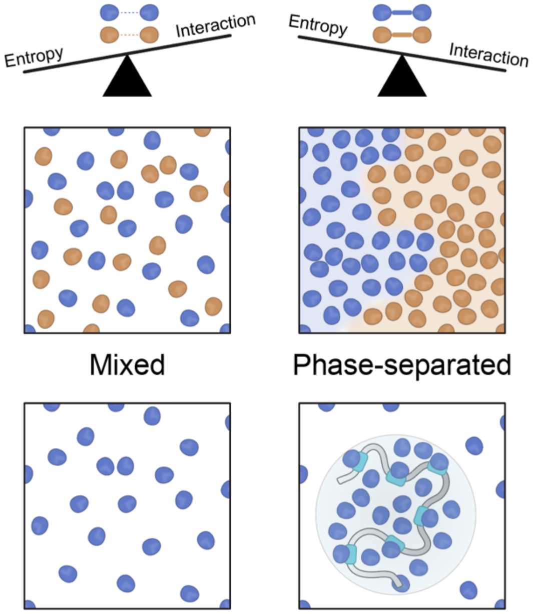


Figure 2. Molecular properties of biomolecules driving condensate formation. (A) Phase diagram of concentration versus valency showing the solubility limit, or saturation concentration (Csat) above which phase separation occurs. Adapted from Brangwynne et., 2013. (B) Depiction of the balance of entropy and enthalpy in LLPS.

Generally, experiments to determine whether a protein of interest forms a condensate often begin by observing morphology of putative condensate-forming proteins in cells, with punctate morphology serving as an early indicator of condensate formation. Morphologically, while many condensates resemble spherical droplets, displaying an effective surface tension minimizing surface area by viscous relaxation to a spherical shape, not all LLPS-mediated condensates are spherical (Hyman et al., 2014). In some cases, for example upon changes in cellular metabolism or under cellular stress, condensates such as the nucleolus have been observed as non-spherical (Brangwynne et al., 2011; Lafontaine et al., 2021).

A variety of liquid properties can be interrogated by microscopy. Puncta are expected to form in a concentration-dependent manner and often exhibit rapid switch-like formation and dissolution upon crossing the concentration threshold for assembly, a hallmark of phase separation (Hyman et al., 2014). Observable properties consistent with LLPS also include fusion and fission of puncta, dripping and wetting of puncta on membranes, and rapid internal rearrangement measured by fluorescence recovery after photobleaching (FRAP) (Brangwynne et al., 2009; Alberti et al., 2017). FRAP experiments can be used to characterize the viscosity of the condensate, which can be affected by a variety of factors including RNA and ATP concentration, and metabolic state (Garcia-Jove Navarro et al., 2019; Patel et al., 2017; Alberti and Hyman, 2021). It is important to note, however, that it is challenging to show unequivocally that a condensate forms by LLPS as material states can be difficult to distinguish in live cells.

Proteins that phase separate in cells are often capable of forming phase-separated droplets in vitro (Alberti et al., 2017). In vitro droplet assays can be used to more closely examine the LLPS capacity of individual condensate components, as well as the partitioning behavior of multiple condensate components. Scaffold proteins are typically expected to phase separate alone in vitro in droplet assays, whereas client proteins may only form droplets in the presence of a scaffold protein, partitioning into the condensates formed by the scaffold (Alberti et al., 2017). Droplets formed in vitro are expected to be concentration-dependent, can undergo fusion and fission, and can be subjected to FRAP analysis to assess the dynamics of components of the droplet.

While LLPS has provided intriguing new models and insights in cell biology, in order to apply the principles of phase transitions and polymer physics to cellular processes, it is important to first perform the aforementioned experiments to assess whether a biomolecular condensate of interest may indeed be formed via LLPS.

Physicochemical environment of condensates: partitioning and molecular grammar

Condensate-promoting features in proteins and nucleic acids

A variety of features of proteins can promote condensate formation as proteins can partake in a range of interactions (Figure 3B). Generally, weak multivalent interactions promote phase separation, but these can come from a variety of sequence features of proteins. Condensates are enriched for proteins with intrinsic disorder, and these regions have been shown for many proteins to be sufficient for condensate formation in vitro and in cells (Banani et al., 2017).

Within intrinsically disordered regions, low complexity sequence (LCS) regions, or stretches of sequence with relatively few different types of amino acids, have been shown to promote condensate formation (Martin and Mittag, 2018; Kato et al,, 2012; Chong, 2018; Nott et al., 2015). Examples of LCSs identified in experimental studies that can promote homotypic or heterotypic condensate formation include acidic patches (Mitrea et al., 2016), basic patches (Li et al., 2020), alanine-rich regions (Basu et al., 2020), proline-rich regions (Dao et al., 2018), pi-pi interacting residues (Vernon et al., 2018), and prion-like domains (Lancaster et al., 2014).

Interactions from structured regions of proteins can also promote phase separation, such as those from modular interacting domains (MIDs) (Li et al., 2012, Banjade and Rosen, 2014). MIDs such as SH3 domains (Li et al., 2012), bromodomains (Sabari et al., 2018), and RNA recognition motifs (RRMs) have been shown to contribute to the formation of condensates (Figure 3B) (Garcia-Jove Navarro et al., 2019).

Nucleic acids can also provide interactions that promote condensate formation. RNA is a prominent component of many cytoplasmic and nuclear condensates, capable of undergoing homotypic condensate formation and serving as a condensate scaffold by providing multiple binding sites for RNA-binding proteins (RBPs) by which these proteins are crowded to sufficiently high concentrations for phase separation (Jain and Vale, 2017; Garcia-Jove Navarro et al., 2019). DNA regulatory elements can also crowd proteins such as transcription factors, which bind to transcription factor binding sites and recruit coactivators to form transcriptional condensates (Shrinivas et al., 2019; Boija et al., 2018; Sabari et al., 2018). The negatively charged nucleic acid backbone can also participate in molecular interactions.

Molecular grammar of condensate formation

Mutational studies have begun to explore the molecular grammar of condensate formation, interrogating how specific amino acids contribute to the formation of specific condensates. A study on FUS family proteins honed in on the importance of tyrosine residues from prion-like domains and arginine residues from RNA-binding domains in promoting FUS phase separation, also highlighting a role for glycine residues in enhancing fluidity and glutamine and serine residues in promoting hardening of these condensates (Wang et al., 2018).

While specific residues may play redundant roles when they appear with interacting partners in other proteins, the types of molecular interactions that can be important for phase separation are typically condensate-specific, as exemplified by studies using perturbants of specific types of interactions on different condensates. For example, while electrostatic interactions contribute prominently to the formation of certain condensates such as FUS condensates, which can be perturbed by changing salt concentration or mutating charged residues, the formation of other condensates, such as transcriptional condensates formed by transcription factor GCN4 and coactivator subunit MED15, seems to be driven largely by hydrophobic interactions (Wang et al., 2018; Boija et al., 2018).

Figure 3

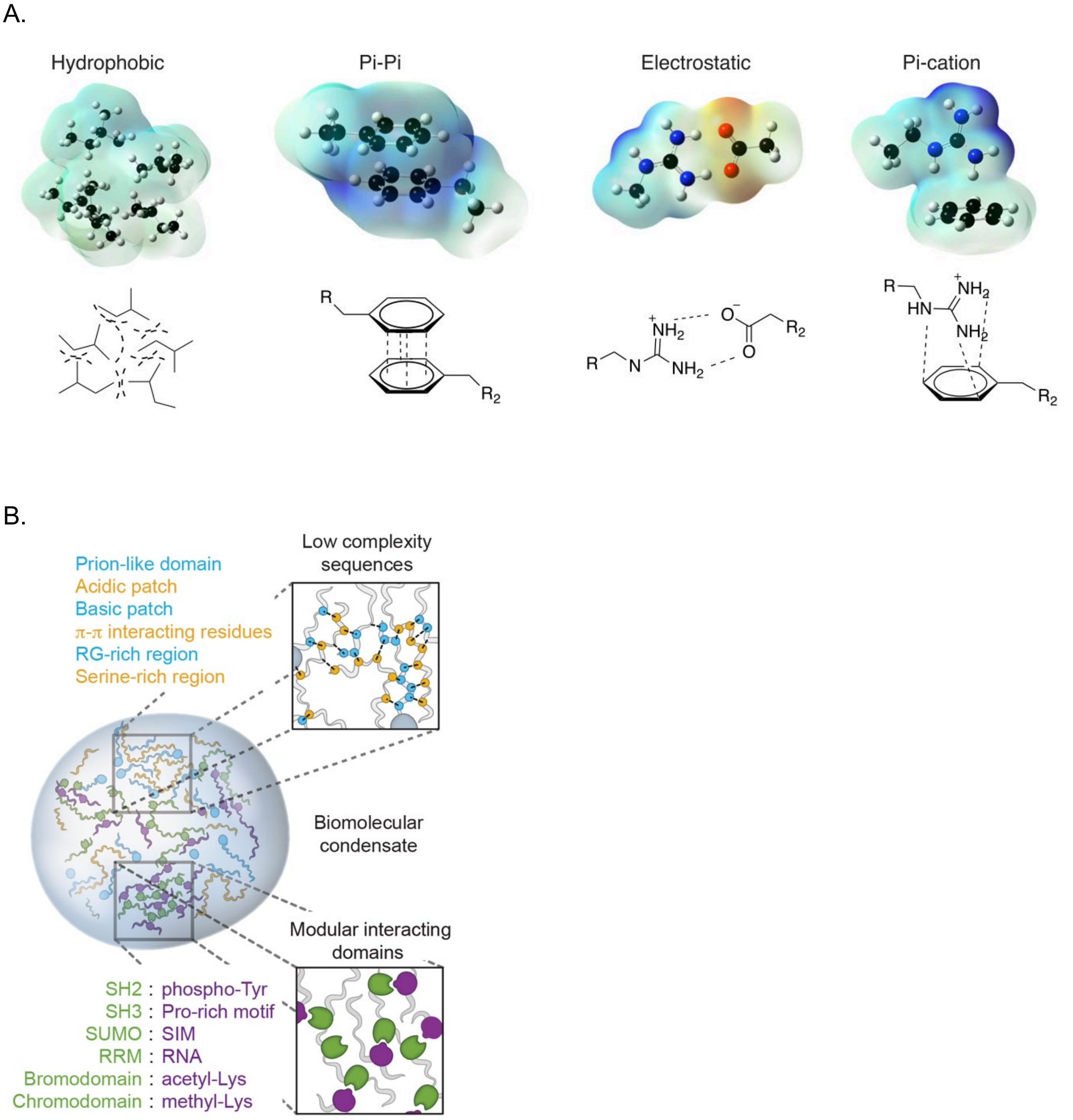


Figure 3. Molecular grammar and condensate partitioning. (A) Interactions that promote condensate formation and govern selective partitioning. Adapted from Boija et al., 2021. (B) Protein sequence features that promote condensate formation.

Differential partitioning of biomolecules into biomolecular condensates

The components of a condensate and the types of molecular interactions they participate in define a physicochemical environment in a condensate (Ditlev et al., 2018). Specific biomolecules such as proteins and nucleic acids can partition preferentially into specific condensates, based upon their physicochemical properties (Figure 3A). Size of the molecule can play a role, as condensates have characteristic mesh sizes that can exclude specific molecules (Hyman et al., 2014; Lyon et al., 2021). The amino acid or nucleic acid sequence of condensate components also seems to play an important role, governing the interactions of these components and whether they are compatible with a given condensate (Figure 3A) (Banani et al., 2016; Ditlev et al., 2018).

In some cases, condensate-forming proteins have been seen to be fully immiscible. Immiscibility of condensate components can keep compartments within a condensate separate, as discussed below for the nucleolus, and can also keep distinct condensates, such as those formed by euchromatin and heterochromatin proteins, apart (Brangwynne et al., 2011, Lafontaine et al., 2021; Shin et al., 2018).

While mutational studies and miscibility studies have begun to elucidate a molecular grammar for specific proteins, there is still much work to be done in order to understand the chemical principles that govern the differential partitioning behavior observed across the diverse condensates in the cell.

Functions of biomolecular condensates

LLPS plays a variety of roles in organizing cellular processes, endowing condensates with “emergent behaviors”. So-called emergent behaviors are a common feature of materials in soft-matter physics by which the material, in this case a liquid-like condensate, displays unique characteristics that represent more than the sum-total of its parts and which can only be fully understood by mesoscale studies (Lyon et al., 2021; Holehouse and Pappu, 2018). Emergent properties of condensates include the concentration and selective compartmentalization of specific proteins and nucleic acids, liquid metastability, and interfacial behaviors (Figure 4A) (Lyon et al., 2021). These properties have provided many insights into the functions of condensates, described in more detail below.

Functional consequences of a dense phase for biochemical reaction components

Increasing reaction rates. Due to the law of mass action, by which the rate of a reaction is proportional to the concentration of the reacting molecules, the propensity of condensates to concentrate substrates and enzymes of biochemical reactions can result in an increase in reaction rates (Banani et al., 2017) (Figure 4B-i). This expectation was tested in a synthetic system in which the SUMOylation enzyme cascade was recruited into engineered condensates. In this system, SUMOylation rates were shown to be increased up to 36-fold in condensates compared to the bulk solution (Peeples and Rosen, 2021).

Figure 4

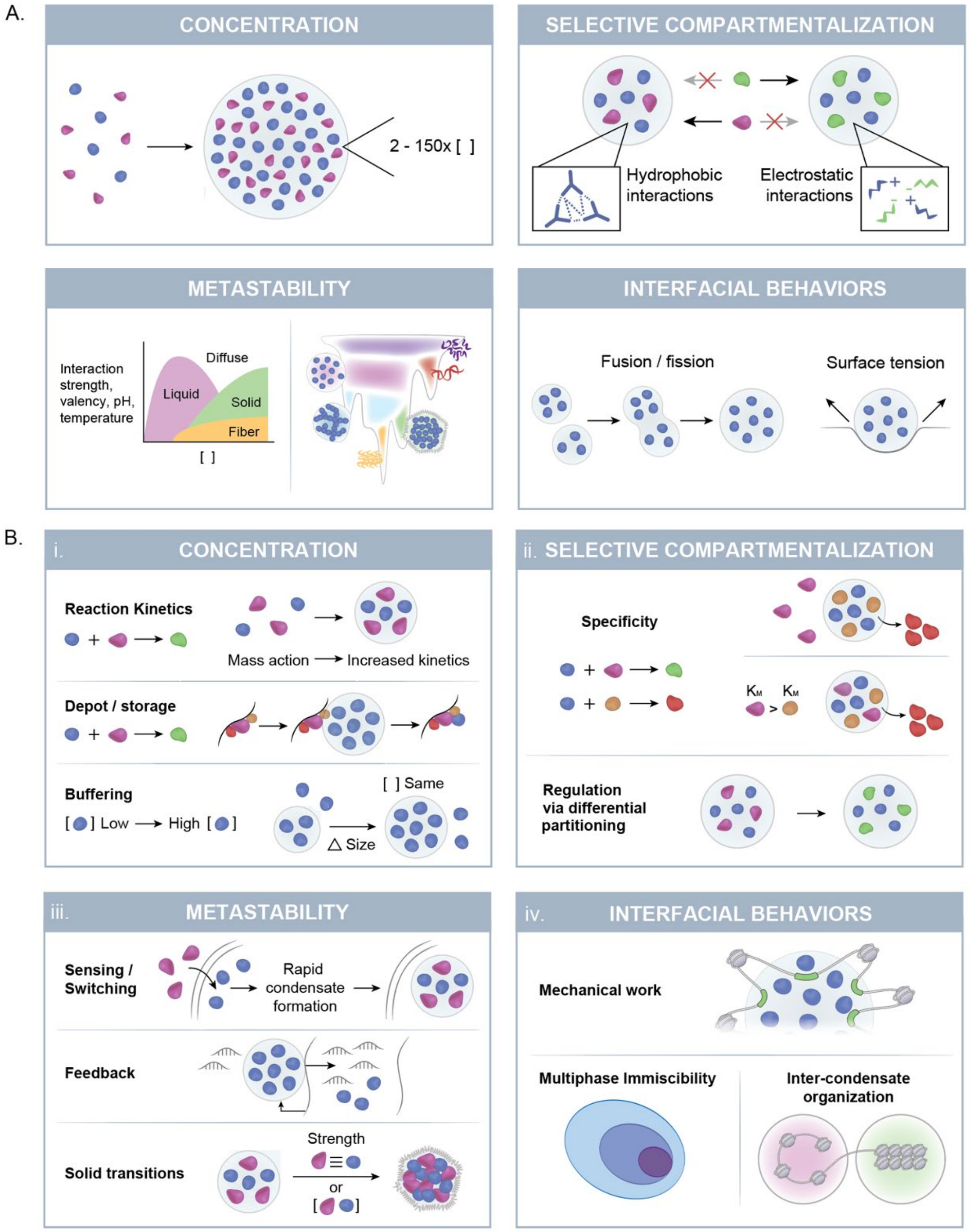


Figure 4. Emergent properties and condensate function. (A) Emergent properties of condensates, including concentration, compartmentalization, metastability, and interfacial behaviors. (B) Functional implications of the emergent properties of condensates.

Decreasing reaction rates. The property of concentration into condensates can also decrease reaction rates or regulatory functions if key substrates or enzymes are concentrated into condensates away from either other reactants or the site where the reaction occurs. For example, viral infection can cause the upregulation of the paraspeckle scaffold NEAT1, which concentrates the protein SFPQ into paraspeckles and depletes this protein from the nucleoplasm, resulting in downregulation of SFPQ targets (Hirose et al., 2014).

Depot functions. In addition to facilitating biochemical reactions internally, biomolecular condensates can also store materials in the cell (Figure 4B-i). For example, RNP granules in the cytoplasm such as stress granules and P bodies can store mRNA until it is later needed for translation (Decker, 2012). The storage of protein in biomolecular condensates can allow for rapid availability of translated proteins involved in processes such as splicing, many components of which are stored in nuclear speckle condensates that can interact directly with sites of transcription (Guo et al., 2019).

Organization of biochemical reactions via selective compartmentalization

Reaction specificity. The selective compartmentalization of specific substrates and enzymes into condensates, based upon the propensity of these biomolecules to phase separate together or partition into specific condensates, allows for additional specificity for biochemical reactions in condensates (Figure 4Bii) (Lyon et al., 2021; Banani et al., 2017; Good et al., 2011; Castellana et al., 2014; Ditlev et al., 2018). Condensate formation of the enzyme ribulose biphosphate carboxylase/ oxygenase (Rubisco) in photosynthetic carbon fixation with specific substrates exemplifies this specificity (Wang et al., 2019). Upon condensate formation, Rubisco is concentrated with its productive carbon dioxide substrate, increasing the enzymatic rate of this productive reaction, while decreasing the rate of a known non-productive side reaction with oxygen (Freeman et al., 2017; Wang et al., 2019). Another demonstration of the reaction specificity endowed by compartmentalization into condensates is the finding in the synthetic SUMOylation cascade system, described above, that increases in reaction rates in condensates are highly dependent upon substrate Km (Peeples and Rosen, 2021).

Partitioning in sequential reactions. Another function that emerges from the selective compartmentalization of biomolecules into condensates is the regulation of sequential reactions by changes in condensate partitioning (Figure 4B-ii). For example, upon phosphorylation of the C terminal domain of RNA Polymerase II, a switch in partitioning from transcriptional condensates into splicing condensates has been observed in vitro and in cells, suggesting that partitioning behavior can facilitate the transition of this enzyme between its role in initiating transcription, and its role in recruiting splicing factors to nascent RNA (Guo et al., 2019).

Metastability in condensate regulation and response to the environment

Sensing and switching. The switch-like formation and dissolution of condensates can allow the cell to rapidly respond to external stimuli and internal changes in expression, degradation or protein modification (Figure 4B-iii) (Söding et al., 2020; Yoo et al., 2019). One such response can be seen in the exquisite temperature sensitivity of mRNA poly(A) binding protein Pab1 in budding yeast. A 10-degree Celsius increase in temperature can accelerate the condensation rate of Pab1 more than 300-fold, with condensation causing a release of many heat shock protein mRNAs that respond to this environmental change (Riback et al., 2017).

Non-equilibrium processes and feedback. Condensate metastability also allows for dynamic responses to the non-equilibrium environment of the cell, resulting from energy-consuming processes and continual flux of substrates and products that form condensates (Brangwynne et al., 2015) (Figure 4B-iii). For example, an RNA-mediated feedback model was recently discovered in transcription by which, at lower concentrations, the RNA created by transcription can promote condensates formed by electrostatic interactions, but at a higher concentration of RNA, repulsive like-charge interactions suppress condensate formation (Henninger et al., 2021; Overbeek and Voorn, 1957; Milin et al., 2018). This reentrant phase behavior, a common feature of complex coacervate models, allows for dynamic feedback adjusting the ability of transcriptional components to concentrate into condensates in response to the transcriptional output of RNA over time.

Liquid-to-solid transitions. The effect of bulk mechanical properties such as viscosity and elasticity of condensates, which change during liquid-to-solid transitions of these metastable liquids and their modulation on the function of these compartments and the biological processes they facilitate is not yet well understood (Lafontaine et al., 2021). While liquid-to-solid transitions of condensates have largely been associated with a pathogenic state in mammalian cells, which seem to prefer to maintain liquid compartments, evidence has emerged that dormancy in bacteria and fungi is associated with a phase transition in which substantial portions of the cytoplasm solidify reversibly and play a protective role of the macromolecules in the cell (Figure 4B-iii) (Parry et al., 2014; Munder et al., 2016; Rabouille and Alberti, 2017). Work on the role of ATP in condensate formation has suggested that ATP functions as a biological hydrotrope (Patel et al,. 2017), and ATP hydrolysis by enzymes such as RNA helicases, as well as RNA granule-remodeling chaperones such as Hsp27 may play a role in actively maintaining condensates in liquid form (Alberti and Dormann, 2019). RNA has also been shown to be capable of affecting the internal viscosity of condensates, as in the case of P body LAF-1 protein in the presence of short polyadenylate RNA (Elbaum-Garfinkle et al., 2015). Lastly, a study optogenetically inducing gelation of the nucleolar scaffold protein nucleophosmin suggested that liquid-like viscosity is important for nucleolar function, finding that increased viscosity of the nucleolus led to the accumulation of unprocessed rRNA and decreased rRNA processing (Zhu et al., 2019).

Interfacial properties in condensate function and organization

Interfacial tension and mechanical work. Condensates have been proposed to be capable of facilitating mechanical work due to forces of interfacial tension and the propensity of liquids to coalesce (Figure 4B-iv) (Lyon et al., 2021). For example, condensate growth or fusion can mechanically exclude chromatin to create regions of low chromatin density or bring together distant genomic sites to facilitate chromatin looping upon coalescence (Shin et al., 2018). Also in support of this claim, recent reports have suggested that membrane-associated condensates formed by endocytic coat and adaptor proteins in clathrin-mediated endocytosis can contribute to making membrane invagination energetically favorable (Bergeron-Sandoval, 2021).

Multi-phase liquid immiscibility. Interfacial tension of liquid-like condensates can organize biological processes in condensates when interfacial forces favor the formation of multiple phases within a condensate over one heterogeneous dense phase (Figure 4B-iv) (Choi et al., 2020; Banani et al., 2017). For example, the nucleolus contains multiple subcompartments that each represent distinct phases that facilitate sequential steps along the ribosome biogenesis pathway, including the fibrillar component (FC), the dense fibrillar component (DFC), and the granular component (GC) (Pederson, 2011; Lafontaine et al., 2021). The immiscibility of these compartments was recapitulated in in vitro droplet assays with key components fibrillarin and nucleophosmin (NPM1) (Feric et al., 2016). Interfacial tension, which favors a spherical shape to minimize surface area, has also been posited to regulate the shape of condensates such as the nucleolus (Lafontaine et al., 2021). Studies of RNA and proteins in the context of interfacial tension point to novel protein functions in the context of this emergent property, including a potential role for proteins like NO145 and Ki-67 at the periphery of the nucleolus as surfactants, and a role for proteins such as MEG-3 in forming clusters that adsorb to the surface of P granules, thereby lowering surface tension and slowing droplet coarsening (Brangwynne et al., 2011; Lafontaine et al., 2021; Folkmann et al., 2021).

Inter-condensate organization via condensate immiscibility. Immiscibility of specific condensates can also play a role in keeping cellular processes or regions with distinct functions separate (Figure 4B-iv). Recent reports have suggested that immiscibility may play a role in separating euchromatin and heterochromatin, both of which have been shown to form phase-separated condensates, thus organizing key regions and functions on the genome. Specifically, BRD4, a transcriptional coactivator, has been shown to be immiscible with the heterochromatin protein HP1a in the CasDrop system, and multiple studies have observed distinct condensate morphologies for euchromatic and heterochromatic regions and proteins (Shin et al., 2018; Li et al., 2020; Klein et al., 2020).

Condensate dysregulation in disease

Just as the study of phase transitions in the cell has created a recent revolution in cell biology, studying how condensates are dysregulated in disease has begun to reveal novel disease mechanisms, paving the way for better treatments for patients (Mullard, 2019; Alberti and Hyman, 2021). LLPS may similarly provide a new framework to evaluate disease states and how they affect biomolecules at mesoscale and the myriad emergent properties of the condensates they form.

Condensate formation depends upon several factors that are known to be altered in disease by mutation or expression changes. These factors include the interactions of condensate components such as proteins and nucleic acids, the concentration of these components, regulators of phase separation, the physicochemical environment (eg. nucleus or cytoplasm) in which a condensate forms, cellular metabolism, and cell homeostasis mechanisms. Thus, in principle, condensate formation could change in a variety of ways in disease, including changes in Csat that alter when condensates are formed, ectopic condensate formation, altered material properties of condensates, or altered composition of condensates (Figure 5A and B) (Alberti and Dormann, 2019).

Pioneering studies in neurological disease have provided the first examples of dysregulated condensate formation in disease and other studies in cancer and infectious disease suggest that condensate dysregulation may be a widespread disease mechanism (Alberti and Hyman, 2021; Tsang et al., 2020). However, it is still not yet known how several mutations cause the dysregulation of condensates and how to predict the effect of a given mutation on a condensate. As discussed in Chapter 2 of this thesis, studies of the molecular grammar of condensate formation could help bridge this gap and provide a foundation for rapid growth in the study of condensates in disease, as has been seen in the study of condensate formation in cell biology with the advent of tools to study condensates.

Pioneering studies on condensates in neurodegenerative disease

Studies in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) have provided proof-of-principle that changes in condensate properties play key roles in disease pathogenesis (Elbaum-Garfinkle and Brangwynne, 2015). Evidence that condensate dysregulation plays a role in these diseases includes findings that ALS-causing mutations in proteins such as FUS and TDP43 cause accumulation of these proteins in stress granule (SG) condensates and change the dynamics, transport and size of these SGs. FUS and TDP-43 also form pathological aggregates in ALS and FTD that contain many stress granule components, suggesting SGs are precursors to these aggregates (Patel et al., 2015; Murakami et al., 2015; Conciella et al., 2016). In vitro evidence that proteins such as FUS form liquid-like droplets that transition to a solid-like state over time, and that this transition is accelerated with disease mutations, suggests that impaired condensate dynamics may play an important role in pathogenesis (Patel et al., 2015). Work in ALS and FTD has also emphasized the role of altered subcellular localization in changing phase behavior. For example, mutations in FUS that impair its ability to localize to the nucleus cause increased condensate formation due to the higher cytoplasmic concentration and decreased RNA levels in the cytoplasm, which usually suppress FUS phase separation (Dormann et al., 2010; Patel et al., 2015; Maharana et al., 2018).

Aberrant phase transitions have also been implicated in neurodegenerative diseases such as Alzheimer Disease (AD) and Huntington's disease (HD). Mutations and abnormal post-translational modifications in AD in microtubule-binding protein Tau shift the equilibrium of Tau toward a phase-separated state and can subsequently promote hardening and aggregation into pathological neurofibrillary tangles (Zhang et al., 2017; Ambadipudi et al., 2017; Wegmann et al., 2018). Similarly, repeat expansions in proteins involved in HD such as polyQ expansions in the Huntingtin protein seem to facilitate the transition of liquid-like assemblies formed by this protein into solid-like assemblies over time that resemble pathological HD aggregates in patients (Peskett et al., 2018).

These studies in neurological disease highlight the importance of maintaining the liquid state of condensates, which is inherently metastable. Further into the two-phase regime, for example at higher concentrations, metastable liquids can turn into solids (Figure 5B) (Brangwynne et al., 2015; Shin and Brangwynne, 2017). Several studies have proposed that the cell has several mechanisms to prevent this solidification and that these quality control mechanisms may be altered in disease or with cellular aging (Alberti and Hyman, 2021). Neurons seem particularly vulnerable to aberrant phase transitions as these cells are post-mitotic and therefore cannot remove pathological solids and aggregates through cell division (Alberti and Dormann, 2019).

Condensates in cancer

Several cellular processes that are dysregulated in cancer are associated with biomolecular condensates. In each of these processes, including transcription, epigenetic regulation, cell signaling, immune signaling, ribosome biosynthesis, degradation, DNA repair, splicing, nuclear transport, DNA replication, storage, autophagy, proteasome function, telomere function, and stress response, several proteins have been implicated in cancer progression via pathogenic alterations such as mutation, expression level change, or differential regulation through post-translational modification (Boija et al., 2021). Indeed, biomolecular condensates have been shown to play a role in several classical hallmarks of cancer, such as sustained proliferative signaling, genome instability and mutation, and avoiding immune destruction (Hanahan and Weinberg, 2011).

Although many condensate-associated proteins are known to be altered in cancer, only a handful of examples of these proteins have been linked experimentally to aberrant phase separation in disease. Such proteins include SPOP, an E3 ligase in which cancer-causing mutations prevent condensation with its substrate and diminish its enzymatic activity (Bouchard et al., 2018), KEAP1, in which mutations affect proteasomal degradation and autophagy by affecting properties of the p62-dependent condensate (Cloer et al., 2018), and nucleoporin, the chromosomal translocation of which establishes oncogenic transcriptional condensates in leukemias (Ahn et al., 20201). Chromosomal translocations that fuse IDRs to various proteins features such as DNA-binding domains have been posited to drive malignancy by promoting aberrant condensate formation via multivalent IDR interactions (Boija et al., 2021). Examples of such fusions include the IDR of EWS and the FLI protein in Ewing's sarcoma or the IDR of FUS and various DNA-binding domains in a specific type of liposarcoma (Chong et al., 2018).

Studies of super-enhancers, clusters of enhancers that have been shown to form transcriptional condensates, in cancer have suggested that aberrant super-enhancer formation at oncogenes is likely a marker of aberrant condensate formation (Hnisz et al., 2013; Hnisz et al., 2017; Sabari et al., 2018; Boija et al., 2018; Cho et al., 2018). Imaging studies remain to be done to visualize condensates at specific loci where super-enhancers of interest form at oncogenes, such as the small binding site insertion seen in T-cell acute lymphoblastic leukemia that causes the aberrant formation of a super-enhancer (Mansour et al., 2016). Some such efforts focusing on the Estrogen Receptor in breast cancer will be discussed in Chapter 3 (Klein et al., 2020).

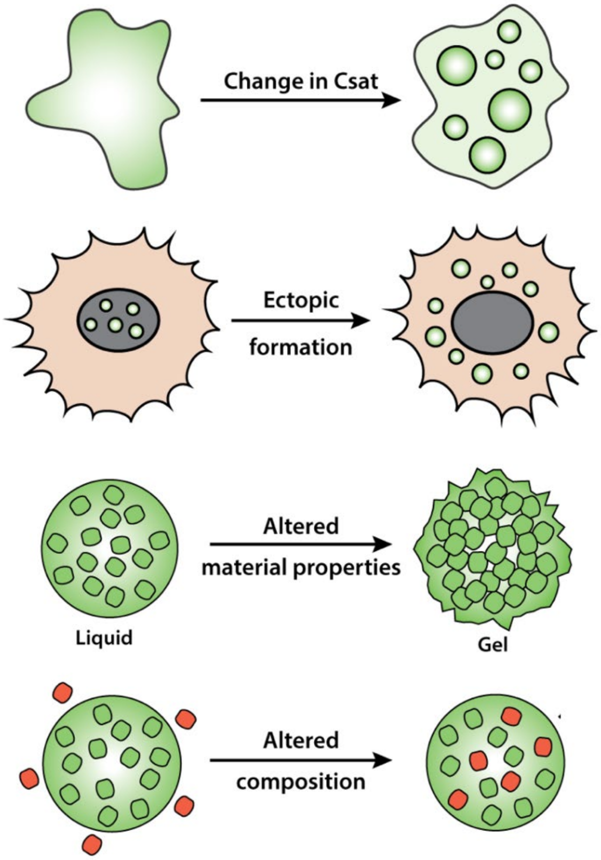
Condensates in infectious disease

Condensates play a role in both the viral life cycle and the immune response from the cell. Pathogens such as vesicular stomatitis virus, SARS-CoV-2 and rabies virus have been seen to form liquid-like compartments in infected cells (Heinrich et al., 2018; Savastano et al., 2020; Nikolic et al., 2017). Several viral factors have been shown to promote condensate formation through features such as protein IDRs, and use these condensates to maximize their ability to propagate in cells (Alberti and Dormann, 2019).

Several biomolecular condensates in the cell play roles in cellular response to viral infection, such as stress granules, to which several components of the innate immune system such as double-stranded RNA-dependent protein kinase R localize (Alberti and Dormann, 2019; Poblete-Durán, 2016). Sensors in the cell such as cGAS that detect viral invasion and replication also form condensates by phase separation (Du and Chen, 2018).

Figure 5

A.



B.

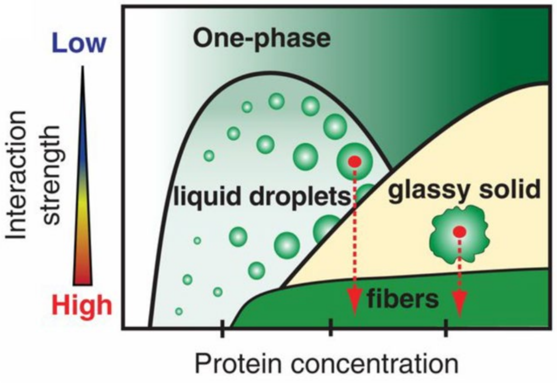


Figure 5. Condensate dysregulation in disease. (A) Potential modes of condensate dysregulation in disease. Adapted from Alberti and Dormann, 2019. (B) Placing pathological condensation in the context of the phase diagram. Adapted from Shin and Brangwynne, 2017.

Viruses can also employ strategies to disrupt the ability of cells to detect and respond to viral infection via condensates. For example, many viruses have evolved to create factors that are capable of inhibiting SG formation or changing SG composition to suppress cellular stress response by interacting with and modulating SG components (Onomoto et al., 2014; Poblete-Durán, 2016). Condensate formation also plays a role in bacterial and fungal infection by facilitating cell dormancy and thereby treatment resistance, as described above (Fisher et al., 2017).

Dysregulation of emergent properties of condensates across disease

Emergent properties of biomolecular condensates have yielded novel insights into their function, and are beginning to provide a framework for understanding the diverse phenotypes of condensate dysregulation. Early studies demonstrating that myriad emergent properties can be dysregulated in disease have suggested that these properties could provide a unifying framework for a deeper understanding of the functional consequences of condensate-dysregulating mutations across diverse diseases. Here, we place condensate dysregulation phenotypes into the context of these emergent mesoscale condensate properties.

<u>Aberrant increases or decreases in concentration via LLPS in disease</u>. A striking example of aberrant concentration of proteins into a condensate was discovered in a study of RAS-MAPK pathway phosphatase SH2 mutations that drive MAPK hyperactivation in developmental syndromes such as Noonan syndrome (NS) and juvenile myelomonocytic leukemias (JMML) (Zhu et al., 2020). While wildtype SHP2 displayed a diffuse localization throughout the cell, NS and JMML mutations were found to cause the formation of liquid-like condensates in cells, to promote LLPS of SHP2 in vitro, and to concentrate mutant SHP2 with wildtype SHP2 and substrates such as DiFMUP in vitro, increasing phosphatase activity and promoting ERK1/2 activation in cells (Zhu et al., 2020). Mutations that disrupt sequence features promoting phase separation have also been shown to disrupt the concentration of molecules into condensates in disease. For example, Rett syndrome mutations in Methyl CpG binding protein 2 (MeCP2), which forms heterochromatic condensates, have been shown to disrupt condensate formation in vitro and to reduce the localization of MeCP2 to heterochromatic condensates in cells (Li et al., 2020).

<u>Altered compartment composition in disease</u>. Mutations in disease can also alter the relative composition of proteins in condensates. An example of altered physicochemical environment and partitioning is the observation of transcriptional condensate unblending in human repeat expansion diseases such as hereditary synpolydactyly (Basu et al., 2020). Specifically, alanine repeat expansion mutations in the HOXD13 IDR enhance condensation, but these condensates are compositionally distinct from wildtype HOXD13 condensates as they display decreased propensity to concentrate coactivators such as Mediator, represented by the MED1 subunit, and BRD4. Global changes in condensate composition across the proteome may also result from mutations of one condensate-forming factor, as in the case of a recent study of chromosomal translocations in acute myeloid leukaemia causing fusion of the IDR-containing N-terminus of Nucleoporin 98 (NUP98) to effectors with roles in gene control, such as KDM5A, NSD1, and HOXA9 (Terlecki-Zaniewicz et al., 2021). This study reported that these fusions form nuclear puncta with proteins known to be involved in biomolecular condensate formation and that tens of proteins were enriched or depleted from cellular condensomes upon expression of NUP98 fusions, as measured by biotinylated isoxazole-mediated condensosome MS. These results suggest that the effects of mutation on both local and global condensate formation and

composition will be important in evaluating the consequences of condensate dysregulation in disease.

Pathological shifts in metastable liquids. Studies in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), discussed in more detail above, have provided proof-of-principle that shifts in metastable liquid condensates toward solids play key roles in disease pathogenesis (Elbaum-Garfinkle and Brangwynne, 2015). Disruption of the switch-like formation and dissolution of condensates in disease has also been hypothesized to lead to decreased dynamic response to stimuli such as stress or signaling molecules (Boija et al., 2021; Alberti and Dormann, 2019). For example, expansion mutations causing the polyGly-Arg and polyPro-Arg C9orf72 toxic dipeptide repeats, the most common cause of ALS and FTD, have been shown to induce the spontaneous assembly of stress granules in the absence of oxidative or osmotic stress, which are rarely disassembled and dramatically impair translation (Lee et al., 2016). While the dysregulation of the switch-like formation and dissolution of signaling pathways known to be involved in controlling cell growth, division, and mobility has not yet been studied in detail, several components of key signaling pathways such as the RAS pathway are mutated in cancer and have recently been shown to form condensates (Boija et al., 2021).

Interfacial behaviors in disease. The consequences of dysregulation of interfacial behaviors, such as interfacial tension and the propensity to undergo fusion and fission, for condensate function have not yet been studied directly in disease. However, as interfacial behaviors can affect condensate morphology and growth, observations of increased size and changes in the shape and number of nucleoli in cancer suggest interfacial behaviors may play a role in nucleolar pathogenesis (Derenzini et al., 2009). Further, the identification of a role for proteins such as Ki-67, which is mutated in about 20% of gliomas, as a biological surfactant localizing to the rim of the nucleolus and to the periphery of mitotic chromosomes during mitosis, suggests surfactant activity, and thus interfacial tension, may be disrupted in disease (Cuylen et al., 2016; Miao et al., 2015).

Condensates and drug development

Motivation for targeting condensates in disease therapy

As condensate formation by LLPS has been shown to be dysregulated in disease, there is growing interest in academia and industry in finding and developing therapeutics that target and modulate condensate formation (Mullard, 2019). Condensate drugs are particularly promising as they could represent a new class of molecules that alter mesoscale properties important for diverse cellular processes and that have a screenable phenotype, namely the formation or dissolution of condensates in vitro and in vivo measured in microscopy imaging studies. A small number of studies have shown that small molecules can disrupt condensates and modulate their properties, such as their propensity to pathologically aggregate (Figure 6A).

Broad-based condensate modulators

How a small molecule affects a given condensate depends upon the interactions that drive LLPS in that condensate. Small molecules such as the aliphatic alcohol 1,6 hexanediol have been used to disrupt weak hydrophobic interactions across many condensate studies (Alberti et al., 2017; Düster et al., 2021). Another broad-based condensate modulator is heparin, a negatively charged molecule that engages in electrostatic interactions and can promote or disrupt condensate formation depending upon the nature of the charged components of the condensate of interest (Babinchak et al., 2020).

While such molecules can be helpful in studying the nature of the interactions driving LLPS of a condensate or the physicochemical environment of a condensate, these molecules are not condensate specific and thus have a broad range of effects in the cell. For example, 1,6 hexanediol has been shown to inactivate kinases and phosphatases in the cell at the concentrations typically used to disrupt condensates, rendering this molecule highly toxic (Düster et al., 2021). Specificity of condensate-modulating molecules is likely to be a pervasive challenge in developing condensate-targeting therapies and will be an important consideration for future studies (Alberti et al., 2017). The role this specificity may play in the toxicity of antisense oligonucleotide-based therapeutics is discussed in Chapter 4 of this thesis.

Small molecules affecting LLPS of stress granule proteins in neurodegenerative disease

Studies focusing on chemical targeting of stress granule proteins, such as FUS, G3BP1 and TDP-43, have identified small molecules that can modulate condensate formation of these proteins in a variety of ways. Two studies have used imaging screens utilizing live cells with fluorescently tagged SG proteins and have evaluated the effect of libraries of small molecules on SG formation. One study suggested that lipoamide and lipoic acid could relieve the effects of ALS-associated FUS mutations in vivo in an experimental systems used to assay ALS phenotypes in *Caenorhabditis elegans* and *Drosophila melanogaster* (Wheeler et al., 2019). The other study demonstrated that several hit compounds prevented the ALS-associated accumulation of TDP-43 in induced pluripotent stem cell-derived motor neurons from ALS patients and suggested that compounds with planar moieties could be a promising class of molecules for targeting condensate formation in ALS/FTD (Fang et al., 2019).

Biphasic modulators of condensate formation

While disrupting condensate formation altogether could be an effective strategy in some diseases, given the important role that many condensates play in normal cellular functions, biphasic modulators of condensate formation- molecules that can disrupt aberrant transitions while preserving functionally important ones- are likely to be critical for condensate-targeting therapies. Two recent studies have identified such biphasic molecules that can target TDP-43.

An in vitro study on the effects of a chemical chaperone, trimethylamine N-oxide (TMAO), known to stabilize protein native folds, on TDP-43 droplet formation indicated that physiologic condensation and pathologic fibrillation are independent processes that can be unlinked with TMAO, suggesting selective targeting of fibrillation may be a promising therapeutic strategy in neurodegenerative disorders (Choi et al., 2018).

Another study focusing on small molecule 4,4'-dianilino-1, 1/-binaphthyl-5,5'-disulfonic acid (bis-ANS) showed that this compound can promote TD-43 LLPS at low concentrations and can disrupt droplets through a reentrant phase transition driven by electrostatic repulsion at high concentrations (Babinchak et al., 2020). It will be important to identify general chemical principles that enable such biphasic modulation of condensates to build upon these promising studies.

Therapeutic targeting of condensate components or processes

Mechanistic studies of condensate dysregulation in disease have suggested several therapeutic strategies that may affect condensate formation by targeting condensate components and cellular processes important for the maintenance of healthy condensate formation.

Small molecules that change parameters such as interaction strength and concentration of important condensate components, such as scaffold proteins, will likely affect condensate formation. For example, the accumulation of hyperphosphorylated TDP-43 in ALS/ FTD aggregates has suggested that targeting enzymes such as poly ADP-ribose polymerase that catalyze the addition of this PTM, which changes the strength of binding interactions of this protein, may be an effective therapeutic strategy (McGurk et al., 2018). Small molecules altering in vitro phase separation of p53, a frequently mutated tumor suppressor, have also been characterized recently, though their therapeutic utility is not yet proven (Lemos et al., 2020).

There is accumulating evidence suggesting the protein quality control machinery (PQC), which comprises molecular chaperones and protein degradation systems, plays an important role in condensate maintenance (Alberti and Dormann, 2019). A striking observation in support of such a role is that in condensate-related diseases and aging, protein homeostasis deteriorates and promotes pathological aggregation of proteins known to form liquid-like condensates in normal physiology. Thus, targeting the PQC to increase or rescue its function may be an effective strategy to mitigate condensate dysregulation (Alberti and Dormann, 2019).

Drug resistance is often caused by adaptations by the cell, such as mutations in a drug target or compensatory mutations in another component of the pathway targeted by the drug. As emergent properties of condensates have yielded insights into the function of many pathways, condensate properties may also uncover novel mechanisms for drug resistance (Figure 5A). The study of condensates in the context of drug resistant cells, as is discussed in part in Chapter 3 of this thesis, could thus be useful for overcoming resistance mechanisms.

Condensate therapies in infectious disease

The emerging role of condensates in infectious disease suggests that therapies targeting biomolecular condensates important for infection of host cells by viruses, bacteria, or fungi may be able to slow or stop these infections (Alberti and Dormann, 2019). In the case of viruses, increasing evidence suggests condensates may play important roles in viral replication and cellular immune response, as discussed above. Therapies to target the condensates formed by viruses to replicate or those that viruses disrupt to evade the immune response could therefore be effective in treating viral infections (Wang et al., 2021). For bacterial and fungal infections, the role of condensates in dormancy transitions could represent a targetable Achilles heel of these pathogens, the disruption of which could lead to decreased resistance to infections (Alberti and Dormann, 2019).

Condensate partitioning of therapeutics

The ability of therapeutics to partition into the condensate in which their target resides will likely be an important consideration in the development of condensate-targeting therapies (Figure 6B). Proteins and nucleic acids partition preferentially into specific condensates based upon the physicochemical environment of the condensates and the properties of the partitioning biomolecules (Ditlev et al., 2018). In Chapter 3 of this thesis, I will discuss our work to show that small molecule therapeutics also partition differentially into biomolecular condensates and that this partitioning behavior affects their pharmacodynamic properties (Klein et al., 2020). Studies on the molecular grammar of condensate formation and further screening of molecules sampling diverse chemical properties will be important for understanding what dictates differential partitioning of therapeutics into biomolecular condensates. Such an understanding could enable the rational design of therapeutics with improved target access and thus higher efficacy.

Figure 6

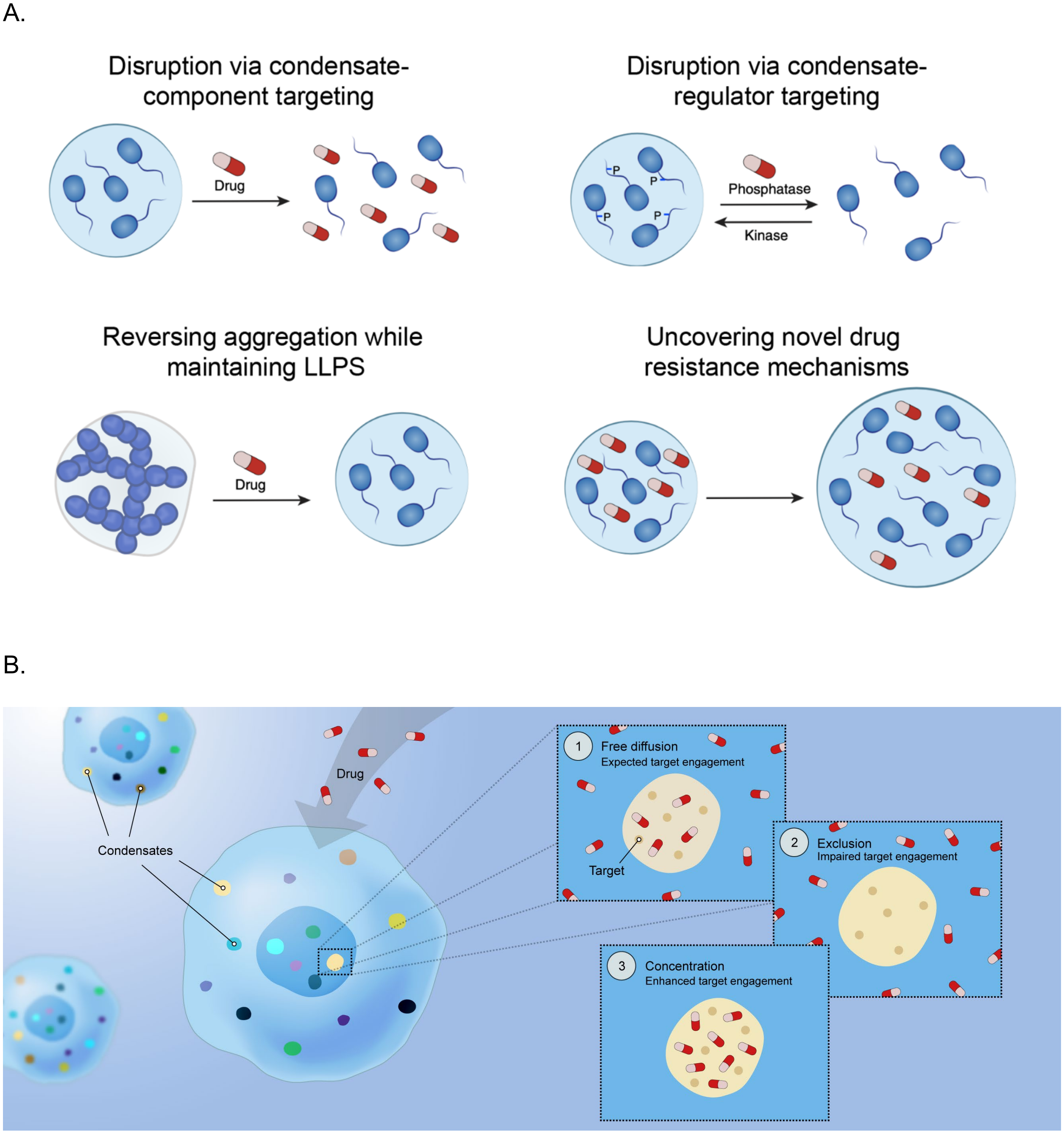


Figure 6. Condensates in drug development. (A) Range of strategies to target condensates in therapy. Adapted from Boija et al., 2021. (B) Potential partitioning behaviors of small molecue therapeutics and implications of differential partitioning of therapeutics into biomolecular condensates for target engagement. Adapted from Klein et al., 2020.

References

Ahn, J. H. *et al.* Phase separation drives aberrant chromatin looping and cancer development. *Nature* 595, 591–595 (2021).

Alberti, S. & Dormann, D. Liquid-Liquid Phase Separation in Disease. *Annu. Rev. Genet.* 53, 171–194 (2019).

Alberti, S., Gladfelter, A., Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Physiol. Behav.* 176, 139–148 (2017).

Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* 22, 196–213 (2021).

Alberts et al. Membrane Structure. *Molecular Biology of the Cell.* 6, 565-596 (2014).

Altmeyer, M. et al. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun*. 6, 8088 (2015).

Ambadipudi S, et al. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun*. 8, 275 (2017).

Babinchak, W. M. *et al.* Small molecules as potent biphasic modulators of protein liquid-liquid phase separation. *Nat. Commun.* 11 (2020).

Banani et al. Compositional control of phase-separated cellular bodies. *Cell*. 166, 651-663 (2016).

Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298 (2017).

Banjade, S. and Rosen, M.K. Phase transitions of multivalent proteins can promote clustering of membrane receptors. *eLife* 3, 3 (2014).

Basu, S., et al. Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* 181, 1062-1079.e30 (2020).

Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842-1855.e16 (2018).

Boija, A., Klein, I. A. & Young, R. A. Biomolecular Condensates and Cancer. *Cancer Cell*. 39, 174–192 (2021).

Bouchard, J. J. *et al.* Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. *Mol. Cell* 72, 19-36.e8 (2018).

Bergeron-Sandoval LP, *et al*. Endocytic proteins with prion-like domains form viscoelastic condensates that enable membrane remodeling. *Proc Natl Acad Sci* 118, 50 (2021).

Beutel, O., et al. Phase Separation of Zonula Occludens Proteins Drives Formation of Tight Junctions. *Cell* 179, 923-936.e11 (2019).

Brangwynne, C. P. Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.* 203, 875–881 (2013).

Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* 11, 899–904 (2015).

Brangwynne, C.P. et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324, 1729-1732 (2009).

Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4334–4339 (2011).

Cahn, JW. and Hilliard, JE. Free energy of a nonuniform system. *The Journal of Chemical Physics*. 28 258–267 (1958).

Cai, D., et al. Phase separation of YAP reorganizes genome topology for long-term YAP target gene expression. *Nat Cell Biol* 21, 1578–1589 (2019).

Case, L.B., Ditlev, J.A., and Rosen, M.K. Regulation of Transmembrane Signaling by Phase Separation. *Annu Rev Biophys* 48, 1–30. (2019).

Castellana, M. et al. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol*. 32, 1011–1018 (2014).

Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415 (2018).

Choi, K. J. *et al.* A Chemical Chaperone Decouples TDP-43 Disordered Domain Phase Separation from Fibrillation. *Biochemistry* 57, 6822–6826 (2018).

Choi, J. M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* 49, 107–133 (2020).

Chong, S., et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361 (2018).

Cloer, E. W. et al. p62-dependent phase separation of patient-derived KEAP1 mutations and NRF2. *Mol. Cell. Biol.* (2018).

Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* 24, 1537–1549 (2016).

Cuylen, S. et al. Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. *Nature* **535**, 308–312 (2016).

Dao, T.P., et al. Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Mol Cell* 69, 965-978.e6 (2018).

De Magistris, P. & Antonin, W. The Dynamic Nature of the Nuclear Envelope. *Curr. Biol.* 28, R487–R497 (2018).

Decker, C. J. & Parker, R. P-Bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb. Perspect. Biol*. 4, a012286 (2012).

Derenzini, M., Montanaro, L. & Trere, D. What the nucleolus says to a tumour pathologist. *Histopathology* **54**, 753–762 (2009).

Ditlev, J. A., Case, L. B. & Rosen, M. K. Who's In and Who's Out—Compositional Control of Biomolecular Condensates. *J. Mol. Biol.* 430, 4666–4684 (2018).

Dormann D, et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *EMBO* J. 29, 2841–2857 (2010).

Du M, Chen ZJ. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* 361, 704–709 (2018).

Düster, R., Kaltheuner, I. H., Schmitz, M. & Geyer, M. 1,6-Hexanediol, commonly used to dissolve liquid-liquid phase separated condensates, directly impairs kinase and phosphatase activities. *J. Biol. Chem.* 296, 100260 (2021).

Elbaum-Garfinkle, S. & Brangwynne, C. P. Liquids, Fibers, and Gels: The Many Phases of Neurodegeneration. *Dev. Cell* 35, 531–532 (2015).

Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7189–7194 (2015).

Fang, M. Y. *et al.* Small-Molecule Modulation of TDP-43 Recruitment to Stress Granules Prevents Persistent TDP-43 Accumulation in ALS/FTD. *Neuron* 103, 802-819.e11 (2019).

Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* 165, 1686–1697 (2016).

Fisher, RA., et al. Persistent bacterial infections and persister cells. *Nat. Rev. Microbiol.* 15, 453–464 (2017).

Flory PJ., Thermodynamics of high polymer solutions. *J. Chem. Phys.* 10, 51-61 (1941).

Freeman Rosenzweig, E. S. et al. The eukaryotic CO2 -concentrating organelle is liquid-like and exhibits dynamic reorganization. *Cell* 171, 148–162.e19 (2017).

Folkmann et al. Regulation of biomolecular condensates by interfacial protein clusters. Science 373 ,1218-1224 (2021).

Forman-Kay, J. D., Ditlev, J. A., Nosella, M. L. & Lee, H. O. What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates? *RNA* 28, 36–47 (2022).

Frottin, F., et al. The nucleolus functions as a phase-separated protein quality control compartment. *Sci New York N Y* 365, 342–347 (2019).

Garcia-Jove Navarro, M. *et al.* RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nat. Commun.* 10, 1–13 (2019).

Gibbs, JW. On the equilibrium of heterogeneous substances. *Scientific Papers* (1961).

Gibson, B., et al. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* 179, 470-484.e21. (2019).

Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332, 680–686 (2011).

Graham, T. Liquid diffusion applied to analysis. *Philosophical Transactions of the Royal Society* 151: 183–224 (1861).

Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548 (2019).

Haas C, Drenth J. Understanding protein crystallization on the basis of the phase diagram. *Journal of Crystal Growth* 196, 388–394 (1999).

Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* 144, 646–674 (2011).

Hardy, WB. On the structure of cell protoplasm. *The Journal of Physiology* 24, 158–210 (1899).

Heinrich, BS et al. Phase transitions drive the formation of vesicular stomatitis virus replication compartments. *mBio* 9, 02290-17 (2018).

Henninger, JE., Oksuz O., et al. RNA-mediated feedback control of transcriptional condensates. *Cell* 184, 207-225 (2021).

Hirose, T. et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell* **25**, 169–183 (2014).

Hnisz, D., et al. A Phase Separation Model for Transcriptional Control. *Cell* 169, 13–23 (2017).

Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* 155, (2013).

Holehouse, A. S. & Pappu, R. V. Functional Implications of Intracellular Phase Transitions. *Biochemistry* 57, 2415–2423 (2018).

Huang, W.Y.C., et al. A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS. *Science* 363, 1098–1103 (2019).

Huggins, ML. Solutions of long chain compounds. *Journal of Chemical Physics* 9, 440 (1941).

Hyman, A. A. & Brangwynne, C. P. Beyond Stereospecificity: Liquids and Mesoscale Organization of Cytoplasm. *Dev. Cell* 21, 14–16 (2011).

Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* 30, 39–58 (2014).

Jain A, Vale RD. RNA phase transitions in repeat expansion disorders. *Nature* 546, 243-247 (2017).

Jiang, H., et al. Phase Transition of Spindle-Associated Protein Regulate Spindle Apparatus Assembly. *Cell* 163, 108–122 (2015).

Kato, M., et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149, 753–767 (2012).

Kilic, S., et al. Phase separation of 53 BP 1 determines liquid-like behavior of DNA repair compartments. *Embo J* 38, 101379 (2019).

King, M.R., and Petry, S. Phase separation of TPX2 enhances and spatially coordinates microtubule nucleation. *Nat Commun* 11, 270 (2020).

Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science* 368, 1386–1392 (2020).

Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* 22, 165–182 (2021).

Lancaster, A.K., et al. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* 30, 2501–2502 (2014).

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. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* 547, 236–240 (2017).

Leduc, S. The mechanism of life (1911).

Lee et al. C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. 167, 774-788 (2016).

Lemos, C. *et al.* Identification of Small Molecules that Modulate Mutant p53 Condensation. *iScience* 23, (2020).

Li, C.H., et al. MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* 1–8 (2020).

Li, P. *et al.* Phase transitions in the assembly of multi-valent signaling proteins. *Nature* 483, 336–340 (2012).

Lu, Y., Wu, T., Gutman, O., Lu, H., Zhou, Q., Henis, Y.I., and Luo, K. Phase separation of TAZ compartmentalizes the transcription machinery to promote gene expression. *Nat. Cell Biol.* 22, 453–464 (2020).

Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* 22, 215–235 (2021).

Maharana S,, et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* 360 918–21 (2018).

Mansour, M. R. *et al.* Mutation of a Noncoding Intergenic Element. *Science* 346, 1373–1377 (2016).

Martin, E.W., and Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry-Us* 57, 2478–2487 (2018).

McGurk L, et al. Nuclear poly(ADP- ribose) activity is a therapeutic target in amyotrophic lateral sclerosis. *Acta Neuropathol. Commun*. 6 (2018).

Miao et al., Expression and mutation analysis of Cyclin A and Ki-67 in glioma and their correlation with tumor progression. Oncol Lett. 10, 1716-1720 (2015).

Milin, A. N. & Deniz, A. A. Reentrant Phase Transitions and Non-Equilibrium Dynamics in Membraneless Organelles. *Biochemistry* 57, 2470–2477 (2018).

Milovanovic, D., Wu, Y., Bian, X., and Camilli, P.D. A liquid phase of synapsin and lipid vesicles. *Science* 361, eaat5671 (2018).

Mitrea, D.M., et al. Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *Elife* 5, e13571 (2016).

Mitrea, D. M. & Kriwacki, R. W. Phase separation in biology; Functional organization of a higher order Short linear motifs - The unexplored frontier of the eukaryotic proteome. *Cell Commun. Signal.* 14, 1–20 (2016).

Mullard, A. Biomolecular condensates pique drug discovery curiosity. *Nat. Rev. Drug Discov.* (2019).

Munder, MC., et al. A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* 5, e09347 (2016).

Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* 88, 678–690 (2015).

Nikolic, J., et al. Negri bodies are viral factories with properties of liquid organelles. Nat. Commun. 8, 58 (2017).

Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* 57, 936–947 (2015).

Onomoto K,et al. Antiviral innate immunity and stress granule responses. *Trends Immunol*. 35, 420–28 (2014).

Overbeek JTG, and Voorn MJ. Phase separation in polyelectrolyte solutions. *Journal of Cellular and Comparative Physiology*. 49, 7–26 (1957).

Pak et al. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol Cell*. 63, 72-85 (2016).

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

Parry, BR., et al. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell* 156, 183–94 (2014).

Patel, A. et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* 162, 1066–1077 (2015).

Patel, A. *et al.* Biochemistry: ATP as a biological hydrotrope. *Science* 356, 753–756 (2017).

Pederson, T. The plurifunctional nucleolus. *Nucleic Acids Res*. 26, 3871–3876 (1998).

Pederson, T. The nucleolus. *Cold Spring Harbor Perspectives in Biology* (2011).

Peeples, W. & Rosen, M. K. Mechanistic dissection of increased enzymatic rate in a phase-separated compartment. *Nat. Chem. Biol.* 17, 693–702 (2021).

Peskett TR, et al. A liquid to solid phase transition underlying pathological huntingtin exon1 aggregation. *Mol. Cell* 70, 588–601.e6 (2018).

Piazza R. Interactions and phase transitions in protein solutions. *Current Opinion in Colloid & Interface Science*. 5, 38–43 (2000).

Poblete-Durán, N., et al. Who regulates whom? An overview ofRNA granules and viral infections. *Viruses* 8 E180 (2016).

Rabouille C, Alberti S. Cell adaptation upon stress: the emerging role ofmembrane-less compart- ments. *Curr. Opin. Cell Biol*. 47, 34–42 (2017).

Riback, J. A. et al. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* 168, 1028–1040.e19 (2017).

Riback, J.A., et al. Composition-dependent thermodynamics of intracellular phase separation. *Nature* 581, 209–214 (2020).

Ruff, K. M., Roberts, S., Chilkoti, A. & Pappu, R. V. Advances in understanding stimulus-responsive phase behavior of intrinsically disordered protein polymers. *J. Mol. Biol*. 430, 4619–4635 (2018).

Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361, 1–17 (2018).

Savastano, A., et al. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun.* 11, (2020).

Schwayer, C., et al. Mechanosensation of Tight Junctions Depends on ZO-1 Phase Separation and Flow. *Cell* 179, 937-952.e18. (2019).

Sheu-Gruttadauria, J., and MacRae, I.J. Phase Transitions in the Assembly and Function of Human miRISC. *Cell* 173, 946-957.e16 (2018).

Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science (80-.).* 357, (2017).

Shin, Y. et al. Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* 175, 1481–1491.e13 (2018).

Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* 75, 549-561.e7 (2019).

Söding, J., Zwicker, D., Sohrabi-Jahromi, S., Boehning, M. & Kirschbaum, J. Mechanisms for Active Regulation of Biomolecular Condensates. *Trends Cell Biol.* 30, 4–14 (2020).

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

Su, X., et al. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* 352, 595–599 (2016).

Terlecki-Zaniewicz et al., Biomolecular condensation of NUP98 fusion proteins drives leukemogenic gene expression. *Nat Struct Mol Biol*. **28**, 190-201 (2021).

Tsang, B., Pritišanac, I., Scherer, S. W., Moses, A. M. & Forman-Kay, J. D. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell* 183, 1742–1756 (2020).

Vernon, R.M., et al. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* 7, e31486 (2018).

Wang, H. et al. Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* 566, 131–135 (2019).

Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174, 688-699.e16 (2018).

Wang, S. *et al.* 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 (2021).

Wegmann, S., et al. Tau protein liquid- liquid phase separation can initiate tau aggregation. *EMBO J*. 37, e98049 (2018).

Wheeler, R. J. *et al.* Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. *bioRxiv* 1–48 (2019).

Wilson, E.B. The structure of protoplasm. *Science* 10, 33–45 (1899).

Woodruff, J.B., et al. The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. *Cell* 169, 1066-1077.e10 (2017).

Yoo, H., Triandafillou, C. & Drummond, D. A. Cellular sensing by phase separation: using the process, not just the products. *J. Biol. Chem.* 294, 7151–7159 (2019).

Zamudio, A. V. *et al.* Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol. Cell* 76, 753-766 (2019).

Zeng, M., Shang, Y., Araki, Y., Guo, T., Huganir, R.L., and Zhang, M. Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity. *Cell* 166, 1163-1175 (2016).

Zhang X, et al. RNA stores tau reversibly in complex coacervates. *PLOS Biol*. 15, e2002183 (2017).

Zhu, L. et al. Controlling the material properties and rRNA processing function of the nucleolus using light. *Proc. Natl Acad. Sci. USA* **116**, 17330–17335 (2019).

Zhu, G. *et al.* Phase Separation of Disease-Associated SHP2 Mutants Underlies MAPK Hyperactivation. *Cell* **183**, 490-502.e18 (2020).

CHAPTER 2:
GENETIC VARIATION ASSOCIATED WITH CONDENSATE DYSREGULATION IN DISEASE

In Press, *Developmental Cell* (2022).

Salman F. Banani[1,2,7], Lena K. Afeyan[1,3,7], Susana W. Hawken[1,3,4,7], 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], Tong Ihn Lee[1], and Richard A. Young[1,3]*

[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
*Correspondence: Richard A. Young (young@wi.mit.edu)

Abstract

A multitude of cellular processes have been shown to 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.

Main Text

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 (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 Hyman, 2021; Boija et al., 2021; 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.

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 manual 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 S10A-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 S10C). 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 (Hughes et al., 2018; Lancaster et al., 2014; Vernon et al., 2018) or by scanning human protein sequences for statistically identified regions of low complexity (Figure S1A-C, Tables S1-3, Supplemental Discussion, *LCS vs. IDR*). 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 S10D).

We then identified the pathogenic mutations that affect condensate-promoting features (Figure 1B). 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 S9, Figure S1H, Table S4, Table S6D, 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 (Supplemental Discussion, *Types of Pathogenic Mutations*). 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 S10F, Methods; see also Supplemental Discussion, *Canonical Models of Protein Dysfunction and Condensate Dysregulation*). 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). 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, Table S10E).

Figure 1

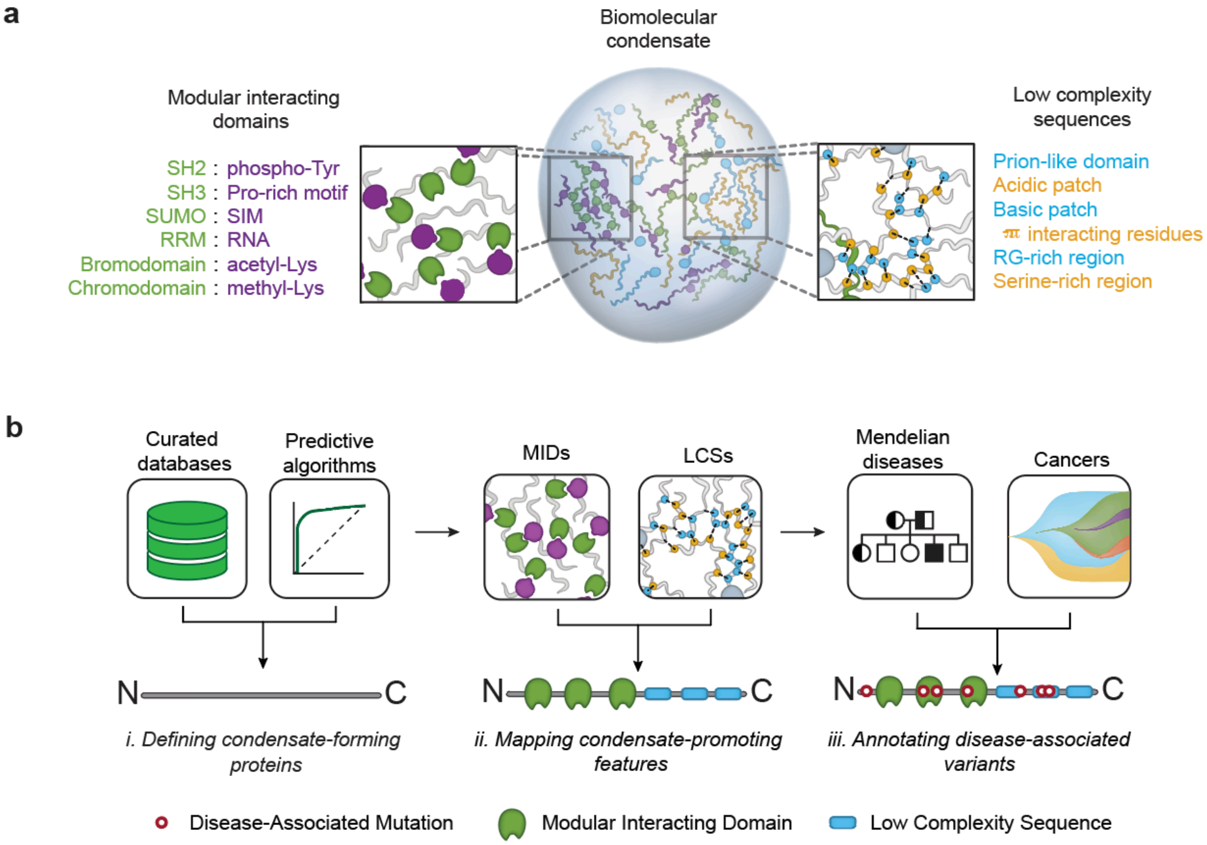


Fig. 1: Mapping mutations to condensate-promoting features in condensate-forming proteins.

a. Multivalent interacting features in proteins that promote biomolecular condensate formation, including low complexity sequences (LCSs, top, blue and yellow) and modular interacting domains (MIDs) with binding partners (bottom, green and purple, respectively), with examples listed.
b. Approach for generating a map of pathogenic mutations that affect condensate-promoting features in known or predicted condensate-forming proteins (see also Extended Data Fig 1a). (i) Curated databases and predictive algorithms were used to define a set of condensate-forming proteins. (ii) MIDs and LCSs, defined as the set of condensate-promoting features, were then mapped across the set of condensate-forming proteins. (iii) Disease-associated variants across Mendelian diseases and cancers were annotated onto this set of proteins, revealing mutations that affect the condensate-promoting features of these proteins.

We mapped these pathogenic variants to the condensate-promoting features annotated within condensate-forming proteins and created 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 S10G). 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 S4). Here, we present an approach to nominate mutations that plausibly contribute to condensate dysregulation, and present a catalog of over 36,000 nominated mutations, with annotations of associated diseases, disrupted condensate-promoting features, and affected condensate-forming proteins for each mutation (Table S10G).

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 (Boija et al., 2021; Cai 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 S10H-I). 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 S10J-K). These results corroborate the hypothesis that condensate dysregulation may be an underlying pathogenic mechanism across a broad spectrum of human diseases.

Figure 2

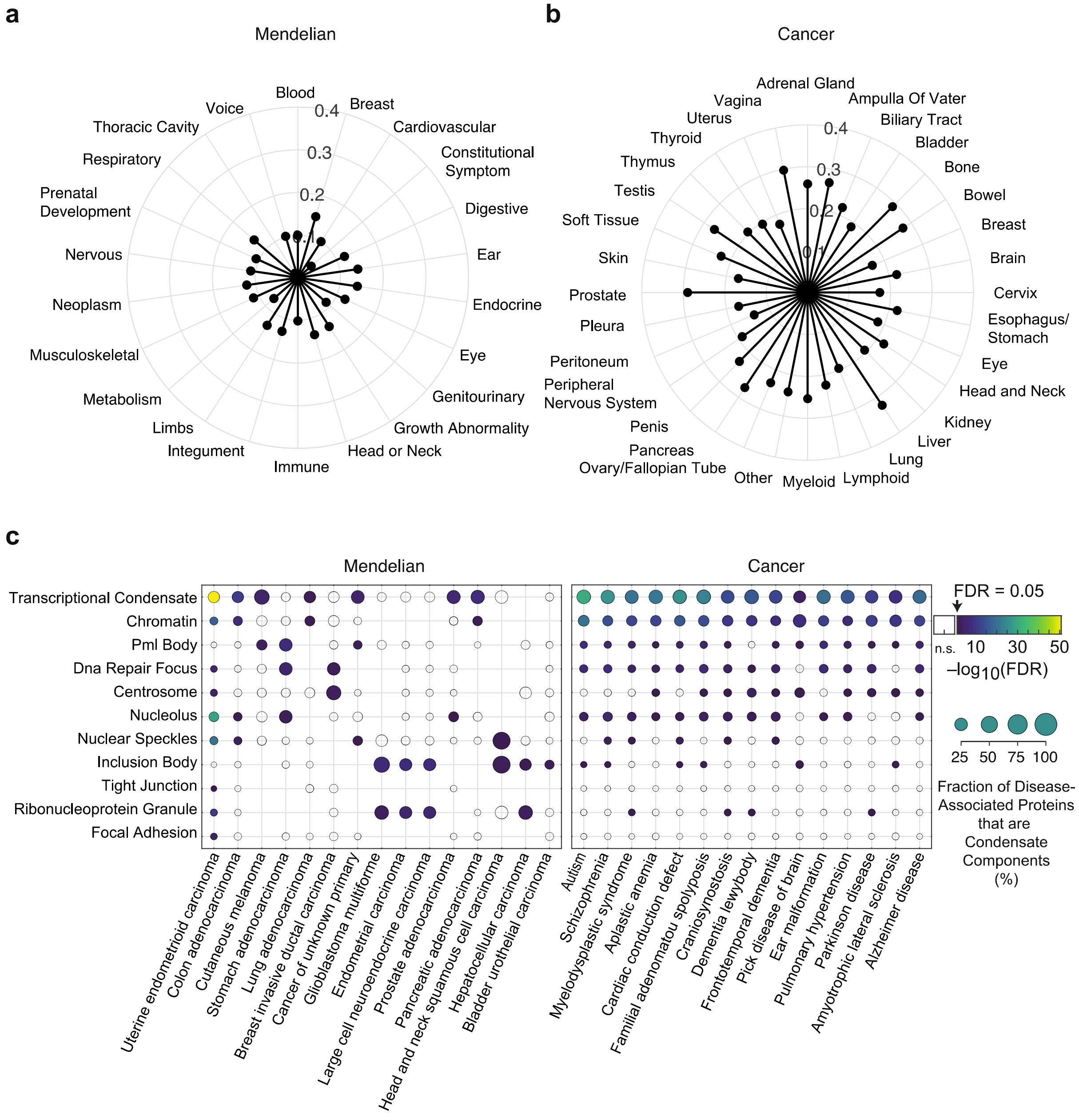


Fig. 2: Condensate dysregulation across the spectrum of disease.

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. 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 (x-axis) and those associated with specific diseases (y-axis) were tested for statistical significance (one-tailed Fisher Exact Test). Selected examples of Mendelian diseases (left) and cancer types (right) are shown (see also Supplementary Table 9j-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 association. Unfilled datapoints correspond to a lack of a statistically significant association. Size of data point is proportional to the fraction of the indicated disease-associated condensate-forming proteins that are components of the indicated condensates.

It is important to note that condensate dysregulation is not the only mechanism by which nominated mutations could contribute to pathogenesis. Several 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 S10L). We do not expect these molecular-scale mechanisms to be mutually exclusive with condensate dysregulation. On the contrary, as intermolecular interactions are important for condensate formation, disruptions of these interactions are likely to affect condensate properties in ways that, when better understood, can deepen our understanding of the mechanistic consequences of pathogenic mutations and how to address these molecular and mesoscale consequences therapeutically (Supplemental Discussion, *Canonical Models of Protein Dysfunction and Condensate Dysregulation*).

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 S5, Supplemental Discussion, *Selection of Candidates and Mutations for Experimental Validation*). 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 S5).

Figure 3

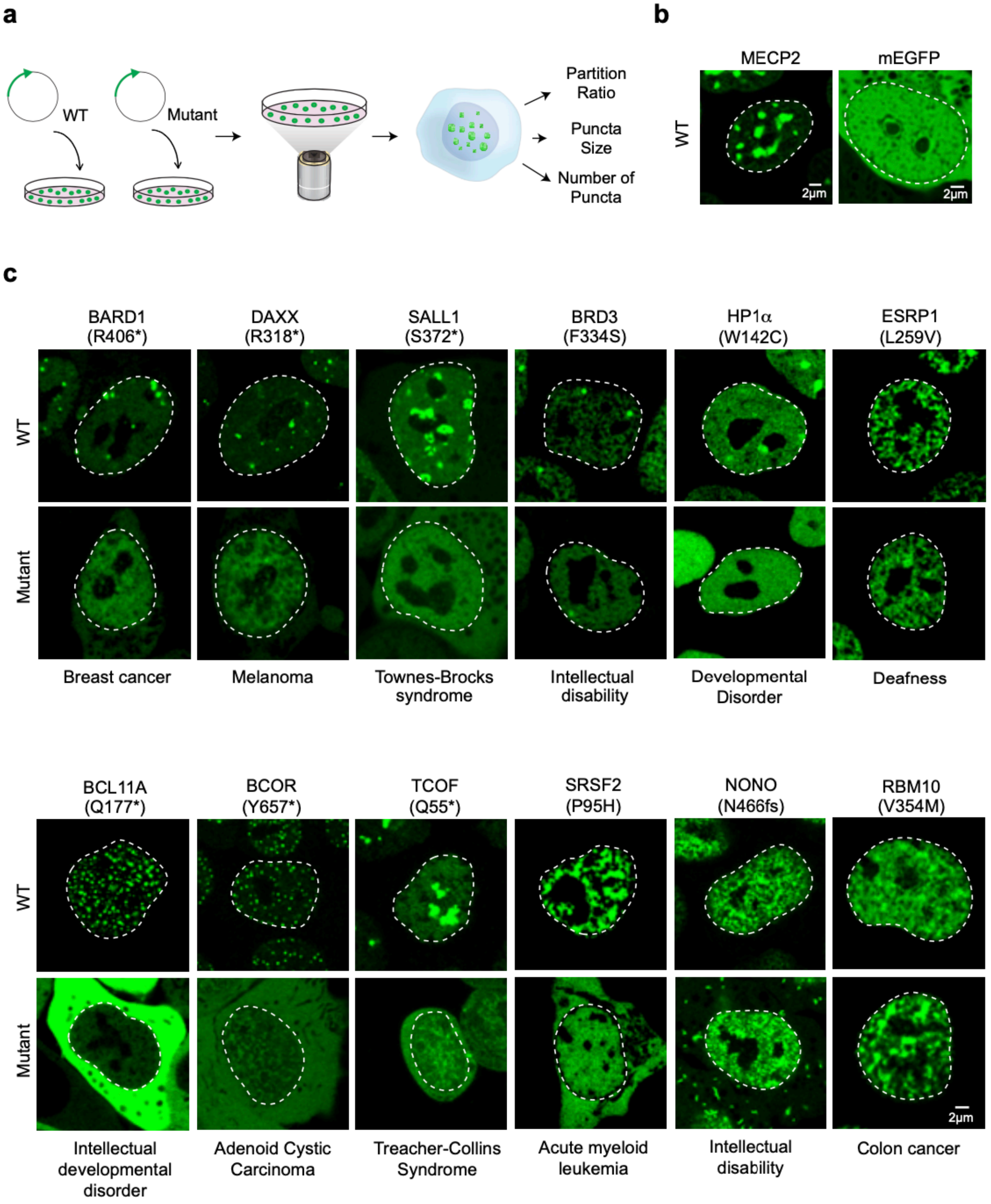


Fig. 3: Pathogenic mutations in condensate-promoting features alter condensate properties in live cells.

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, TCOF, HP1a, SRSF2, ESRP1. Specific mutations that were tested along with their associated disease are indicated adjacent to the images.

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 S5). Mutations were selected to represent the approximate proportion of mutation types and affected condensate-promoting features in the catalog (Figure S3D, Supplemental Discussion, *Selection of Candidates and Mutations for Experimental Validation*). 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, Table S6). In 11 of 13 cases, the mutations caused significant reductions in partitioning of the protein into condensates. In the remaining two cases, the mutations enhanced the ability of the proteins to associate with condensates and, in at least one case, to form puncta in other cellular locations (Figure 3C, Figure S4B, Figure S4D). 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 an 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.

Figure 4

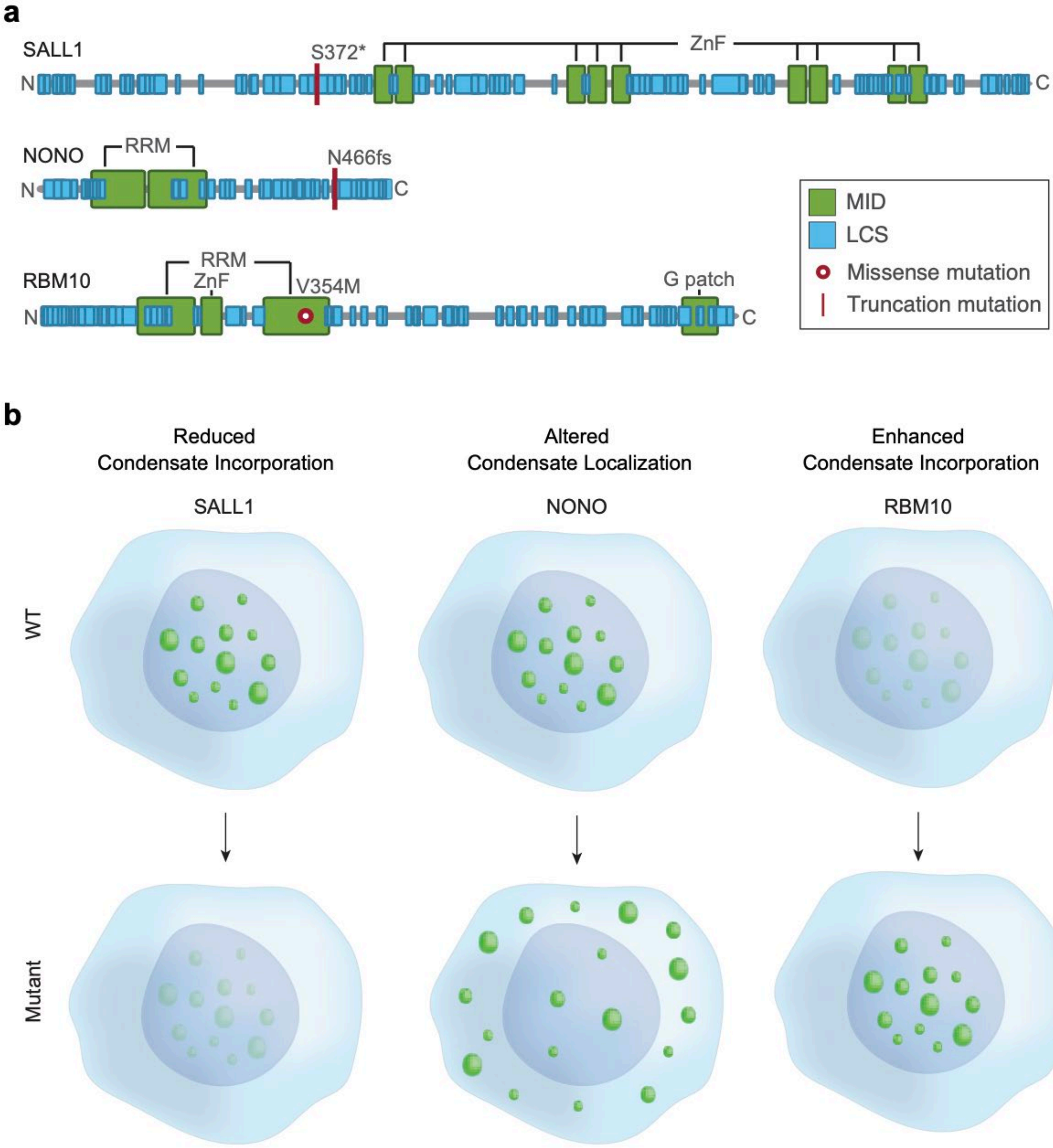


Fig. 4: Mutations in condensate-promoting features cause diverse condensate dysregulation phenotypes.

a. Protein schematics of SALL1, NONO, and RBM10, depicting locations of condensate promoting features (MIDs and LCSs) and a representative pathological mutation that was experimentally tested for condensate dysregulation (Fig. 3c).
b. 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 (right), and altered condensate localization (middle). A protein candidate where with which phenotypes were observed (Fig. 3c) are listed.

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.

We note, however, that there are limitations with the approach described here and the resulting catalog. 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, 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.

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.

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. 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 (R.A.Y.), 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 5F32CA254216 (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.). 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.

Data and code availability

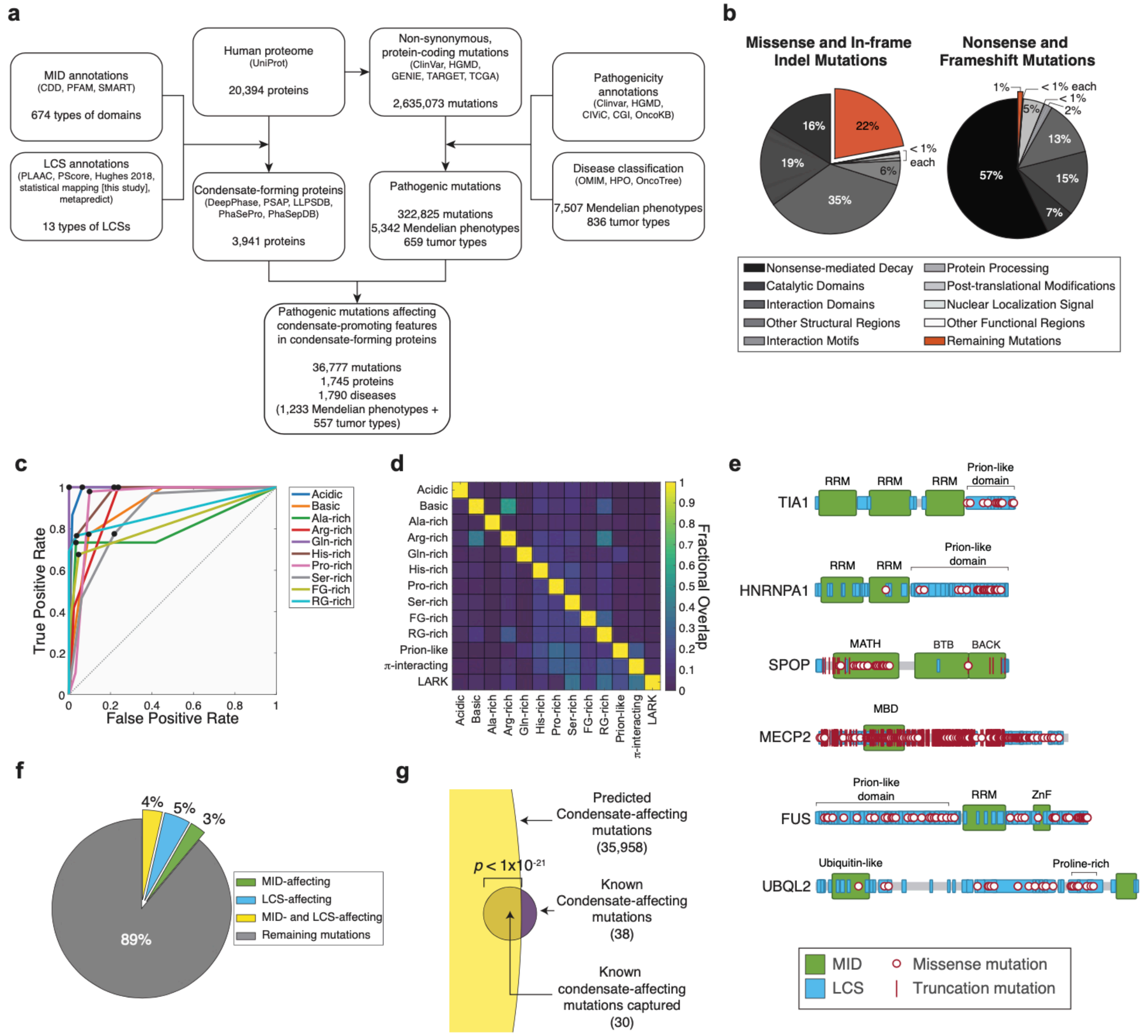
Datasets used in this study are available from the sources cited in Methods. Processed data used in the study is available in the Supplemental Tables. Code used for analyses is available via https://github.com/sb02457/.

Declaration of interests

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

Extended Data Figures

Extended Data Figure 1



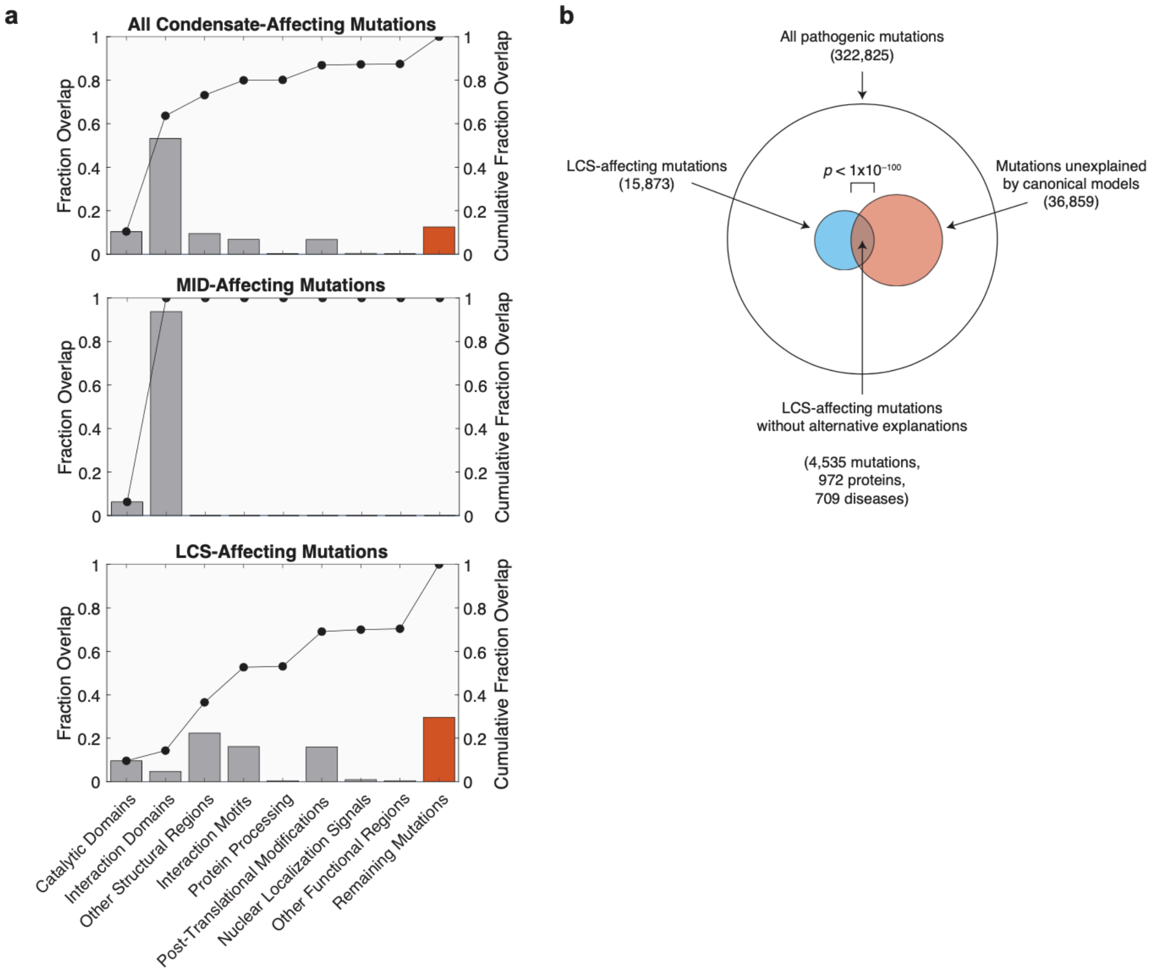
Extended Data Fig. 1: Mapping mutations to condensate-promoting features in condensate-forming proteins.

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. Proportion of pathogenic missense and in-frame indel mutations (left) or nonsense and frameshift mutations (right) that affect protein features associated with protein properties altered in canonical structure-function-based models of protein dysfunction. These features include catalytic domains, interaction domains, other structural regions, protein regions involved in protein processing, post-translational modification sites, nuclear localization and export signals, as well as other annotated regions involved in various protein functions (see Methods). For pathogenic truncation mutations, the proportion predicted to elicit NMD are also indicated. For mutations classifiable in multiple categories, starting with nonsense-mediated decay and moving counter-clockwise, mutations counted in one category were not counted again in subsequent categories. Proportion of mutations that were not mapped to any of these features are shown in red. See also Fig. 1a.

c. ROC curves for optimizing cutoffs for mapping indicated LCSs across the proteome, benchmarked against a set of validated LCSs (Supplementary Table 2). Black dots represent position on the curve corresponding to the optimal cutoff.
d. 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. See also Supplementary Table 3).
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. Fraction of pathogenic mutations that affect MIDs, LCSs, or both of these types of condensate-promoting features within putative condensate forming proteins (see also Supplemental Table 9c,g).
g. Overlap of predicted condensate-affecting mutations with known condensate-affecting mutations curated from the literature (see also Supplementary Table 4).

Extended Data Figure 2

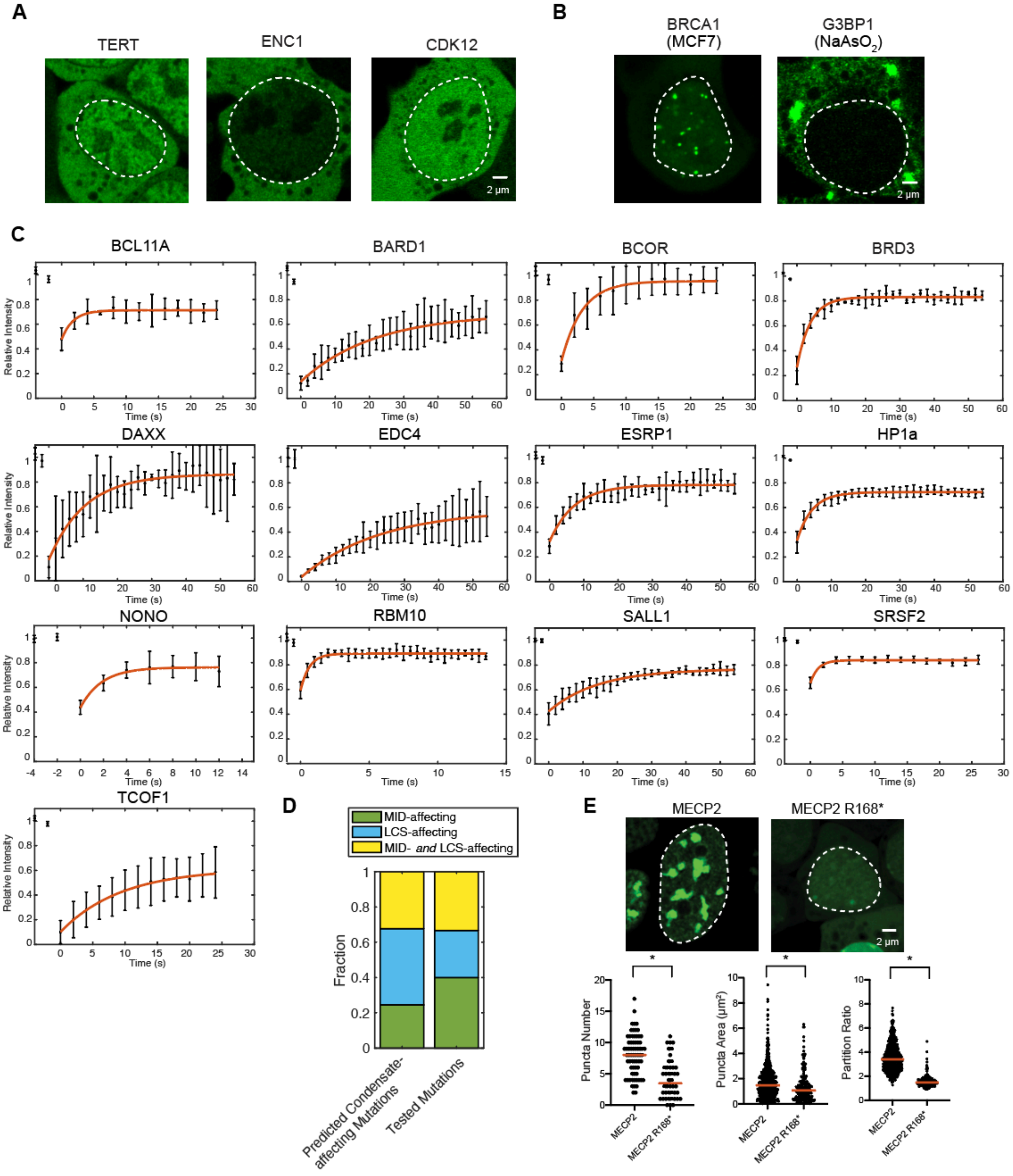


Extended Data Fig. 2: Condensate dysregulation across the spectrum of disease.

a. Proportion of all (top) predicted condensate-affecting mutations, those affecting MIDs (middle), or those affecting LCSs (bottom) (Extended Data Fig 1c.) that also affect protein features associated with canonical models of protein dysfunction (Methods, Fig. 1a). For mutations classifiable in multiple categories, starting with catalytic domains and moving left to right, mutations counted in one category were not counted again in subsequent categories. Proportion of mutations that were not mapped to any of these features are shown in red, showing that many mutations affecting LCSs (as opposed to those affecting MIDs) are not explained by alternative canonical models of protein dysfunction. Black dots indicate cumulative fraction moving from left to right.

b. Overlap of pathogenic LCS-affecting mutations and those mutations unexplained by canonical models or “remaining mutations” showing a significant overlap.

Extended Data Figure 3

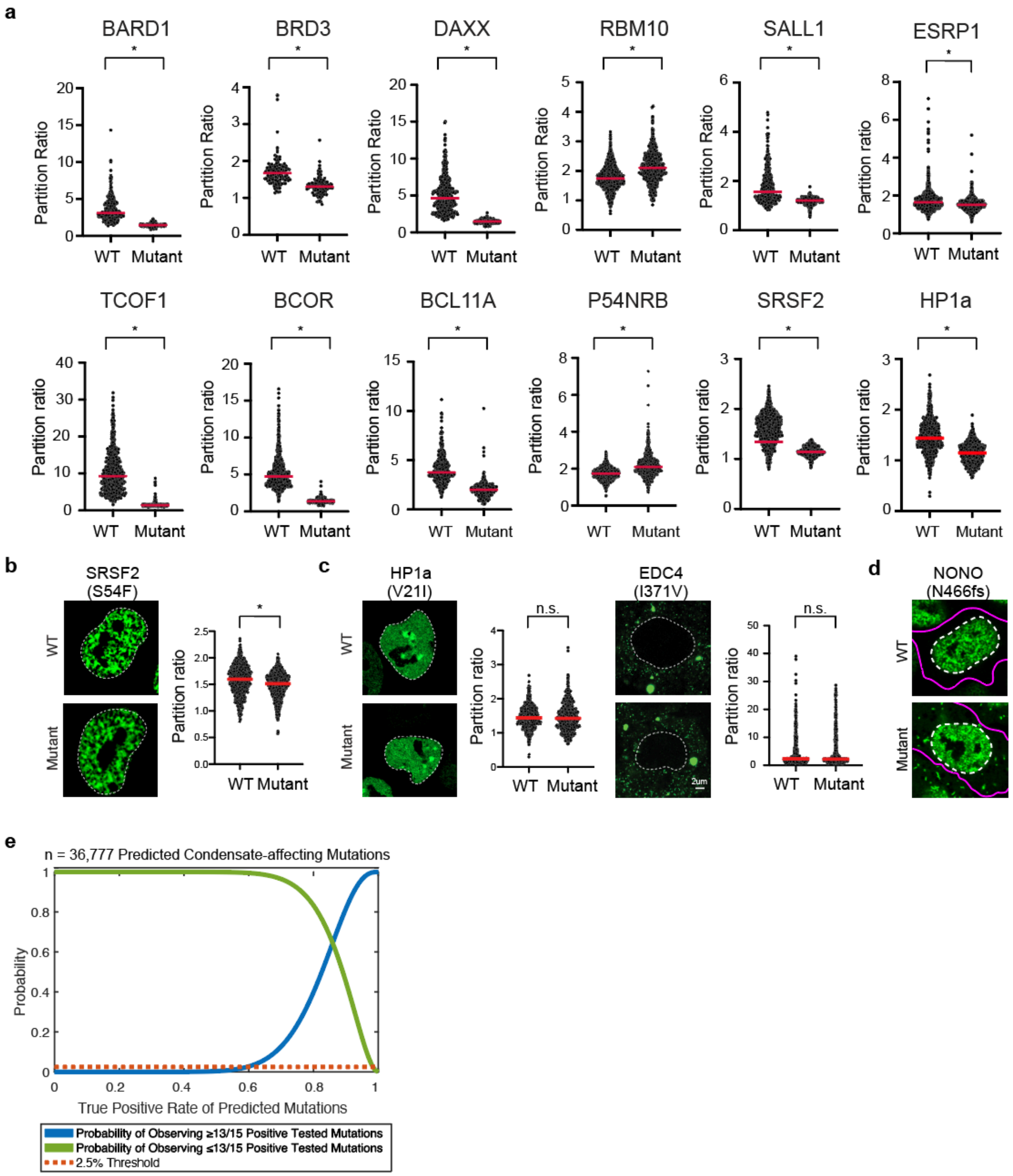


Extended Data Fig. 3: Pathogenic mutations in condensate-promoting features alter condensate properties in live cells.

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 Supplementary Table 5.

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. 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. *, P-value < 0.0001.

Extended Data Figure 4



Extended Data Fig. 4: Mutations in condensate-promoting features cause diverse condensate dysregulation phenotypes.

a. Quantification of partition ratios of wild-type and mutant mEGFP-tagged cell lines for indicated candidates. *, P-values < 0.0001. See also Supplementary Table 6.

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. Red dashed line indicates probability of 2.5% and denotes an estimate for the 95% confidence interval for the true catalog accuracy to be between 60-98%.

Supplementary Information

Supplemental Discussion

LCS vs. IDR

IDRs refer to regions in proteins that lack defined 3D structures (Das and Pappu, 2013; Lee et al., 2014), which are thought to be involved in diverse cellular and molecular functions (Babu et al., 2012; Zarin et al., 2019) and are known to be mutated in a broad range of diseases (Lu et al., 2015; Midic et al., 2009; Pajkos et al., 2011; Uversky et al., 2001, 2008). IDRs are also thought to be important drivers of phase separation (Banani et al., 2017; Choi et al., 2020; Elbaum-Garfinkle et al., 2015; Lin et al., 2015; Molliex et al., 2015; Nott et al., 2015; Patel et al., 2015; Shin and Brangwynne, 2017) through weak, multivalent, homo- or heterotypic interactions mediated by specific amino acid side chains found clustered or dispersed along the IDR (Brangwynne et al., 2015; Choi et al., 2020). In general, the term *low complexity sequence* is used to describe any region within a protein where the amino acid diversity is much less than the possible diversity of all 20 types of residues. In this study, we use *LCSs* specifically to refer to the types of low complexity sequences found in IDRs that are known to be involved in condensate formation (Table S1). Given that not all IDRs are thought to be capable of phase separating, we expect the LCS content of a protein defined in this way to be a much more specific indicator of condensate forming ability than IDR content.

Types of Pathogenic Mutations

Among the full set of pathogenic mutations, only a subset of mutation types was evaluated for their effect on condensate-promoting features. We chose to evaluate only the mutation types where the effect of the mutation on the condensate-promoting feature at the protein level could reasonably be predicted. These mutation types included missense, truncating (frameshift and nonsense), in-frame deletions, and in-frame insertions. Together these mutations accounted for over 98% of pathogenic variants.

Notably, we did not evaluate mutations that affect splicing (0.5% of pathogenic mutations; splice region, splice donor, and splice acceptor mutations). Such mutations could result in alternate or novel isoforms of the protein. Alternative splicing is a crucial regulatory mechanism for many genes (Barbosa-Morais et al., 2012; Ellis et al., 2012; Merkin et al., 2012) that can specifically affect IDRs (Buljan et al., 2012; Romero et al., 2006; Smith et al., 2020), and thus dysregulation thereof in a condensate-forming protein may affect phase separation and condensate properties (Batlle et al., 2020; Gueroussov et al., 2017; Tsang et al., 2020). However, the relevance of alternate isoforms of all proteins is not comprehensively understood, and protein feature annotations such as those for conserved domains or IDR predictions are not readily available for noncanonical, alternate isoforms in the data sources we used. For these reasons, we chose in this study to focus on protein isoforms considered to be the canonical isoforms (Consortium et al., 2020) which represent the best annotated isoforms across all proteins. These isoforms are selected based on criteria such as prevalence, similarity to homologs, and in the absence of other information, sequence length. This approach also omits analysis of novel isoforms (e.g. those resulting from retained introns), which can result in the introduction of premature stop codons in the transcript, leading to protein truncation. However, the absolute quantity of these mutations (≤0.5% of all pathogenic mutations) is negligible in comparison the truncation mutations that were included in the study (51% of all pathogenic mutations).

The remaining 0.6% of pathogenic mutations were not evaluated for a variety of reasons, including that they introduced confounding effects due to loss of the whole protein (start codon loss mutations, 0.3% of pathogenic mutations); that they represented complex changes to the protein sequence (e.g. deletion-insertions, 0.2% of pathogenic mutations); or that they were too infrequent among pathogenic mutations to represent a general mechanism of condensate dysregulation.

Canonical Models of Protein Dysfunction and Condensate Dysregulation

We analyzed the extent to which mutations in the catalog predicted to affect condensates could also be explained by additional mechanisms that are often considered in the conventional interpretations of how disease-associated mutations disrupt protein function and confer pathogenicity (Methods, Figure S2F, Table S7, Table S10L). These mechanisms include altered protein fold, catalytic activity, ligand interactions, and post-translational modifications. As we discuss below, many of these additional mechanisms affect molecular-scale protein properties that are known to influence condensate properties at the mesoscale. Moreover, protein features encoding these molecular-scale properties may overlap with and/or influence the properties of condensate-promoting features—for example, MIDs may participate in typical protein-protein interactions or contribute to condensation by contributing to multivalent interactions, and post-translational modifications may occur withing LCSs and regulate the interaction propensity of the LCS. Thus, we do not expect these additional mechanisms to be mutually exclusive with condensate dysregulation. Indeed, many of the pathogenic mutations in the literature known to disrupt condensates can simultaneously be explained by effects on molecular-scale protein properties. We consider several of these canonical models below and discuss how their disruption could affect condensates.

Truncation mutations that elicit NMD were considered to confound the loss of a condensate-promoting feature with the loss of the whole protein and therefore excluded from our analyses. However, we note that NMD is expected to reduce protein levels within the cell and therefore has the potential to reduce condensate formation, as decreases in protein concentration can in principle cause concentrations to fall below the threshold necessary to drive condensation. This may especially be a salient consideration for condensate-forming proteins involved in diseases driven by haploinsufficiency, where sensitivity to reduced protein dosage may reflect ineffective condensate formation.

Interaction motifs (e.g. SLiMs (Kumar et al., 2019)) are interaction motifs that often bind to MIDs and can therefore contribute to canonical protein-protein interactions but also to the valency of multivalent proteins and therefore contribute to condensate formation(Banani et al., 2016; Bouchard et al., 2018; Hastings and Boeynaems, 2021; Li et al., 2012; Su et al., 2016). Therefore, disease-associated mutations in SLiMs (Uyar et al., 2014) could conceivably also lead to condensate dysregulation.

NLSs direct a protein’s subcellular localization to the nucleus. NLSs within condensate-forming protein can regulate condensate formation by ensuring proper localization of a condensate-forming protein (Guo et al., 2018; Hofweber et al., 2018; Yoshizawa et al., 2018). Mutations in NLSs of a condensate-forming protein could therefore mislocalize the protein and the corresponding condensate (Schneider et al., 2020). NESs could conceivably affect condensates in analogous ways.

PTMs can serve as docking sites for interacting domains, and thus can regulate condensate formation by contributing to multivalent interactions (Dao et al., 2018; Gibson et al., 2019;

Hofweber and Dormann, 2019; Hofweber et al., 2018; Kim et al., 2019; Li et al., 2012; Monahan et al., 2017; Nott et al., 2015; Qamar et al., 2018; Rai et al., 2018; Saito et al., 2019; Su et al., 2016; Yasuda et al., 2020) or can regulate the localization of a molecule in a condensate (Guo et al., 2019; Kwon et al., 2013). In some systems, PTMs may serve as the predominant source of multivalent interactions (Banani et al., 2016; Sun et al., 2018) where they serve as the cognate ligands for MIDs. PTM sites are also known to be frequently mutated in disease (Reimand et al., 2015) and could dysregulate condensates if affected within condensate-forming proteins.

Because condensate formation and properties arise as a result of weak, multivalent interactions, it necessarily requires that molecular-scale protein properties that contribute to interactions are intact. In principle, any molecular scale property, including structural stability and catalysis (e.g. of a kinase that deposits a phosphate group important for a condensate-promoting interaction), that affects homotypic and heterotypic interactions can directly or indirectly affect condensate properties at the mesoscale. We therefore propose that condensate dysregulation does not exclude 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.

Selection of Candidates and Mutations for Experimental Validation

We performed experimental validation by selecting 25 proteins from the catalog (Table S5). Candidate proteins in the panel 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 (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.

The 25 proteins were initially tested for punctate localization when ectopically expressed with a GFP tag in mESCs, our chosen experimental system. 12 of these proteins (48%) did not show punctate localization in mESCs (Figure S3A, Table S5). 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 result suggests that mESCs in resting state may not provide a conducive environment to observe condensation for all proteins. Indeed, a subset of these candidates did exhibit punctate localization in other cell types or under stressed conditions (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. Given these considerations, we assume that statistical analyses performed to compute the predictive accuracy of the full catalog (Figure S4E) provide reasonable estimates.

Table S1. LCSs associated with condensate formation.

LCS	Example(s)	References
Acidic patch	NPM1	(Mitrea et al., 2016)
Basic patch	MECP2	(Li et al., 2020)
Alanine-rich region	HOXA13, HOXD13, RUNX2	(Basu et al., 2020)
Arginine-rich region	SURF6	(Mitrea et al., 2018)
Glutamine-rich region	HTT, Whi3	(Peskett et al., 2018; Zhang et al., 2015)
Histidine-rich region	HBP-1, DYRK1a	(Gabryelczyk et al., 2019; Lu et al., 2018)
Proline-rich region	UBQLN2	(Dao et al., 2018)
Serine-rich region	MED1	(Sabari et al., 2018)
FG-rich region	NUP98	(Frey et al., 2006; Schmidt and Görlich, 2015)
RG-rich region	LAF-1, DDX4, FUS	(Elbaum-Garfinkle et al., 2015; Nott et al., 2015; Wang et al., 2018)
Prion-like domain	FUS, HNRNPA1	(Lancaster et al., 2014; Martin et al., 2020; Wang et al., 2018)
pi-pi interacting residues	various	(Vernon et al., 2018)
LARKS	FUS, HNRNPA1	(Hughes et al., 2018)

List of LCSs within IDRs that have been associated with condensate formation. Corresponding proteins studies in which the role of the LCS type was demonstrated are also listed. FG, phenylalanine-glycine; RG, arginine-glycine; LARKS, low-complexity, aromatic-rich kinked segments.

Table S2. Optimization of LCS mapping against a set of benchmark proteins.

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	HOXD13	57-71	(Basu et al., 2020)	18-39,46-75,92-126	0.80	1.0	0.73	0.03
Arginine-rich 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, 330-331, 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
Glutamine-rich region	Gln	HTT	17-40	(Peskett et al., 2018)	15-41, 49-68, 498-506, 593-601	1.00	0.5	1.0	3×10^{-4}
Histidine-rich region	His	DYRK1a	590-616	(Lu et al., 2018)	531-541, 596-622, 648-656	0.97	0.1	1.0	0.22
Proline-rich region	Pro	UBQLN2	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, 1021-1029, 1077-1150, 1163-1171, 1223-1285, 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, 113-153, 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

Protein regions corresponding to previously characterized LCSs identified in prior studies are listed alongside regions identified by our computational mapping approach (Methods). 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). TPR, true positive rate. FPR, false positive rate. Refs., references.

Table S3. Proteome-wide LCS mapping statistics.

LCS	Total # of regions identified	Median Size (residues) [IQR]	Total # Proteins identified
Acidic patch	42,599	7 [5,9]	12,555
Basic patch	38,983	7 [5,9]	12,945
Alanine-rich region	507	6 [5,7]	446
Arginine-rich region	49,038	8 [6,9]	14,355
Glutamine-rich region	5,427	6 [5,7]	3,418
Histidine-rich region	79,539	9 [8,10]	16,890
Proline-rich region	82,779	8 [6,11]	16,202
Serine-rich region	95,575	8 [6,11]	17,219
FG-rich region	31,212	7 [6,8]	12,116
RG-rich region	78,056	8 [6,9]	16,768
Prion-like domain	361	119 [83,182]	361
pi-interacting residues	9,636	2 [1,5]	2,119
LARKS	659	12 [8,19]	324

Summary statistics of mapping of LCS regions across the proteome (Methods). IQR, interquartile range.

Table S4. Pathological mutations known to affect condensates.

Protein	Diseases	Mutations	References	Notes for uncaptured mutations
ANXA11	Amyotrophic lateral sclerosis	**Asp40Gly**, **Gly38Arg**, **His390Pro**, **Arg456His**	(Nahm et al., 2020)	
αSYNUCLEIN	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; Valentin-Vega 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
UBQLN2	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 S5. Selected protein candidates and mutations used in experimental tests.

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
HP1α	(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*.

Table S6. Summary statistics from experimental tests.

Metric	Protein	Median WT	Median Mutant	*p*-value
Partitioning Ratio	BRD3	1.687	1.418	<0.0001*
	RBM10	1.745	2.101	<0.0001*
	BARD1	3.109	1.454	<0.0001*
	DAXX	4.648	1.494	<0.0001*
	SALL1	1.559	1.209	<0.0001*
	MECP2	3.394	1.478	<0.0001*
	BCL11a	3.751	1.996	<0.0001*
	BCOR	4.728	1.374	<0.0001*
	NONO	1.734	2.098	<0.0001*
	ESRP1	1.646	1.508	<0.0001*
	TCOF	9.245	1.377	<0.0001*
	SRSF2 (Ser54Phe)	1.598	1.513	<0.0001*
	SRSF2 (Pro95His)	1.621	1.147	<0.0001*
	HP1α (Trp142Cys)	1.44	1.147	<0.0001*
	HP1α (Val21Ile)	1.44	1.425	0.2326
	EDC4	2.303	2.214	0.574
Number of Puncta	BRD3	3	3	0.9629
	RBM10	15.5	20.5	0.8193
	BARD1	8	4	0.0004*
	DAXX	6	7	0.7214
	SALL1	4	7	0.1137
	MECP2	8	3.5	<0.0001*
	BCL11a	11	4	0.0028
	BCOR	10	5.5	0.0001*
	NONO	27	9.5	<0.0001*
	ESRP1	9	8.5	0.8586
	TCOF	5	9	0.0769

	SRSF2 (Ser54Phe)	14	20	0.9674
	SRSF2 (Pro95His)	14	11.5	0.4042
	HP1α (Trp142Cys)	7	4	0.0615
	HP1α (Val21Ile)	7	4	0.004
	EDC4	5	6	0.1257
Puncta area (μm^2)	BRD3	0.3058	0.2903	0.735
	RBM10	0.2994	0.3284	0.0748
	BARD1	0.3938	0.2105	<0.0001*
	DAXX	0.3684	0.2867	0.0015
	SALL1	0.4155	0.1796	<0.0001*
	MECP2	1.468	1.071	<0.0001*
	BCL11a	0.4192	0.2404	<0.0001*
	BCOR	0.4664	0.2323	<0.0001*
	NONO	0.2994	0.2504	0.0002*
	ESRP1	0.2667	0.2522	0.272
	TCOF	0.7277	0.245	<0.0001*
	SRSF2 (Ser54Phe)	0.2758	0.2522	0.0039
	SRSF2 (Pro95His)	0.2758	0.2277	0.0007*
	HP1α (Trp142Cys)	0.2922	0.2885	0.9005
	HP1α (Val21Ile)	0.2922	0.2486	0.0088
	EDC4	1.925	1.36	0.0065

Summary statistics for image analyses. Metrics include median partition ratio, median number of condensates per cell, and median area of condensates per candidate. Mann Whitney U test was used for all statistical tests. * Indicates p-value < Bonferroni adjusted α-value. Bonferroni adjusted α-value = 0.00119.

Table S7. Additional mechanisms to interpret mutations that alter condensate properties in cells.

Protein	Mutation	MID	LCS	Catalytic Domain	Interaction Domain	Structural Element	Interaction Motif	Protein Processing	PTM	NLS/ NES	Other Functional
MECP2	Arg186Ter		•								
BARD1	Arg406Ter	•			•	•	•		•	•	
BCL11A	Gln177Ter	•	•		•	•	•		•	•	
BCOR	Tyr657Ter	•	•		•	•	•		•	•	•
BRD3	Phe334Ser	•				•	•				
HP1α	Trp142Cys	•			•	•					
DAXX	Arg318Ter		•			•	•		•	•	•
ESRP1	Leu259Val	•					•				
NONO	Asn466fs		•						•	•	
RBM10	Val354Met	•			•						
SALL1	Ser372Ter	•	•		•		•		•	•	
SRSF2	Ser54Phe	•	•		•						
SRSF2	Pro95His		•								
TCOF	Gln55Ter		•						•	•	

Black dot (•) indicates protein features predicted to be affected by corresponding mutation. Listed mutations were experimentally demonstrated to confer condensate dysregulation (Figure S3C, Figure S4B). Features listed correspond to condensate-promoting properties as well as additional molecular scale properties typically considered in the conventional interpretation of how mutations disrupt protein function. See also Supplemental Discussion, *Canonical Models of Protein Dysfunction and Condensate Dysregulation*.

Table S8. List of TCGA and TARGET studies obtained from cBioPortal.

Identifier in cBioPortal	Study Name	Consortium
ucec_tcga_pan_can_atlas_2018	Uterine Corpus Endometrial Carcinoma	TCGA
skcm_tcga_pan_can_atlas_2018	Skin Cutaneous Melanoma	
coadread_tcga_pan_can_atlas_2018	Colorectal Adenocarcinoma	
luad_tcga_pan_can_atlas_2018	Lung Adenocarcinoma	
stad_tcga_pan_can_atlas_2018	Stomach Adenocarcinoma	
lusc_tcga_pan_can_atlas_2018	Lung Squamous Cell Carcinoma	
blca_tcga_pan_can_atlas_2018	Bladder Urothelial Carcinoma	
brca_tcga_pan_can_atlas_2018	Breast Invasive Carcinoma	
hnsc_tcga_pan_can_atlas_2018	Head and Neck Squamous Cell Carcinoma	
cesc_tcga_pan_can_atlas_2018	Cervical Squamous Cell Carcinoma	
gbm_tcga_pan_can_atlas_2018	Glioblastoma Multiforme	
lihc_tcga_pan_can_atlas_2018	Liver Hepatocellular Carcinoma	
ov_tcga_pan_can_atlas_2018	Ovarian Serous Cystadenocarcinoma	
lgg_tcga_pan_can_atlas_2018	Brain Lower Grade Glioma	
esca_tcga_pan_can_atlas_2018	Esophageal Adenocarcinoma	
prad_tcga_pan_can_atlas_2018	Prostate Adenocarcinoma	
paad_tcga_pan_can_atlas_2018	Pancreatic Adenocarcinoma	
kirp_tcga_pan_can_atlas_2018	Kidney Renal Papillary Cell Carcinoma	
kirc_tcga_pan_can_atlas_2018	Kidney Renal Clear Cell Carcinoma	
sarc_tcga_pan_can_atlas_2018	Sarcoma	
thca_tcga_pan_can_atlas_2018	Thyroid Carcinoma	
acc_tcga_pan_can_atlas_2018	Adrenocortical Carcinoma	
ucs_tcga_pan_can_atlas_2018	Uterine Carcinosarcoma	
laml_tcga_pan_can_atlas_2018	Acute Myeloid Leukemia	
dlbc_tcga_pan_can_atlas_2018	Diffuse Large B-Cell Lymphoma	
thym_tcga_pan_can_atlas_2018	Thymoma	
meso_tcga_pan_can_atlas_2018	Mesothelioma	
kich_tcga_pan_can_atlas_2018	Kidney Chromophobe	
tgct_tcga_pan_can_atlas_2018	Testicular Germ Cell Tumors	
chol_tcga_pan_can_atlas_2018	Cholangiocarcinoma	
pcpg_tcga_pan_can_atlas_2018	Pheochromocytoma and Paraganglioma	
uvm_tcga_pan_can_atlas_2018	Uveal Melanoma	
wt_target_2018_pub	Pediatric Wilms' Tumor	TARGET
all_phase2_target_2018_pub	Pediatric Acute Lymphoid Leukemia	
aml_target_2018_pub	Pediatric Acute Myeloid Leukemia	
nbl_target_2018_pub	Pediatric Neuroblastoma	
rt_target_2018_pub	Pediatric Rhabdoid Tumor	

Study identifiers and associated cancer types are shown.

Table S9. Definitions of pathogenicity used in this study.

Disease Class	Variant Database	Pathogenicity Assessment	Designation	Number of Variants
Mendelian	ClinVar	ClinVar Clinical Significance	“Pathogenic”	70,860
			“Likely pathogenic”	36,177
	HGMD	HGMD Variant Class	“DM”	179,537
			“DM?”	50,007
Cancer	GENIE	CGI	All	2,615
		CIViC	All	379
		OncoKB	“Oncogenic”	1,734
			“Likely oncogenic”	50,795
			“Predicted oncogenic”	1,349
	TCGA and TARGET	CGI	All	1,266
		CIViC	All	211
		OncoKB	“Oncogenic”	603
			“Likely oncogenic”	13,129
			“Predicted oncogenic”	541

Definitions of pathogenicity used in this study. Mendelian and cancer variants were obtained from indicated data sources and assessed as described in Methods, and pathogenic variants were defined using the indicated designations in the source datasets or the indicated variant knowledgebases. Terms in quotes represent designations as indicated in the source datasets.

Table S10. Results from analyses reported in this study.

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. The file can be found here: https://drive.google.com/drive/folders/1xQjbffM8lTPWNG0WfU7m5lhlrt2gSQTH?usp=sharing.

Materials and Methods

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 S10A). 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 S10B.

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 S10D and Table S10L.

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 S2F and Table S7, 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 S2F and Table S7 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 S10C.

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 S2). 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 S2). 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 S3, and Table S10D.

Mapping 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, Table S8). 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 or of variant class for HGMD variants (Figure S1A). For cancer variants, pathogenicity 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 (Chakravarty et al., 2017). Statistics for these pathogenicity designations are shown in Table S9. The resulting 322,825 pathogenic mutations analyzed are shown in Table S10E.

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 they represented complex changes to the protein sequence (e.g. deletion-insertions, 0.2% of pathogenic mutations) (Supplemental Discussion, *Types 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 S10F 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 S10G.

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 or 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 S10H. 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.

mESC Cell Culture and Cell Line Generation

V6.5 murine embryonic stem cells (mESCs) were a gift from R. Jaenisch. 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).

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. 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 hygromcin (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 v.3.6.9. A Python package, *cellpose*, 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.

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

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.

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.

Statistical Analyses

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.

Data Availability

Relevant data supporting the findings of this study are available within the paper and Supplementary Information. Source datasets used for analyses are indicated within the Methods.

Code Availability

Code used for analysis can be obtained from the authors upon request.

References

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

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.

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.

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.

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.

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.

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.

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.

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.

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-.

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.

Bienz, M. (2020). Head-to-Tail Polymerization in the Assembly of Biomolecular Condensates. Cell *182*, 799–811.

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-.

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.

Boija, A., Klein, I.A., and Young, R.A. (2021). Biomolecular condensates and cancer. Cancer Cell *39*, 174–192.

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.

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.

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.

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.

Cai, D., Liu, Z., and Lippincott-Schwartz, J. (2021). Biomolecular Condensates and Their Links to Cancer Progression. Trends Biochem Sci.

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.

Case, L.B., Ditlev, J.A., and Rosen, M.K. (2019). Regulation of Transmembrane Signaling by Phase Separation. Annu Rev Biophys *48*, 1–30.

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.

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.

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.

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.

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.

Consortium, T.A.P.G. (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov *7*, 818–831.

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.

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.

Du, M., and Chen, Z. (2018). DNA-induced liquid phase condensation of cGAS activates innate immune signaling. Science *361*, eaat1022.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

King, M.R., and Petry, S. (2020). Phase separation of TPX2 enhances and spatially coordinates microtubule nucleation. Nat Commun *11*, 270.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Letunic, I., Khedkar, S., and Bork, P. (2020). SMART: recent updates, new developments and status in 2020. Nucleic Acids Res *49*, 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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Milovanovic, D., Wu, Y., Bian, X., and Camilli, P.D. (2018). A liquid phase of synapsin and lipid vesicles. Science *361*, eaat5671.

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.

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-.

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.

Nedelsky, N.B., and Taylor, J.P. (2019). Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease. Nat Rev Neurol *15*, 272–286.

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.

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.

Pawson, T., and Nash, P. (2003). Assembly of Cell Regulatory Systems Through Protein Interaction Domains. Science *300*, 445–452.

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.

Ramaswami, M., Taylor, J.P., and Parker, R. (2013). Altered Ribostasis: RNA-Protein Granules in Degenerative Disorders. Cell *154*, 727–736.

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.

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.

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.

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.

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.

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.

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.

Sheu-Gruttadauria, J., and MacRae, I.J. (2018). Phase Transitions in the Assembly and Function of Human miRISC. Cell *173*, 946-957.e16.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*, eaaf4382.

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*.

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.

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.

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.

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.

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.

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.

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*.

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.

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.

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.

Viny, A.D., and Levine, R.L. (2020). Drug modulation by nuclear condensates. Sci New York N Y *368*, 1314–1315.

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.

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.

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.

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.

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.

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.

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.

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*.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res *21*, 564–578.

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.

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.

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.

Supplemental References

Babu, M.M., Kriwacki, R.W., and Pappu, R.V. (2012). Versatility from Protein Disorder. Science *337*, 1460–1461.

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.

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.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Çolak, R., et al. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science *338*, 1587–1593.

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.

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.

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.

Brangwynne, C.P., Tompa, P., and Pappu, R.V. (2015). Polymer physics of intracellular phase transitions. Nat Phys *11*, 899–904.

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.

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.

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.

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.

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.

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.

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.

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.

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

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C.-H., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. P Natl Acad Sci Usa *112*, 7189–7194.

Ellis, J.D., Barrios-Rodiles, M., Çolak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M., et al. (2012). Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. Mol Cell *46*, 884–892.

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.

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.

Frey, S., Richter, R.P., and Görlich, D. (2006). FG-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties. Science *314*, 815–817.

Gabryelczyk, B., Cai, H., Shi, X., Sun, Y., Swinkels, P.J.M., Salentinig, S., Pervushin, K., and Miserez, A. (2019). Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides. Nat Commun *10*, 5465.

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.

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.

Guo, L., Kim, H.J., Wang, H., Monaghan, J., Freyermuth, F., Sung, J.C., O'Donovan, K., Fare, C.M., Diaz, Z., Singh, N., et al. (2018). Nuclear-Import Receptors Reverse Aberrant Phase Transitions of RNA-Binding Proteins with Prion-like Domains. Cell *173*, 677-692.e20.

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.

Hastings, R.L., and Boeynaems, S. (2021). Designer condensates: a toolkit for the biomolecular architect. J Mol Biol 166837.

Hofweber, M., and Dormann, D. (2019). Friend or foe—Post-translational modifications as regulators of phase separation and RNP granule dynamics. J Biol Chem *294*, 7137–7150.

Hofweber, M., Hutten, S., Bourgeois, B., Spreitzer, E., Niedner-Boblenz, A., Schifferer, M., Ruepp, M.-D., Simons, M., Niessing, D., Madl, T., et al. (2018). Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. Cell *173*, 706-719.e13.

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.

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.

Kim, T.H., Tsang, B., Vernon, R.M., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2019). Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. Science *365*, 825–829.

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.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-Regulated Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains. Cell *155*, 1049–1060.

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.

Lee, R., Buljan, M., Lang, B., Weatheritt, R., Daughdrill, G., Dunker, A., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D., et al. (2014). Classification of Intrinsically Disordered Regions and Proteins. Chem Rev *114*, 6589–6631.

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.

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.

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.

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.

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.

Lu, H.-C., Chung, S.S., Fornili, A., and Fraternali, F. (2015). Anatomy of protein disorder, flexibility and disease-related mutations. Frontiers Mol Biosci *2*, 47.

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.

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.

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.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science *338*, 1593–1599.

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.

Midic, U., Oldfield, C.J., Dunker, A.K., Obradovic, Z., and Uversky, V.N. (2009). Protein disorder in the human diseasome: unfoldomics of human genetic diseases. Bmc Genomics *10*, S12.

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.

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.

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.

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.

Monahan, Z., Ryan, V.H., Janke, A.M., Burke, K.A., Rhoads, S.N., Zerze, G.H., O'Meally, R., Dignon, G.L., Conicella, A.E., Zheng, W., et al. (2017). Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. Embo J *36*, 2951–2967.

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.

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.

Pajkos, M., Mészáros, B., Simon, I., and Dosztányi, Z. (2011). Is there a biological cost of protein disorder? Analysis of cancer-associated mutations. Mol Biosyst *8*, 296–307.

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.

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.

Qamar, S., Wang, G., Randle, S.J., Ruggeri, F.S., Varela, J.A., Lin, J.Q., Phillips, E.C., Miyashita, A., Williams, D., Ströhl, F., et al. (2018). FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. Cell *173*, 720-734.e15.

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.

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.

Reimand, J., Wagih, O., and Bader, G.D. (2015). Evolutionary Constraint and Disease Associations of Post-Translational Modification Sites in Human Genomes. Plos Genet *11*, e1004919.

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.

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.

Saito, M., Hess, D., Eglinger, J., Fritsch, A.W., Kreysing, M., Weinert, B.T., Choudhary, C., and Matthias, P. (2019). Acetylation of intrinsically disordered regions regulates phase separation. Nat Chem Biol *15*, 51–61.

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.

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.

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.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*, eaaf4382.

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*.

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.

Sun, D., Wu, R., Zheng, J., Li, P., and Yu, L. (2018). Polyubiquitin chain-induced p62 phase separation drives autophagic cargo segregation. Cell Res *28*, 405–415.

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.

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.

Uversky, V.N., Iakoucheva, L.M., and Dunker, A.K. (2001). eLS.

Uversky, V.N., Oldfield, C.J., and Dunker, A.K. (2008). Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. Annu Rev Biophys *37*, 215–246.

Uyar, B., Weatheritt, R.J., Dinkel, H., Davey, N.E., and Gibson, T.J. (2014). Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? Mol Biosyst *10*, 2626–2642.

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.

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.

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.

Yasuda, S., Tsuchiya, H., Kaiho, A., Guo, Q., Ikeuchi, K., Endo, A., Arai, N., Ohtake, F., Murata, S., Inada, T., et al. (2020). Stress- and ubiquitylation-dependent phase separation of the proteasome. Nature *578*, 296–300.

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.

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.

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*.

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

Zhang, H., Elbaum-Garfinkle, S., Langdon, E.M., Taylor, N., Occhipinti, P., Bridges, A.A., Brangwynne, C.P., and Gladfelter, A.S. (2015). RNA Controls PolyQ Protein Phase Transitions. Mol Cell *60*, 220–230.

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.

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

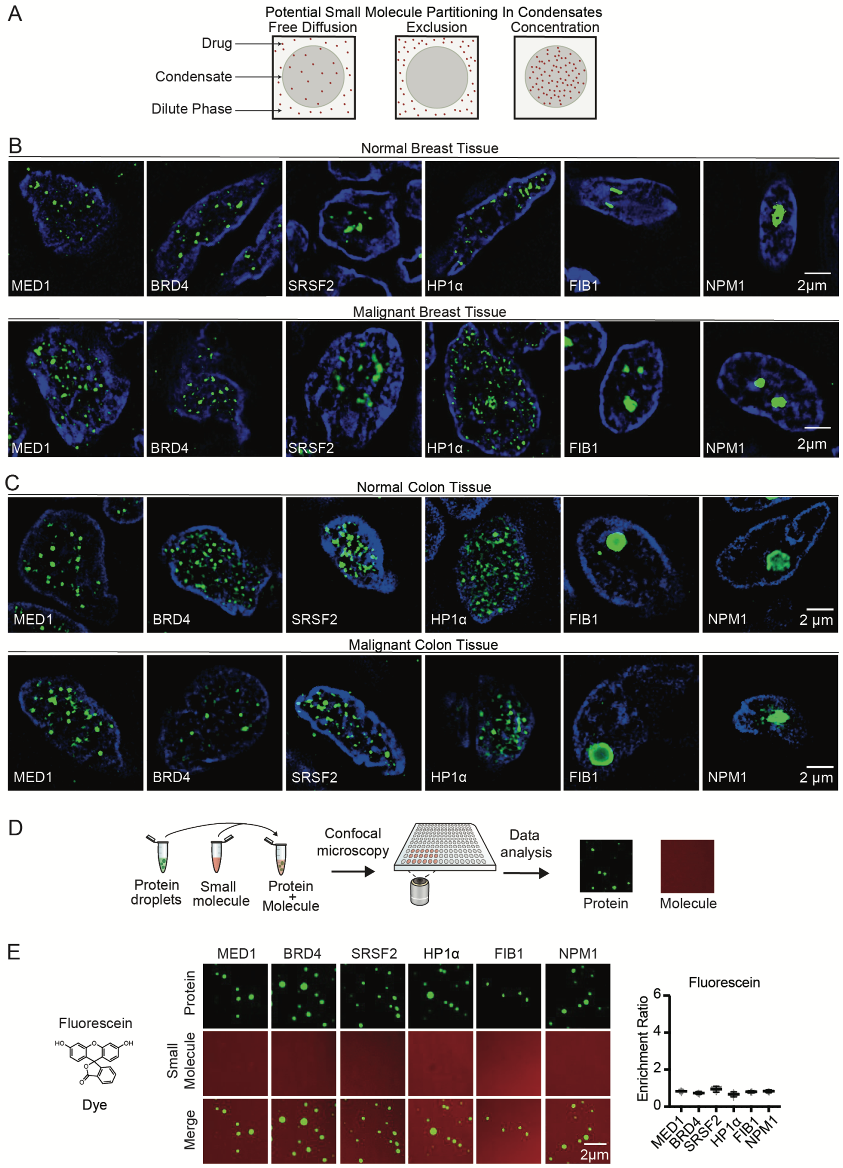


Fig. 1. Nuclear condensates in human tissue and in vitro. (A) Model illustrating potential behaviors of small molecules in nuclear condensates. (B-C) Immunofluorescence of scaffold proteins of various nuclear condensates in tissue biopsies from benign and malignant human breast (B), and benign and malignant 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

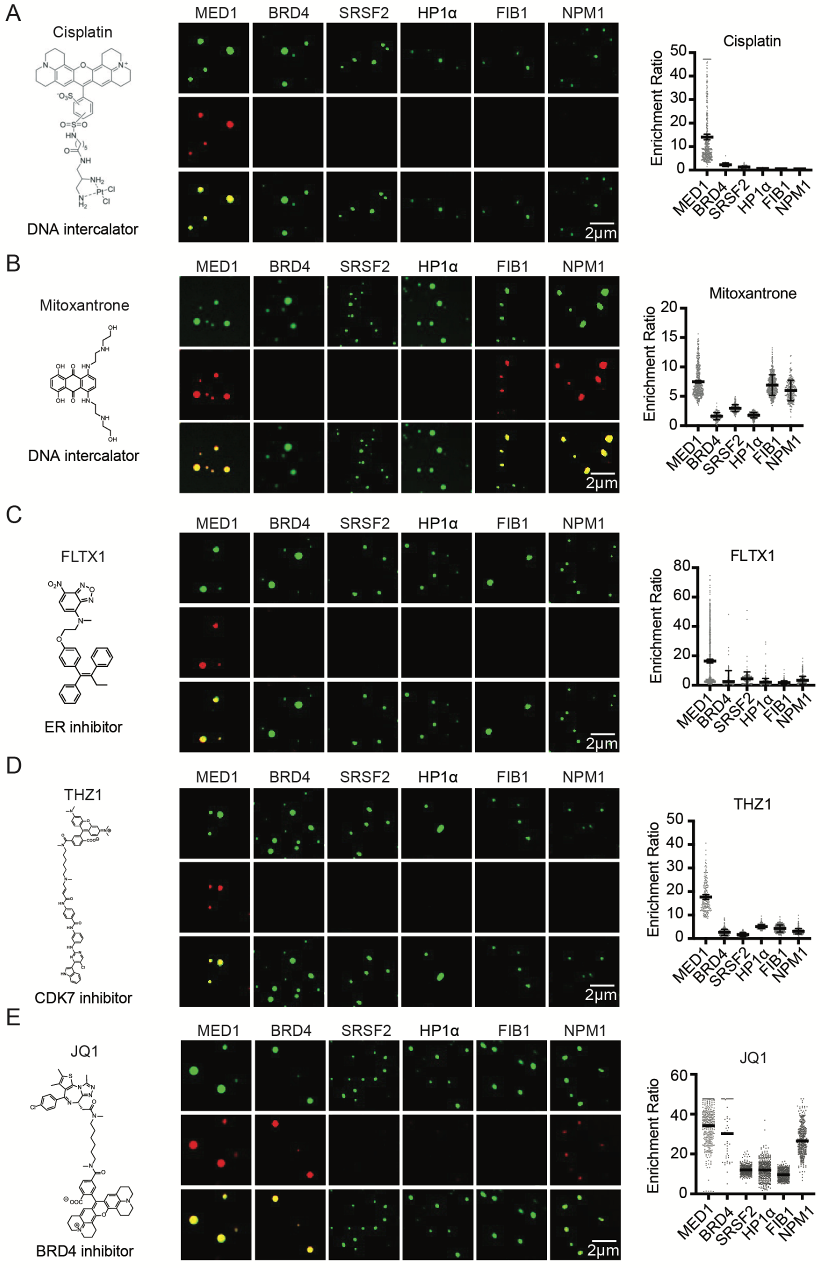


Fig. 2. The partitioning behavior of small molecule drugs in nuclear condensates in a droplet assay. Six nuclear condensates formed in 125mM NaCl and 10% PEG, with 10μM protein treated with either (A) 5μM Cisplatin-TMR, (B) 50 μM Mitoxantrone, (C) 100μM FLTX1, (D) 5μM THZ1-TMR, or (D) 1μM JQ1-ROX imaged 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

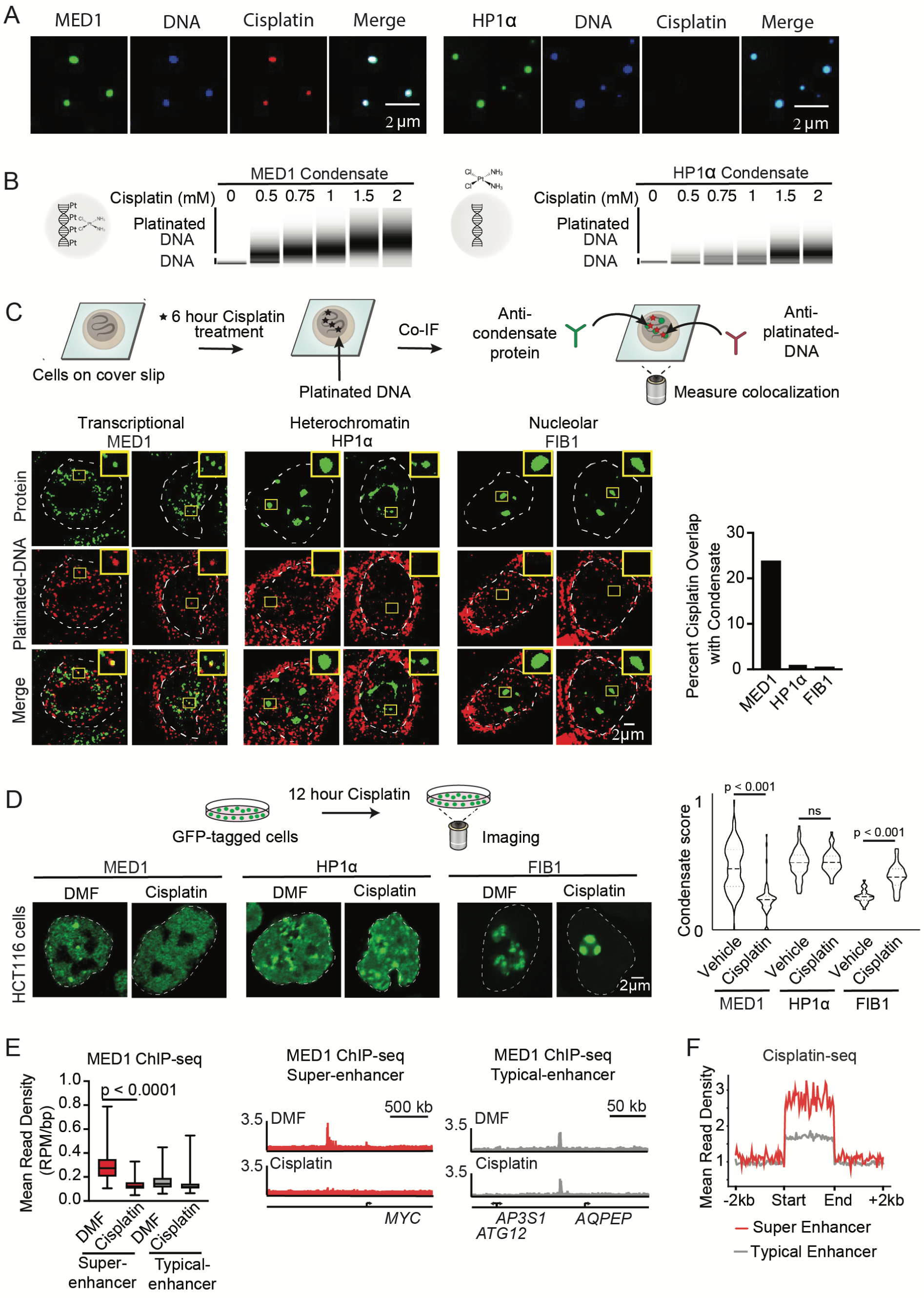


Fig. 3. Small molecule concentration within condensates influences drug activity. (A) In vitro droplet assay of MED1 and HP1α condensates formed in 125mM NaCl and 10% PEG, 5nM of 450bp DNA, 10μM MED1, and 5μM cisplatin-TR, imaged at 150x on a confocal fluorescent microscope (see also Figure S15). (B)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

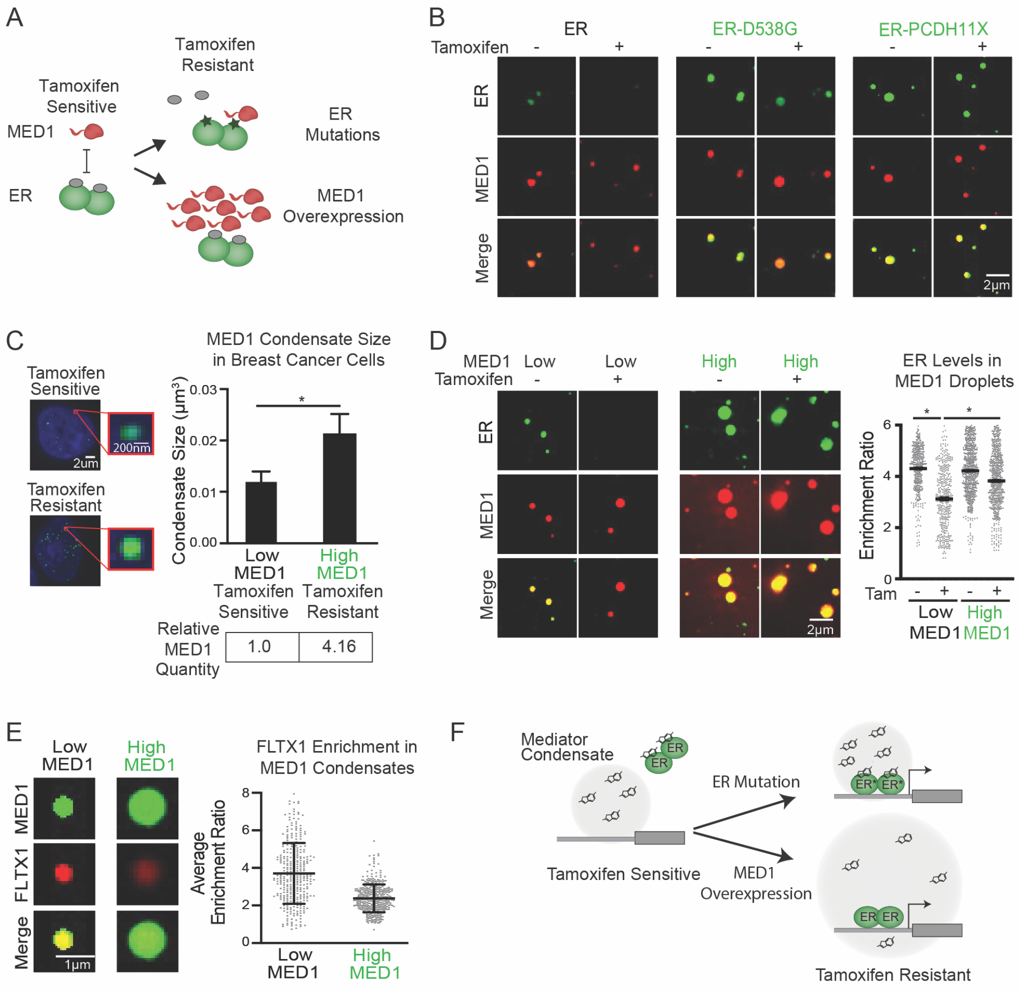


Fig. 4. Tamoxifen action and resistance in MED1 condensates. (A) Schematic showing tamoxifen resistance by ER mutation and MED1 overexpression in breast cancer. (B) In vitro droplets assay of the indicated form of GFP-tagged ER in the presence of estrogen, +/- 100µM tamoxifen. Droplets are formed in 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).

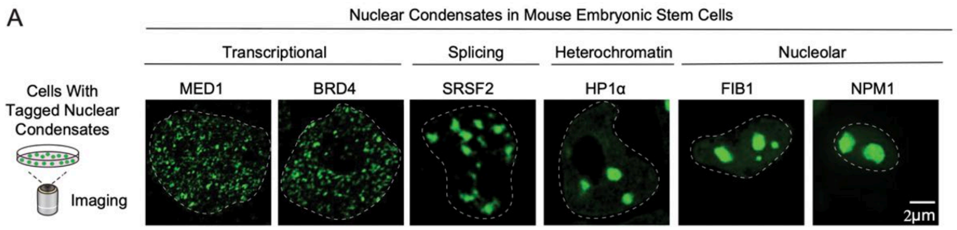
Our results show that drugs partition selectively into condensates, that this can occur through physicochemical properties that exist independent of their molecular targets, and that cells can develop resistance to drugs through condensate altering mechanisms. This may explain the surprising observation 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	

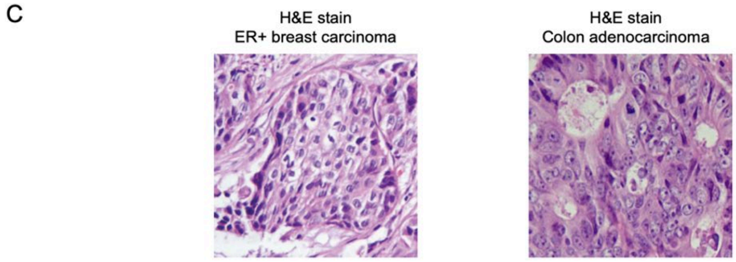
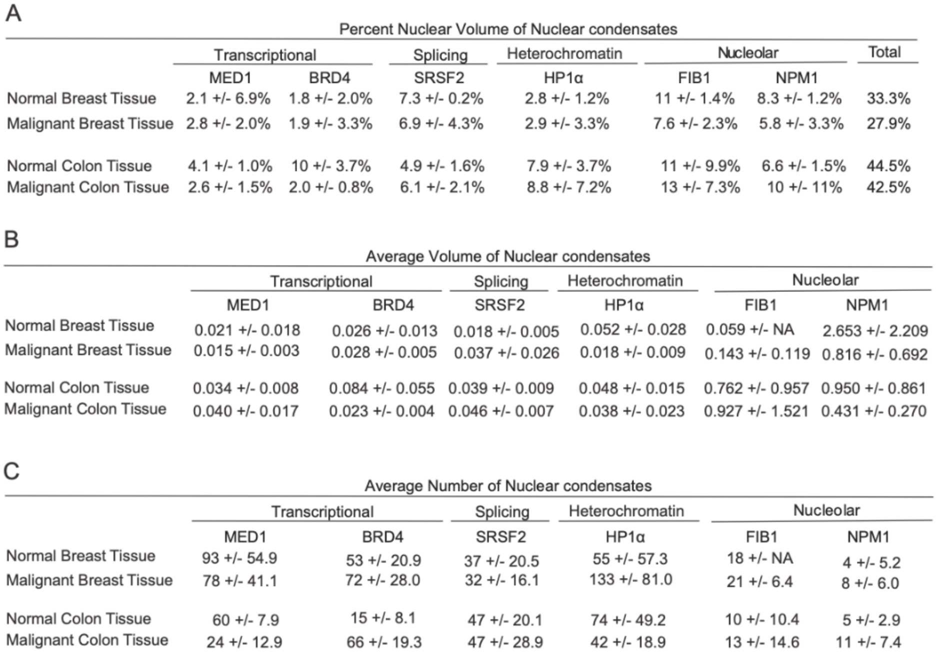


Figure S1. Nuclear condensates in cell lines and human tumor tissue. (A) Mouse embryonic stem cells expressing either endogenously mEGFP-tagged proteins (MED1, BRD4, SRSF2), mCherry-tagged proteins (HP1α) or transfected with constructs expressing GFP-tagged proteins (NPM1, FIB1) were imaged by confocal 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

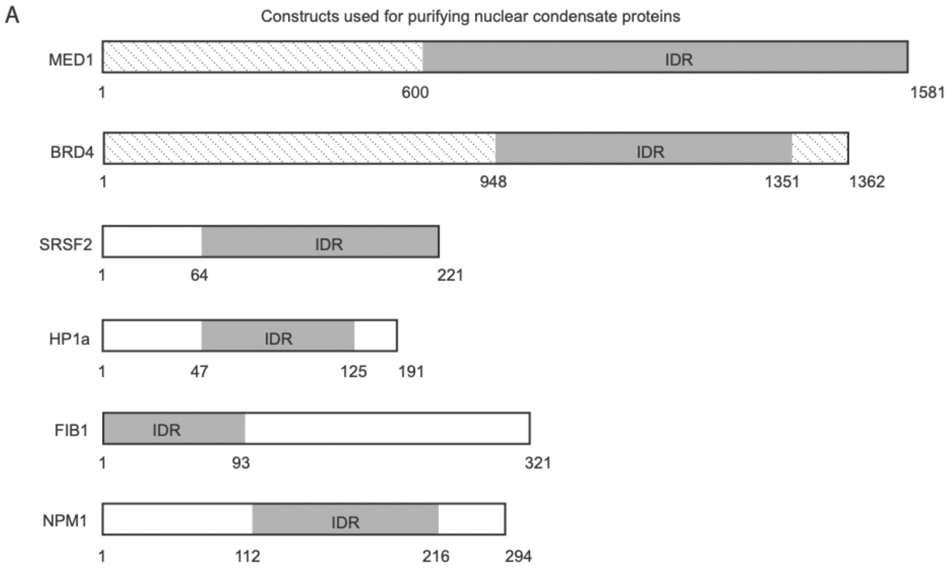
C

Average Number of Nuclear condensates

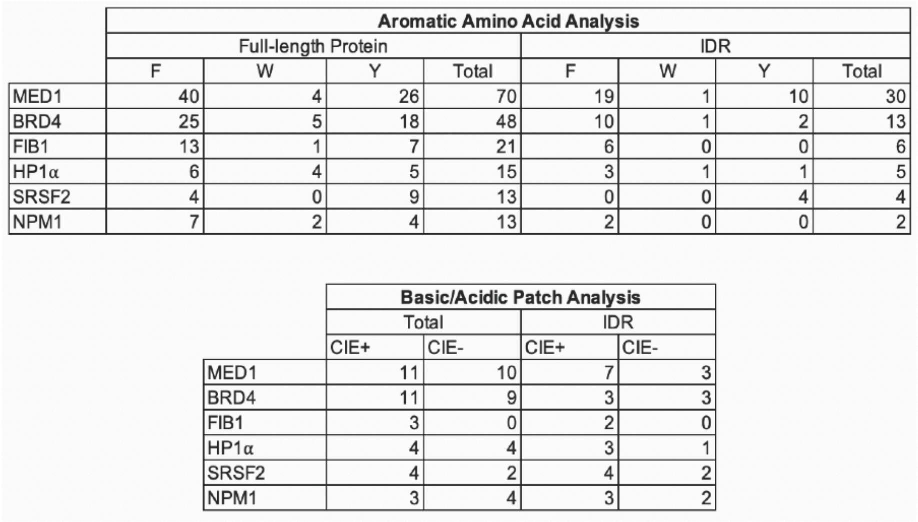
	Transcriptional		Splicing	Heterochromatin	Nucleolar	
	MED1	BRD4	SRSF2	HP1α	FIB1	NPM1
Normal Breast Tissue	93 +/- 54.9	53 +/- 20.9	37 +/- 20.5	55 +/- 57.3	18 +/- NA	4 +/- 5.2
Malignant Breast Tissue	78 +/- 41.1	72 +/- 28.0	32 +/- 16.1	133 +/- 81.0	21 +/- 6.4	8 +/- 6.0
Normal Colon Tissue	60 +/- 7.9	15 +/- 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



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

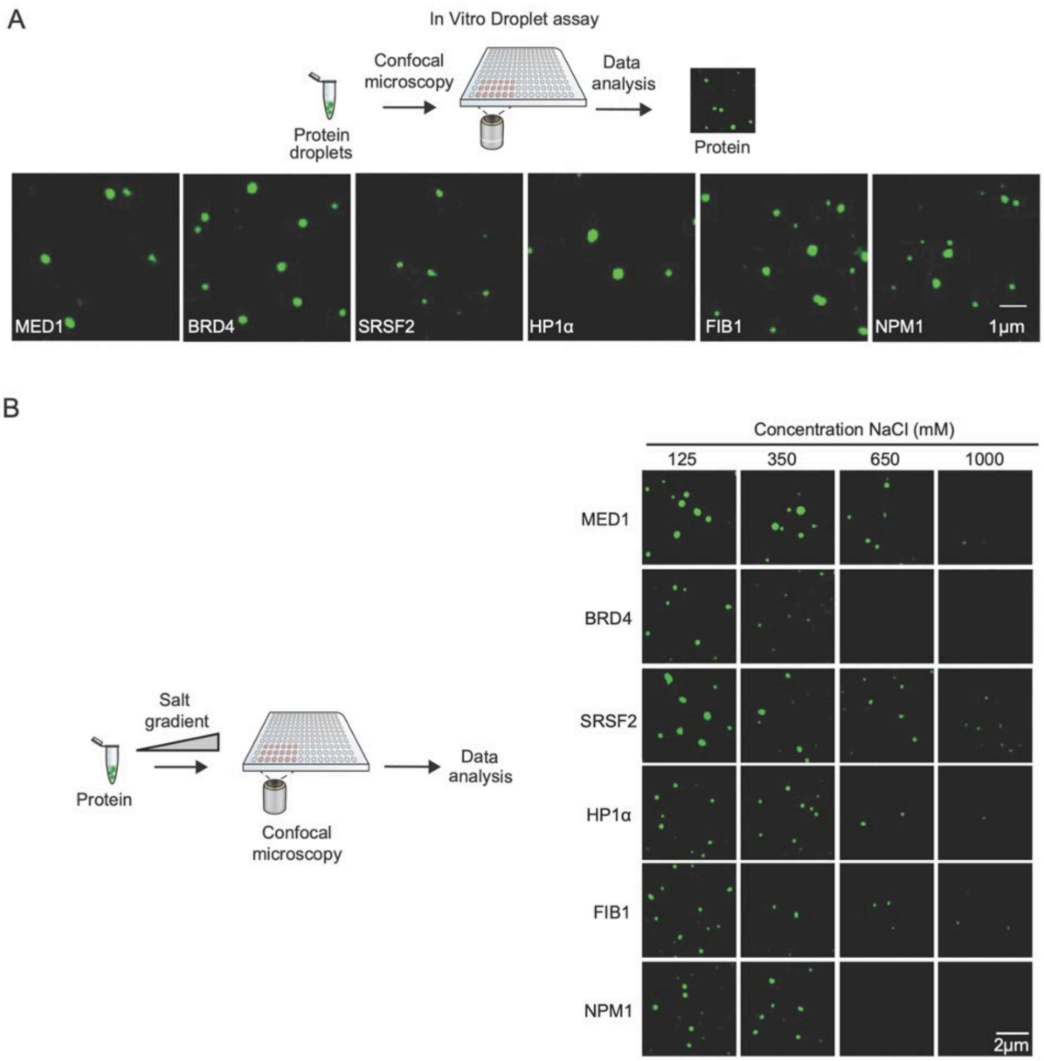
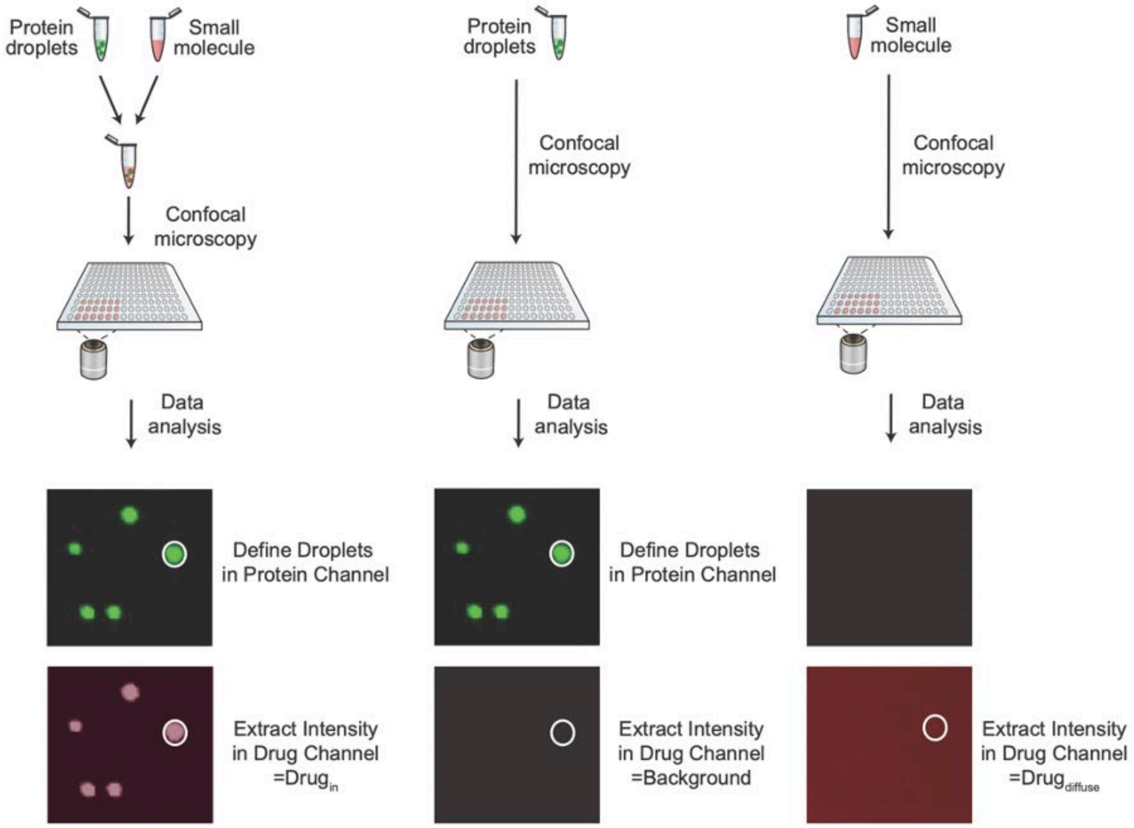


Figure S4. In vitro droplets of condensate forming proteins. (A) Confocal microscopy of in vitro droplet formation assays of the indicated GFP-tagged protein in 125mM NaCl and 10% PEG. MED1 and BRD4 proteins are the IDR portion only. (B) Confocal microscopy images of MED1, BRD4, SRSF2, HP1α, FIB1, and 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

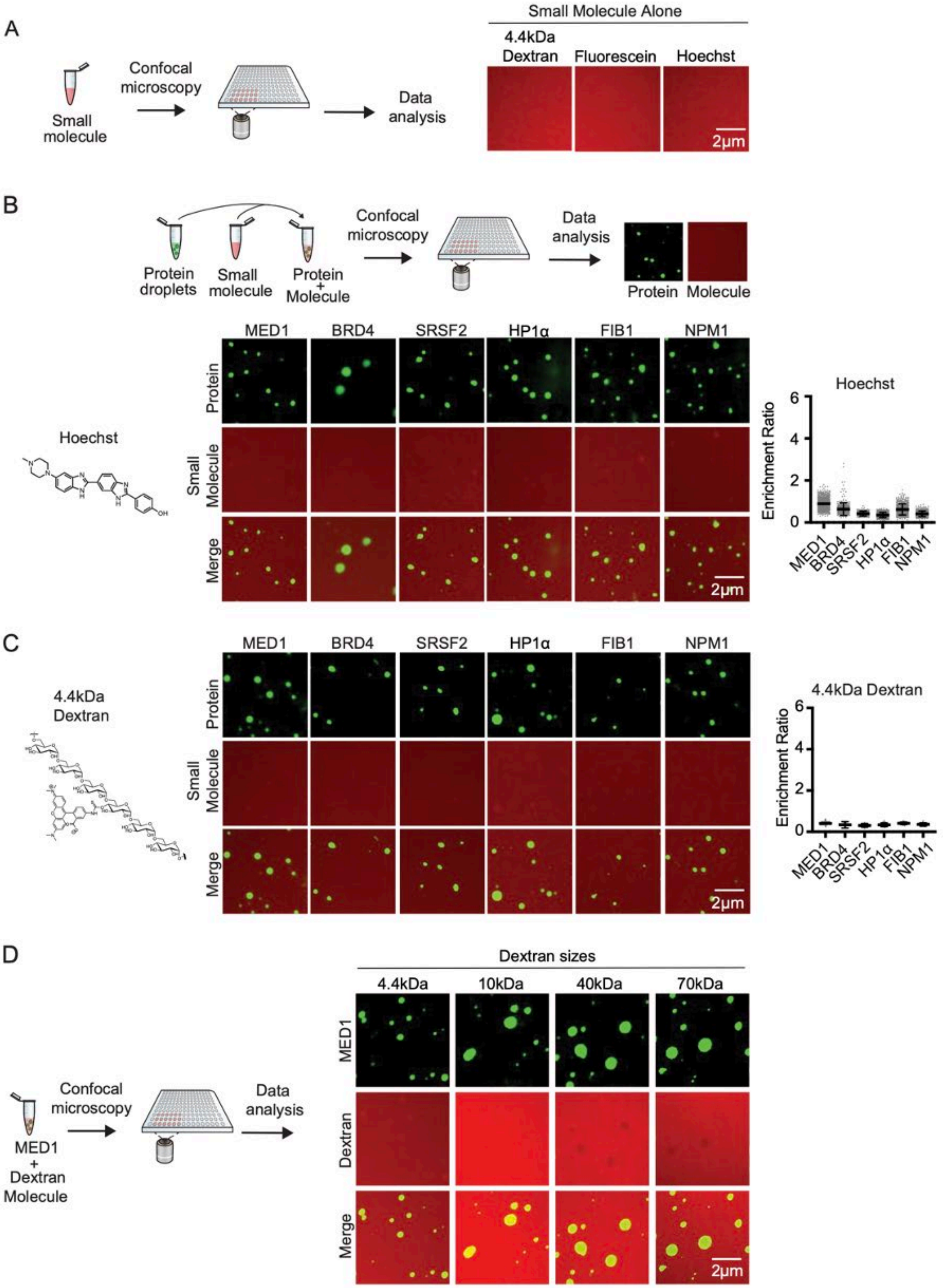

A
Small molecule
Confocal microscopy
Data analysis
Small Molecule Alone
4.4kDa Dextran
Fluorescein
Hoechst
2μm
B
Protein droplets
Small molecule
Protein + Molecule
Confocal microscopy
Data analysis
Protein
Molecule
MED1
BRD4
SRSF2
HP1α
FIB1
NPM1
Hoechst
Protein
Small Molecule
Merge
2μm
Hoechst
Enrichment Ratio
6
4
2
0
MED1
BRD4
SRSF2
HP1α
FIB1
NPM1
C
4.4kDa Dextran
MED1
BRD4
SRSF2
HP1α
FIB1
NPM1
Protein
Small Molecule
Merge
2μm
4.4kDa Dextran
Enrichment Ratio
6
4
2
0
MED1
BRD4
SRSF2
HP1α
FIB1
NPM1
D
MED1 + Dextran Molecule
Confocal microscopy
Data analysis
Dextran sizes
4.4kDa
10kDa
40kDa
70kDa
MED1
Dextran
Merge
2μm

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

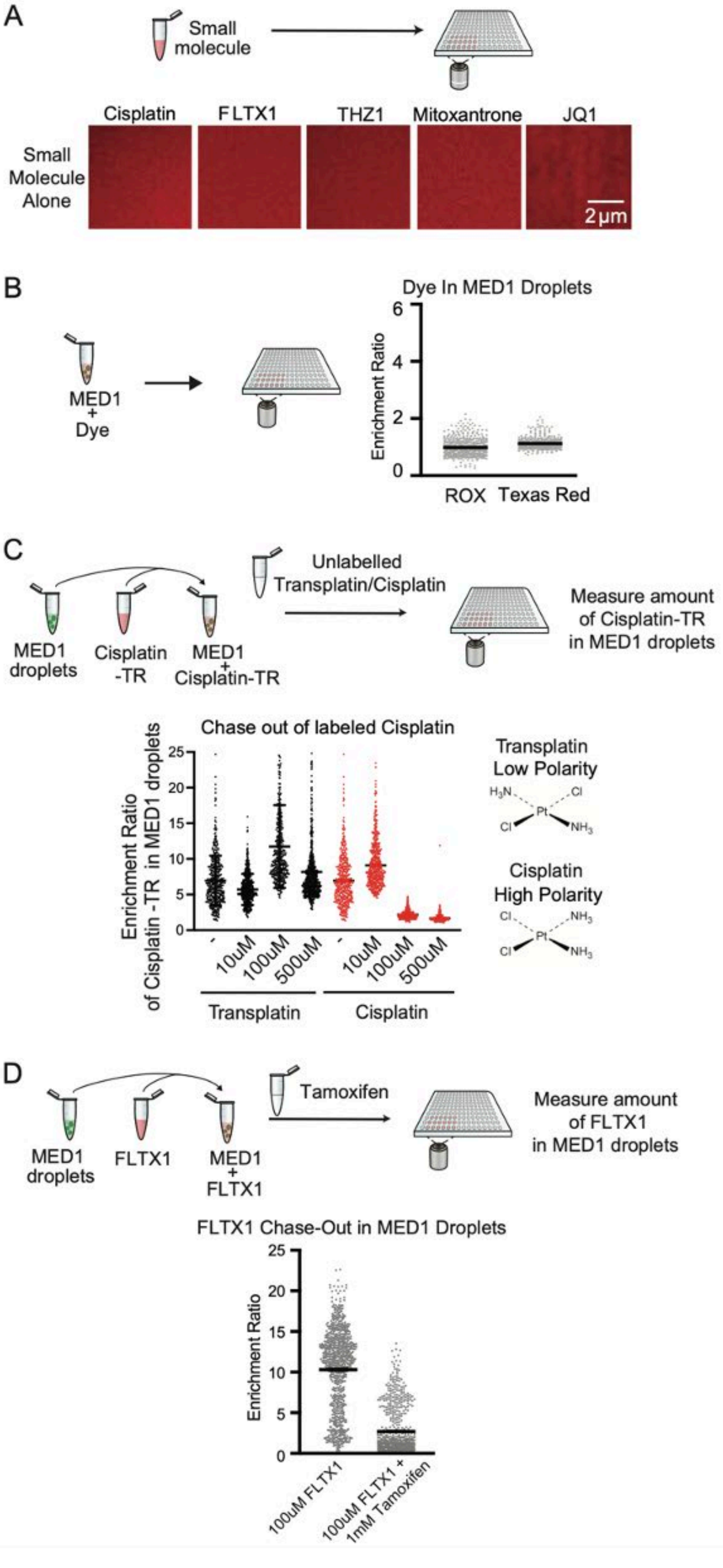
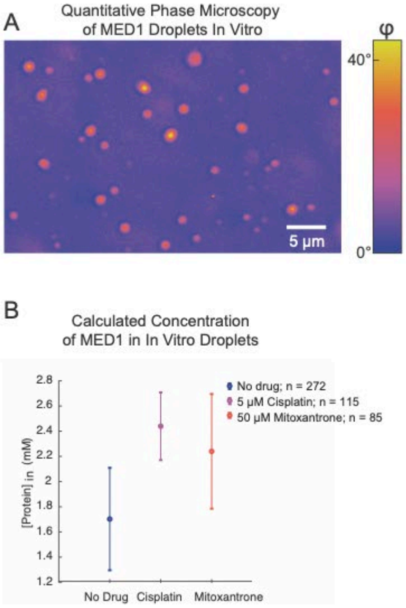

A
Small molecule
Cisplatin
FLTX1
THZ1
Mitoxantrone
JQ1
Small Molecule Alone
2 µm
B
MED1 + Dye
Dye In MED1 Droplets
Enrichment Ratio
0
2
4
6
ROX
Texas Red
C
MED1 droplets
Cisplatin -TR
MED1 + Cisplatin-TR
Unlabelled Transplatin/Cisplatin
Measure amount of Cisplatin-TR in MED1 droplets
Chase out of labeled Cisplatin
Enrichment Ratio of Cisplatin -TR in MED1 droplets
0
5
10
15
20
25
-
10uM
100uM
500uM
-
10uM
100uM
500uM
Transplatin
Cisplatin
Transplatin Low Polarity
H3N
Cl
Pt
Cl
NH3
Cisplatin High Polarity
Cl
NH3
Pt
Cl
NH3
D
MED1 droplets
FLTX1
MED1 + FLTX1
Tamoxifen
Measure amount of FLTX1 in MED1 droplets
FLTX1 Chase-Out in MED1 Droplets
Enrichment Ratio
0
5
10
15
20
25
100uM FLTX1
100uM FLTX1 + 1mM Tamoxifen

Figure S7. Properties of small molecule drugs, not their fluorescent moiety, govern partioning 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

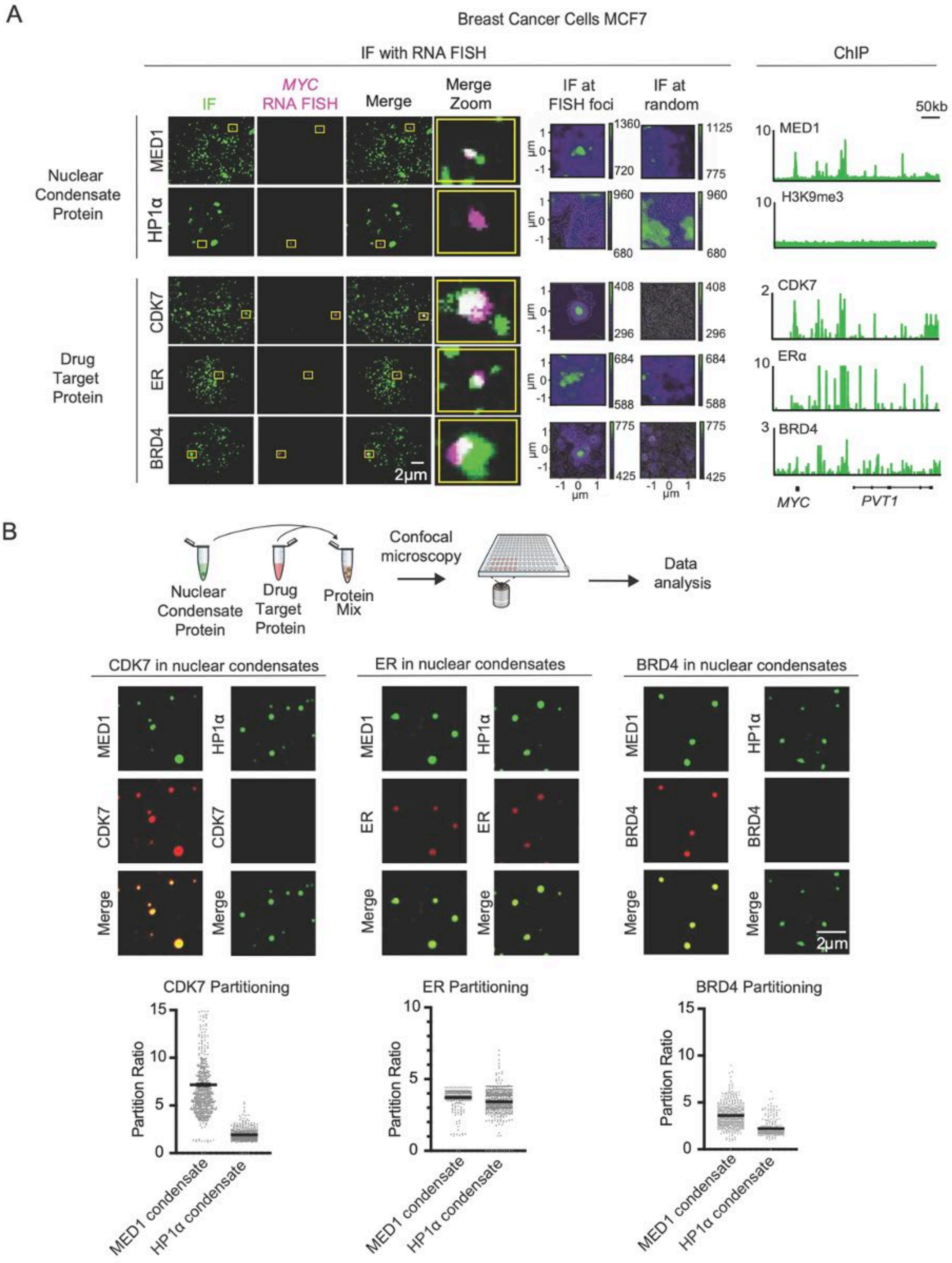


Figure S9. Association of drug targets with transcriptional condensates. (A) Immunofluorescence of MED1, HP1α, CDK7, ER, and BRD4 together with MYC *RNA* FISH. Consistent with the finding that MED1, a marker of transcriptional condensates, is present in puncta at the MYC oncogene, CDK7, ER, and BRD4 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

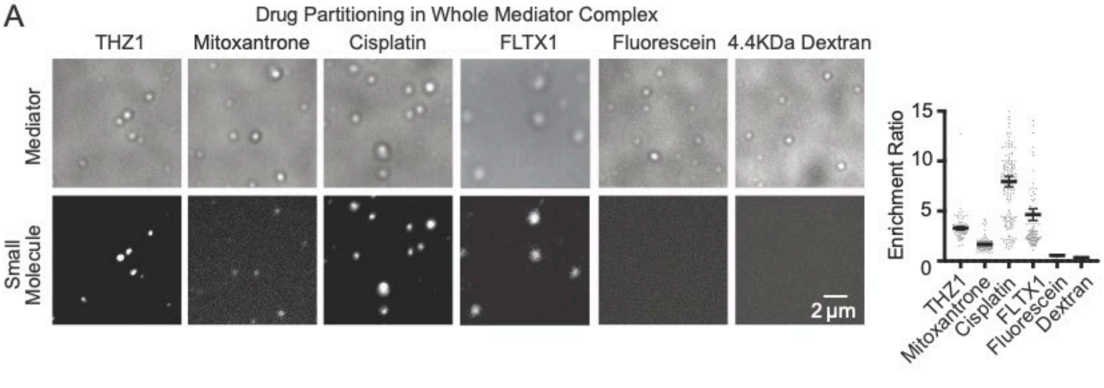


Figure S10. Partitioning behavior of various small molecule drugs in whole Mediator complex. Confocal microscopy images of drugs (THZ1, mitoxantrone, cisplatin, FLTX1, fluorescein, and 4.4kDa dextran) in whole mediator complex condensates. Mediator was imaged in brightfield while the small molecule was 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

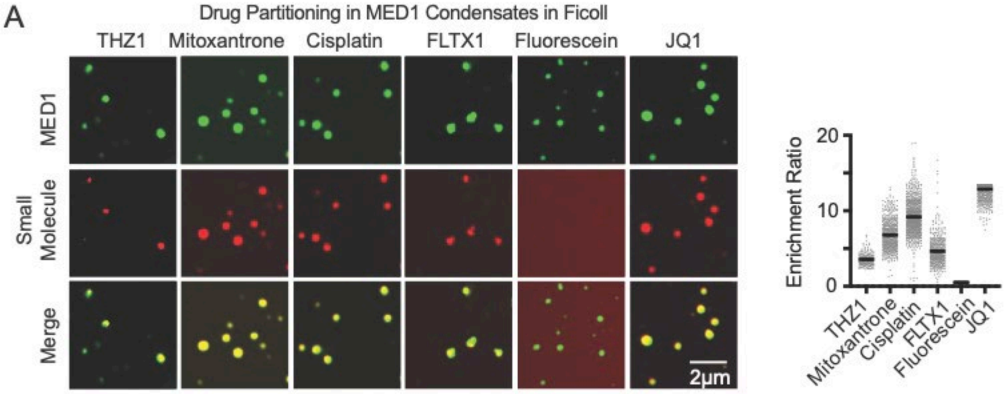


Figure S11. Partitioning behavior of various small molecule drugs into MED1 condensates formed in ficoll. Confocal microscopy images of small molecule drugs (THZ1, mitoxantrone, cisplatin, FLTX1, fluorescein, and JQ1) concentration behavior in MED1 condensates in the presence of 125mM NaCl and 20% ficoll. 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

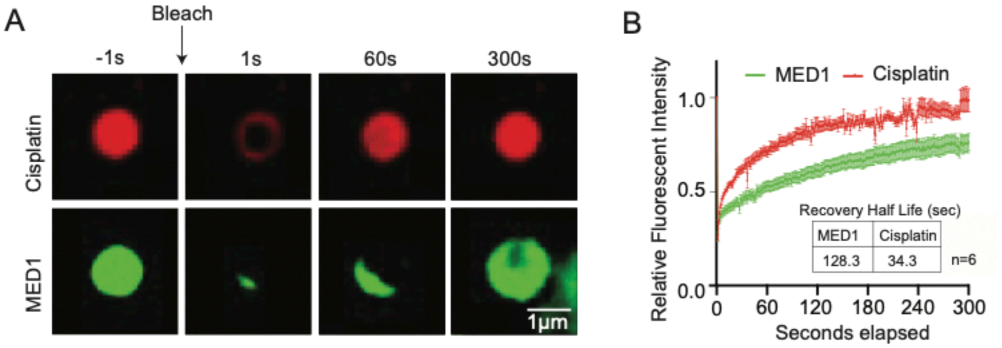


Figure S12. Cisplatin molecules are highly mobile in MED1 droplets. (A) Confocal microscopy images showing fluorescence recovery after photobleaching (FRAP) of TR-cisplatin and MED1 in condensates formed in the presence of 125mM NaCl and 10% PEG with 5µM TR-cisplatin and 10µM protein. (B) Quantification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

Figure S13

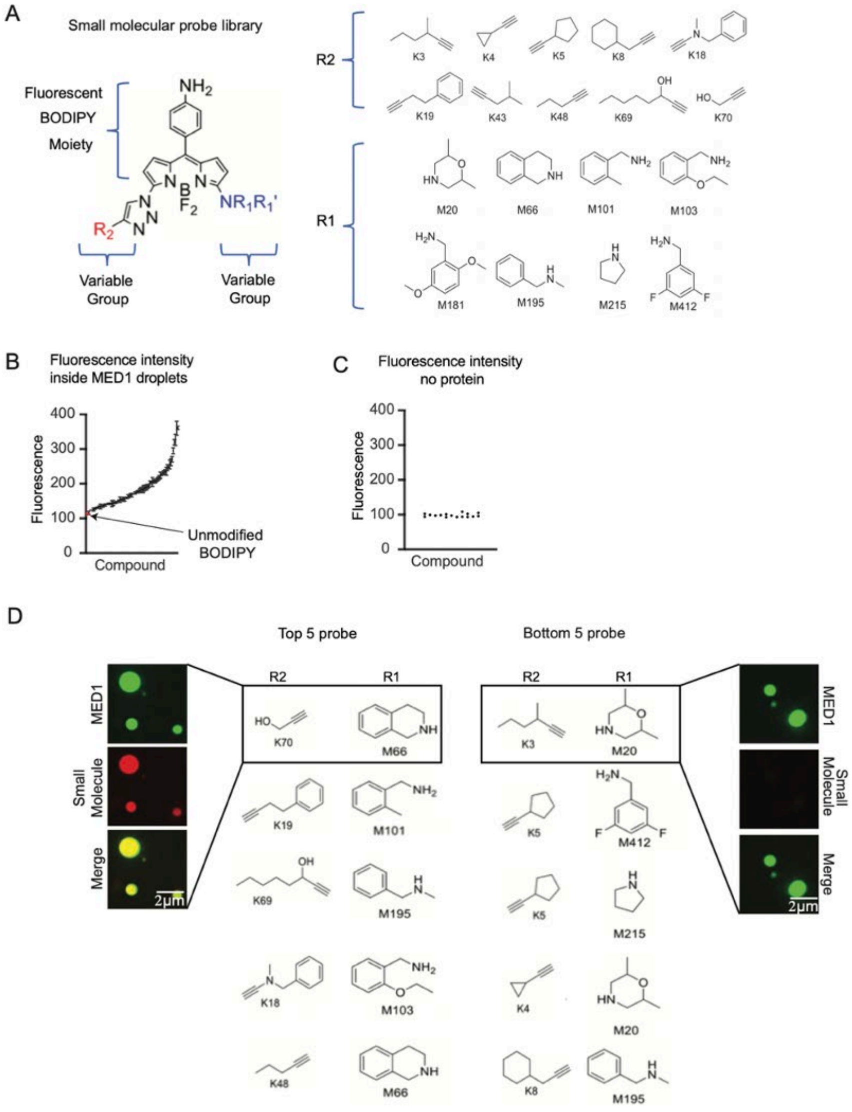


Figure S13. Specific chemical moieties govern concentration in MED1 condensates. (A) Depiction of small molecule boron-dipyrromethene (BODIPY) library. (B) Fluorescence intensity of probe library in MED1 droplets measured by confocal microscopy. Experiments were performed in 125mM NaCl and 10% PEG, with 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

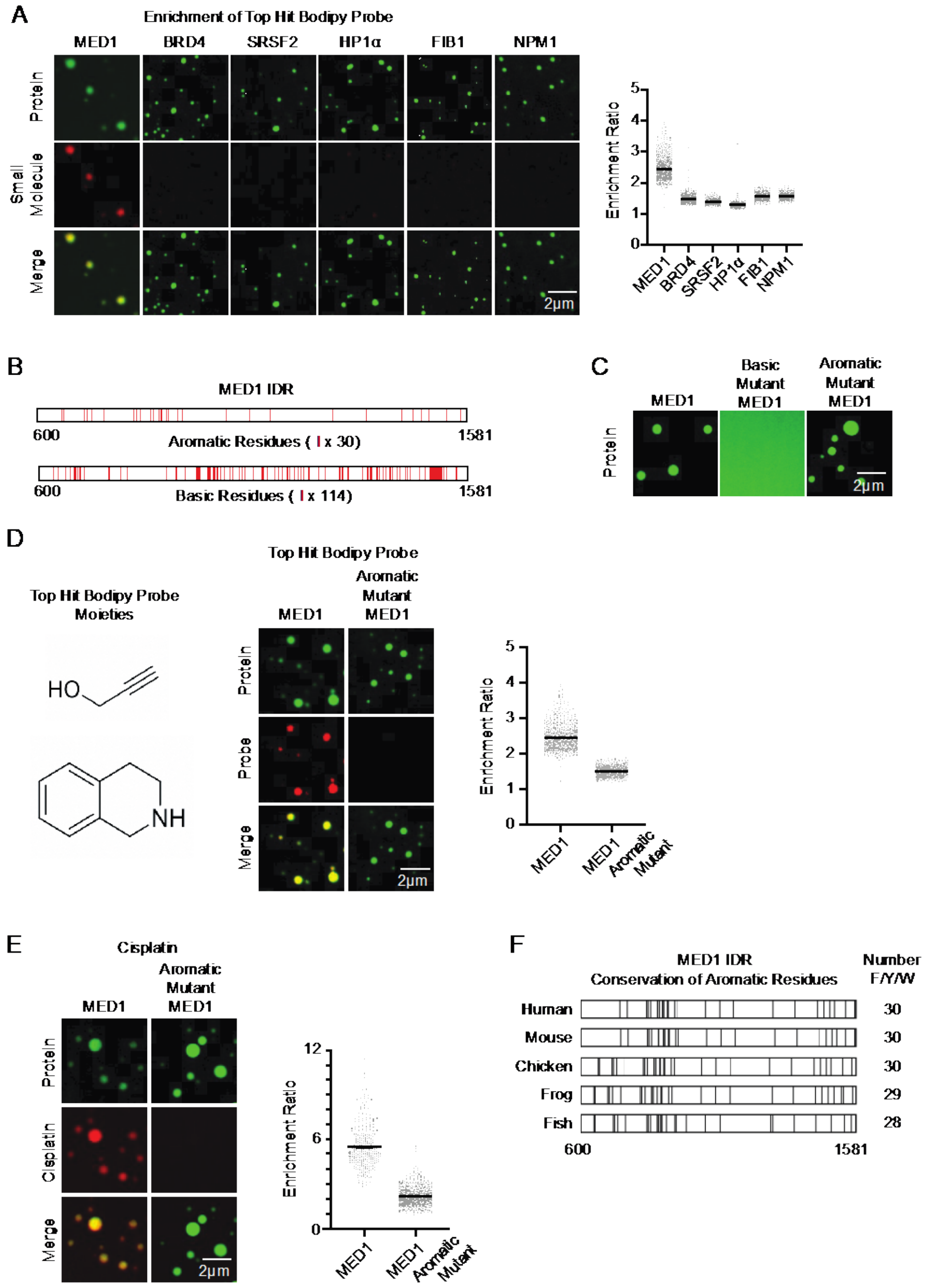

A
Enrichment of Top Hit Bodipy Probe
MED1
BRD4
SRSF2
HP1α
FIB1
NPM1
Protein
Small Molecule
Merge
2µm
Enrichment Ratio
0
1
2
3
4
5
MED1
BRD4
SRSF2
HP1α
FIB1
NPM1
B
MED1 IDR
600
Aromatic Residues (| x 30)
1581
600
Basic Residues (| x 114)
1581
C
MED1
Basic Mutant MED1
Aromatic Mutant MED1
Protein
2µm
D
Top Hit Bodipy Probe Moieties
HO
NH
Top Hit Bodipy Probe
MED1
Aromatic Mutant MED1
Protein
Probe
Merge
2µm
Enrichment Ratio
0
1
2
3
4
5
MED1
MED1 Aromatic Mutant
E
Cisplatin
MED1
Aromatic Mutant MED1
Protein
Cisplatin
Merge
2µm
Enrichment Ratio
0
6
12
MED1
MED1 Aromatic Mutant
F
MED1 IDR
Conservation of Aromatic Residues
Number F/Y/W
Human
30
Mouse
30
Chicken
30
Frog
29
Fish
28
600
1581

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

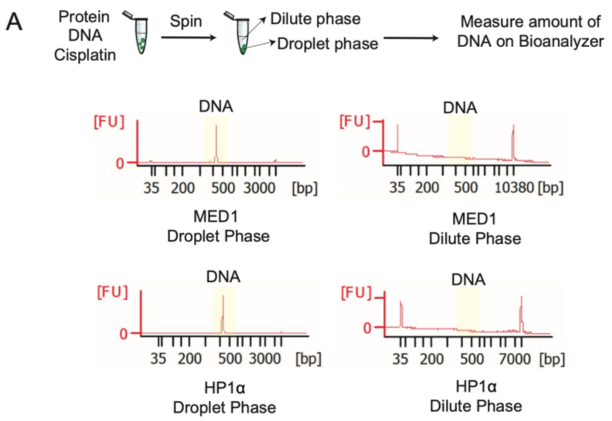


Figure S15. DNA can be compartmentalized and concentrated in nuclear condensates. (Top) Schematic of droplet assay showing protein, DNA, and cisplatin mixed in droplet forming conditions, then spun down to separate the droplet phase from the dilute phase. The amount of DNA in the two phases is subsequently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

Figure S16

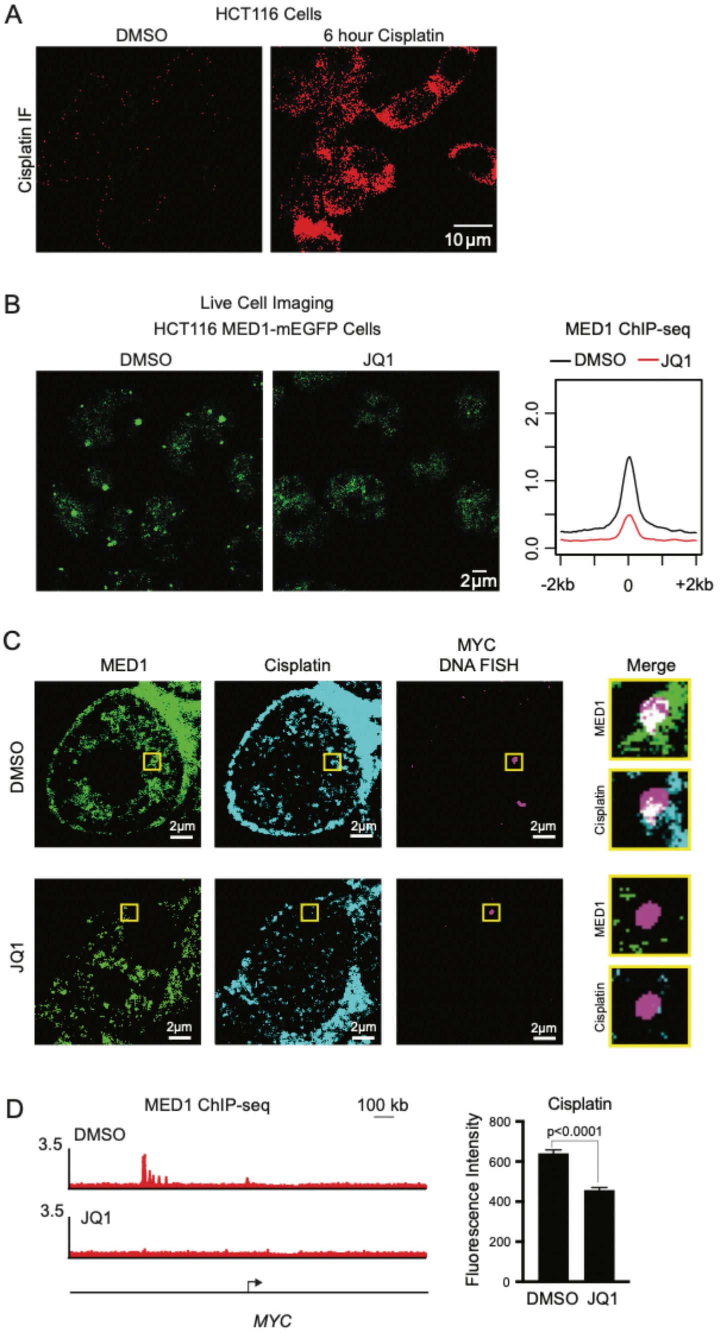


Figure S16. Concentration of small molecules in specific condensates can influence target engagement. (A) HCT116 cells were treated with DMSO or 50μM cisplatin for 6 hours followed by cisplatin immunofluorescence. The antibody only recognizes platinated DNA in cells treated with cisplatin, supporting 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

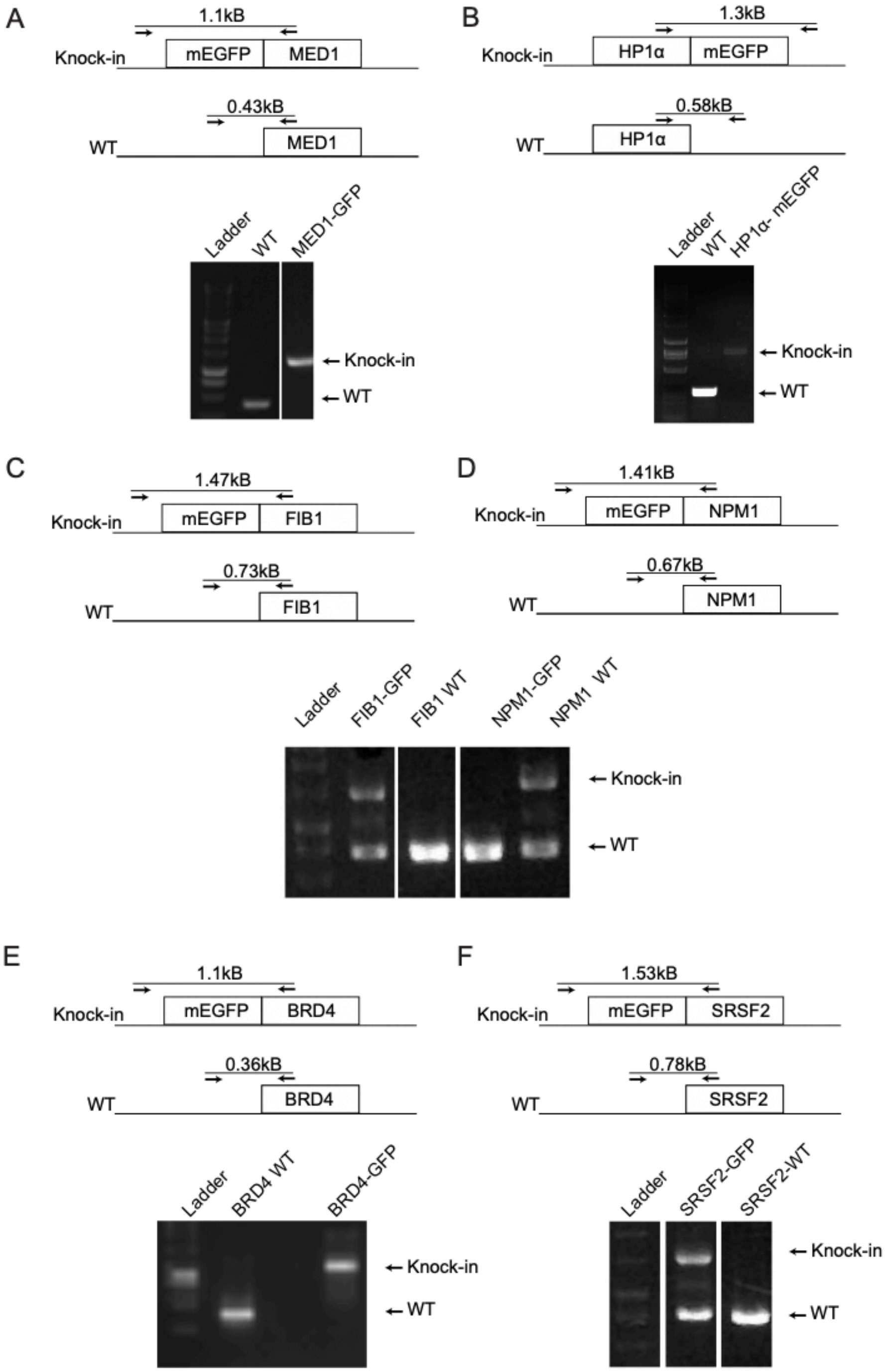


Figure S17. Genotyping of endogenously tagged cell lines. Schematic image and genotyping agarose gel showing mEGFP tagged (A) MED1, (B) HP1α, (C) FIB1 (D) NPM1, (E) BRD4, and (F) SRSF2 in HCT116 colon cancer cells.

Figure S18

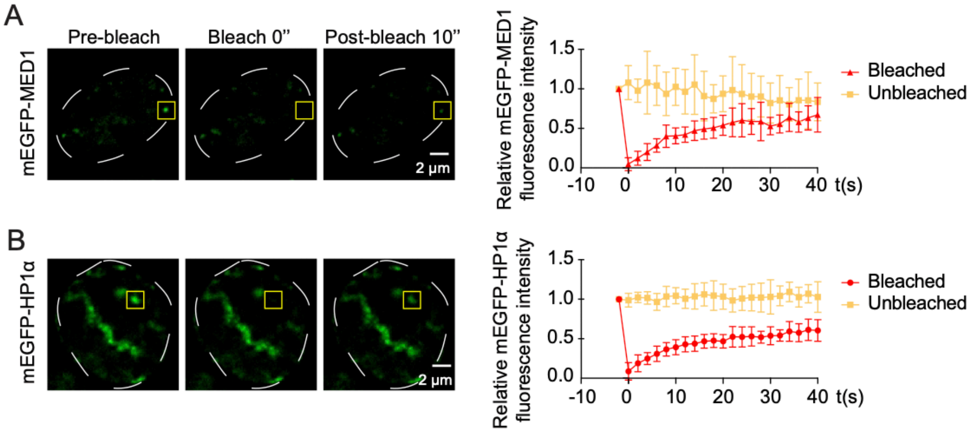


Figure S18. Nuclear condensates in cells are highly dynamic. FRAP of mEGFP-tagged (A) MED1 and (B) HP1α in HCT116 cell lines (error bars represent SEM) (n=7).

Figure S19

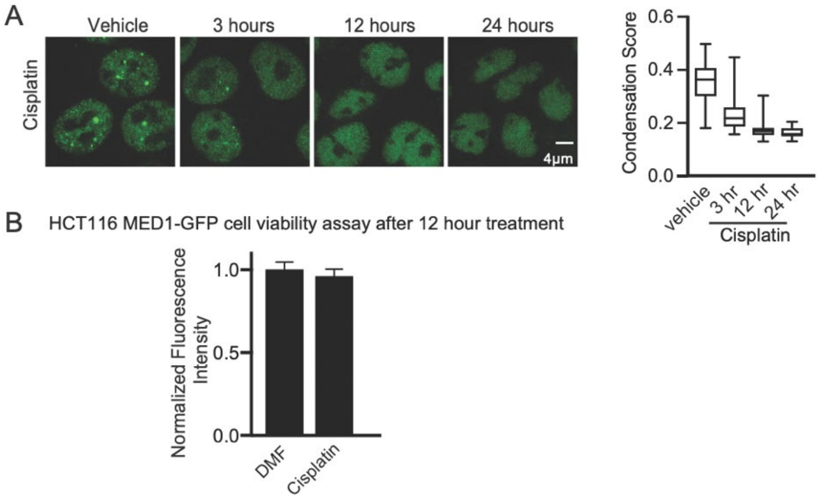


Figure S19. Dissolution of MED1 condensates in cells upon prolonged cisplatin treatment. (A) HCT116 cells endogenously GFP-tagged MED1 treated with DMF or 50µM cisplatin for 3, 6, or 12 hours. Quantification shown to the right, error bars are SD. (B) Cell viability assay of HCT116 cells expressing GFP-MED1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

Figure S20

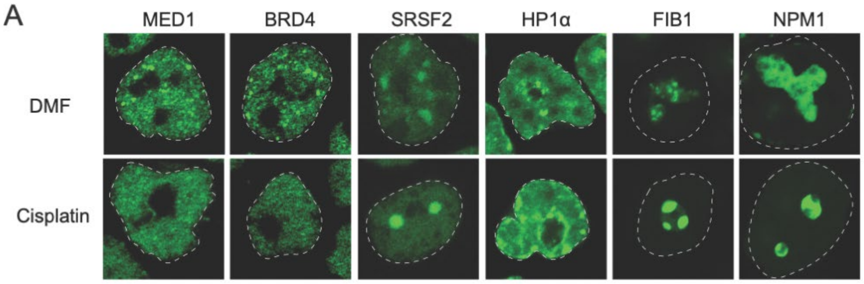


Figure S20. Effect of cisplatin on various nuclear condensates. (A) HCT116 cells bearing either endogenously GFP-tagged MED1, BRD4, HP1α, FIB1, NPM1, or SRSF2 treated with 50μM cisplatin for 12 hours. Cisplatin specifically disrupts MED1 and BRD4 condensates, consistent with cisplatin and BRD4 being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

Figure S21

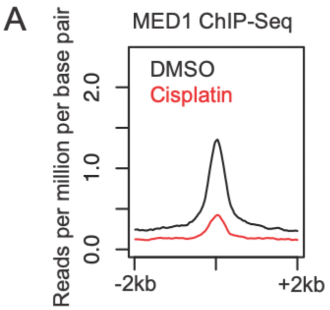


Figure S21. Decreased MED1 genomic occupancy upon cisplatin treatment. Graph shows MED1 ChIP-seq after 6 hours of DMSO or 50µM cisplatin treatment, MED1 genomic levels are reduced after cisplatin treatment.

Figure S22

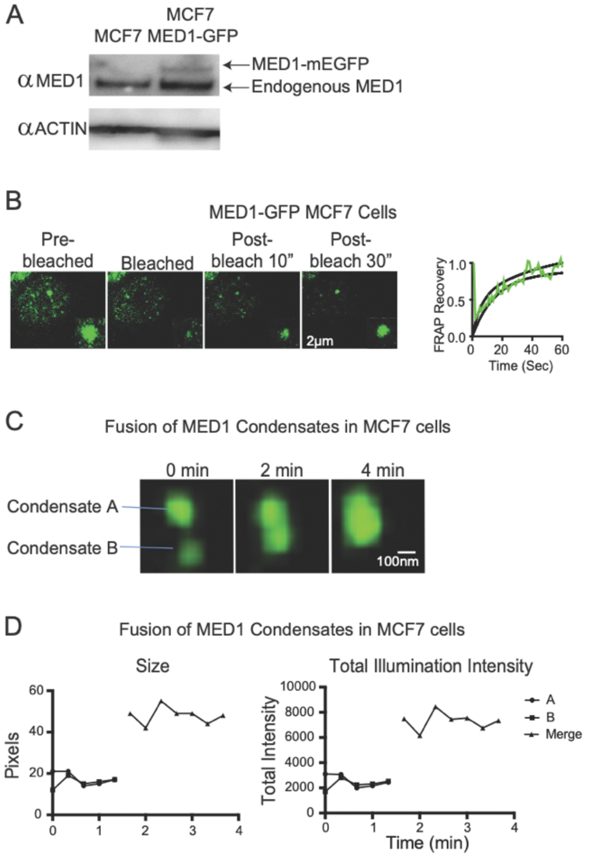


Figure S22. Characterization of MED1 condensates in MCF7 cells. (A) Western blot of MED1 in MCF7 cells and MCF cells infected with MED1-mEGFP lentiviral vector. (B) FRAP of MED1-mEGFP in MCF7 cells expressing this fusion protein by virtue of a lentiviral vector. Quantification shown to the right, black 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

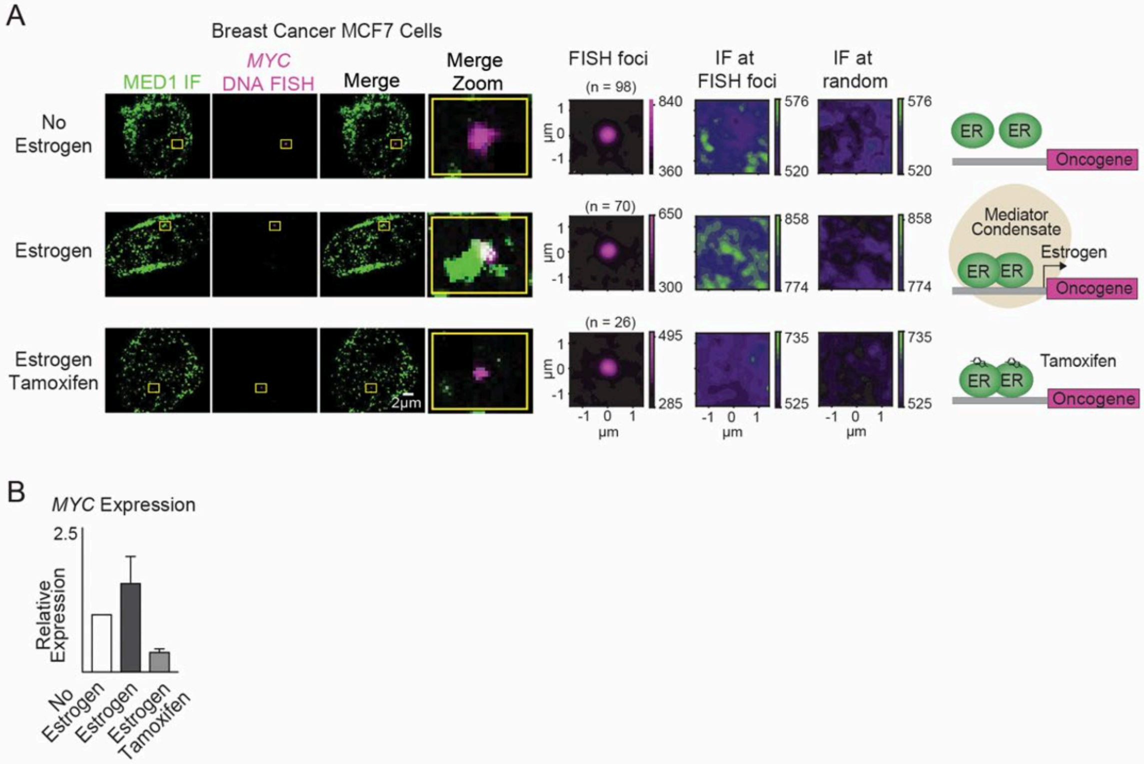


Figure S23. Estrogen and tamoxifen dependent MED1 condensate formation at the *MYC* oncogene. (A) DNA FISH and immunofluorescence in estrogen-starved MCF7 cells treated with 100nM estrogen or 100nM estrogen and 5µM tamoxifen for 24 hours. Average image analysis and random image analysis shown to the 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

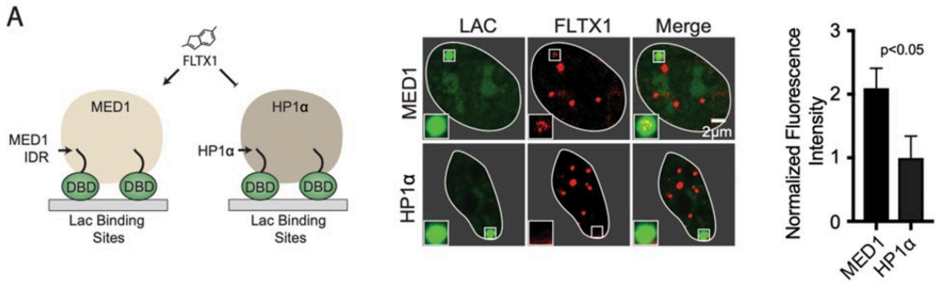


Figure S24. FLTX1 concentrates in MED1 condensates in cells. (Left) Schematic of MED1 or HP1α tethered to the LAC array in U2OS cells generating a MED1 or HP1α condensate. (Middle) Representative images of isolated U2OS cell nuclei with either MED1 or HP1α tethered to the LAC array exposed to FLTX1. 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

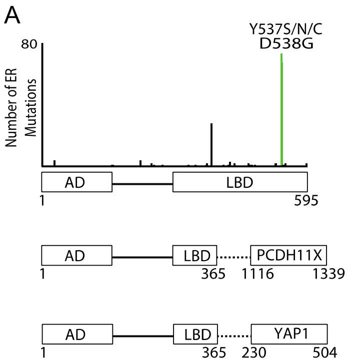


Figure S25. Patient derived hormonal therapy resistant mutations of ESR1. Plot of ER mutation frequency derived from a 220 patient set from the cBioPortal database showing locations of ER point mutations with hotspots at 537 and 538 (*68*).

Figure S26

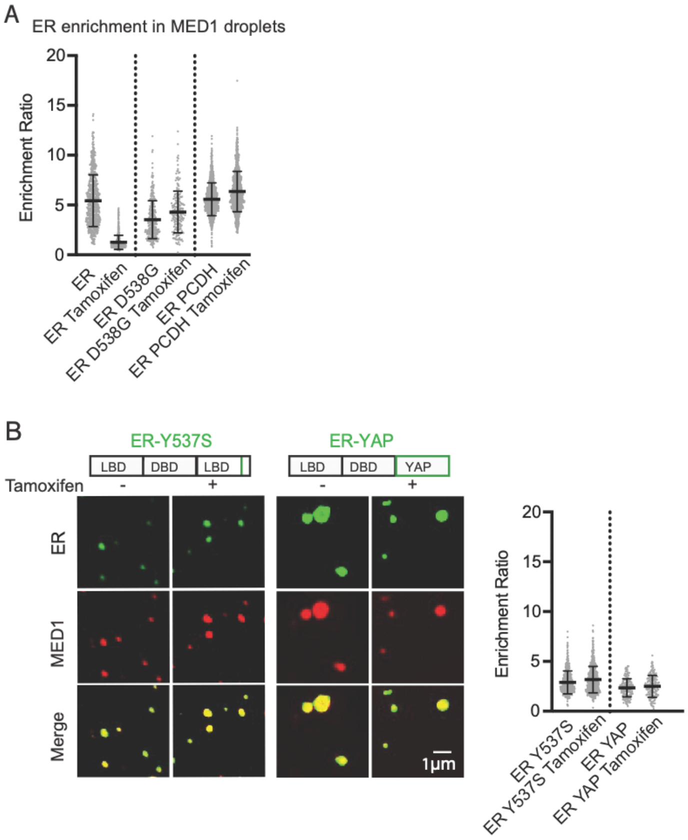
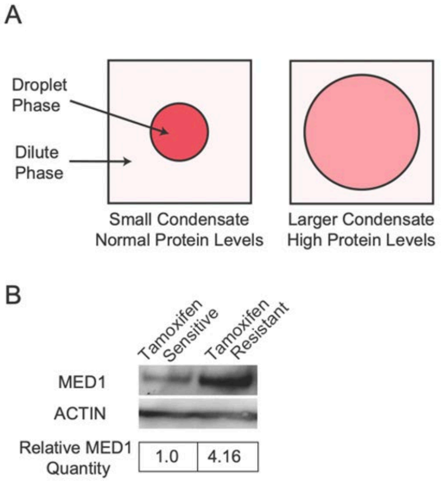


Figure S26. Enrichment ratios of ER and ER mutants in MED1 droplets. (A) Quantification of ER or ER mutant enrichment ratios in MED1 droplets in the presence of either estrogen or estrogen and tamoxifen. (B) (Left) Representative images of ER mutants partitioning in MED1 droplets, enrichment ratios shown 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

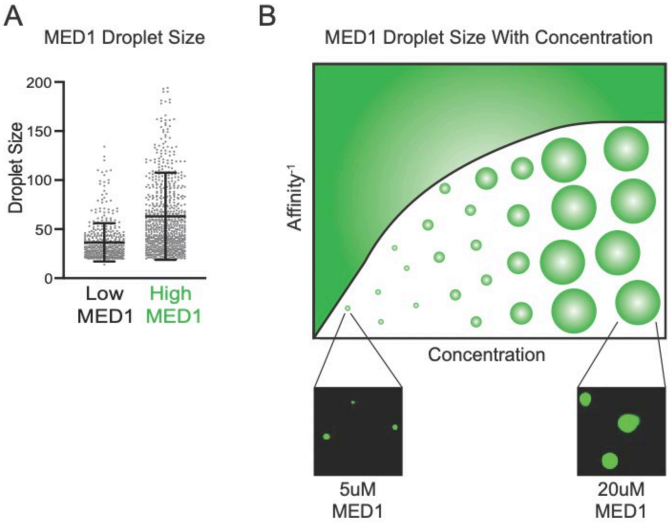


Figure S28. MED1 condensates increase in size with increasing MED1 concentration. (A) Droplet size in pixels from in vitro droplet assays performed with either 5μM (Low) or 20μM (High) MED1-GFP in 125mM NaCl and 10% PEG. Quantification shown to the right, error bars represent SD. (B) Schematic phase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

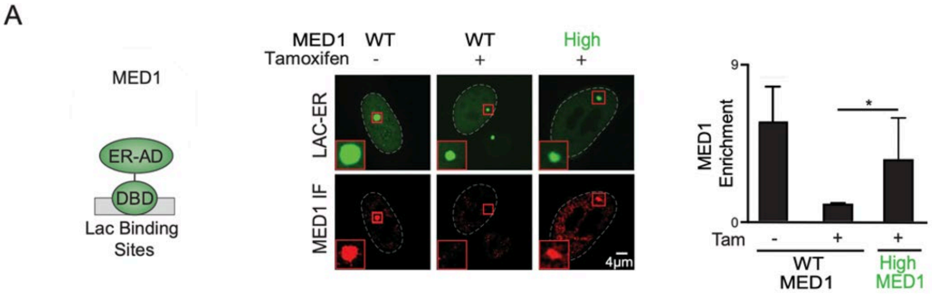
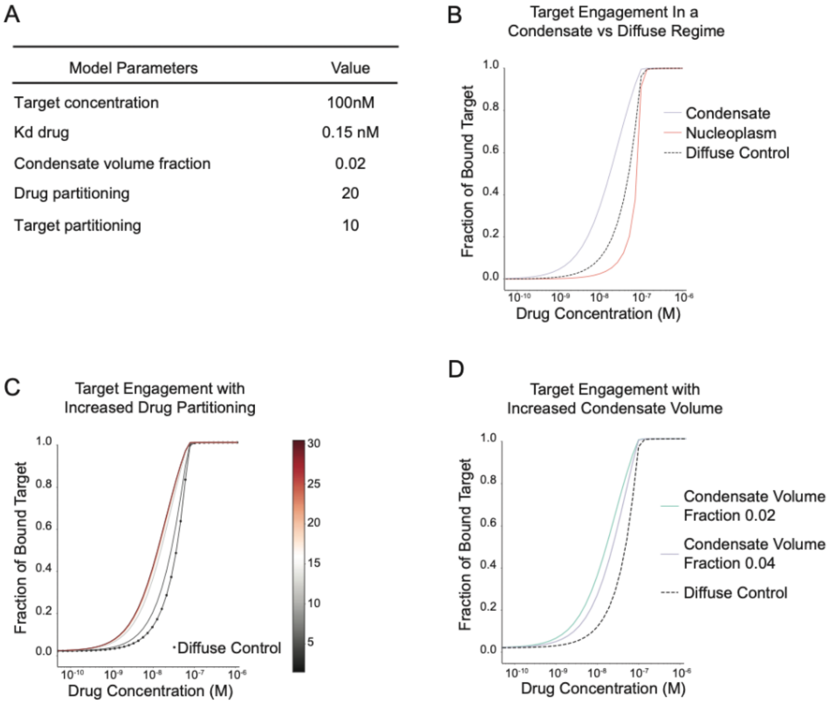


Figure S29. MED1 Condensation at the Lac Array. (A) (Left) Schematic of the Lac array assay. U2OS cells bearing 50,000 copies of the Lac binding site are transfected with a construct expressing the Lac DNA binding domain (DBD) to the estrogen receptor ligand binding domain (LBD). When the transcriptional 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



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.

<u>FRAP of in vitro droplets with drug</u>

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.

<u>Calculating drug enrichment ratios</u>

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.

<u>Chromatin immunoprecipitation (ChIP) and sequencing</u>

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=4*π*·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

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

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

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

which can be simplified and rearranged to yield

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

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

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

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,

$$A = \mathrm{Log10}(I0/I) = \varepsilon cL \qquad (5)$$

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

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

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

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

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

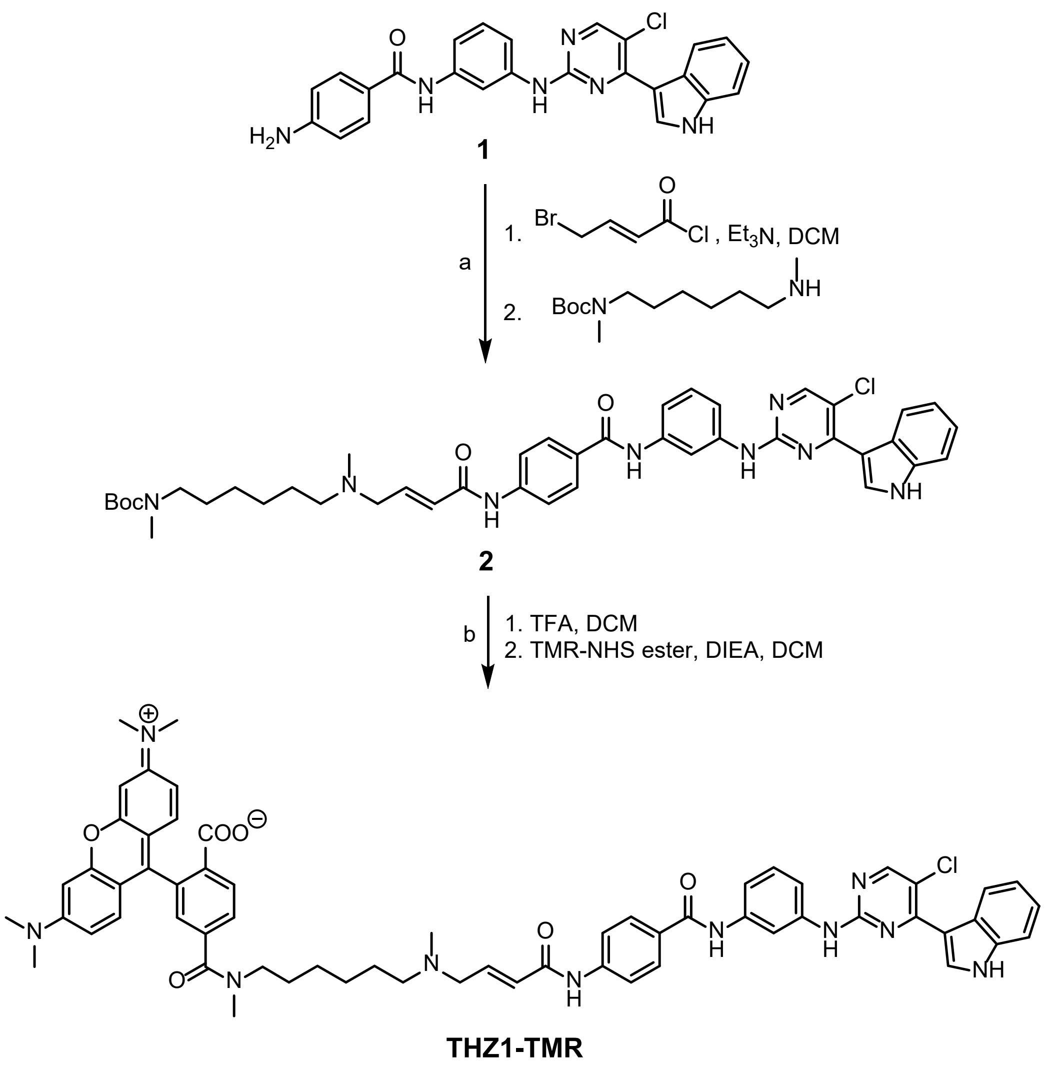
To generate MCF7 mEGFP-MED1 cells, a lentiviral construct containing the full length MED1 with a N-terminal mEGFP fusion connected by a 10 amino acid GS linker was cloned, containing a puromycin selection marker. Lentiviral particles were generated in HEK293T cells. 250,000 MCF7 cells were plated in 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

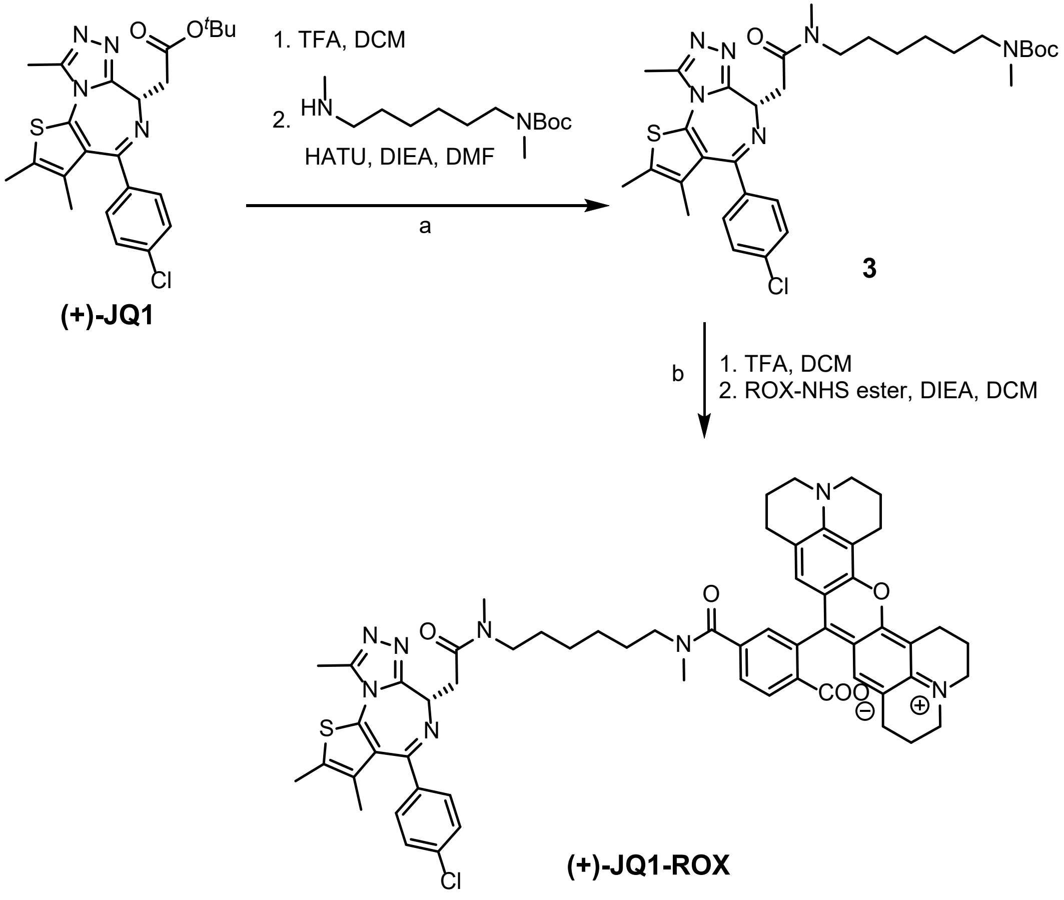
Unless otherwise noted, reagents and solvents were obtained from commercial suppliers and were used without further purification. Mass spectra were obtained on a Waters Micromass ZQ instrument. Preparative HPLC was performed on a Waters Sunfire C18 column (19 mm × 50 mm, 5 μM) using a gradient of 15−95% 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:
CONDENSATE PARTITIONING OF ANTISENSE OLIGONUCLEOTIDE THERAPEUTICS AND THERAPEUTIC INDEX

Lena K. Afeyan[1, 2], Susana Wilson Hawken[1, 2, 3], Richard Young[1, 2].

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[3]Program of Computational & Systems Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

Abstract

Nucleic acids are key components and regulators of condensates, and the basis for various therapeutic compounds. Here we use therapeutic antisense oligonucleotides (ASOs) as a tool to investigate the ability of these nucleic acid polymers to selectively partition into specific condensates. Our results show that ASOs are concentrated selectively into specific protein condensates in vitro independent of the presence of target RNA. In cells, our results suggest that ASOs form liquid-like puncta, can localize with known markers of biomolecular condensates and may selectively disrupt specific biomolecular condensates over others. We identify oligonucleotide modifications that modulate selective condensate partitioning of ASOs, have been shown to alter subcellular localization of ASOs, and may thus influence the efficacy or toxicity of these compounds. Ultimately, these results suggest that oligonucleotides can selectively concentrate in specific condensates, that this partitioning can be modified with specific chemistries, and that rules may emerge along these lines to enable optimization of the therapeutic index of nucleic acid drugs.

Introduction

Biomolecular condensates concentrate and compartmentalize diverse nucleic acids and proteins to facilitate myriad cellular processes (Hyman et al., 2014). Nucleic acids are a key component of biomolecular condensates as they can serve as scaffolds for RNA and DNA-binding proteins, crowding these components to sufficient concentrations for condensate formation, and providing weak multivalent interactions such as electrostatic interactions of the nucleic acid backbone (Garcia-Jove Navarro et al., 2019; Shrinivas et al., 2019). One poignant example of the role of nucleic acids emerges in recent studies of transcriptional condensates, in which DNA provides binding sites for transcription factors that can crowd these proteins, and enhancer- and promoter-associated RNAs, as well as pre-mRNA, can promote condensate formation at low concentrations, and disrupt condensate formation at higher concentrations (Figure 1A) (Shrinivas et al., 2019; Henninger et al., 2020).

Our recent work, presented in Chapter 2 of this thesis, has suggested that small molecules can partition differentially into biomolecular condensates and that differential partitioning behavior can shape therapeutic activity (Klein et al., 2020). Differential partitioning has been hypothesized to arise from the physicochemical environment within condensates, which dictates which molecules are concentrated into specific condensates over others. Whether differential partitioning extends to other therapeutic modalities, such as nucleic acid-based therapeutics, is not yet known.

Antisense oligonucleotides (ASOs) are short synthetic nucleic acids that can be used as therapeutics or laboratory tools to modulate the expression of RNA targets in the nucleus and cytoplasm (Swayze and Bhat, 2008; Liang et al., 2017). ASOs have represented a promising therapeutic modality for several decades, as they are easily programmable to target a variety of RNAs, they can be taken up by the central nervous system, allowing for the treatment of neurological diseases, and they can modulate RNA expression and modification in a variety of ways (Roberts et al., 2020). The study of these nucleic acid-based therapeutics in the context of biomolecular condensates, based upon the critical role for their RNA targets in condensates, as well as the ability of DNA and RNA to modulate condensate properties, may provide novel insights for the improvement of this class of therapies.

ASOs are typically 8-30 nucleotides long and can be made of both RNA and DNA nucleotides. Most ASOs currently used are modified with phosphorothioate (PS) nucleotide linkages (Figure 1B), which are more nuclease-resistant and hydrophobic than phosphate (PO) linkages (Brown et al., 1994; Crooke et al., 2020). ASOs with PS linkages (PS-ASOs) are traditionally categorized into one of two classes based upon their mechanism of action. The first, called a gapmer (Figure 1B), relies upon RNase H1 degradation to target RNAs and must contain a central deoxynucleotide region, which is typically flanked at each end by modified ribonucleotides to increase stability (Figure 1C) (Shen and Corey, 2019). The second utilizes steric blocking to modulate processes such as translation or splicing of RNA targets and tends to consist of modified or unmodified ribonucleotides alone (Figure 1D) (Shen and Corey, 2019). Several nucleotide modifications have been developed to improve the therapeutic properties of ASOs and other nucleic acid-based therapies. For example, modifications at the 2’-ribose, such as O-methoxyethyl (MOE), fluoro (F), locked nucleic acid (LNA), and constrained ethyl (cEt) confer improved RNA affinity, nuclease resistance and stabilization of duplex interactions (Figure 1B) (Khvorova et al., 2017).

Figure 1

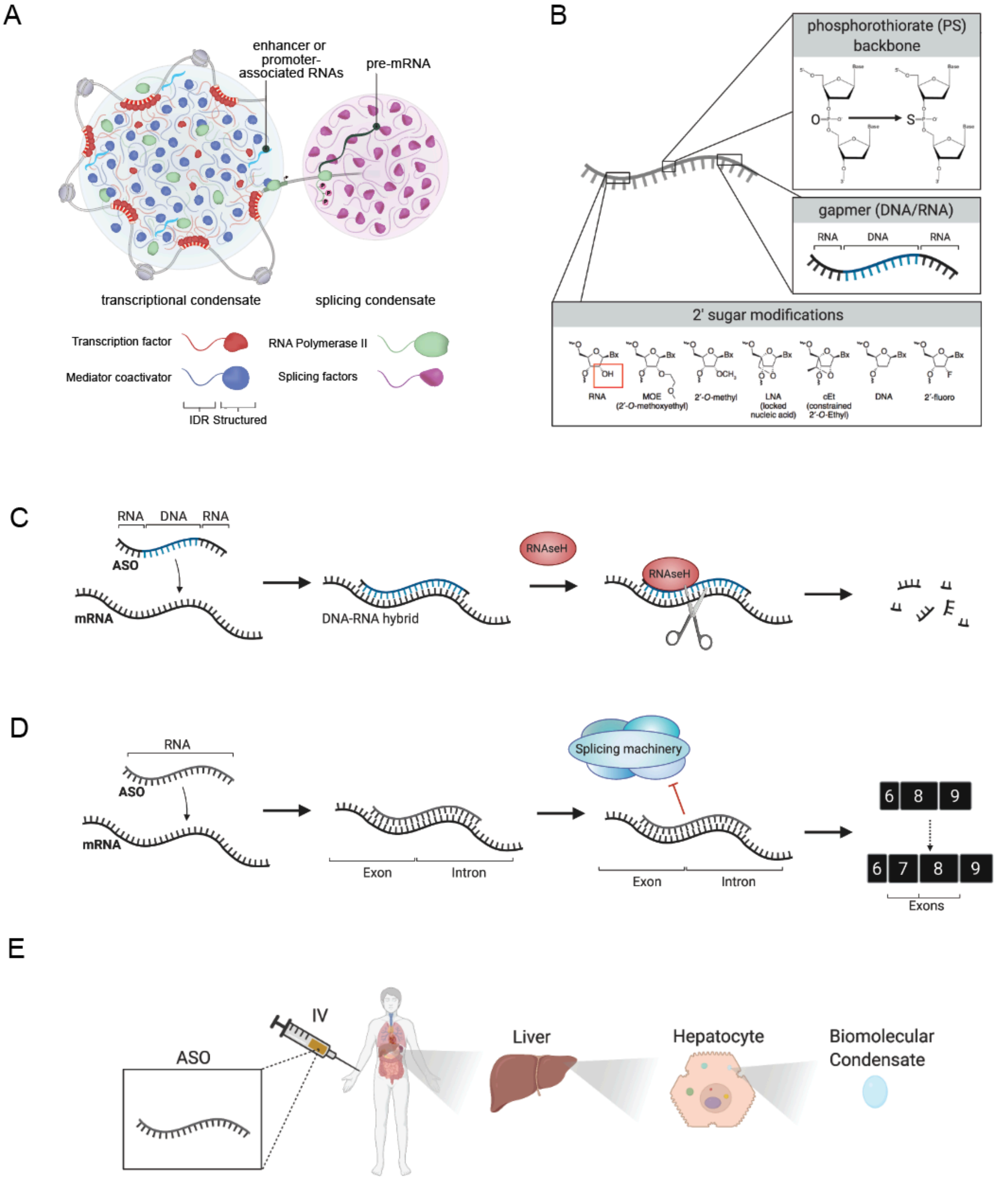


Figure 1. Properties of antisense oligonucleotides (ASOs). (A) Diagram of the roles of nucleic acids in transcriptional and splicing condensates. DNA can serve as a scaffold for transcription

factors to bind, crowding these molecules to levels sufficient to promote condensate formation (Shrinivas et al., 2019). RNAs produced in transcription, including enhancer and promoter-associated RNAs and pre-mRNA, may promote or disrupt condensate formation (Henninger et al., 2020) depending upon their levels. (B) ASO modifications in the development of ASOs as therapeutics. Backbone modifications, changes in DNA versus RNA content for the design of gapmers, and 2’ sugar modifications can alter the therapeutic properties of ASOs. (C) RNA degradation mechanism of ASO action. ASOs hybridize to target RNA sequences and the hybridization of DNA nucleotides of the gapmer facilitates the recruitment of RNase H1, which cleaves the RNA target. (D) Steric blocking mechanism of ASO action. ASOs hybridize to the RNA target and block protein binding, such as the binding of the splicing machinery, disrupting the function of these proteins in processes such as splicing. (E) Zoom in model of the localization of ASOs from intravenous infusion to concentration within biomolecular condensates.

Despite the promise of ASOs as a therapeutic strategy, the relatively narrow therapeutic window, namely the range of concentrations between the lowest dose that is effective to the highest dose that is not toxic, of many ASOs is a critical issue that must be improved to realize the promise of this class of molecules (Frazier, 2015). At concentrations required to provide effective target knockdown, several ASOs have shown substantial toxicity in preclinical models and in the clinic, including inflammation, nephrotoxicity, hepatotoxicity and thrombocytopenia (Frazier, 2015; Shen et al., 2019). An increased hepatotoxicity has been observed with specific nucleotide modifications, such as the cET, fluoro, and LNA modifications (Swayze et al., 2007; Kakiuchi-Kiyota et al., 2016; Kamola et al., 2017; Dieckmann et al., 2018; Shen et al., 2019). While this toxicity can be explained, in part, by off-target hybridization, several studies have suggested that subcellular localization of ASOs and interactions with proteins may play an important role in both the efficacy and the toxicity of these drugs (Kakiuchi-Kiyota et al., 2014; Shen et al., 2019).

Upon entering the body through modes of administration such as intravenous or intrathecal injection, ASOs traffic to a variety of organs and are known to build up at especially high levels in the liver and kidney (Figure 1E) (Crooke et al., 2017). ASOs enter cells through a variety of mechanisms, such as clathrin- and caveolin-dependent receptor-mediated uptake, which are thought to be among the most productive mechanisms by which cells internalize ASOs (Crooke et al., 2017). Once in the cell, the subcellular localization of PS-ASOs and where these ASOs carry out their mechanisms of action is relatively poorly understood.

While ASOs have not yet been studied in the context of biomolecular condensate partitioning, several lines of evidence suggest condensates may play a role in ASO activity. Early studies visualizing fluorescently-labeled PS-ASOs in cells showed that PS-ASOs form spherical puncta, termed “PS Bodies”, in the nucleus of diverse cell types that are concentration-dependent, and disappear and reappear over the cell cycle (Lorenz et al., 1998). The formation of PS Bodies was also found in this early study to be a feature of PS-ASOs but not PO-ASOs, and was independent of the presence of the target RNA of the ASO in the cell or the sequence of the ASO tested (Shin and Brangwynne, 2017). These findings are consistent with a model in which PS-ASOs partition into or form condensates (Figure 1E). Pull-down experiments have also suggested that PS-ASOs interact with several condensate-forming proteins, including paraspeckle protein p54nrb, nucleolar protein NPM1, and P body protein DDX6 (Liang et al., 2015). Immunofluorescence experiments further indicated that proportion of PS Bodies overlap with puncta formed by these condensate-forming proteins in cells (Shen et al., 2014; Shen et al., 2018; Vickers and Crooke, 2016). Lastly, sequence-matched ASOs with different chemical modifications have been shown to have different toxicity and subcellular distribution (Crooke et al., 2017; Shen et al., 2019).

These observations led us to hypothesize that ASOs partition into biomolecular condensates and may affect condensate formation in cells, and that modifications known to shape their efficacy and toxicity alter their condensate partitioning behavior. Here, we examine several markers of biomolecular condensates and asses the ability of FDA-approved ASOs to partition into the condensates formed by these proteins in vitro. We also examine the subcellular distribution of PS-ASOs with respect to known scaffold proteins of diverse nuclear and cytoplasmic biomolecular condensates and examine how these ASOs affect condensate formation of these proteins. Lastly, we examine the effect of various 2' ribose chemical modifications on the partitioning behavior of condensates.

Our preliminary findings suggest that ASOs can partition into condensates in vitro independent of the presence of their RNA target. This partitioning behavior changes upon chemical modification of the ASOs. While certain published findings on the overlap of ASOs with markers of biomolecular condensates have not yet been reproducible (Liang et al., 2014; Shen et al., 2019), our in-cell experiments have suggested that specific PS-ASOs may disrupt condensate formation of certain condensates in the cell. Further studies, considered in more detail in the Discussion and Future Directions section, will be needed to determine the implications of the differential partitioning behavior of PS-ASOs observed in vitro for ASO therapeutic activity. However, preliminary studies have suggested that the continued evaluation of ASOs in the context of our evolving understanding of condensates and how they are affected by therapeutic molecules may yield insights that can improve the therapeutic window of these drugs.

Results

In vitro partitioning behavior of FDA-approved ASOs into condensates

We selected five proteins that are known scaffolds of a range of biomolecular condensates and have been used as markers of these condensates in vitro and in cells (Klein et al., 2020; Banani et al., 2017). The specific proteins and condensates - MED1 (transcriptional), SRSF2 (splicing), p54nrb (paraspeckle), NPM1 (nucleolus), DDX6 (P body)- were chosen based upon published literature on the condensate localization of common RNA targets of ASOs such as eRNAs, intronic pre-mRNA sequences and splicing junctions (transcriptional, splicing), as well as the published overlap observed of PS Bodies with known condensates (paraspeckles, nucleolus, P bodies) (Sabari et al., 2018; Guo et al., 2019; Crooke et al., 2017).

We used an assay developed in Chapter 3 of this thesis to measure the partitioning behavior of PS-ASOs of interest in vitro into droplets formed by the selected condensate markers (Klein et al., 2020). Specifically, we produced and purified recombinant, green fluorescent protein-labeled versions of each of the five marker proteins and confirmed the ability of these proteins to form droplets in an in vitro droplet assay. We then added 5' Cy5-labeled PS-ASOs of interest to these droplets and assayed the extent to which these ASOs partitioned into the droplets formed by each of the marker proteins, measuring a partition ratio for the ASO into each condensate.

We began by testing whether FDA-approved ASOs partition into condensates. We selected two FDA-approved ASOs, Spinraza and Kynamro, that each function through one of the two main mechanisms for ASO action. Spinraza is a 2'MOE-modified 18-nucleotide PS-ASO that functions by steric blocking, targeting SMN2 for exon inclusion to treat spinal muscular atrophy (Wurster and Ludolph, 2018). Kynamro is a 20-nucleotide PS-ASO gapmer with 2'MOE-modified RNA wings that targets ApoB for RNaseH1-mediated degradation to treat homozygous familial hypercholesterolemia (Crooke and Geary, 2012). We found that upon introducing these fluorescently-labeled PS-ASOs into the droplet formation assay preformed with MED1, SRSF2, p54nrb, NPM1 and DDX6, that both of these FDA-approved ASOs concentrated differentially into each of the droplets, with the highest partition ratio recorded for MED1, followed by SRSF2 and p54nrb (Figure 1A,B). We also noted that these two FDA-approved ASOs displayed similar relative partitioning behavior among the five condensates tested in vitro. These data suggest that the differential partitioning behavior observed of small molecules (Klein et al., 2020) can also be observed in vitro with two clinically relevant nucleic acid-based therapeutics.

Characterizing PS-ASO distribution in cells

To begin to test the hypothesis that PS-ASOs form condensates in cells, several PS-ASOs were transfected into live cells and their distribution was characterized. The PS-ASOs were transfected at 60nM using lipofectamine, a concentration and transfection-mode used in several published studies to recapitulate cellular uptake in animal models such as mice (Liang et al., 2014; Shen et al., 2019). First, 60nM of a Cy5-labeled 2'MOE gapmer PS-ASO targeting PTEN, used in several studies to visualize subcellular localization of ASOs and hereafter referred to as "A3", was transfected for 12 hours into HCT116 cells with lipofectamine. Visualization by live-cell confocal microscopy demonstrated that the PS-ASOs were concentrated into discrete spherical puncta, largely in the nuclei of cells (Figure 3A) (Bailey et al., 2017; Wang et al., 2019). Nuclei displayed a range in puncta size (~100-600nM per punctum) and number (~2-10 puncta/nucleus), which correlated with the amount of ASO in the nucleus (Figure 3A). We also confirmed these results in liver cancer cells (HEPG2) that were fixed with paraformaldehyde in a manner consistent with previously published studies (Figure S1A) (Liang et al., 2014; Shen et al., 2015). These results recapitulate published observations of puncta or "PS Body" formation by this PS-ASO (Lorenz et al., 1998; Shen et al., 2015).

In contrast to the well-studied PS-ASO A3, when Spinraza and Kynamro were transfected with lipofectamine, very few cells displayed substantial enrichment of ASO in the nucleus, but puncta formation was observed in the few cells with high fluorescence intensity of ASO in the nucleus (Figure S1B). This observation was consistent with the finding in published studies that puncta formation is concentration dependent (Lorenz et al., 1998). Further, while buildup of ASO in the nucleus, and thus puncta formation, have been shown to be positively correlated with ASO activity, the role that formation of puncta in the nucleus plays in ASO activity is not yet understood (Crooke et al., 2017). Thus, the implications of the decreased buildup and puncta formation observed in transfections of Spinraza and Kynamro are not yet clear. We attempted to increase transfection and uptake in the nucleus by testing another mode of transfection and higher concentrations of ASO. Kynamro and Spinraza were transfected with the RNAiMAX protocol, and similar observations of relatively low enrichment in the nucleus of cells, with a few examples of cells with high nuclear concentration and concomitant puncta formation (Figure S1C). These results suggest that while ASOs may transfect and build up in the nucleus with variable efficiency, at sufficient nuclear concentrations, several ASOs display a propensity to form puncta.

Next, we assayed whether the puncta formed by PS-ASOs displayed liquid-like behavior consistent with liquid-like condensates. We subjected several puncta formed by the PS-ASO A3 in live HCT116 cells to fluorescence recovery after photobleaching (FRAP) analysis. We measured the fluorescence intensity over time and saw that PS-ASO puncta display rapid FRAP recovery consistent with liquid-like dynamic rearrangement (Figure 3B).

Figure 2

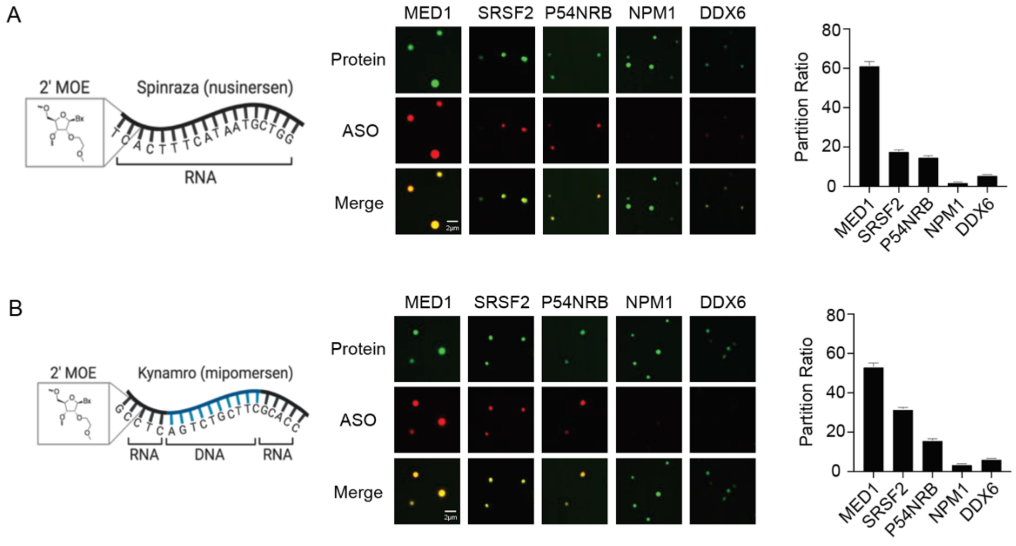


Figure 2. FDA-approved ASOs display differential partitioning in condensates in vitro. (A) In vitro droplet assay showing the behavior of Cy5-labeled Spinraza (nusinersen) in the presence of five protein condensates formed in 100mM NaCl and 10% PEG with 100nM protein and 100nM ASO imaged at 150X on a confocal fluorescent microscope. Representative images of protein (green), ASO (red), and merge are shown. Quantification of enrichment of the drug is shown on the right. Error bars represent standard error of the mean (SEM). (B) In vitro droplet assay showing the behavior of Cy5-labeled Kynamro (mipomersen) in the presence of five protein condensates formed in 100mM NaCl and 10% PEG with 100nM protein and 100nM ASO imaged at 150X on a confocal fluorescent microscope. Representative images of protein (green), ASO (red), and merge are shown. Quantification of enrichment of the drug is shown on the right. Error bars represent SEM.

Evaluating overlap of puncta formed by PS-ASOs with known condensate markers

To determine whether PS-ASOs partition into known biomolecular condensates, as was seen in vitro, puncta were visualized in cells in concert with fixed-cell immunofluorescence of condensate markers. We began by aiming to recapitulate published studies showing that PS Bodies overlap with markers of nuclear and cytoplasmic bodies, shown in recent years to be phase-separated biomolecular condensates (Crooke et al., 2017). To do so, we transfected 60nM of A3 for 12 hours into HEPG2 cells, as liver cells are among the cell types ASOs build up in at the highest levels in the human body (Crooke et al., 2017). We then performed immunofluorescence on these cells for the condensate markers studied in vitro, namely MED1, SRSF2, NPM1, p54nrb, DDX6. Confocal microscopy images indicated that PS-ASO puncta did not appear to overlap substantially with each of these markers, a handful isolated puncta among hundreds observed notwithstanding (Figure S2A). It is important to note, here, that published observations of colocalization have not provided quantification for colocalization events. These results may thus be consistent with published findings if overlap is rare. In the context of our in vitro results, however, differential partitioning behavior of A3 into these condensates, with such rare instances of overlap, was infeasible to discern and quantify reliably. Where available, we tested additional antibodies for markers to increase the likelihood that the lack of overlap was not due to issues of antibody quality and, while these efforts are ongoing, the colocalization results have thus far remained consistent.

As an orthogonal approach to circumvent potential antibody-based limitations, we examined colocalization in live cells expressing fluorescently-tagged proteins. Based upon work in Chapter 3 of this thesis, several scaffold proteins that serve as markers of biomolecular condensates such as transcriptional condensates (MED1), splicing condensates (SRSF2) and the nucleolus (NPM1) were tagged at their endogenous loci in HCT116 cells with green fluorescent protein (GFP). Additional markers of interest, p54nrb and DDX6, were tagged with GFP and stably integrated into HCT116 cells for evaluation. Transfection of 60nM A3 for 12 hours into these five cell lines with tagged condensate markers, followed by live-cell confocal microscopy with an LSM880 microscope with Airyscan processing to improve resolution, also did not display substantial overlap between PS-ASO puncta and puncta formed by these condensate markers (Figure S2C,D). To control for the possibility that colocalization changes over time, we confirmed that these results remained consistent over several timepoints, transfecting A3 for 3 hours, 6 hours, 12 hours, and 24 hours (data not shown). Notably, we also performed immunofluorescence and live-cell imaging for the one protein identified in studies thus far to overlap substantially with PS Bodies by immunofluorescence, the chaperone TCP1B complex component chaperone, and also did not detect overlap (Figure S2B,D) (Liang et al., 2014). In interpreting these results, it is important to note that a persistent problem in the study of condensates in cells is that several condensates, such as transcriptional condensates, are small enough to approach the lower limits of microscope resolution, which can make subtle enrichments difficult to detect (Alberti et al., 2017). Thus, it remains possible that PS-ASOs build up in transcriptional condensates, where RNA targets are known to exist as they are being transcribed to levels that affect their therapeutic activity, but may not yet be detectable by fluorescence microscopy.

Although RNA targets in exonic sequences may be present at several locations in the cell as mRNAs traffic in the nucleus and cytoplasm, intronic sequences are typically present mainly during transcription and splicing for many RNAs (Fazal et al., 2019). Thus, intronic target sequences would be expected to be enriched within transcriptional and splicing condensates. We examined two PS-ASOs that target intronic RNA sequences in NEDD4 and ATXN10 in order to determine whether the puncta formed by these ASOs overlap with transcriptional or splicing condensates, as these condensates may be enriched for the targets of these ASOs (Lai et al., 2020). For both of these intron-targeting PS-ASOs, we detected substantial overlap between PS-ASO puncta and splicing condensates, marked by SRSF2 (Figure 3C, Figure S2E). Several instances colocalization or proximity to transcriptional condensates, marked by MED1, were also detected, but the quantification of this overlap varied between repeated iterations of the experiment (Figure S2F). This variation may result from effects that the ASOs have on the condensates being imaged, as nucleic acids are known to be capable of both enhancing and disrupting condensate formation.

Figure 3

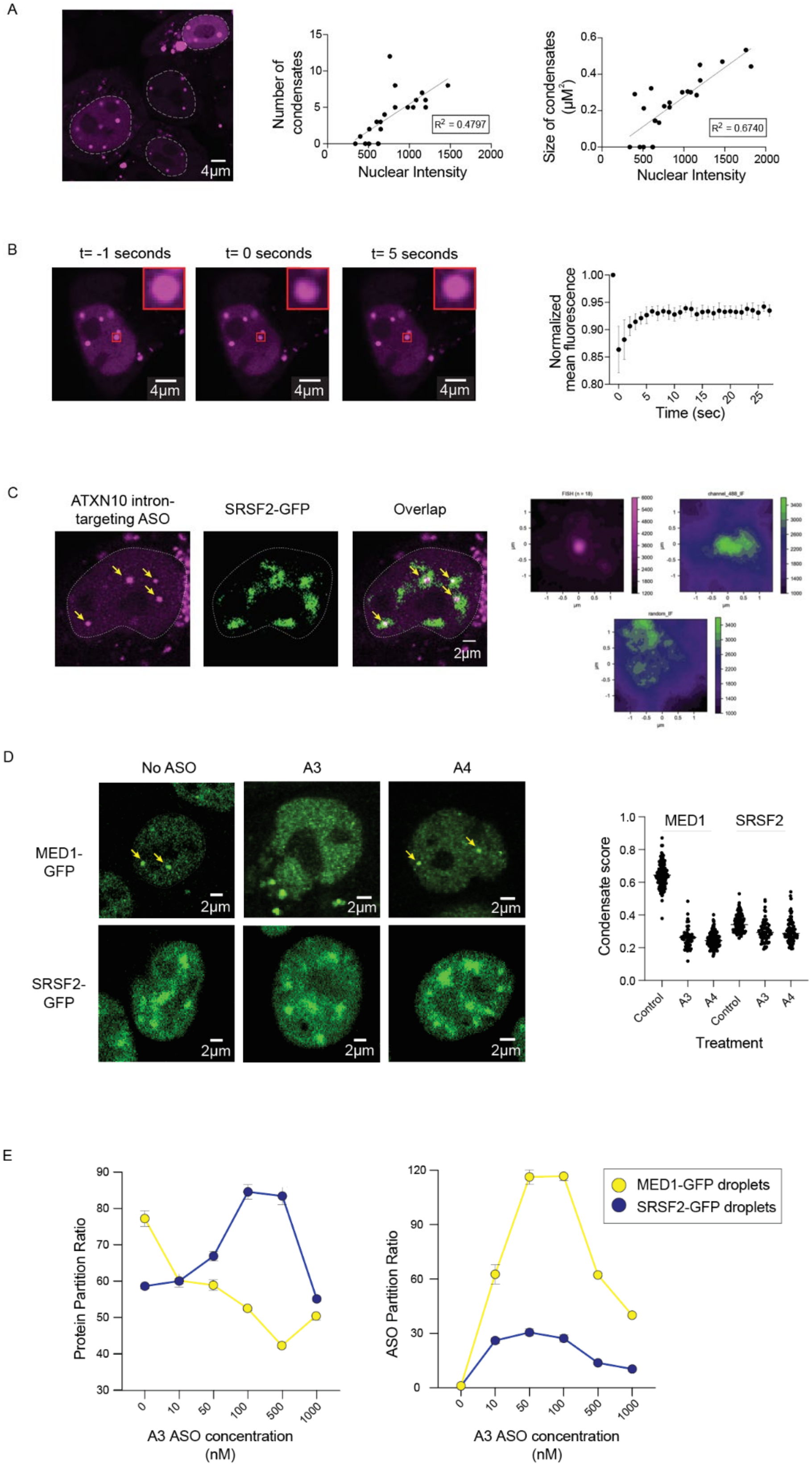

A
4µm
Number of condensates
15
10
5
0
0
500
1000
1500
2000
Nuclear Intensity
R² = 0.4797
Size of condensates (µM²)
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
Nuclear Intensity
R² = 0.6740
B
t= -1 seconds
t= 0 seconds
t= 5 seconds
4µm
4µm
4µm
Normalized mean fluorescence
1.00
0.95
0.90
0.85
0.80
0
5
10
15
20
25
Time (sec)
C
ATXN10 intron-targeting ASO
SRSF2-GFP
Overlap
2µm
FISH (n = 18)
channel_488_IF
random_IF
µm
D
No ASO
A3
A4
MED1-GFP
SRSF2-GFP
2µm
2µm
2µm
2µm
2µm
2µm
Condensate score
1.0
0.8
0.6
0.4
0.2
0.0
MED1
SRSF2
Control
A3
A4
Control
A3
A4
Treatment
E
Protein Partition Ratio
90
80
70
60
50
40
30
0
10
50
100
500
1000
A3 ASO concentration (nM)
ASO Partition Ratio
120
90
60
30
0
0
10
50
100
500
1000
A3 ASO concentration (nM)
MED1-GFP droplets
SRSF2-GFP droplets

Figure 3. Condensate formation, colocalization, and effects of PS-ASOs in cells. (A) Representative confocal microscopy image of PS-ASO A3 transfected into HCT116 cells at 60nM for 16 hours, imaged at 63X on a confocal fluorescent microscope. Nuclei are outlined with dashed white lines. Quantification of number (middle) of condensates and size (right) of condensates of cells with varying nuclear intensities is shown. Nuclei labeled with white dashed lines. (B) Fluorescence recovery after photobleaching (FRAP) analysis of PS-ASO puncta formed by A3 transfected into HCT116 cells at 60nM for 16 hours, imaged at 63X. Representative images of puncta at -1 seconds, 0 seconds, and 5 seconds post-bleaching. Red box shows zoom in of bleached punctum. Quantification of normalized mean fluorescence intensity over time is quantified on the right (N=3 puncta). Error bars represent SEM. (C) Confocal microscopy image of Cy5-labeled ATXN10 intron-targeting PS-ASO (magenta) transfected at 60nM for 16 hours into HCT116 cells expressing GFP-tagged SRSF2 (green), imaged at 63X. Yellow arrows highlight examples of ASO puncta overlap with SRSF2 puncta. Nuclei labeled with white dashed lines. Overlap analysis of PS-ASO puncta with SRSF2 signal shown to the right. (D) Confocal microscopy images of cells expressing GFP-tagged MED1 or SRSF2 (green) transfected with A3 or A4 PS-ASOs at 60nM for 16 hours, imaged at 63X. Yellow arrows highlight larger MED1 puncta that appear disrupted upon treatment with A3. Nuclei labeled with white dashed lines. Condensate score analysis for MED1 and SRSF2 shown on the right. (E) Partition ratio of indicated protein (left) or ASO (right) in droplets formed by MED1 (yellow) or SRSF2 (blue) in 100mM NaCl and 10% PEG with 100nM protein and 100nM ASO imaged at 150X on a confocal fluorescent microscope with a range of concentrations of A3 PS-ASO added to the droplets. Error bars represent SEM.

To explore whether ASOs do indeed affect transcriptional and splicing condensates and whether these effects vary across these condensates, we visualized and quantified condensate formation of MED1 and SRSF2 in live cells upon transfection with A3 for 16 hours. This experiment suggested that A3 disrupts condensate formation, quantified by a condensate score developed (Klein et al., 2020), for transcriptional condensates but not splicing condensates (Figure 3D). We noted that A3 seems to disrupt the formation of larger MED1 condensates, while smaller MED1 condensates still appear. However, the functional differences in larger and smaller MED1 condensates are, as of yet, not well understood, so it is not clear what the functional consequences of the disruption of these large transcriptional condensates is. We confirmed that similar observations of condensate disruption of MED1 condensates could be observed at an earlier time point, 4 hours of treatment, although at this early timepoint, fewer cells had built up ASO in the nucleus, limiting the ability to quantify a condensate score (Figure S3A). It is also important to note that ASO treatment may affect splicing condensates in ways other than those visible by visualizing the puncta formed by SRSF2, a marker of splicing condensates. Further, the degree to which the modulation of any biomolecular condensate contributes to the efficacy or toxicity of ASOs is not yet known. Nonetheless, these results suggest that PS-ASOs may affect specific condensates differently in cells.

Based upon these observations, we sought to better understand these potential differential condensate modulation effects of PS-ASOs on transcriptional and splicing condensates in vitro. We performed a droplet assay in which droplets formed by fluorescently-labeled recombinant MED1 or SRSF2 were treated with a range of concentrations of A3 ranging from 10nM-1uM ASO and imaged by confocal microscopy. Partition ratios were then calculated for both the protein and the ASO. The partition ratios of MED1 and SRSF2 indicated that A3 affects MED1 and SRSF2 condensate formation differently, at the selected protein concentration, across the ASO concentration range tested (Figure 3E). The addition of increasing concentrations of ASO increasingly disrupted MED1 condensates, up until 1uM, where partitioning increased over the 500nM treatment. SRSF2 displayed the opposite trend, with increasing concentrations of ASO promoting SRSF2 droplet formation to a peak at 100nM followed by disruption of droplet formation at 1uM ASO. Notably, the behavior of SRSF2 droplets in response to treatment with ASO is consistent with the re-entrant phase separation observed in complex coascervation, a type of liquid-liquid phase separation mediated by electrostatic interactions between polyelectrolytes with opposite charges (Srivastava et al., 2016). In complex coascervates, low levels of RNA can enhance condensate formation, whereas high levels can disrupt condensate formation, as a function of the charge balance of interacting components. The enrichment of positively charged residues such as arginines in SRSF2 (serine/arginine-rich splicing factor 2) could account for this behavior in response to the addition of increasing concentrations of a negatively charged nucleic acid (Sapra et al., 2009). Examination of the partition ratios of the ASO at each of these concentrations indicated that MED1 and SRSF2 displayed a similar pattern of increases in partitioning between 10-50nM, followed by decreases up to 1uM ASO. However, consistent with the differential partitioning of PS-ASOs observed in Figure 2, the partitioning of A3 into MED1 was 2-4X higher at each concentration (Figure 3E). These results suggest that ASOs display differential modulation of condensates in vitro and in vivo, and that electrostatic interactions may contribute to these effects.

Characterizing the role of 2' modifications on differential partitioning

As 2' modifications of ASOs differ in their chemical properties and have been shown to affect ASO efficacy and toxicity, we hypothesized that 2' modifications may affect the partitioning behavior of ASOs into condensates. 2' modifications differ in their hydrophobicity, which, among other chemical characteristics, has been shown to affect condensate partitioning (Shen et al., 2019; Roberts et al., 2020). To interrogate the effect of 2' modifications on the condensate partitioning behavior of PS-ASOs, we examined A3, which contains 2'MOE modifications, and two ASOs with the same sequence as A3, differing only in their 2' modifications, containing either 2'F (hereafter called A4) or 2'LNA (hereafter called A5) modifications on the RNA wings of the gapmers. These three ASOs have been characterized in cellular toxicity assays and in rodents, and their hydrophobicity and hepatotoxicity were found to be correlated (ordered from least to most hydrophobic and toxic: A3, A5, A4) (Figure 4A) (Shen et al., 2018; Shen et al., 2019).

These three sequence-matched ASOs were added to droplets formed by each of the five condensate markers we had previously examined in vitro and their partitioning behavior was measured in an in vitro droplet assay. Results of this assay indicated that 2' modifications altered the condensate partitioning behavior of these ASOs in different ways depending on the condensate marker (Figure 4B-D). For example, for MED1, SRSF2, and p54nrb, the relative partition ratios from highest to lowest were 2'MOE>2'LNA> 2'F, whereas for NPM1 and DDX6, the order was 2'LNA>2'F>2'MOE. These results suggest that the chemical differences endowed by the 2' modifications alter the propensity of these PS-ASOs to partition into different condensates in a manner that is dictated by the physicochemical properties of the condensate.

Figure 4

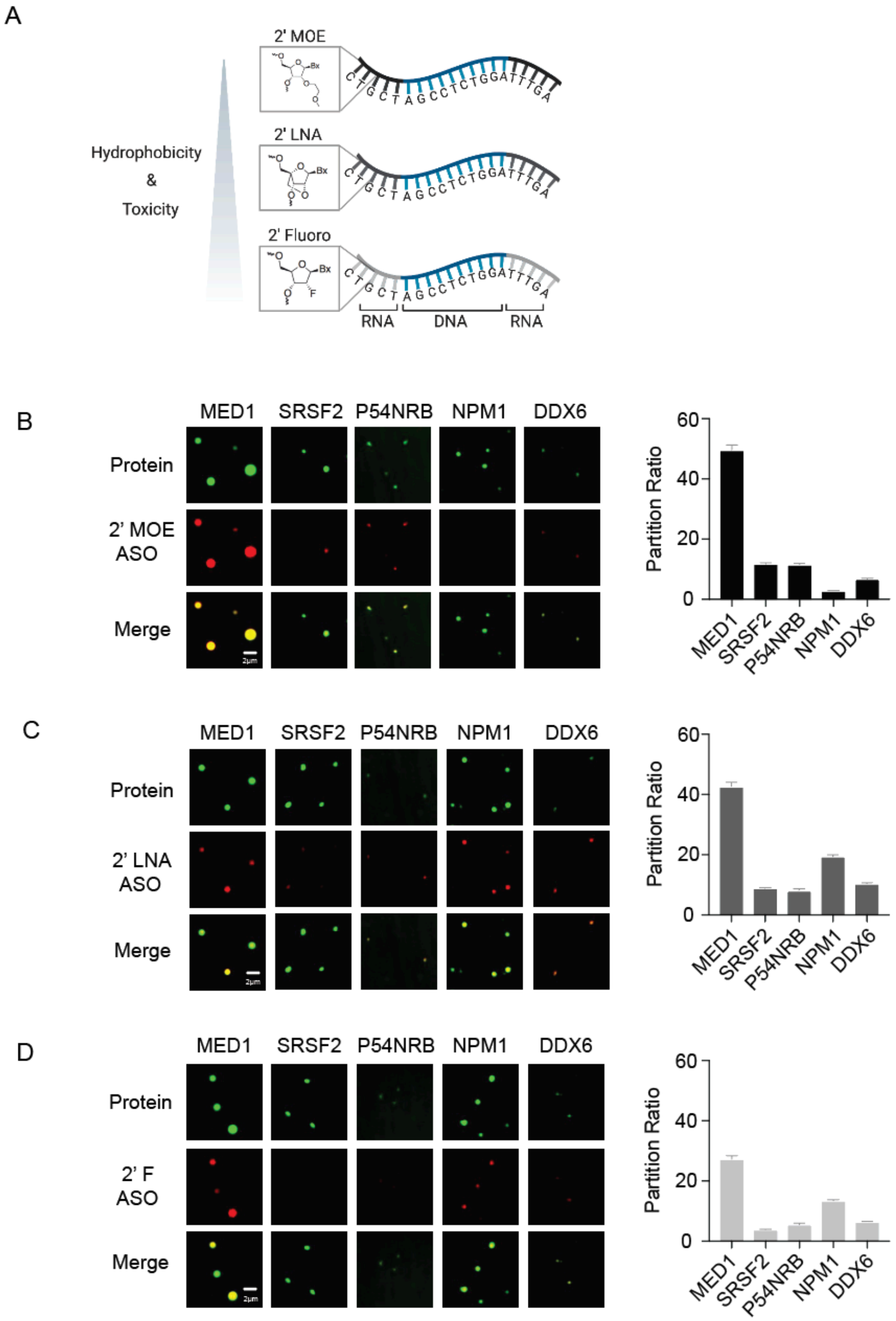


Figure 4. 2' Modifications alter the partitioning behavior of PS-ASOs in vitro. (A) Relative toxicity and hydrophobicity of sequence-matched PS-ASOs with 2'MOE (A3), 2'LNA (A5), or 2'Fluoro (A4) modifications. (B-D) In vitro droplet assays showing the behavior of Cy5-labeled sequence-matched PS-ASOs in the presence of droplets formed by MED1, SRSF2, P54NRB, NPM1, and DDX6 in 100mM NaCl and 10% PEG with 100nM protein and 100nM ASO imaged at 150X on a confocal fluorescent microscope. Representative images of protein (green), ASO (red), and

merge are shown. Quantification of enrichment of the drug is shown on the right. Error bars represent SEM.

Preliminary studies of this set of ASOs with different 2’ modifications in cells have suggested that while these ASOs form puncta, similarly to A3, these puncta do not seem to overlap substantially with markers of condensates (data not shown). We have also tested the effect of two of these ASOs, the most and least toxic (A4 and A3, respectively) on transcriptional and splicing condensates in live cells transfected with 60nM ASO for 16 hours and found that condensate score analysis suggests similarly disrupted MED1 condensate formation, while SRSF2 condensate scores did not change (Figure 3D). Upon examination of MED1 puncta in cells, these ASOs may display differential effects, as A3 appeared to disrupt large MED1 puncta, while these puncta remained present in A4-transfected cells. It is possible that A4 exerts its MED1 condensate effects, as seen in the decreased condensate score, by disrupting smaller MED1 condensates, but as these small MED1 condensates near the limit of resolution of the microscope, this possibility will be difficult to visualize definitively. We also confirmed that cellular phenotypes of MED1 condensate disruption appeared similar in cells transfected with A3 and A4 for a shorter time period, 4 hours (Figure S3A). These preliminary results suggest that 2’ modifications may alter condensate effects of ASOs, but what, if any, role condensate modulation in cells plays in ASO efficacy and toxicity is not yet known.

Discussion and Future Directions

Here, we present ongoing studies of the role of condensate partitioning and modulation on the therapeutic efficacy and toxicity of ASOs. Several lines of evidence from previous studies inspired the hypothesis that biomolecular condensates may shape the therapeutic window of ASOs, suggesting that their future study may enable improvement of these nucleic acid-based therapeutics. Such evidence includes the findings that the differential partitioning of small molecule cancer therapeutics shapes their therapeutic activity, that PS-ASOs form concentration-dependent puncta in cells, and that 2’ nucleotide modifications that contribute to their efficacy and toxicity can affect their subcellular distribution (Shen et al., 2014; Liang et al., 2015; Shen et al., 2015; Shen et al., 2018; Vickers and Crooke, 2016).

Preliminary evidence testing the condensate partitioning behavior of FDA-approved ASOs and ASOs with diverse 2’ chemical modifications in vitro and in vivo has provided early support for our hypothesis. However, limitations of resolution for imaging in cells and minimal overlap of ASOs with known biomolecular condensates have presented challenges for evaluating the role of condensate partitioning in the function of ASOs.

Ongoing studies will aim to address these challenges in a variety of ways. As mentioned above, we will continue to test additional antibodies to evaluate overlap. We will also continue to test other ASOs containing various 2’ modifications. One toxic modification in particular, the constrained ethyl modification, while thus far difficult to procure, has demonstrated the most drastic changes in subcellular localization, localizing to the nucleolus and disrupting the localization of condensate proteins such as p54nrb (Shen et al., 2019). Partitioning behavior with respect to the nucleolus is an attractive phenotype that could overcome limitations of the scale of several condensates studied thus far as nucleoli are orders of magnitude larger than transcriptional condensates (Lafontaine et al., 2021; Forman-Kay et al., 2022)

A key goal for future studies will be to probe how, if at all, condensate partitioning and modulation effects uncovered thus far shape the therapeutic activity of ASOs. Several models for the role of condensate formation by PS-ASOs could be consistent with what we and others have observed experimentally thus far. For example, PS Bodies may be sites of action for ASO degradation or steric blocking activity, they may be sites of storage, allowing for protection and release of ASO over time, or PS Bodies may concentrate most of the ASO in the nucleus while a small proportion of ASOs actually exert their effects on target RNAs or other condensates outside of PS Bodies. Further, while experiments examining the effect of ASOs on condensate formation of markers of transcriptional and splicing condensates have suggested that ASOs modulate condensate differentially, what role this modulation plays in the therapeutic window of ASOs is not yet clear.

To gain insights that enable the design of more efficacious and less toxic ASOs, we will need to be able to alter the condensate partitioning and modulation of ASOs to hone in on their impact on ASO therapeutic window. We are currently developing artificial in-cell condensate systems to this end such as condensates tethered to a lac array. Specifically, we will aim to test whether the enrichment of PS-ASOs into condensate formed by a variety of proteins that ASOs interact with, such as p54nrb and DDX6, decreases or increases ASO RNA degradation activity, measured by qPCR. The ability to modulate PS Bodies chemically could also allow us to test the effects of decreased PS Body formation on ASO activity and toxicity, discerning between the models discussed above.

Another challenge in this endeavor is that the contents of PS Bodies, other than PS-ASOs, is not known. Although one protein, TCP1B, has been suggested to overlap with PS Bodies, we have not yet found this result to be reproducible. A better understanding of what, if any, proteins are concentrated with PS-ASOs into PS Bodies would greatly improve the study of how these bodies affect ASO activity.

ASOs may also prove to be useful tools for probing and altering condensate properties in cells. Their ability to target diverse RNAs, in concert with their potential role in affecting proteins that form condensates, once better understood, could be leveraged to target a wide variety of condensate components. Ultimately, the study of condensate partitioning and modulation of ASOs could enable more precise use of these and other nucleic acid-based therapeutics to target RNAs and potentially a wide variety of condensates in lab and in the clinic.

Supplementary Figures

Figure S1

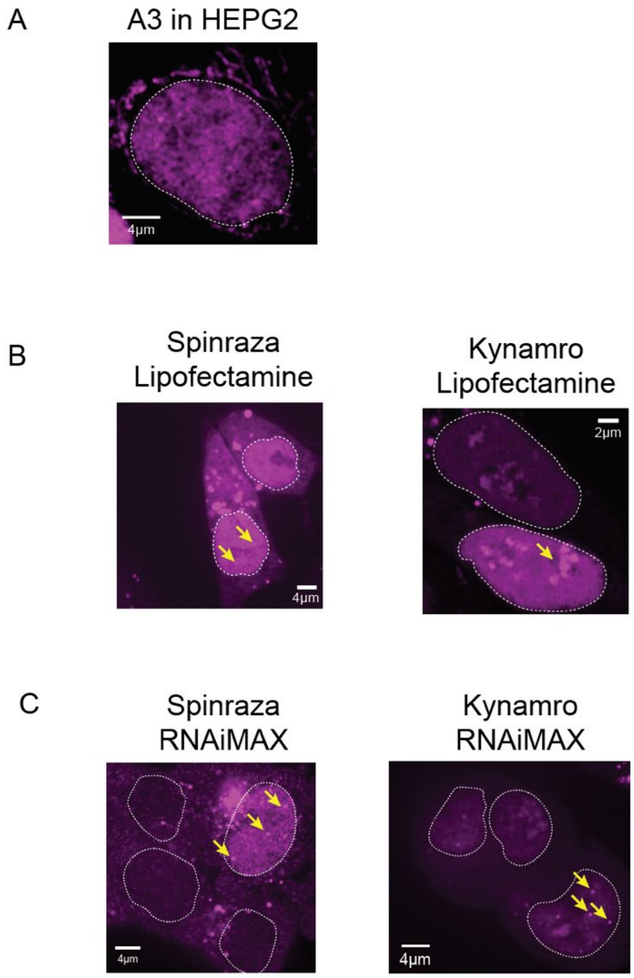


Figure S1. PS-ASO puncta formation in cells. (A) Confocal microscopy images of fixed HEPG2 cells, transfected with Cy5-labeled PS-ASO A3 at 60nM with Lipofectamine 3000 for 16 hours. Nucleus labeled with white dashed line. Images were collected with RPI Spinning Disk confocal microscopy at 100X magnification. (B) Confocal microscopy images of Cy5-labeled Spinraza or Kynamro (magenta) transfected at 60nM with Lipofectamine 3000 into HCT116 cells for 16 hours. Nuclei labeled with white dashed lines. Yellow arrows highlight several ASO puncta formed. Live-cell imaging was done with RPI Spinning Disk confocal microscopy (left) or Zeiss LSM 880 confocal microscopy with Airyscan processing (right). (C) Confocal microscopy images of Cy5-labeled Spinraza or Kynamro (magenta) transfected at 500nM with RNAiMAX into HCT116 cells. Nuclei labeled with white dashed lines. Yellow arrows highlight several ASO puncta formed. Live-cell imaging was done with RPI Spinning Disk confocal microscopy.

Figure S2

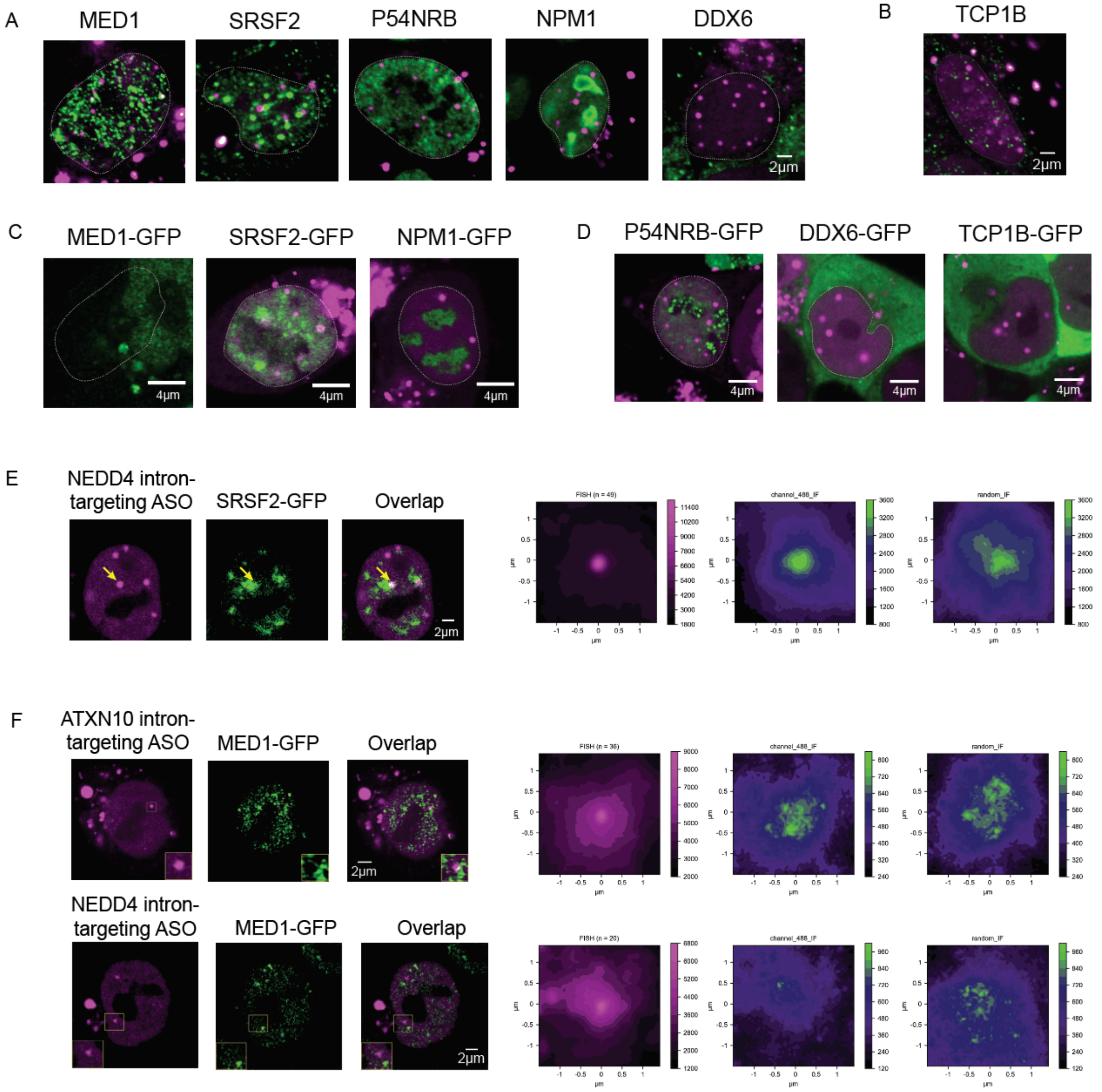


Figure S2. PS-ASO puncta colocalization in cells. (A) Confocal microscopy images of Cy5-labeled A3 (magenta) transfected at 60nM into HEPG2 cells for 16 hours, with immunofluorescence for indicated protein markers MED1, SRSF2, P54NRB, NPM1 and DDX6 (green). Nuclei labeled with white dashed lines. PS-ASO puncta did not demonstrate substantial overlap with tested markers. (B) Confocal microscopy images of Cy5-labeled A3 (magenta) transfected at 60nM into HEPG2 cells for 16 hours, with immunofluorescence for TCP1B (green). Nucleus labeled with white dashed lines. Nuclear PS-ASO puncta did not demonstrate overlap with this protein. (C) Confocal microscopy images of Cy5-labeled A3 (magenta) transfected into HCT116 cells endogenously expressing GFP tagged proteins MED1, SRSF2, and NPM1 (green). Nuclei labeled with white dashed lines. PS-ASO puncta did not demonstrate substantial overlap with tested markers. (D) Confocal microscopy images of Cy5-labeled A3 (magenta) transfected into HCT116 cells stably expressing GFP-tagged proteins P54NRB, DDX6, and TCP1B (green). Nuclei labeled with white dashed lines. PS-ASO puncta did not demonstrate substantial overlap with tested markers. (E) Confocal microscopy images of Cy5-labeled NEDD4 intron-targeting PS-ASO (magenta) transfected at 60nM for 16 hours into HCT116 cells expressing GFP-tagged SRSF2 (green). Overlap analysis of PS-ASO puncta with SRSF2 signal shown to the right. Nuclei labeled with white dashed lines. (F) Confocal microscopy images of Cy5-labeled ATXN10 intron-targeting PS-ASO (top) or NEDD4 intron-targeting PS-ASO (bottom) (both magenta) transfected into HCT116 cells expressing GFP-tagged MED1 (green). Overlap analysis of PS-ASO puncta with MED1 signal shown to the right. Nuclei labeled with white dashed lines.

Figure S3

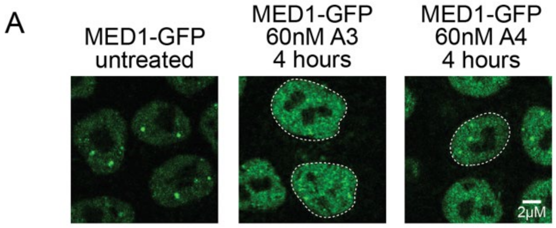


Figure S3. Effects of PS-ASOs on MED1 condensates in live cells. (A) Representative confocal microscopy images of HCT116 cells expressing MED1-GFP untreated (left), or treated with 60nM A3 (middle) or A4 (right) for 4 hours. White dashed lines highlight nuclei with enrichment of ASO signal in the 640 channel, as some cells did not take up ASO. Early evidence of cellular phenotypes of disrupted MED1 condensate formation observed and quantified at the 16-hour treatment time point can be seen in these 4-hour treatment images.

Materials and Methods

Procurement and preparation of ASOs

ASOs were synthesized by Integrated DNA Technologies with a 5'Cy5 label. ASOs were purified by HPLC and underwent Na+ salt exchange. Lyophilized ASOs were resuspended in nuclease-free water to stocks of 100uM and stored at -80C. The following ASO sequences were ordered:

Spinraza-
/5Cy5//i2MOErT/*/i2MOErC/*/i2MOErA/*/i2MOErC/*/i2MOErT/*/i2MOErT/*/i2MOErT/*/i2MOErC/*/i2MOErA/*/i2MOErT/*/i2MOErA/*/i2MOErA/*/i2MOErT/*/i2MOErG/*/i2MOErC/*/i2MOErT/*/i2MOErG/*/32MOErG/

Kynamro-
/5Cy5//i2MOErG/*/i2MOErC/*/i2MOErC/*/i2MOErT/*/i2MOErC/*A*G*T*/iMe-dC/*T*G* /iMe-dC/*T*T*/iMe-dC/*/i2MOErG/*/i2MOErC/*/i2MOErA/*/i2MOErC/*/32MOErC/

A3-
/5Cy5//i2MOErC/*/i2MOErT/*/i2MOErG/*/i2MOErC/*/i2MOErT/*A*G*C*C*T*C*T*G*G*A*/i2MOErT/*/i2MOErT/*/i2MOErT/*/i2MOErG/*/32MOErA/

A4-
/5Cy5/+C*+T*+G*+C*+T*A*G*C*C*T*C*T*G*G*A*+T*+T*+T*+G*+A

A5-
/5Cy5/i2FC*i2FU*i2FG*i2FC*i2FU*A*G*C*C*T*C*T*G*G*A*i2FU*i2FU*i2FU*i2FG*i2FA

NEDD4 Intronic ASO-
/5Cy5//i2MOErC/*/i2MOErG/*/i2MOErT/*/i2MOErT/*/i2MOErG/*T*A*C*G*C*A*A*T*G*C*/i2MOErA/*/i2MOErG/*/i2MOErG/*/i2MOErA/*/32MOErT

ATXN10 Intronic ASO-
/5Cy5//i2MOErG/*/i2MOErT/*/i2MOErC/*/i2MOErT/*/i2MOErT/*A*T*T*C*T*C*T*A*T*C*/i2MOErA/*/i2MOErA/*/i2MOErC/*/i2MOErG/*/32MOErC/

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 mEGFP 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 and 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 described. 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.

In vitro droplet assay with ASO

Recombinant GFP 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) at a working concentration of 100mM NaCl and 10% PEG and mixed. The Cy5-labeled ASO of interest was then added to this mixture at 5X the indicated final concentration of the ASO and the solution was mixed and incubated at room temperature in the dark for 1 hour. The solution was then loaded onto glass bottom 384 well plate (Cellvis P384-1.5H-N) and imaged with an Andor confocal microscope with a 150x objective. Images presented are of droplets settled on the glass coverslip. For each experiment at least 15 images were taken, typically capturing about 50-200 droplets for analysis in total. Images were post-processed using Fiji Is Just ImageJ (https://fiji.sc/).

In vitro droplet assay analysis

Partition ratios were calculated as previously published (Sabari et al., 2018, Henninger et al., 2020). In short, to analyze in vitro droplet assay images, custom MATLAB scripts were written to identify droplets formed by the indicated protein. Intensity threshold were set based on the peak of the histogram and size thresholds (2 pixel radius). Droplets were then defined as a region of interest in FIJI based upon the protein channel, and the mean fluorescence of the ASO or protein channel within the region of interest was determined. To calculate the partition ratios, the mean fluorescence intensity within the region of interest was divided by the background mean fluorescence intensity of the rest of the channel, excluding the region of interest. This represents the relative enrichment of the fluorescent molecule of interest, such as the Cy5-labeled ASO, in the in vitro protein droplets.

Cell culture

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)) at 37°C with 5% CO_2 in a humidified incubator. HepG2 cells (ATCC HB-8065™) were cultured in EMEM (ATCC 30-2003) supplemented with 10% FBS (Sigma Aldrich, F4135) at 37°C with 5% CO_2 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 and dissociate cell clumps. To ensure proper cell dissociation, cells were incubated with TrypLE at 37°C with 5% CO_2 in a humidified incubator for 5 minutes. TrypLE was quenched with DMEM or EMEM, depending on the cells, supplemented with 10% FBS and cells were plated in new tissue culture-grade plates.

Cell line generation

Cell lines with mEGFP-tagged MED1, SRSF2, or NPM1 integrated at the endogenous locus were generated using the CRISPR/Cas9 system. 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 mEGFP. 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.

Stable cell lines were generated for HCT116 cells with mEGFP-tagged p54nrb, DDX6, or TCP1B by cloning WT and mutant gene sequences using NEBuilder HiFi DNA Master Mix (NEB E2621S) into a doxycycline-inducible, N-terminal mEGFP-tagged expression construct with a hygromycin-resistance gene, which was integrated into HCT116 cells using the PiggyBac transposon system. 0.5 x 10^6 wildtype HCT116 cells 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 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)). The next day, fresh media was added supplemented with 500 µg/mL hygromcin (ThermoFisher) for selection. Selection media was exchanged every day and un-transfected control cells were monitored to assess selection.

ASO transfection

The 100uM stocks of ASO were diluted in water to 60uM. HCT116 cells were plated at 100k cells per well in 8-well chamber slides (Life Technologies # 155409PK) and transfected for a final concentration of 60nM with Lipofectamine 3000 (Invitrogen L3000) at 3uL lipofectamine/mL for 16 hours, unless otherwise indicated. HEPG2 cells were plated at 50k cell/well and incubated for 2 days at 37C before transfection at 60nM ASO with Lipofectamine 3000 (Invitrogen L3000).

Imaging for live cells

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). mEGFP-labeled proteins or proteins subjected to immufluorescence were visualized in the 488 channel, whiel Cy5-labeled ASOs were visualized in the 640 channel. Images were processed using Fiji is Just ImageJ (FIJI). 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 v.3.6.9. A Python package, *cellpose*, was used 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 threshold cutoff of 3 standard deviations above the mean of the image intensity.

Live cell 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 nucleus, the fluorescence signal in the GFP channel from the z-stacks (corresponding to either MED1 or SRSF2) was maximally-projected. 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. Finally, to derive the condensation score, these values were subtracted from 1.

FRAP of ASOs

FRAP analysis was performed on LSM880 Airyscan microscope with 488 nm laser. Photobleaching was performed by defining and exposing a region of interest around a punctum to 100% laser power. Images were collected every 1 second. FIJI was used to calculate intensity values in 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 normalized mean fluorescence intensities were plotted over time.

Immunofluorescence

HEPG2 cells were plated in 24-well plate at 50k cells per well. Cells were then incubated at 37°C for 2 days then transfected with the ASO of interest at 60nM as described above. After 16 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 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). The following primary antibodies were used:

MED1- Abcam ab64965
SRSF2- Abcam ab11826
P54NRB- Abcam ab227014
NPM1- Thermo Fisher 32-5200
DDX6- Thermo Fisher PA555012

IF RNA FISH 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=4*π*·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.

References

Alberti, S., Gladfelter, A., Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Physiol. Behav.* **176**, 139–148 (2017).

Bailey, J. K., Shen, W., Liang, X. H. & Crooke, S. T. Nucleic acid binding proteins affect the subcellular distribution of phosphorothioate antisense oligonucleotides. *Nucleic Acids Res.* **45**, 10649–10671 (2017).

Brown, D. A. et al. Effect of phosphorothioate modification of oligodeoxynucleotides on specific protein binding. *J. Biol. Chem.* **269**, 26801–26805 (1994).

Crooke, S. T. & Geary, R. S. Clinical pharmacological properties of mipomersen (Kynamro), a second generation antisense inhibitor of apolipoprotein B *Br J of ClinPharmacol* **76**, 269-276 (2012).

Crooke, S. T., Vickers, T. A. & Liang, X. Phosphorothioate modified oligonucleotide – protein interactions. **48**, 5235–5253 (2020).

Crooke, S. T., Wang, S., Vickers, T. A., Shen, W. & Liang, X. H. Cellular uptake and trafficking of antisense oligonucleotides. *Nat. Biotechnol.* **35**, 230–237 (2017).

Dieckmann, A. et al. A sensitive in vitro approach to assess the hybridization- dependent toxic potential of high affinity gapmer oligonucleotides. *Mol. Ther. Nucleic Acids* **10**, 45–54 (2018).

Fazal, F. M. *et al.* Atlas of Subcellular RNA Localization Revealed by Resource Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473-490.e26 (2019).

Forman-Kay, J. D., Ditlev, J. A., Nosella, M. L. & Lee, H. O. What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates? *Rna* **28**, 36–47 (2022).

Frazier, K. S. Antisense Oligonucleotide Therapies : The Promise and the Challenges from a Toxicologic Pathologist's Perspective. *Toxicologic Pathology* **43**, 78–89 (2015).

Garcia-Jove Navarro, M. *et al.* RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nat. Commun.* **10**, 1–13 (2019).

Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).

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. RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225.e24 (2021).

Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).

Kamola, P. J. et al. Strategies for in vivo screening and mitigation of hepatotoxicity associated with antisense drugs. *Mol. Ther. Nucleic Acids* **8**, 383–394 (2017).

Kakiuchi-Kiyota, S., Whiteley, L. O., Ryan, A. M. & Mathialagan, N. Development of a method for profiling protein interactions with LNA- modified antisense oligonucleotides using protein microarrays. *Nucleic Acid Ther.* **26**, 93–101 (2016).

Kakiuchi-Kiyota, S. et al. Comparison of hepatic transcription profiles of locked ribonucleic acid antisense oligonucleotides: evidence of distinct pathways contributing to non-target mediated toxicity in mice. *Toxicol. Sci.* **138**, 234–248 (2014).

Khvorova, A. & Watts, J. K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **35**, 238–248 (2017).

Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science* **368**, 1386–1392 (2020).

Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021).

Lai, F., Damle, S. S. & Ling, K. K. Directed RNase H Cleavage of Nascent Transcripts Article Directed RNase H Cleavage of Nascent Transcripts Causes Transcription Termination. *Mol. Cell* **77**, 1032-1042.e5 (2020).

Liang, X. H., Shen, W., Sun, H., Prakash, T. P. & Crooke, S. T. TCP1 complex proteins interact with phosphorothioate oligonucleotides and can co-localize in oligonucleotide-induced nuclear bodies in mammalian cells. *Nucleic Acids Res.* **42**, 7819–7832 (2014).

Liang, X. H., Sun, H., Shen, W. & Crooke, S. T. Identification and characterization of intracellular proteins that bind oligonucleotides with phosphorothioate linkages. *Nucleic Acids Res.* **43**, 2927–2945 (2015).

Liang, X., Sun, H., Nichols, J. G. & Crooke, S. T. RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Mol. Ther.* **25**, 2075–2092 (2017).

Lorenz, P., Baker, B. F., Bennett, C. F. & Spector, D. L. Phosphorothioate antisense oligonucleotides induce the formation of nuclear bodies. *Mol. Biol. Cell.* **9**, 1007–1023 (1998).

Roberts, T. C., Langer, R. & Wood, M. J. A. Advances in oligonucleotide drug. *Nat. Rev. Drug Discov.* **19**, 673-694 (2020).

Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, 1–17 (2018).

Sapra, A. K. *et al.* SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo. *Mol. Cell* **34**, 179–190 (2009).

Shen, X. & Corey, D. R. Chemistry , mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res* **46**, 1584–1600 (2019).

Shen, W. et al. Acute hepatotoxicity of 2' fluoro-modified 5-10-5 gapmer phosphorothioate oligonucleotides in mice correlates with intracellular protein binding and the loss of DBHS proteins. *Nucleic Acids Res.* **46**, 2204–2217 (2018).

Shen, W. *et al.* Chemical modification of PS-ASO therapeutics reduces cellular protein-binding and improves the therapeutic index. *Nat. Biotechnol.* **37**, (2019).

Shen, W., Liang, X. H. & Crooke, S. T. Phosphorothioate oligonucleotides can displace NEAT1 RNA and form nuclear paraspeckle-like structures. *Nucleic Acids Res.* **42**, 8648–8662 (2014).

Shen, W., Liang, X. H., Sun, H. & Crooke, S. T. 2'-Fluoro-modified phosphorothioate oligonucleotide can cause rapid degradation of P54nrb and PSF. *Nucleic Acids Res.* **43**, 4569–4578 (2015).

Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science.* **357**, (2017).

Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* **75**, 549-561.e7 (2019).

Srivastava, S., and Tirrell, M.V. Polyelectrolyte complexation. *Advances in Chemical Physics*. 499–544 (2016).

Swayze, E. E. & Bhat, B. in *Antisense Drug Technology—Principles, Strategies, and Applications* 2nd edn (ed. Crooke, S. T.) 143–182 (CRC Press, 2008).

Swayze, E. E. et al. Antisense oligonucleotides containing locked nucleic acid improve potency but cause significant hepatotoxicity in animals. *Nucleic Acids Res.* **35**, 687–700 (2007).

Vickers, T. A. & Crooke, S. T. Development of a quantitative BRET affinity assay for nucleic acid-protein interactions. *PLoS ONE* **11**, e0161930 (2016).

Wang, Y., Shen, W., Liang, X. & Crooke, S. T. Phosphorothioate Antisense Oligonucleotides Bind P-Body Proteins and Mediate P-Body Assembly. *Nucleic Acid Therapeutics* **29**, 343–358 (2019).

Wurster, C. D. & Ludolph, A. C. Nusinersen for spinal muscular atrophy. *Ther Adv Neurol Disord* **11**, 1-3 (2018).

CHAPTER 5:
FUTURE DIRECTIONS AND DISCUSSION

The study of biomolecular condensates and liquid-liquid phase separation in the cell has already begun to inspire novel mechanistic and therapeutic hypotheses. As a result, condensate dysregulation has emerged as a diverse set of novel disease mechanisms affecting various emergent mesoscale properties of these membraneless organelles (Tsang et al., 2020). Concepts of biphasic modulation and differential partitioning of therapeutics have suggested condensates may represent tunable, screenable drug targets (Choi et al., 2018; Babinchak et al., 2020; Klein et al., 2020). However, several open questions remain to realize the promise of the study of condensates for developing improved and novel therapies for patients.

Condensates in disease

Further mechanistic study of condensate dysregulation mutations

The map of condensate dysregulation mutations created in Chapter 2 of this thesis could serve as a foundation for myriad future studies further exploring what properties and functions of condensates are disrupted by a mutation of interest and how this disruption drives disease. We identified over 36,000 pathogenic mutations that are predicted to cause condensate dysregulation phenotypes in over 1,200 congenital diseases and 550 cancers. Thus, research across a spectrum of diseases is likely to gain mechanistic hypotheses from the study of the role of condensate dysregulation in pathogenesis.

Future work could entail exploring how condensate dysregulation manifests as a result of a specific mutation. For a given mutation, asking whether a predicted condensate dysregulating mutation affects emergent properties such as the partitioning of the mutated component, the ability of other components to form the condensate, or the material properties of the condensate of interest could suggest how to therapeutically target this dysregulation. For example, for condensate dysregulating mutations in a transcriptional protein, future work could focus on assaying whether transcriptional condensates are able to form across the genome or at specific loci, and how the condensate dysregulation phenotype observed affects functional output of transcriptional condensates, such as the recruitment of RNA Polymerase II and the transcription of mRNA. Such studies are likely to yield further disease mechanistic insights into mutations that have not yet been studied, such as variants considered to be of unknown significance, as well as into mutations that have been studied in the context of their disruption of known molecular-scale mechanisms discussed in Chapter 2.

Because many factors that are important for protein function, such as interaction strength and concentration, are altered in disease, it will be important to distinguish the contribution of condensate dysregulation to disease mechanism. In other words, molecular-scale disruptions are likely to have mesoscale consequences on condensates, but discerning how these mesoscale consequences drive pathogenesis will be important. Experiments in which condensate principles are used to phenocopy or rescue disease phenotypes may be helpful to this end. For example, if a truncating mutation removes condensate-promoting features from a condensate-forming protein and thus disrupts the condensate it forms, a phenocopying experiment may disrupt the condensate via chemical modulation to show that condensate dysregulation plays a role in the downstream pathogenesis. In this case, rescuing the disease phenotype by re-introducing condensate-promoting features, such as by creating a chimera of the mutated protein with condensate-promoting features different in sequence but capable of similar interactions, would suggest that the loss of these features contributes to pathogenesis.

Missense mutations and condensate dysregulation

Several of the first examples of mutations identified that cause condensate dysregulation phenotypes were missense mutations in neurological disease, but we still lack general principles to predict the effect of missense mutations on condensate-forming proteins (Alberti and Dormann, 2019). In the striking condensate hardening and aggregation phenotypes caused by missense mutations in proteins such as FUS in ALS, point mutations are capable of drastically altering interactions of proteins and their propensity to form condensates that undergo a liquid to solid transition (Patel et al., 2015; Murakami et al., 2015; Conicella et al., 2016). Studies of the molecular grammar of condensates have identified weak multivalent interactions across stretches of residues, such as low complexity sequence features (LCSs), to be important for promoting condensate formation (Banani et al., 2017). However, while tools currently exist to predict the effect of missense mutations on interactions and 3D folding of structured protein regions such as MIDs, how missense mutations in a single residue in stretches of LCSs can have such drastic condensate dysregulation effects is not yet understood. Also, when missense mutations are predicted to affect aspects of 3D structure and function, the extent to which these effects alter condensate formation is also poorly understood.

In Chapter 2 of this thesis, we mapped condensate-promoting features across condensate-forming proteins and defined putative condensate-dysregulating mutations as those that fall within these features. For example, missense mutations that fall in an LCS known to promote condensate formation, such as an acidic patch in an IDR, would be nominated as condensate-dysregulating. Among the experimentally tested examples, we noted that although nearly 90% of our predicted mutations caused condensate dysregulation phenotypes, all mutations tested that did not appear to display condensate dysregulation phenotypes were missense mutations. Although these mutations may affect condensates in ways other than the morphology assayed in mESCs, these data suggest our predictions of the effect of truncating mutations in removing several amino acids in condensate-promoting features may be more accurate than that of missense mutations, based upon our current understanding.

Further mutational studies sampling different types of missense mutations in LCSs, placing these mutations into the context of the interactions important for the condensate formed by the protein of interest, and assaying the resulting phenotype after mutation, will be important for developing a stronger understanding of the effect of missense mutations on condensates. Such studies could enable the development of a set of principles that could improve our ability to predict the effect of missense mutations on condensates, thus enabling researchers to hone in on which patient populations are most likely to benefit from condensate-targeting therapeutics.

Leveraging emergent properties of condensates to study condensate dysregulation

While several reviews have posited that condensate dysregulation is likely to be a common feature across many diseases such as neurological disease, cancers, and infectious diseases, there is much work to be done in understanding how the emergent properties of condensates and the diverse functions they facilitate, outlined in Chapter 1, are affected by or contribute to pathogenesis (Tsang et al., 2020; Alberti and Hyman, 2021). Emergent properties of LLPS-mediated condensates in the cell include the concentration and compartmentalization of specific proteins and nucleic acids, liquid metastability, and interfacial behaviors (Lyon et al., 2021). The study of these emergent behaviors of condensates, integrating theories of polymer physics, nonequilibrium thermodynamics, advanced microscopy techniques, and molecular biology, has yielded fundamental insights into their functions in cell biology (Hyman et al., 2011; Brangwynne et al., 2015; Holehouse and Pappu 2018, Choi et al., 2020). Emergent properties can provide a useful framework through which to the consequences of dysregulated condensates in disease and how function may be restored via therapeutic modulation of condensates.

Many current studies of condensate dysregulation in disease have shown such phenotypes as altered condensate formation or dissolution, changes in material properties such as liquid to solid transitions, and altered composition of condensates (Alberti and Dormann, 2019). Furthering these studies to understand how these changes in condensates affect the functions they perform in the cell will be an important next step in forming more sophisticated disease mechanistic and therapeutic hypotheses. For example, the mutation of a specific protein in a condensate could have myriad consequences beyond typical phenotypic readouts such as the formation of puncta in cells. While puncta formation or condensate morphology can be helpful places to start, further mechanistic studies should examine the functional consequences of condensate dysregulation, such as changes to the condensate physicochemical environment, internal organization, and interfacial behaviors such as interfacial tension or coalescence. Such studies will be important for determining how condensate dysregulation affects condensate functions discussed in Chapter 1, such as regulating biochemical reaction kinetics and specificity, generating mechanical force, and providing rapid and reversible responses to external stimuli (Lyon et al., 2021).

Condensates in drug development

Partitioning behavior of other drug modalities

Examining the role of condensate partitioning of therapeutic modalities beyond small molecules will be an important future direction for the work described in this thesis. As described in Chapter 3 of this thesis, small molecules were found to partition differentially into biomolecular condensates in vitro and in cells and this partitioning was found to shape therapeutic activity (Klein et al., 2020). Whether other drug modalities display differential partitioning is not yet known and may similarly change how subcellular distribution of other modalities is understood and manipulated for better target engagement. Several drug modalities, including complex macromolecules, protein-based drugs, and DNA and RNA-based therapeutics, play increasingly important roles in the treatment of disease. As nucleic acids, proteins, and metabolites can be concentrated in specific compartments based upon their physicochemical environment and the types of interactions they participate in, condensate partitioning is likely to play an important role in the pharmacodynamics of these drug modalities (Brangwynne et al., 2015).

In Chapter 4 of this thesis, I describe early efforts to examine the role of condensate partitioning in shaping the therapeutic window for antisense oligonucleotide (ASO)-based therapeutics. In this work, we found that ASOs display differential partitioning in vitro independent of the presence of their target RNA, and that chemical modifications that are known to change efficacy and toxicity alter this partitioning. ASOs were also seen to form liquid-like condensates in cells, which overlap with markers of biomolecular condensates in specific cases. Further understanding of how subcellular distribution affects the efficacy of ASOs will be an important future direction for this work. To this end, future studies could ask whether the condensates formed by ASOs represent sites of function or buildup of nonfunctional ASO, and if perturbed, whether ASO condensate formation changes knock-down activity or cellular toxicity.

Physicochemical determinants of drug partitioning

Further study of the physicochemical determinants of drug partitioning into condensates will be important for leveraging the findings in Chapters 3 and 4 of this thesis to develop more efficacious therapies. Specifically, it will be important to characterize what aspects of the physicochemical environment of specific condensates dictate drug partitioning behavior and how this physicochemical environment is established. Molecular grammar studies of condensates have suggested that certain proteins, such as scaffold proteins, and the interactions they participate in, may determine the biochemical milieu of condensates that dictates which proteins, nucleic acids, and small molecules will become concentrated in the condensate versus which will be excluded or pass through via free diffusion (Wang et al., 2018; Klein et al., 2020). Work in this thesis has identified, using a screen of diverse chemical groups partitioning into MED1 condensates, that specific side-chain interactions appear important in dictating which chemical groups concentrate into MED1 condensates. However, the molecular determinants of drug partitioning across diverse condensates remains uncharacterized.

Future studies could work to identify these molecular determinants in vitro and in cells by observing the partitioning behavior across many different types of molecules, using different scaffold proteins or combinations of proteins to represent specific condensates and determining to what extent these mirror in cell behaviors. Machine learning may play a role here in taking small molecule partitioning data from in vitro screens and deciphering the physicochemical determinants that are important in shaping this partitioning. Such work could define a chemical grammar, in addition to the molecular grammar of proteins in condensates, establishing rules that drug developers could use to increase the likelihood that a molecule of interest concentrates in the condensate in which its target is found. In order to recapitulate in cell behaviors in vitro, it will be important to sample a variety of condensate scaffolds for a given condensate of interest to increase the likelihood that screening information is relevant to the in-cell physicochemical environment of the condensate of interest. Sampling diverse chemical features will also be important in these efforts.

Ultimately, the study of condensate partitioning could enable the development of specific chemical groups that bias molecules toward specific condensates, thus increasing their efficacy by concentrating molecules with their targets. Modulation of condensates, discussed in more detail below, may emerge from the same chemical groups as those that shape partitioning, or it may be possible to create therapeutics or chemical tools that have specific domains that target them to specific condensates that can be combined with other features that affect the properties of these condensates or bind to specific targets.

Modulating condensates: physicochemical properties and the potential for biphasic modulation

While a handful of drugs have been observed to modulate condensate formation, as described in Chapter 1, our ability to predict the effect of specific molecules on condensates is rudimentary (Choi et al., 2018; Babinchak et al., 2020). The ability to predict the effect of mutations on condensates was enabled by a foundational understanding of sequence features that can promote condensate formation. Similarly, future study of the interactions that shape the propensity to form condensates and undergo liquid to solid transitions may enable rational design of small molecules that modulate these propensities. Whether these interactions emerge from the same proteins and nucleic acids is not yet well understood, but preliminary evidence has suggested that specific biomolecules may play distinct roles in condensates, such as the role of RNA in modulating viscosity and the role of specific proteins in modulating surface tension (Elbaum-Garfinkle et al., 2015; Brangwynne et al., 2011; Lafontaine et al., 2021). Developing biphasic condensate modulators by striking a balance between disrupting interactions that drive liquid to solid transitions and maintaining the ability to form liquid condensates will be important for maintaining important condensate functions in the cell while limiting the destructive effects of solid aggregates (Babinchak et al., 2020).

An added layer of complexity in this endeavor will be to consider how mutations in disease alter the chemical features of condensates and how these specific changes may be therapeutically targeted. Identification of these condensate changes may reveal vulnerabilities in diseased cells that healthy cells do not display and that can be targeted. Such targeting may arise from a traditional focus on the specific mutated proteins as drug targets, or if condensate dysregulation plays a role in pathogenesis, function may be restored by targeting condensate properties, as opposed to the specific mutated protein. The latter strategy may be appealing for proteins that are difficult to target, such as those that do not have binding pockets small molecules can easily be designed to bind and inhibit.

Developing screens for condensate therapies

In order to target condensates, it will be important to assay not only the partitioning behavior of therapeutics, but also their ability to change properties of condensates of interest such as the propensity to phase separate or undergo liquid to solid transitions. While work in Chapter 3 of this thesis established assays that can be used for measuring condensate partitioning, future efforts must include the development of high-throughput tools for screening molecules with diverse phenotypic effects on condensates.

In vitro tools could include assays to measure changes in droplet formation, measured by size and number of droplets, or liquid properties, measured, for example, by fluorescence recovery after photobleaching or by assaying droplet morphology to distinguish condensates from pathological aggregates. In-cell screening approaches may include fluorescently tagged disease proteins in disease-relevant cell lines and associated downstream functional assays for condensates of interest (Alberti et al., 2017).

A persistent challenge in these efforts will be to observe partitioning behavior and/or the effect of therapeutics on smaller condensates, as several condensates of interest such as transcriptional condensates, in which RNA targets of condensates can be found, approach the limit of resolution of several microscopes (Lyon et al., 2021; Alberti et al., 2017). A possible solution to this challenge could be to use engineered condensate systems such as the optodroplet or Lac array systems for creating artificial condensates more amenable to microscopy and to convert these systems into screenable in-cell platforms.

Condensate specificity and the therapeutic window

The effect of therapeutics on condensates could be important not only for on-target activity, but also for toxicity. For example, a drug of interest may target not only the protein or nucleic acid of interest, but may also cause toxicity by disrupting other condensates. Screening methods described above may enable the study of condensate-mediated drug toxicity as the effect of a drug on a condensate of interest can be assayed for dysfunction similarly to a mutated protein of interest. Condensates could also play a role in exacerbating off-target binding events if a drug of interest concentrates into a condensate in which an off-target binding partner is found, even if the drug binds at a lower Kd to this partner than the intended drug target. Studies of chemical and molecular grammar relating to drug partitioning and condensate-targeting could allow for prediction of potential condensate-associated toxicity, and could be an important companion to on-target activity screening. Thus, it will be important to consider the activity of a broad range of therapeutics as they relate to diverse condensates in the cell.

Condensates and drug resistance

Our studies of the action of tamoxifen in ER+ breast cancer in Chapter 3 of this thesis suggested that condensates may also play a role in drug resistance (Klein et al., 2020). We posited that, in the context of MED1 overexpression in tamoxifen-resistance, tamoxifen may be diluted within transcriptional condensates such that its efficacy is decreased. The degree to which this may represent a general mechanism for drug resistance is not yet known, and further studies of overexpression of transcriptional or other condensate components could answer this question. Condensates may also play a role in other resistance mechanisms the cell develops via mutation and expression changes, such as drug efflux or degradation mechanisms, presenting alternate condensate-focused therapeutic strategies for countering these resistance mechanisms. One example of the role of condensates in drug resistance has emerged in the resistance of bacteria and fungi developed via liquid to solid transitions in the dormant state (Fisher et al., 2017). This could represent a targetable phenomenon that could re-sensitize infected hosts to drugs targeting these pathogens (Alberti and Dormann, 2019).

Concluding thoughts

The study of condensates in disease and drug development upon which this thesis focuses, while at its early stages, will yield insights that are both fundamental to the biophysics and chemical biology of condensates, and translatable to the clinic. Both of these areas will require tool development, including more advanced microscopy and condensate assays in vitro and in vivo to allow for more sophisticated readouts for the emergent properties of condensates and how they can be disrupted in disease and therapeutically modulated. Further study of the mechanisms by which condensate dysregulation drives pathogenesis and the physicochemical determinants of drug partitioning and modulation of condensates can provide a foundation for new therapeutic strategies for diseases with limited treatment options, a new class of condensate-disrupting drugs, and general principles for drug development.

References

Alberti, S. & Dormann, D. Liquid-Liquid Phase Separation in Disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).

Alberti, S., Gladfelter, A., Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Physiol. Behav.* **176**, 139–148 (2017).

Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **22**, 196–213 (2021).

Babinchak, W. M. *et al.* Small molecules as potent biphasic modulators of protein liquid-liquid phase separation. *Nat. Commun.* **11**, (2020).

Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4334–4339 (2011).

Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).

Choi, K. J. *et al.* A Chemical Chaperone Decouples TDP-43 Disordered Domain Phase Separation from Fibrillation. *Biochemistry* **57**, 6822–6826 (2018).

Choi, J. M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).

Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **24**, 1537–1549 (2016).

Elbaum-Garfinkle, S. & Brangwynne, C. P. Liquids, Fibers, and Gels: The Many Phases of Neurodegeneration. *Dev. Cell* **35**, 531–532 (2015).

Fisher, RA., et al. Persistent bacterial infections and persister cells. *Nat. Rev. Microbiol.* **15**, 453–64 (2017).

Holehouse, A. S. & Pappu, R. V. Functional Implications of Intracellular Phase Transitions. *Biochemistry* **57**, 2415–2423 (2018).

Hyman, A. A. & Brangwynne, C. P. Beyond Stereospecificity: Liquids and Mesoscale Organization of Cytoplasm. *Dev. Cell* **21**, 14–16 (2011).

Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science* **368**, 1386–1392 (2020).

Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021).

Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).

Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88**, 678–690 (2015).

Patel, A. et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell **162**, 1066–1077 (2015).

Tsang, B., Pritišanac, I., Scherer, S. W., Moses, A. M. & Forman-Kay, J. D. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell* **183**, 1742–1756 (2020).

Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).