Enhancers and phase separation in the control of gene expression

By

John C. Manteiga

B.S., University of Massachusetts Amherst (2011)

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

February 2020

Signature of Author:_____

Department of Biology
October 21st, 2019

Certified by:_____

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

Accepted by:_____

Steve Bell
Professor of Biology
Co-Chair, Biology Graduate Committee

Enhancers and phase separation in the control of gene expression

By

John C. Manteiga

Submitted to the Department of Biology on November 15th, 2019 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the Massachusetts Institute of Technology

ABSTRACT

Gene regulation underlies the control of cell identity, development, and disease. Transcription of genes is regulated by DNA elements called enhancers, which are bound by transcription factors and coactivators, leading to the recruitment of RNA polymerase II and the production of RNA. Enhancers are thought to loop to specific gene promoters to stimulate transcription, but the mechanisms that cause enhancers to selectively loop to specific gene promoters is not well understood. In this thesis, I first describe new insights into enhancer-promoter loop specificity from studies examining the mechanisms that allow tumor-specific super-enhancers to loop to the *MYC* oncogene in diverse cancer types (Schuijers and Manteiga et al., 2018). While conducting these studies, it was proposed that super-enhancers and the factors associated with them form liquid-liquid phase-separated condensates. Following this proposal, I contributed to collaborative studies that strongly supported this model (Boija et al., 2018; Sabari et al., 2018, see Appendix I and II of this thesis). This model of transcription led me to ask how key transcriptional components could be recruited into super-enhancer condensates. I performed studies showing that the interaction of RNA polymerase II with these condensates involves the large heptapeptide repeat of the C-terminal domain (CTD) of the enzyme. Furthermore, these studies provided evidence that phosphorylation of the CTD, which is associated with the initiation to elongation transition, weakens these interactions, thus facilitating the transition of RNA polymerase II into different condensates involved in co-transcriptional splicing of the nascent transcript (Guo and Manteiga et al., 2019). These studies provide new insights into the mechanisms of enhancer-promoter interaction, roles for the RNA polymerase II CTD in the enzyme's partitioning into nuclear condensates, and a role for phosphorylation in switching the nuclear condensate partitioning behavior of RNA polymerase II.

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

ACKNOWLEDGEMENTS

Many people shaped my experiences while at MIT and deserve great thanks. Beginning on the scientific side, I am extremely grateful to all the members of the Young lab. I learned so much from being part of this group and you all made coming to work every day worth it. I'd like to give Jurian Schuijers a special thanks for taking me under his wing when I first joined the lab. Alicia, Jon, Ben, Abe, and many others were also great friends and scientists. Rick deserves a special thanks for taking me into the lab, and for creating an environment where incredible research can happen. I learned so much about how to plan, execute, and communicate science under Rick's guidance. I am also grateful to my thesis committee of Phil Sharp and Laurie Boyer, who always asked the right questions and helped keep me on the right course. There were also many people outside of MIT who supported me along the way. To my friends: Lucas, Nick, Jason, the North Andover gang, the UMass crew, the Boston Spikeball group, I want to say thanks for the fun times and sometimes greatly needed distraction from research. Of course, none of this would have been possible without my loving family: Mom, Dad, Carrie, and Matt. Finally, I want to thank my wonderful fiancé Aly. From meeting at UMass to moving to Boston for our PhDs, Aly has always been by my side tackling life's challenges with me. She makes every day better and I am so glad we get to get married this year.

STATEMENT ON WORK PRESENTED

Chapter 1
I wrote Chapter 1 with input and minor edits from Rick Young.

Chapter 2
I performed experiments and analyses in Figures 2 and 4 and contributed to the design of the others. Jurian Schuijers and Rick Young wrote the manuscript with input from me and the other authors.

Chapter 3
I either performed or was directly involved in the design of all experiments and analyses. Rick Young, myself, and Eric Guo wrote the manuscript with input from the other authors.

Chapter 4
I wrote Chapter 4 with input and minor edits from Rick Young.

Appendix I
I was heavily involved in the development of the analyses performed on the immunofluorescence data and provided input in the writing of the manuscript.

Appendix II
I performed a subset of the immunofluorescence experiments and provided input in the writing of the manuscript.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Overview

The regulation of gene expression underlies the control of cell identity, development, and disease. Gene expression is the process by which DNA is transcribed into an RNA intermediate which is then translated into a protein product. All cells in the human body encode the same information in their DNA, yet there are hundreds of different cell types which differ vastly in form and function. Each cells' identity is ultimately the result of different patterns of gene expression, and these patterns must be tightly controlled for an organism to develop properly. Indeed, dysregulated gene expression is associated with many diseases including cancer (Lee and Young, 2013). This thesis will focus on the regulation of transcription by enhancers, the discovery that enhancers form phase-separated condensates, and the mechanisms by which RNA polymerase II enters and exits condensates associated with transcription and RNA processing. In the overview section, I will briefly describe our current understanding of subjects related to these topics and how I came to study them before going into greater detail in the rest of the introduction.

Transcription is the process by which DNA is transcribed into RNA by RNA polymerase. In humans, there are ~20,000 genes, but only a portion of these are transcribed in any one cell type. Cells achieve this selective transcription by regulating the recruitment and activity of RNA polymerase at gene promoters. The proteins responsible for the recruitment of RNA polymerase to specific genes are called transcription factors (TFs). TFs function by binding to specific DNA sequences and interacting with another family of proteins called coactivators which ultimately recruit and activate RNA polymerase (Lambert et al., 2018). Each cell type expresses a unique combination of TFs that recognize distinct DNA sequences such that RNA polymerase is recruited to different genes in different cell types (Buganim et al., 2013). In addition to binding DNA sequences in gene promoters, TFs also bind sequences located distally from promoters in a cell-type specific manner. These distal sequences came to be called "enhancers" because they could enhance the transcription of a gene (Banerji et al., 1981). Early studies of enhancers revealed that they activate transcription by coming into close proximity to the promoters they regulate via the formation of a DNA loop between them (Levine et al., 2014).

Given that enhancers can exist distal to gene promoters, some key questions in the field of transcription have been "How do enhancers activate the right genes?" and "What factors are responsible for enhancer-promoter loop formation?" In the first part of my thesis work, I sought to address these questions by studying enhancer-promoter interactions in the *MYC* locus. *MYC* is an oncogene that is overexpressed in many cancer types, where its transcription is driven by different sets of particularly strong enhancers called super-enhancers. These super-enhancers are bound by an exceptionally high density of TFs and coactivators, and usually control the expression of genes that play prominent roles in cell identity (Hnisz et al., 2013; Whyte et al., 2013). As will be discussed in detail in Chapter 2, we discovered that *MYC* utilizes a single promoter-proximal site bound by the protein CTCF to interact with all the distinct super-enhancers found in different cancer types (Schuijers and Manteiga et al., 2018).

While completing the *MYC* enhancer-promoter interaction study, it was proposed that super-enhancers (like those driving *MYC* expression) and the factors associated with them form liquid-liquid phase-separated compartments (Hnisz et al., 2017). Liquid-liquid phase separation (LLPS) is the process by which molecules in a solution condense and form a separate, dense phase. An everyday example of LLPS can be observed in the behavior of oil and vinegar based salad dressings. When thoroughly mixed, the oil is fairly evenly distributed throughout the solution. Over

time however, the oil will demix (phase separate) from the surrounding solution and form droplets. The concentration of oil within these droplets is far greater than the concentration of oil in the surrounding solution. While the molecular details differ, a conceptually similar process occurs in cells with the end result being the formation of a droplet, often referred to as a condensate, containing high concentrations of proteins and other components necessary for a given biochemical reaction. While the oil particles phase separate due to hydrophobic interactions, condensate formation in cells is often driven by transient, multivalent interactions between the intrinsically disordered regions (IDRs) of proteins that make up the condensate. A condensate model for transcription was very attractive because it could explain the extreme cooperativity observed in super-enhancer formation and dissolution, elucidate the mechanisms by which TFs activate genes, and explain a number of other conundrums in the field of transcription.

Following the proposal that super-enhancers form condensates, I contributed to a lab-wide effort to provide experimental evidence in support of this model. These studies (Boija et al., 2018; Sabari et al., 2018, included as Appendix I and II of this thesis) as well as studies by our colleagues (Cho et al., 2018; Chong et al., 2018) were ultimately successful and strongly supported the idea that enhancers form condensates. These condensates form when TFs are locally concentrated at super-enhancers and recruit large numbers of coactivator molecules through transient, multivalent interactions between the unstructured activation domain of the TFs and unstructured IDR regions of the coactivators. One proposed function of these enhancer condensates is to concentrate the machinery necessary for transcription such that the reaction can occur quickly and robustly.

A key question that remained to be answered following the discovery that super-enhancers form condensates was “How do factors necessary for transcription partition into and become concentrated in the condensates?” In my final thesis project, I sought to answer this question for arguably the most important factor in transcription: RNA polymerase. Because IDRs with multivalent interaction capability seem to be important features of proteins found in condensates, we hypothesized that the RNA Polymerase II (RNAPII) C-terminal domain (CTD) allows RNAPII to enter condensates formed by enhancers. The CTD of RNAPII is a long unstructured IDR essential for transcription that contains repeats with the consensus peptide sequence YSPTSPS (Harlen and Churchman, 2017). These repeats serve as multivalent interaction platforms for various proteins through the course of transcription. During transcription initiation the CTD is mostly unphosphorylated. Escape from the promoter and transition into elongation is associated with phosphorylation of specific CTD residues. Because phosphorylation changes the physicochemical properties of proteins, we further hypothesized that CTD phosphorylation could cause eviction of RNAPII from condensates associated with transcription initiation thus enabling elongation to occur.

During transcript elongation, the nascent pre-mRNA of most genes is co-transcriptionally spliced (Herzel et al., 2017) and the phosphorylated CTD interacts with proteins involved in splicing (Ebmeier et al., 2017). These proteins form a scaffold that positions the core spliceosome on the nascent RNA to perform the splicing reaction (Jeong, 2017). It has been long known that many splicing proteins exist in punctate structures in the nucleus (Spector and Lamond, 2011) which have recently been proposed to be condensates (Banani et al., 2017; Kim et al., 2019). These observations led us to finally hypothesize that in addition to enabling eviction from condensates associated with transcription initiation, CTD phosphorylation could also allow RNAPII and the nascent transcript to partition into condensates associated with splicing to enable more robust co-transcriptional splicing. In Chapter 3, I will present detailed evidence supporting these hypotheses.

The control of gene expression is a complex process involving multiple reactions that may occur within nuclear condensates. Transcription is one of these reactions, and is controlled by the binding of TFs to promoters and enhancers. Enhancers interact with promoters through the formation of a DNA loop between them, and these interactions can sometimes be facilitated by the protein CTCF. During transcription initiation, RNAPII enters transcriptional condensates via interactions mediated through its intrinsically disordered CTD. Phosphorylation of the CTD may allow escape from these condensates and entry into elongation. During elongation the phosphorylated CTD may facilitate entry into different condensates associated with pre-mRNA splicing.

Transcriptional control of gene expression and enhancers

Generation of an RNA transcript is a key step in the expression of a gene. First studied in bacteria, this reaction is carried out by the enzyme RNA polymerase (RNAP). Early studies showed that transcription by RNAP could be controlled by the binding of repressors or activators to cis-acting DNA elements (Englesberg et al., 1965; Gilbert and Muller-Hill, 1966; Jacob and Monod, 1961; Ptashne, 1967). Repressors prevented the recruitment of RNAP to genes while activators enabled recruitment (Eron and Block, 1971; Ippen et al., 1968; Scaife and Beckwith, 1966). Later studies revealed similar principles in eukaryotic cells (Dynan and Tjian, 1983; Engelke et al., 1980; McKnight and Kingsbury, 1982; Payvar et al., 1981), cementing this general model in place. In eukaryotes, there are three different DNA directed RNA polymerases that transcribe different types of RNAs (Roeder and Rutter, 1969). RNAPII transcribes messenger RNA of protein coding genes, while RNAPI and RNAPIII transcribe ribosomal RNA and tRNA, respectively. In this section I will describe transcription by RNAPII and how it is regulated by cis-acting elements known as enhancers. There are multiple reviews focusing on transcription by the other polymerases (Geiduschek and Kassavetis, 2001; Goodfellow and Zomerdijk, 2013; Russell and Zomerdijk, 2006; Vannini and Cramer, 2012).

Classic views of transcription involve a series of protein-DNA and protein-protein binding events that ultimately result in the recruitment and activation of RNAPII at a gene. The location of a gene in a stretch of DNA is specified by a cis-regulatory element known as the core promoter, which spans approximately 30 base pairs upstream of the transcription start site and contains sequence elements including the TATA box and initiator elements. These sequences are bound by proteins called general transcription factors which serve to position RNAPII at the transcription start site (Kadonaga, 2012; Lenhard et al., 2012). Indeed, without general transcription factors, in vitro transcription by RNAPII occurs at non-specific locations on the DNA template (Matsui et al., 1980; Samuels et al., 1982; Sawadogo, 1990; Sayre et al., 1992; Weil et al., 1979; Young, 1991). The recruitment and activation of RNAPII at specific genes is controlled by proteins called transcription factors, which bind to promoter-proximal DNA elements as well as distally located DNA elements called enhancers (Buecker and Wysocka, 2012; Bulger and Groudine, 2011; Heinz et al., 2015; de Laat and Duboule, 2013; Levine, 2010; Maston et al., 2006; Ren and Yue, 2015). Transcription factors bind to these elements with their structured DNA binding domains, and interact with another family of proteins called coactivators often via interactions mediated by their unstructured trans-activation domains (Lambert et al., 2018; Ptashne, 1988). Coactivators ultimately recruit and activate RNAPII. A key coactivator is the Mediator complex, which is composed of ~30 subunits and acts as an interface between RNAPII and the activating transcription factors (reviewed in Allen and Taatjes, 2015; Malik and Roeder, 2010). Together, RNAPII, the general transcription factors, and the Mediator complex form the RNAPII holoenzyme, which can perform site specific transcription that is responsive to activators (reviewed in Greenblatt, 1997; Koleske

and Young, 1995; Lee and Young, 2000). Finally, eukaryotic DNA is packaged into nucleosomes, and some of the binding events described above require the clearing of nucleosomes to make the DNA accessible (Kornberg and Lorch, 1991; Workman and Buchman, 1993). The clearing of nucleosomes may be mediated by a special class of transcription factors called pioneer factors, which are capable of binding to nucleosome-bound DNA, and recruit nucleosome remodelers like the SWI/SNF complex (Zaret and Carroll, 2011).

Enhancers have emerged as key regulators of the spatiotemporal expression of genes. A defining feature of enhancers is that they can exist at a distance from the promoters of the genes they regulate. The first DNA element capable of activating gene expression from a distance was discovered in the polyomavirus simian virus 40 genome (Benoist and Chambon, 1981). When this element was transplanted to a reporter construct, it was also able to *enhance* the expression of a β-globin reporter gene from multiple kilobases away, and was thus termed an “enhancer” (Banerji et al., 1981). Following this discovery, many more instances of enhancers were found, and a few, including the β-globin locus control region (Bulger et al., 2002; Grosveld et al., 1993; Martin et al., 1996), the Drosophila eve locus (Small et al., 1992; Stanojevic et al., 1991), and the interferon-β enhanceosome (Maniatis et al., 1998) came to be model loci for the study of enhancers. Study of these model loci revealed enhancers tend to be a few hundred base pairs in size, and contain binding sites for various transcription factors that bind in a combinatorial fashion to ultimately recruit coactivators and RNAPII (Levine and Tjian, 2003).

An immediate question that arose following the discovery of enhancers was “How do enhancers activate transcription if they exist at a distance from the gene they regulate?” A number of models were proposed (Ptashne, 1986), but only a looping model was consistent with various observations that had been made about enhancers. In this now widely accepted model, enhancers activate transcription by coming into close proximity to the promoters they regulate via the formation of a DNA loop between them, thus allowing the enhancer to deliver the TFs and coactivators bound to it to the site of gene transcription. Early support for this looping model came from a series of studies in prokaryotic systems which I will summarize here. First, DNA loops mediated by the binding of proteins to the ends of a DNA construct could be observed directly via electron microscopy (Griffith et al., 1986). Second, enhancer activity was dependent on periodicity; regulation was disrupted when a non-integral number of helical turns was introduced between an enhancer and promoter. Over the short distances (<500 bp) tested in the experiment, it was argued that the energy required for a loop to form between proteins on the opposite face of a DNA helix was prohibitively high, and thus prevented the enhancer from interacting with the promoter (Hahn et al., 1984; Hochschild and Ptashne, 1986). Finally, when an enhancer and promoter were separated across two different DNA constructs, the enhancer construct could only activate gene expression if it was physically tethered to the promoter construct, suggesting that proximity of the two sequences, which could be achieved by looping, was required (Dunaway and Droge, 1989; Müller et al., 1989). Together, these studies led to the general acceptance of looping as the mechanism by which enhancers control promoters.

Given that enhancers loop to genes from a distance, it was unclear how enhancer-promoter interactions were controlled such that enhancers target the right genes. Our current understanding of this question invokes a model in which the genome is structured into a series of hierarchical domains that constrain enhancer-promoter interaction such that enhancers can only target promoters within their own domain (Fig. 1a) (Gibcus and Dekker, 2013). At the highest level, chromosomes appear to occupy distinct spaces in the nucleus that have been termed chromosome territories (Cremer and Cremer, 2010). Detection of more granular domains has been greatly facilitated by the development of chromosome conformation capture based techniques to map DNA interactions in cells (Dekker, 2002). Using these techniques, it was shown

that chromosomes are divided into a series of alternating active and inactive compartments approximately 3 megabases in size. The active compartments tend to physically interact more with other active compartments than inactive compartments and vice versa (Lieberman-aiden et al., 2009). These compartments are then further divided into ~1 mb regions of high internal interaction frequency called topologically associated domains (TADs) (Dixon et al., 2012; Nora et al., 2012). Within TADs there are loops of DNA termed insulated neighborhoods (Dowen et al., 2014; Hnisz et al., 2016a; Ji et al., 2016). Insulated neighborhoods are typically several hundred kilo bases in size, contain 2-3 genes on average, and are formed by the interaction of two DNA sites bound by the transcription factor CTCF and reinforced by the cohesin complex. Enhancers found within insulated neighborhoods typically only interact with genes located in the same neighborhood. This hierarchical organization appears to be functionally important in the control of enhancer-promoter interactions because when the boundaries of insulated neighborhoods or TADs are disrupted, inappropriate enhancer-promoter interactions and aberrant gene expression occurs (Hnisz et al., 2016b; Ji et al., 2016; Liu et al., 2016; Lupiáñez et al., 2015). Thus chromosomes occupy distinct territories within the nucleus, and are subdivided into active and inactive compartments which are further parsed into TADs and insulated neighborhoods which constrain enhancer-promoter interactions.

Figure 1

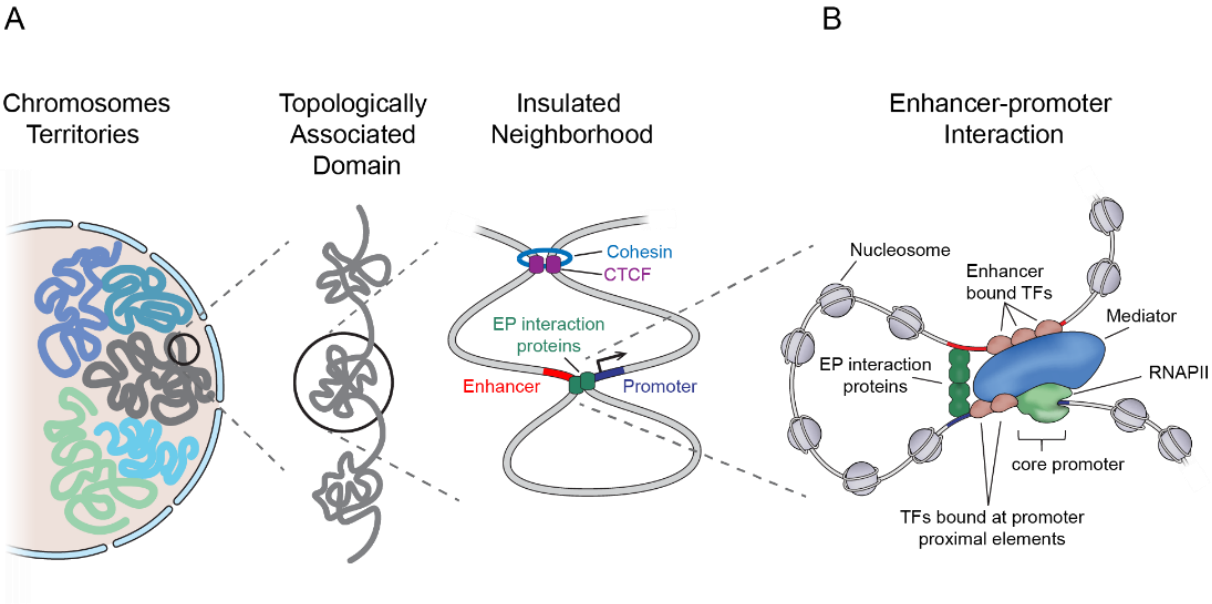


Figure 1. Enhancers regulate transcription through DNA looping constrained by genome structure.
(A) The hierarchy of the 3D genome. Individual chromosomes tend to interact with themselves more than other chromosomes, giving rise to chromosome territories. Chromosomes fold into topologically associating domains which contain insulated neighborhoods. Insulated neighborhood loops are anchored by CTCF and cohesin proteins, and contain enhancer-promoter (EP) loops mediated by proteins including YY1 and CTCF.
(B) Classic model of transcriptional control and enhancer-promoter looping. Transcription factors bind to enhancers and promoter-proximal regions and recruit coactivators including the Mediator complex. Mediator integrates the signals of these activating transcription factors by making contacts with RNAPII bound to the core promoter. RNAPII is positioned at the core promoter by general transcription factors (not shown). Distal enhancers bound by TFs and coactivators are brought into close proximity to the promoter via formation of a loop mediated in part by enhancer-promoter interaction proteins including YY1 and CTCF. These binding events occur on nucleosome-free DNA.

With the genome structure based constraint model in place, a new goal in the enhancer field has been to identify the factors responsible for bringing enhancers and promoters together within the same domain. This question has been partially addressed by combining information from genome wide DNA interaction data with genome wide protein occupancy (ChIP-seq) data. Proteins that occupy both enhancers and promoters and sit at the anchors of DNA loops are candidate mediators of enhancer-promoter interactions (Fig. 1a, b). Some cell-type specific proteins have been implicated in mediating enhancer-promoter interactions (Deng et al., 2012; Lee et al., 2017), but so far there is only one protein, YY1, that has been reported to mediate enhancer-promoter looping generally in all cell types (Weintraub et al., 2018). In Chapter 2 of this thesis, I will describe the mechanisms of enhancer-promoter interaction in the *MYC* locus in cancer (Schuijers and Manteiga et al., 2018). Interestingly, we found that enhancers interact with the *MYC* promoter through CTCF based interactions across many cancer types, despite CTCF typically being thought of as a mediator of Insulated Neighborhood loop formation.

Thus transcription is a multistep process that can be controlled on many levels. RNA polymerase II binds to promoters with the help of general transcription factors. Transcription factors bind to the promoter-proximal region and enhancers and recruit coactivators such as the Mediator complex that enhance RNAPII recruitment and activity. Enhancers play a major role in the control of transcription and interact with promoters via the formation of loops mediated by proteins including YY1 and CTCF (Fig. 1b). When all of this machinery comes together, transcription can occur.

Condensates compartmentalize cellular biochemistry

It was recently proposed that enhancers and the proteins that associate with them form liquid-liquid phase-separated compartments (Hnisz et al., 2017). This proposal was made in the midst of a revelation that many biochemical reactions enhance reaction rates by concentrating the necessary reaction components in “membraneless organelles” through the process of liquid-liquid phase separation (Banani et al., 2017). In this section I will describe the principles of biological phase separation in general before discussing phase separation in transcription in the next section.

A key problem in biology is how the densely packed cellular space is organized such that complex biochemical reactions occur at the right place and at the right time. One solution to this problem is to regulate the localization of reaction components; bringing reactants together at high concentrations increases reaction rates, while keeping them separate can inhibit reactions. Organelles achieve this function through compartmentalizing factors within generally impermeable membranes. However, there are many processes that appear to occur in compartments containing high concentrations of factors in the absence of a membrane (Banani et al., 2017). Recent discoveries have revealed that many of these membraneless compartments form through the process of liquid-liquid phase separation. The resulting phase-separated droplets, termed “biomolecular condensates”, can concentrate reaction components while retaining desirable features for biochemical reactions such as rapid diffusion within the condensate and dynamic exchange of factors with the surrounding environment. There are many excellent reviews describing condensate biology and the many processes it contributes to (Banani et al., 2017; Hyman and Simons, 2012; Hyman et al., 2014; Shin and Brangwynne, 2017), as well as the methods used to characterize new systems where condensation is suspected to be involved (Alberti et al., 2019). Here, I will discuss key features of condensates, including how they form and how they can be regulated. I will also describe contemporary examples of processes

governed by phase-separated condensates relating to development and the regulation of gene expression.

Despite the diversity of phase-separated compartments in the cell, the principles that govern their formation and dissolution are shared. There are two key features that determine if biomolecules will form a condensate or remain in the diffuse phase: valency and concentration. Valency refers to the ability of a biomolecule to interact with other biomolecules. A factor with high valency is one that can make multiple points of contact with multiple other factors, such as a protein or nucleic acid with many independent binding sites for partner factors. When interacting multivalent molecules reach some threshold concentration, it becomes thermodynamically favorable for them to form a network of interactions and condense (Cohen and Benedek, 1982; Flory, 1942, 1953; Huggins, 1941; Stockmayer, 1952), with molecules of higher valencies requiring lower concentrations to condense and vice-versa (Fig. 2a). Condensation controlled by these two features occurs in a switch-like fashion; small changes in valency or concentration determine the presence or absence of condensates, providing a means to regulate these structures (Banani et al., 2017). One popular model of phase separation terms the multivalent factors that form the condensate as “scaffolds”, which can then selectively partition and concentrate other factors termed “clients” into the condensate based on their chemical properties (Banani et al., 2016). Concrete examples demonstrating these principles can be observed in condensates formed by proteins involved in actin polymerization and signaling (Banjade and Rosen, 2014; Li et al., 2012; Su et al., 2016). In these systems, proteins with multiple SRC homology 3 (SH3) domains interact with partners that have multiple proline-rich motifs (PRMs). These multivalent interactions between scaffold proteins lead to condensate formation which occurs at lower threshold concentrations as the valency of interaction is increased by increasing the number of SH3 domains or PRMs. The biochemical reactions of signal propagation and actin polymerization are both enhanced in conditions where condensate formation occurs, likely due to the concentration of client proteins necessary for the reactions or the exclusion of inhibitory factors. In these examples, interactions are mediated by the structured SH3 protein domain. However, interactions mediated by structured domains are not required for phase separation, as many phase-separated systems rely on interactions between unstructured regions of proteins called intrinsically disordered regions (IDRs).

IDRs are regions of proteins that lack a defined 3D structure. Much of the human proteome is intrinsically disordered, and proteins known to participate in phase separation are even further enriched for IDRs (Banani et al., 2017; Xie et al., 2007). IDRs often have lower than average sequence complexity and are enriched in a small number of amino acids. These amino acids are often distributed in patches throughout the IDR, and this patchy distribution is important for the ability to phase separate (Jiang et al., 2015; Lin et al., 2015; Molliex et al., 2015; Nott et al., 2015). Multivalent intermolecular interactions between these patches, sometimes referred to as “stickers”, which are separated by flexible spacer regions, leads to phase separation in a conceptually similar manner as discussed above (Fig. 2b) (Banjade et al., 2015; Harmon et al., 2017; Holehouse and Pappu, 2018; Wang et al., 2018). The key difference is that the IDR based interactions are shorter lived and not structurally ordered, consistent with the dynamic nature of condensates. As will be discussed in the next section of the introduction, condensate formation by transcriptional proteins is primarily mediated through IDR based interactions.

Figure 2

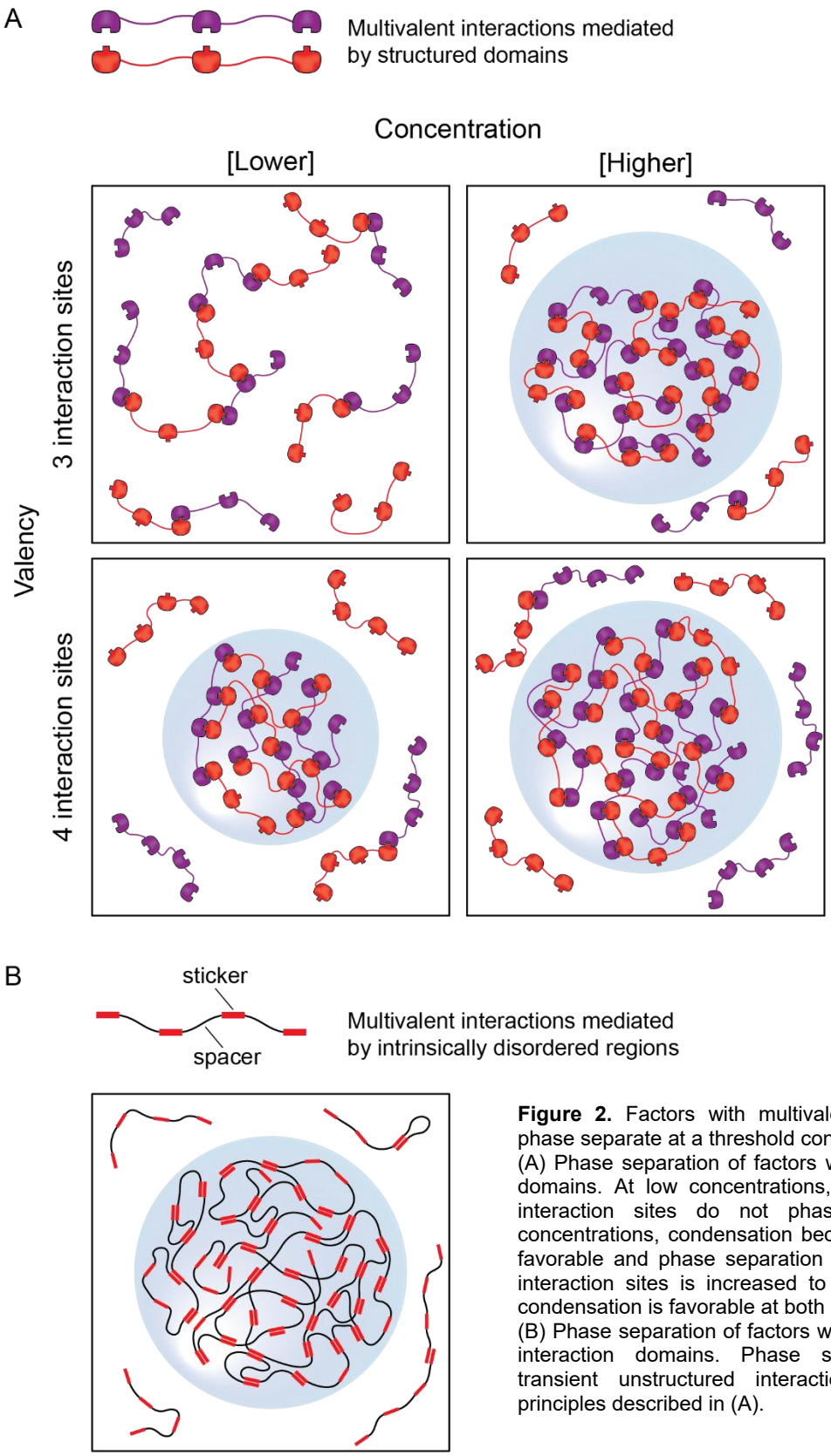


Figure 2. Factors with multivalent interaction capability phase separate at a threshold concentration.
(A) Phase separation of factors with structured interaction domains. At low concentrations, factors with 3 potential interaction sites do not phase separate. At higher concentrations, condensation becomes thermodynamically favorable and phase separation occurs. If the number of interaction sites is increased to 4 (increase in valency), condensation is favorable at both concentrations.
(B) Phase separation of factors with intrinsically disordered interaction domains. Phase separation mediated by transient unstructured interactions follows the same principles described in (A).

A key feature of any biological process is the ability to regulate it. The dependence of condensation on valency and concentration provide simple means to exert regulation; condensate formation can be controlled by changing the number of interaction sites (altering valency), or by changing the concentration of component molecules. A common way to change the valency of a protein is through post-translational modification. There are many examples where PTMs like phosphorylation or methylation serve to create or destroy sites of interaction to control condensate formation and dissolution (Banjade and Rosen, 2014; Boehning et al., 2018; Li et al., 2012; Mikhaleva and Lemke, 2018; Milovanovic et al., 2018; Nott et al., 2015; Rai et al., 2018; Su et al., 2016; Wippich et al., 2013). Factor concentration can be modified by changing expression levels, degradation rates, or localization. Many phase-separated systems control local concentrations without modulating total cellular concentrations by tethering large numbers of component molecules together in close proximity through binding to RNA or DNA (Berry et al., 2015; Chung et al., 2011; Kaiser et al., 2008; Mao et al., 2011; Ries et al., 2019; Shevtsov and Dundr, 2011). As will be discussed in the next section, condensates formed by transcriptional proteins are regulated in part through this mechanism. Interestingly, nucleic acids have also been shown to inhibit rather than promote condensate formation (Maharana et al., 2018), showing the importance of specific nucleic acid sequence and context. Finally, valency and effective concentration can theoretically each be modulated by small molecules to disrupt or promote condensate formation. ATP is known to act as a hydrotrope to inhibit phase separation (Patel et al., 2017), and it is possible small molecule drugs used in the clinic preferentially localize to or dissolve condensates. The interplay between small molecules and condensates will be examined further in Chapter 4 of this thesis.

Condensate biology has been shown to play important roles in development and the regulation of gene expression. A seminal study which cemented phase separation as a fundamental process in biology focused on the behavior of P granules during *Caenorhabditis elegans* development (Brangwynne et al., 2009). The *C. elegans* one-cell embryo contains bodies composed of protein and RNA known as P granules. Before the first cell division, these P granules segregate towards the posterior side of the embryo. Cell division then gives rise to a P granule-containing progenitor germ cell and a non-P granule-containing somatic sister cell. This study showed that P granules are liquid-liquid phase-separated compartments, and that their segregation occurs through selective condensation only on the posterior side of the cell. The P granules exhibited all of the properties expected of a phase-separated droplet, including circular shape, fusion and fission upon exposure to shear force, whetting on the nuclear membrane surface, fast recovery of components following photo-bleaching, and measured viscosities and surface tensions typical of liquid droplets. A later study (Saha et al., 2016) revealed that phase separation is encouraged on the posterior end of the cell by mRNA that binds to key P granule components and reduces the concentration required for condensation, likely by increasing the valency of interaction by linking multiple P granule components together. On the anterior side of the cell, this mRNA is sequestered by a polarity protein. These findings primed the field to discover that phase separation is a ubiquitous process with roles throughout many aspects of biology.

A number of processes involved in the regulation of gene expression have been shown to function in part through condensates. The repression of gene expression and chromatin compaction is one of these processes. Both HP1α (Larson et al., 2017; Strom et al., 2017) and the PRC1 component protein CBX2 (Plys et al., 2019) have been shown to phase separate in vitro and in vivo. In both cases these condensates are able to compartmentalize the expected ligands including nucleosomes. Phase separation of both HP1α and CBX2 is enhanced by phosphorylation on specific residues, reiterating the importance of post translational modification in the regulation of condensate formation. In the case of CBX2, the same regions of the protein that are required for nucleosome compaction are also required for phase separation, suggesting

that the two processes are functionally linked. These findings that both constitutive and facultative heterochromatin components exist as dynamic condensates are especially intriguing given the classic view of heterochromatin as a static, highly compact structure. Condensates also play roles in the positive regulation of gene expression. The nucleolus is a bonafide condensate (Brangwynne et al., 2011) that compartmentalizes high levels of rRNA transcription by RNAPI and ribosome biogenesis. The establishment of DNA interactions both between enhancers and promoters (Shin et al., 2018) and insulators (Hansen et al., 2018) may be in part controlled by condensation. Finally, as will be discussed in the next section of the introduction and in Chapter 3, the transcription of protein coding genes likely begins in condensates formed by enhancers which can recruit RNA polymerase II.

Thus condensates compartmentalize a broad range of biochemical reactions. Their formation occurs when proteins or nucleic acids with multivalent interaction capability reach a threshold concentration. These multivalent interactions can be mediated by structured or unstructured domains. Condensate formation and properties can be regulated in a number of ways, with the post-translational modification of component molecules being a common axis of control. Many important processes involved in development and the regulation of gene expression function in part through condensates.

Super-enhancers form condensates

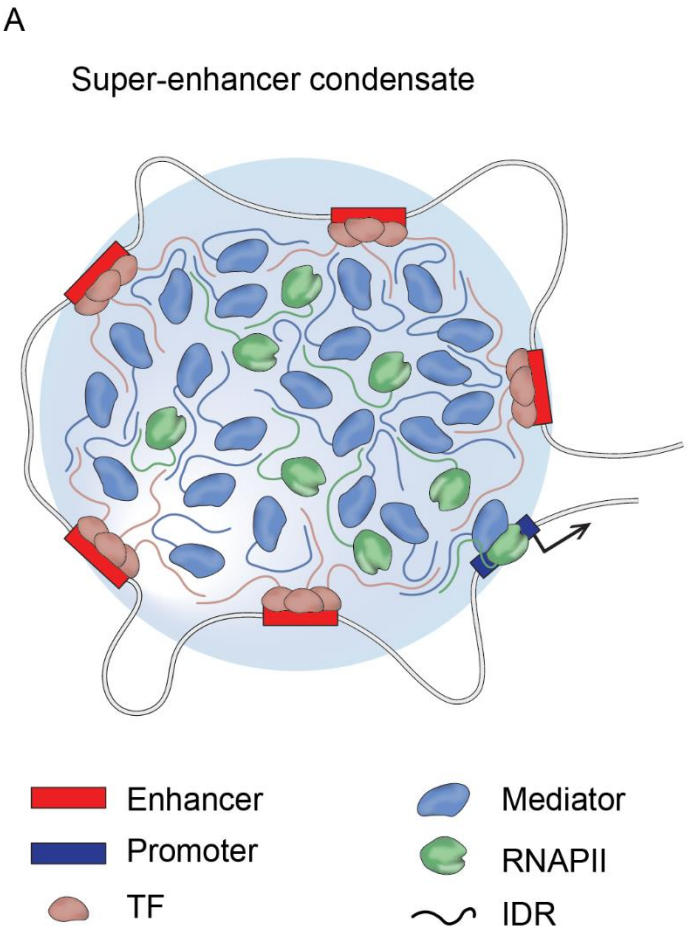
The molecular interactions that govern transcription have been the subject of intense study for the past several decades, yet there are a number of conundrums in the field that have yet to be fully explained. The recent realization that transcription may occur in condensates formed by enhancers and their associated factors has provided unifying answers to many of these questions. In this section I will outline some of these conundrums, describe how a condensate model of transcription can explain them, and detail studies leading up to and directly showing that some enhancers form phase-separated condensates.

One unanswered question relates to the peculiar behavior of some enhancers in disease. In certain cancers, relatively minor mutations in DNA lead to the formation of super-enhancers that then drive the transcription of oncogenes (Abraham et al., 2017; Mansour et al., 2014). Super-enhancers are large groups of closely spaced enhancers bound by a high density of factors, and usually control the expression of the genes most important in defining cell identity (Hnisz et al., 2013; Whyte et al., 2013). This phenomenon was mysterious because small mutations that create a binding site for just one additional transcription factor would not be expected by themselves to cause the formation of a massive super-enhancer containing an exceptionally high density of factors. Phase separation can explain this phenomena because the formation of condensates is switch-like. A small increase in the concentration of factors on the DNA from just one additional factor binding could feasibly bring the system above the concentration at which phase separation would occur, leading to a large increase in the concentration of factors in the region. Similarly, there are a number of examples where therapeutics have preferential effects on super-enhancers as compared to typical enhancers (Chapuy et al., 2013; Chipumuro et al., 2014; Kwiatkowski et al., 2014; Lovén et al., 2013; Wang et al., 2015a) . While the drugs targets may be present across all enhancers, reducing the targets effective concentration or valency may have greater effects on super-enhancers that rely on condensation to impart some of their activity as compared to typical enhancers where this may not be the case.

Phase separation can also explain a number of features of transcription factors relating to their structure and function. TFs typically consist of a DNA binding domain (DBD) and a separate activation domain (Brent and Ptashne, 1985; Keegan et al., 1984). While the mechanisms that underlie TF DBD function are well documented, comparatively little is known about activation domains. The DBDs of many TFs have been characterized at atomic resolution, and TFs are typically classified by the features of their DBDs (Fulton et al., 2009; Vaquerizas et al., 2009). DBDs can be composed of basic helix-loop-helix, helix-turn-helix, zinc-coordinating, or basic-leucine zipper DNA-binding structures. These DBDs bind specific DNA sequence motifs, and the motifs preferred by hundreds of different transcription factors have been described (Hume et al., 2015; Jolma et al., 2013; Khan et al., 2018). With the DBDs binding to specific genomic locations, TF activation domains serve to recruit co-activators and activate gene expression (Allen and Taatjes, 2015; Juven-Gershon and Kadonaga, 2010; Malik and Roeder, 2010; Plaschka et al., 2016; Reiter et al., 2017; Soutourina, 2017). In contrast to DBDs, most activation domains are intrinsically disordered and therefore cannot be crystallized due to lack of structure. Activation domains are therefore categorized based on their amino acid compositions or theoretical shapes. Many activation domains are enriched in acidic, proline, serine/threonine, or glutamine residues, and may adopt shapes described as acid blobs, peptide lassos, or negative noodles (Mitchell and Tjlan, 1989; Paul B. Sigler, 1988; Roberts, 2000; Staby et al., 2017; Triezenberg, 1995). Despite the differences in amino acid composition and shape, many activation domains interact with the same sets of coactivators (Allen and Taatjes, 2015; Avantaggiati et al., 1996; Eckner et al., 1996; Green, 2005; Hansen et al., 2007; Yin and Wang, 2014; Yuan et al., 1996) and activation domains are often functionally interchangeable between TFs (Godowski et al., 1988; Hope and Struhl, 1986; Jin et al., 2016; Lech et al., 1988; Ransone et al., 1990; Struhl, 1988; Tora et al., 1989). These observations are not compatible with traditional lock-and-key models of protein-protein interaction. Consistent with these concepts, the activation domain of the TF GCN4 has recently been described to interact with the Mediator complex at multiple sites and in multiple orientations (Brzovic et al., 2011; Jedidi et al., 2010; Tuttle et al., 2018; Warfield et al., 2014). As described in the previous section, multivalent interactions between unstructured proteins can lead to the formation of phase-separated condensates. A model in which unstructured TF activation domains form condensates with various coactivators through many possible different interactions with IDRs can explain all of these observations.

Further, dynamics of enhancer and RNAPII interaction with genes suggests a condensate model for transcription. A single enhancer was shown to be able to activate expression of two different distally located genes simultaneously (Fukaya et al., 2016; Heist et al., 2019), and large clusters containing up to 80 molecules of RNAPII could be detected at active genes (Cho et al., 2016; Cisse et al., 2013). Neither of these observations are compatible with the classic model of enhancer-promoter interaction and gene activation as depicted in Fig. 1b. However, these observations are explained by a model in which enhancers form condensates containing large numbers of RNAPII that can be shared between multiple genes simultaneously. I will now detail a series of studies providing experimental evidence for the condensate model of transcriptional control (Fig. 3a).

Figure 3



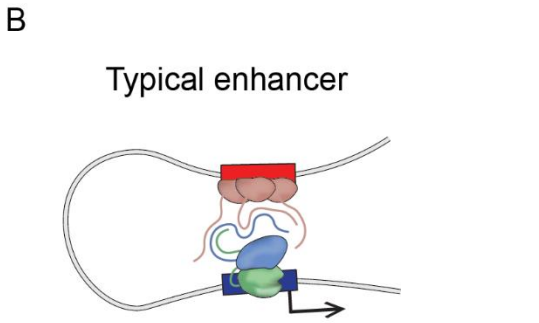


Figure 3. A condensate model for transcriptional control.
(A) Super-enhancers can form condensates. The concentration of factors recruited to super-enhancers exceeds that required for phase separation. Multivalent interactions occur between the disordered activation domains of TFs and the IDRs of coactivators including Mediator. The resulting condensate can concentrate clients including RNAPII, partially through interactions mediated by the intrinsically disordered RNAPII C-terminal domain. Multiple genes can access this large pool of transcriptional components simultaneously to rapidly recruit and subsequently fire RNAPII from promoters.
(B) Typical enhancers may not form condensates. The concentration of factors recruited to typical enhancers may not reach the threshold required for phase separation.

Key components of super-enhancers were first predicted (Hnisz et al., 2017) and later shown (Sabari et al., 2018, see Appendix I) to form condensates. Mediator and BRD4 are coactivators with essential roles in transcription initiation and elongation, and constitute some of the most highly enriched factors found at super-enhancers. The first indication that these factors are compartmentalized in condensates came from microscopy experiments examining their distribution and physical properties in cells. MED1 (a subunit of the Mediator complex) and BRD4 were found to occur in circular puncta throughout the cell nucleus which frequently colocalized with genes regulated by super-enhancers. Puncta indicate a high density of factors, as would be expected in a condensate. These puncta exhibited physical properties that would be expected of phase-separated condensates, including fast recovery following photobleaching, sensitivity to small molecules known to perturb condensates, and the ability to fuse. When purified, the IDRs of these factors were able to form phase-separated droplets in vitro. Further, MED1 droplets formed in nuclear extract were able to concentrate and compartmentalize transcriptional components including RNAPII, implicating concentration of factors as a function of transcriptional condensates. A co-published study corroborated many of these results with highly quantitative data, and showed additionally that Mediator condensates transiently interact with active genes (Cho et al., 2018), possibly contributing to transcriptional bursting. Collectively, these studies showed that coactivators exist in condensates in cells, but the determinants of condensate formation remained to be revealed.

A simple model to explain selective formation of transcriptional condensates is one where the binding of TFs to enhancers and subsequent recruitment of coactivators increases the local concentration of those factors past the point at which condensation occurs. Both TFs and coactivators could contribute to condensate formation through multivalent interactions between their unstructured regions. Under this model, condensates would only form at select genomic loci with sufficient TF binding and coactivator recruitment (Fig. 3a, b), allowing transcriptional condensates to form in different locations depending on the cell type or signaling state. Multiple recent studies have provided strong evidence supporting this model (Boija et al., 2018; Shrinivas et al., 2019). The coactivator puncta localized to super-enhancer-associated genes described above were found to also colocalize with puncta of transcription factors. Acute degradation of these TFs led to a loss of the coactivator puncta (Boija et al., 2018, see Appendix II), as would be expected if condensation depended on local increases in concentration enabled by TF binding. In vitro, the intrinsically disordered activation domains of various TFs were able to form droplets together with Mediator, and mutated TFs that failed to form droplets with Mediator in vitro also failed to strongly activate transcription in vivo (Boija et al., 2018). A follow up study showed that the number of TFs bound to a model enhancer DNA construct controlled the concentration at which coactivators would phase separate (Shrinivas et al., 2019). In computational models and in vitro experiments, higher numbers of TF binding sites allowed the Mediator coactivator to form droplets at lower concentrations. In cells, increasing the number of TF binding sites past a critical point lead to a switch-like increase in transcriptional output as detected by a luciferase reporter assay. These results suggest that just one additional TF binding site is capable of inducing condensation and greatly increasing transcriptional output, essentially modeling what may occur in cancers where minor mutations lead to super-enhancer formation (Abraham et al., 2017; Mansour et al., 2014). Together, these studies show that transcriptional condensate formation is controlled through modulating the local concentrations of component factors at specific locations on the genome. Transcription factors achieve this control by binding to specific enhancers, recruiting coactivators, and contributing to condensate formation via multivalent interactions mediated by their intrinsically disordered activation domains.

Following these discoveries, it remained to be determined how factors necessary for transcription partition into and are concentrated in super-enhancer condensates. In Chapter 3 of this thesis, this question will be discussed in detail for RNA polymerase II (Guo and Manteiga et al., 2019). Briefly, we found that the RNAPII C-terminal domain allows RNAPII to enter condensates formed by super-enhancers. The CTD of RNAPII is a long unstructured IDR essential for transcription that contains repeats with the consensus peptide sequence YSPTSPS (reviewed in Eick and Geyer, 2013; Harlen and Churchman, 2017; Hsin and Manley, 2012; Jeronimo et al., 2013). These repeats serve as multivalent interaction platforms for various proteins through the course of transcription. During transcription initiation, the CTD contacts a number of coactivators including the Mediator complex (Robinson et al., 2016), and upon phosphorylation by CDK7 (a part of the general transcription factor TFIIH) these contacts are disrupted, allowing RNAPII to escape from the promoter (Wong et al., 2014). We found that these phosphorylation events also cause eviction of RNAPII from condensates associated with transcription initiation thus enabling elongation to occur.

Co-transcriptional splicing may occur in splicing factor condensates

During transcript elongation, the nascent pre-mRNA of most genes is co-transcriptionally spliced (Herzel et al., 2017) and the phosphorylated CTD of RNAPII interacts with proteins involved in splicing (Ebmeier et al., 2017). In Chapter 3 of this thesis I will present evidence that CTD phosphorylation allows RNAPII and the nascent transcript to enter condensates associated with splicing to enable more robust co-transcriptional splicing. In this section, I will introduce the process of splicing by describing its discovery and the key steps and factors involved, discuss the interplay between splicing and transcription, and finally discuss the evidence suggesting the splicing reaction also occurs in condensates.

When a transcript is spliced, non-coding sequences called introns are removed from within the transcript while coding or regulatory sequences called exons are ligated together. In some cases, certain exons are also removed to generate an alternate transcript. This ability to include different sets of exons in a gene through alternative splicing greatly increases the diversity of possible gene expression patterns and provides yet another point at which gene expression can be controlled (Baralle and Giudice, 2017; Lee and Rio, 2015; Wang et al., 2015b). The realization that large segments of the transcript are removed before the mature transcript is generated came from a series of experiments examining the hybridization of mature transcripts from Adenovirus with the Adenovirus DNA genome. Direct imaging of the hybridization with electron microscopy revealed that the mature RNA transcript was complementary to multiple discrete segments of the DNA genome, with regions of non-complementary DNA interspersed between each segment (Berget et al., 1977; Chow et al., 1977). The authors correctly postulated that the complementary and intervening regions were transcribed as a single unit, and that the intervening segments were then removed. These studies paved the way for the discovery and characterization of the factors that carry out this removal.

Splicing involves hundreds of factors, and there are many excellent reviews describing the process in great detail (Braunschweig et al., 2013; Burge et al., 1999; Herzel et al., 2017; Sperling, 2017; Wang and Burge, 2008; Will and Luhrmann, 2011). To summarize, the removal of an intron from between two flanking exons involves two consecutive transesterification reactions (Domdey et al., 1984; Moore et al., 1993; Padgett et al., 1984; Ruskin et al., 1984). Key sequence elements on the pre-mRNA transcript include the 5’ and 3’ splice sites (which demarcate the junctions between the intron and the upstream and downstream exons, respectively), and the branch point (BP), which occurs 18-40 nucleotides upstream of the 3’ splice site (Fig. 4a) (Burge et al., 1999; Linz et al., 1985; Newman et al., 1985). In the first step of the reaction, the 2’OH group of the branch point Adenosine carries out a nucleophilic attack on the 5′ splice site, resulting in cleavage at the 5’ splice site and ligation of the 5′ end of the intron to the branch point adenosine. The resulting looped intron is known as the lariat (Domdey et al., 1984; Padgett et al., 1984). In the second step, the free 3’OH of the upstream exon attacks the 3’ splice site, resulting in ligation of the upstream and downstream exons and release of the intron lariat (Fig. 4a). This process is carried out by a multi-megadalton ribonucleoprotein complex called the spliceosome (Brody and Abelson, 1985). Core components of the spliceosome include the U1, U2, U5, and U4/U6 small nuclear ribonucleoproteins (snRNPs), each of which is composed of a small nuclear RNA and a number of other proteins (Bringmann et al., 1983; Lerner et al., 1980; Riedel et al., 1987; Steitz et al., 1983; Weinberg and Penman, 1968; Will and Luhrmann, 2011; Wise et al., 1983).

Akin to how transcription factors and coactivators control RNAPII, there are a host of regulatory factors that control the positioning and activity of the catalytic spliceosome components. Splicing is enhanced when certain RNA binding proteins (often of the SR protein family) bind to exonic or intronic splicing enhancer sequences to form a scaffold on which the spliceosome can assemble

(Blencowe, 2000; Cáceres and Hurst, 2013; Long and Caceres, 2009; Wang et al., 2012). Conversely, splicing is suppressed when exonic or intronic splicing silencers are bound by proteins that prevent spliceosome assembly (Carstens et al., 2015; Cartegni et al., 2002). Over the course of the reaction, the spliceosome and auxiliary RNA binding proteins adopt multiple conformations, with ATPases aiding in the transitions between these conformations (Burgess et al., 1990; Christian et al., 2013; Pena et al., 2016; Schwer and Guthrie, 1991). Further adding to the complexity, the splicing of most genes occurs co-transcriptionally (Fig. 4b) (Beyer et al., 1981; Osheim et al., 1985; Perales and Bentley, 2009; Wu et al., 1991; Wuarin and Schibler, 1994), and multiple aspects of transcription and splicing feedback on each other (Almada et al., 2013; Bieberstein et al., 2012; Braunschweig et al., 2013; Manley, 2002; Moore and Proudfoot, 2009; Naftelberg et al., 2015; Oesterreich et al., 2011; Saldi et al., 2016). During transcript elongation, the CTD of RNAPII is phosphorylated. Many proteins involved in splicing interact with the phosphorylated CTD and the CTD may serve to recruit splicing factors into close proximity to the nascent RNA transcript (Fig. 4b) (Ebmeier et al., 2017; Harlen et al., 2016). Interestingly, progressive truncation of the RNAPII CTD causes splicing defects before transcription is lost (McCracken et al., 1997).

There are a number of reasons to believe the splicing reaction may occur in nuclear condensates. Given the complexity of splicing, one might expect it to have slow reaction kinetics, but recent studies indicate that ligation of exons occurs almost immediately following synthesis by RNAP II (Alpert et al., 2017; Carrillo Oesterreich et al., 2016). This high efficiency may be enabled in part by the concentration of factors necessary for splicing in condensates. Indeed, the splicing machinery as a class is highly enriched in intrinsic disorder, and many RNA binding proteins have been previously shown to phase separate (Banani et al., 2017; Herzel et al., 2017). Further, much of the splicing machinery, including the snRNPs and many RNA binding proteins, can be observed in large membraneless nuclear compartments called nuclear speckles (Beck, 1961; Lerner et al., 1981; Spector and Lamond, 2011; Spector et al., 1983). These compartments are classically defined by one of their components, an SR protein called SRSF2. Recent work suggests that speckles are phase-separated condensates; they exhibit dynamic properties of internal-external rearrangement by FRAP and frequently fuse in response to certain types of cellular stress (Kim et al., 2019; Kruhlak et al., 2000; Misteli and Phair, 2000).

It has been unclear as to whether these speckle condensates contribute to active splicing. Many in the field believe speckles are primarily storage sites for the splicing machinery (reviewed in Lamond and Spector, 2003). However, there is also a large body of work showing that many highly transcribed genes in the cell localize very close to speckles, probably to ensure robust splicing (Chen and Belmont, 2019; Chen et al., 2018; Hu et al., 2009; Khanna et al., 2014; Moen et al., 2004; Quinodoz et al., 2018; Shopland et al., 2003; Xing, 1995). Part of this discrepancy may come from the fairly loose definition of a speckle as any punctate structure in the nucleus that stains positive for SRSF2. It is possible that some SRSF2 puncta represent storage speckles, while others represent sites of highly active transcription and splicing (Huang and Spector, 1996). In Chapter 3, I will provide evidence that condensates composed of splicing factors do associate with active genes, and that the ability of the transcribing RNAPII, and thereby the nascent transcript, to interact with these condensates depends in part on the RNAPII CTD phosphorylation state.

Figure 4

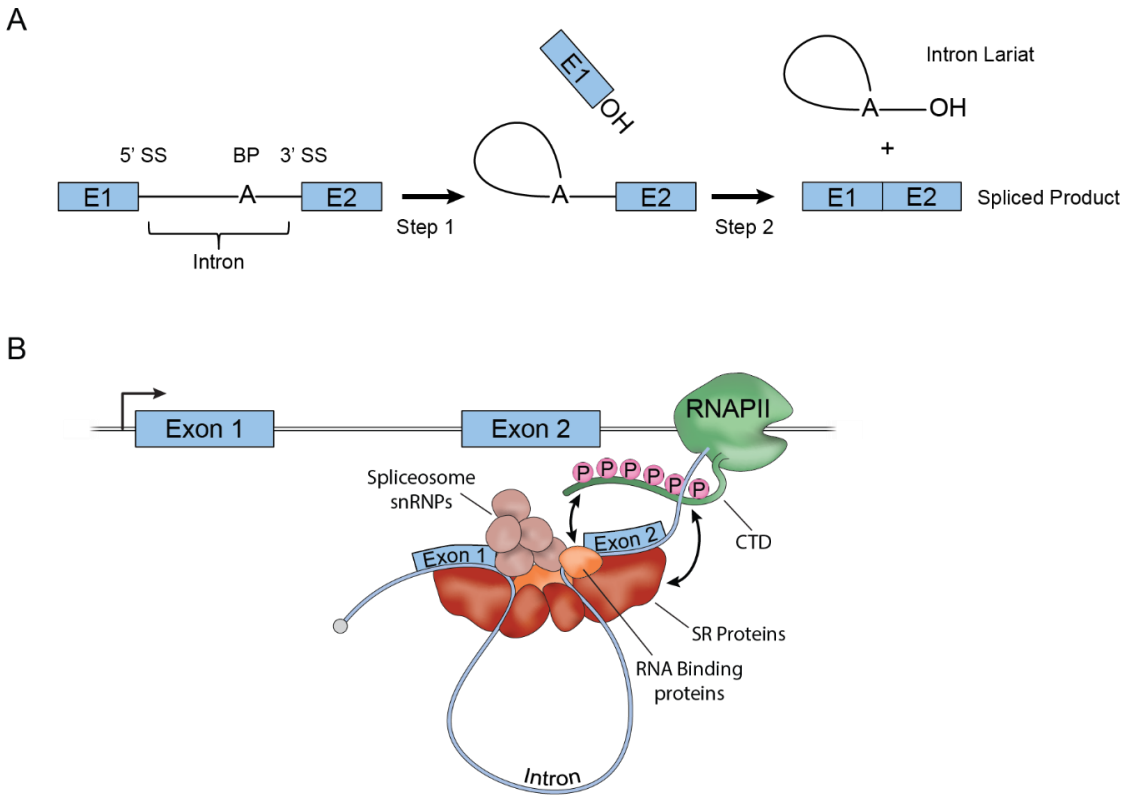


Figure 4. mRNA splicing steps and regulation.
(A) Basic steps of the splicing reaction. SS: splice site, BP: branch point, E1: exon 1, E2: exon 2.
(B) Splicing occurs co-transcriptionally. As a gene is transcribed, SR proteins and other RNA binding proteins bind to the nascent transcript, bringing exons into close proximity and aiding in the recruitment of the spliceosome such that the reaction outlined in (A) can occur. The phosphorylated RNAPII CTD interacts with SR proteins and RNA binding proteins to aid in their recruitment to the nascent transcript.

In this introduction I have outlined how gene expression is controlled by enhancers and phase separation. Transcription is regulated by enhancers, which are bound by TFs and coactivators that serve to recruit RNAPII to gene promoters. Distal enhancers come into close proximity to promoters through the formation of DNA loops which are constrained by a hierarchy of genome structures. Recent evidence has shown that enhancers form biomolecular condensates, which serve to concentrate and compartmentalize biochemical reactions. During transcript elongation, the pre-mRNA of most genes is spliced, and the splicing machinery can also be found in condensates. In Chapter 2 of this thesis, I will describe how various oncogenic enhancers located throughout the *MYC* locus come into close proximity with the *MYC* promoter through CTCF-CTCF mediated interactions. In Chapter 3, I will present evidence that the interaction of RNA polymerase II with enhancer condensates involves the C-terminal domain of the enzyme. Furthermore, I will provide evidence that phosphorylation of the CTD weakens these interactions, facilitating the transition of RNAPII into different condensates involved in splicing of the nascent transcript. In Chapter 4, I will conclude by describing some of the interesting directions in which the condensate field may be headed. Specifically, I will discuss our evolving understanding of the molecular rules that govern condensate formation and client incorporation, and how these rules may be utilized in the development of novel therapeutics.

References

Abraham, B.J., Hnisz, D., Weintraub, A.S., Kwiatkowski, N., Li, C.H., Li, Z., Weichert-Leahey, N., Rahman, S., Liu, Y., Etchin, J., et al. (2017). Small genomic insertions form enhancers that misregulate oncogenes. Nat. Commun. *8*, 14385.
Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell *176*, 419–434.
Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.
Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360–363.
Alpert, T., Herzel, L., and Neugebauer, K.M. (2017). Perfect timing: splicing and transcription rates in living cells. Wiley Interdiscip. Rev. RNA *8*, 1–12.
Avantaggiati, M.L., Carbone, M., Graessmann, A., Nakatani, Y., Howard, B., and Levine, A.S. (1996). The SV40 large T antigen and adenovirus E1a oncoproteins interact with distinct isoforms of the transcriptional co-activator, p300. EMBO J. *15*, 2236–2248.
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 Biol. *18*, 285–298.
Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a B-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.
Banjade, S., and Rosen, M.K. (2014). Phase transitions of multivalent proteins can promote clustering of membrane receptors. Elife *3*, 1–24.
Banjade, S., Wu, Q., Mittal, A., Peeples, W.B., Pappu, R. V., and Rosen, M.K. (2015). Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. Proc. Natl. Acad. Sci. *112*, E6426–E6435.
Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. Nat. Rev. Mol. Cell Biol. *18*, 437–451.
Beck, J. (1961). Variations in the morphological patterns of “autoimmune” nuclear fluorescence. Lancet *1*.
Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304–310.
Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. *74*, 3171–3175.
Berry, J., Weber, S.C., Vaidya, N., Haataja, M., and Brangwynne, C.P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. Proc. Natl. Acad. Sci. *112*, E5237–E5245.
Beyer, A.L., Bouton, A.H., and Miller, O.L. (1981). Correlation of hnRNP structure and nascent transcript cleavage. Cell *26*, 155–165.
Bieberstein, N.I., Oesterreich, F.C., Straube, K., and Neugebauer, K.M. (2012). First exon length controls active chromatin signatures and transcription. Cell Rep. *2*, 62–68.
Blencowe, B.J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci *25*.
Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T.-K., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., et al. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat. Struct. Mol. Biol. *25*, 833–840.
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.
Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science (80-.). *324*, 1729–1732.
Brangwynne, C.P., Mitchison, T.J., and Hyman, A.A. (2011). Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc. Natl. Acad. Sci. *108*, 4334–4339.
Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic Integration of Splicing within Gene Regulatory Pathways. Cell *152*, 1252–1269.
Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729–736.
Bringmann, P., Rinke, J., Appel, B., Reuter, R., and Lührmann, R. (1983). Purification of snRNPs U1, U2, U4, U5 and U6 with 2,2,7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1)RNP antisera. EMBO J. *2*.
Brody, E., and Abelson, J. (1985). The “Spliceosome”: yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science (80-.). *228*.
Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., et al. (2011). The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15 Using a Simple Protein Interface Forming a Fuzzy Complex. Mol. Cell *44*, 942–953.
Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: Lessons from genomics. Trends Genet. *28*, 276–284.
Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat. Rev. Genet. *14*, 427–439.
Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.
Bulger, M., Sawado, T., Schübeler, D., and Groudine, M. (2002). ChIPs of the β-globin locus: unraveling gene regulation within an active domain. Curr. Opin. Genet. Dev. *12*, 170–177.
Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of Precursors to mRNAs by the Spliceosomes.
Burgess, S., Couto, J.R., and Guthrie, C. (1990). A putative ATP binding protein influences the fidelity of branchpoint recognition in

yeast splicing. Cell *60*, 705–717.
Cáceres, E.F., and Hurst, L.D. (2013). The evolution, impact and properties of exonic splice enhancers. Genome Biol. *14*.
Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell *165*, 372–381.
Carstens, R.P., McKeehan, W.L., and Garcia-Blanco, M.A. (2015). An Intronic Sequence Element Mediates Both Activation and Repression of Rat Fibroblast Growth Factor Receptor 2 Pre-mRNA Splicing. Mol. Cell. Biol. *18*, 2205–2217.
Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: Exonic mutations that affect splicing. Nat. Rev. Genet. *3*, 285–298.
Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G.M., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and Characterization of Super-Enhancer-Associated Dependencies in Diffuse Large B Cell Lymphoma. Cancer Cell *24*, 777–790.
Chen, Y., and Belmont, A.S. (2019). Genome organization around nuclear speckles. Curr. Opin. Genet. Dev. *55*, 91–99.
Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., Steensel, B. van, Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J Cell Biol jcb.201807108.
Chipumuro, E., Marco, E., Christensen, C.L., Kwiatkowski, N., Zhang, T., Hatheway, C.M., Abraham, B.J., Sharma, B., Yeung, C., Altabef, A., et al. (2014). CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer. Cell *159*, 1126–1139.
Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T., et al. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife *5*, 1–31.
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 (80-.). *361*, 412–415.
Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science (80-.). *361*, eaar2555.
Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5’ ends of Adenovirus 2 messenger RNA. Cell *12*, 1–8.
Christian, H., Urlaub, H., Ficner, R., Fourmann, J., Schmitzova, J., Boon, K., Fabrizio, P., and Lu, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. Genes Dev. *23*, 413–428.
Chung, I., Leonhardt, H., and Rippe, K. (2011). De novo assembly of a PML nuclear subcompartment occurs through multiple pathways and induces telomere elongation. J. Cell Sci. *124*, 3603–3618.
Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. Science (80-.). *341*, 664–667.
Cohen, R.J., and Benedek, G.B. (1982). Equilibrium and kinetic theory of polymerization and the sol-gel transition. J. Phys. Chem. *86*.
Cremer, T., and Cremer, M. (2010). Chromosome Territories. Cold Spring Harb. Perspect. Biol. *2*, a003889–a003889.
Dekker, J. (2002). Capturing Chromosome Conformation. Science (80-.). *295*, 1306–1311.
Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.
Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.
Domdey, H., Apostol, B., Lin, R.J., Newman, A., Brody, E., and Abelson, J. (1984). Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. Cell *39*, 611–621.
Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.
Dunaway, M., and Droge, P. (1989). Transactivation of the Xenopus rRNA gene promoted by its enhancer. Nature *341*.
Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79–87.
Ebmeier, C.C., Erickson, B., Allen, B.L., Allen, M.A., Kim, H., Fong, N., Jacobsen, J.R., Liang, K., Shilatifard, A., Dowell, R.D., et al. (2017). Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. Cell Rep. *20*, 1173–1186.
Eckner, R., Yao, T.P., Oldread, E., and Livingston, D.M. (1996). Interaction and functional collaboration of p300/CBP and bHLH proteins in muscle and B-cell differentiation. Genes Dev. *10*, 2478–2490.
Eick, D., and Geyer, M. (2013). The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. Chem. Rev. *113*, 8456–8490.
Engelke, D.R., Ng, S.Y., Shastry, B.S., and Roeder, R.G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell *19*, 717–728.
Englesberg, E., Irr, J., Power, J., and Lee, N. (1965). Positive control of enzyme synthesis by gene C in the L-arabinose system. J. Bacteriol. *90*, 946–957.
Eron, L., and Block, R. (1971). Mechanism of Initiation of Repression of In Vitro Transcription of the Lac Operon of Escherichia coli. Proc. Natl. Acad. Sci. *68*, 1828–1832.
Flory, P.J. (1942). Thermodynamics of High Polymer Solutions. J. Chem. Phys. *10*.
Flory, P.J. (1953). Principles of Polymer Chemistry. Cornell Univ. Press.
Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.
Fulton, D.L., Sundararajan, S., Badis, G., Hughes, T.R., Wasserman, W.W., Roach, J.C., and Sladek, R. (2009). TFCat: The curated catalog of mouse and human transcription factors. Genome Biol. *10*.
Geiduschek, E.P., and Kassavetis, G.A. (2001). The RNA polymerase III transcription apparatus. J. Mol. Biol. *310*, 1–26.
Gibcus, J.H., and Dekker, J. (2013). The Hierarchy of the 3D Genome. Mol. Cell *49*, 773–782.
Gilbert, W., and Muller-Hill, B. (1966). Isolation of the Lac Repressor. Proc. Natl. Acad. Sci. *56*, 1891–1898.
Godowski, P.J., Picard, D., and Yamamoto, K.R. (1988). Signal transduction and transcriptional regulation by glucocorticoid receptor-LexA fusion proteins California, San Francisco 94143-0448. Science (80-.). *241*, 812–816.

Goodfellow, S.J., and Zomerdijk, J.C.B.M. (2013). Basic Mechanisms in RNA Polymerase I Transcription of the Ribosomal RNA Genes. In Subcell Biochem, pp. 211–236.
Green, M.R. (2005). Eukaryotic transcription activation: Right on target. Mol. Cell *18*, 399–402.
Greenblatt, J. (1997). RNA polymerase II holoenzyme and transcriptional regulation. Curr. Opin. Cell Biol. *9*, 310–319.
Griffith, J., Hochschild, A., and Ptashne, M. (1986). DNA loops induced by cooperative binding of λ repressor. Nature *322*, 750–752.
Grosveld, F., Antoniou, M., Berry, M., Boer, E., Dillon, N., Ellis, J., Fraser, P., Hurt, J., Imam, A., Meijer, A., et al. (1993). The regulation of human globin gene switching. Cold Spring Harb. Symp. Quant. Biol. *58*.
Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J., 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.
Hahn, S., Dunn, T., and Schleif, R. (1984). Upstream repression and CRP stimulation of the Escherichia colil-arabinose operon. J. Mol. Biol. *180*, 61–72.
Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Darzacq, X., and Tjian, R. (2018). An RNA-binding region regulates CTCF clustering and chromatin looping. BioRxiv 495432.
Hansen, S.K., Zhou, S., Andresen, J.M., Oliner, J.D., and Tjian, R. (2007). SREBP transcriptional activity is mediated through an interaction with the CREB-binding protein. Genes Dev. *10*, 2903–2911.
Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol.
Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep. *15*, 2147–2158.
Harmon, T.S., Holehouse, A.S., Rosen, M.K., and Pappu, R. V. (2017). Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. Elife *6*, 1–31.
Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154.
Heist, T., Fukaya, T., and Levine, M. (2019). Large distances separate coregulated genes in living Drosophila embryos. Proc. Natl. Acad. Sci. *116*, 15062–15067.
Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat. Rev. Mol. Cell Biol. *18*, 637–650.
Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell Resour. *155*.
Hnisz, D., Day, D.S., and Young, R.A. (2016a). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell *167*, 1188–1200.
Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science (80-.). *351*, 1454–1458.
Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.
Hochschild, A., and Ptashne, M. (1986). Cooperative binding of λ repressors to sites separated by integral turns of the DNA helix. Cell *44*, 681–687.
Holehouse, A.S., and Pappu, R. V. (2018). Functional Implications of Intracellular Phase Transitions. Biochemistry *57*, 2415–2423.
Hope, I.A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of Yeast. Cell *46*, 885–894.
Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.
Hu, Y., Kireev, I., Plutz, M., Ashourian, N., and Belmont, A.S. (2009). Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. J. Cell Biol. *185*, 87–100.
Huang, S., and Spector, D.L. (1996). Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. J. Cell Biol. *133*, 719–732.
Huggins, M.L. (1941). Solutions of long chain compounds. J. Chem *9*.
Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: New tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. *43*, D117–D122.
Hyman, A.A., and Simons, K. (2012). Beyond Oil and Water--Phase Transitions in Cells. Science (80-.). *337*, 1047–1049.
Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-Liquid Phase Separation in Biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.
Ippen, K., Miller, J.H., Scaife, J., and Beckwith, J. (1968). New controlling element in the lac operon of E. coli. Nature *217*, 825–827.
Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.
Jedidi, I., Zhang, F., Qiu, H., Stahl, S.J., Palmer, I., Kaufman, J.D., Nadaud, P.S., Mukherjee, S., Wingfield, P.T., Jaroniec, C.P., et al. (2010). Activator Gcn4 employs multiple segments of med15/Gal11, including the KIX domain, to recruit mediator to target genes in vivo. J. Biol. Chem. *285*, 2438–2455.
Jeong, S. (2017). SR Proteins: Binders, Regulators, and Connectors of RNA. Mol. Cells *40*, 1–9.
Jeronimo, C., Bataille, A.R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. Chem. Rev. *113*, 8491–8522.
Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., et al. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell *18*, 262–275.
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.
Jin, W., Wang, L., Zhu, F., Tan, W., Lin, W., Chen, D., Sun, Q., and Xia, Z. (2016). Critical POU domain residues confer Oct4 uniqueness in somatic cell reprogramming. Sci. Rep. *6*, 1–12.
Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.
Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. *339*, 225–229.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip. Rev. Dev. Biol. *1*, 40–51.
Kaiser, T.E., Intine, R. V., and Dundr, M. (2008). De Novo Formation of a Subnuclear Body. Science (80-.). *322*, 1713–1717.
Keegan, L., Gill, G., and Ptashne, M. (1984). Separation of DNA Binding from the Function of a Eukaryotic Regulatory Protein. Science (80-.). *231*, 699–704.
Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., Van Der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46*, D260–D266.
Khanna, N., Hu, Y., and Belmont, A.S. (2014). HSP70 Transgene Directed Motion to Nuclear Speckles Facilitates Heat Shock Activation. Curr. Biol. *24*, 1138–1144.
Kim, J., Han, K.Y., Khanna, N., Ha, T., and Belmont, A.S. (2019). Nuclear speckle fusion via long-range directional motion regulates speckle morphology after transcriptional inhibition. J. Cell Sci. *132*, jcs226563.
Koleske, A.J., and Young, R.A. (1995). The RNA polymerase II holoenzyme and its implications for gene regulation. Trends Biochem. Sci. *20*, 113–116.
Kornberg, R.D., and Lorch, Y. (1991). Irresistible force meets immovable object: Transcription and the nucleosome. Cell *67*, 833–836.
Kruhlak, M.J., Bazett-Jones, D.P., Lever, M.A., Hendzel, M.J., Fischle, W., and Verdin, E. (2000). Reduced mobility of the alternate splicing factor (ASF) through the nucleoplasm and steady state speckle compartments. J. Cell Biol. *150*, 41–51.
Kwiatkowski, N., Zhang, T., Rahl, P.B., Abraham, B.J., Reddy, J., Ficarro, S.B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B., et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. Nature *511*, 616–620.
de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.
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.
Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: A model for nuclear organelles. Nat. Rev. Mol. Cell Biol. *4*, 605–612.
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.
Lech, K., Anderson, K., and Brent, R. (1988). DNA-bound Fos proteins activate transcription in yeast. Cell *52*, 179–184.
Lee, T.I., and Young, R. a (2000). Transcription of Eukaryotic Protein- Coding Genes. Annu. Rev. Genet. 77–137.
Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.
Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu. Rev. Biochem. *84*, 291–323.
Lee, J., Krivega, I., Dale, R.K., and Correspondence, A.D. (2017). The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. CellReports *19*, 2490–2502.
Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. *13*, 233–245.
Lerner, E.A., Lerner, M.R., Janeway, C.A., and Steitz, J.A. (1981). Monoclonal antibodies to nucleic acid-containing cellular constituents: probes for molecular biology and autoimmune disease. Proc. Natl. Acad. Sci. *78*, 2737–2741.
Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S., and Steitz, J.A. (1980). Are snRNPs involved in splicing? Nature *283*.
Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.
Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147–151.
Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping Back to Leap Forward: Transcription Enters a New Era. Cell *157*, 13–25.
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.
Lieberman-aiden, E., Berkum, N.L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science (80-.). *326*, 289–294.
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.
Linz, R.-J., Newmang, A.J., Cheng, S.-C., and Abelson, J. (1985). Yeast mRNA splicing in vitro. J. Biol. Chem. *260*, 14780–14792.
Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247.e17.
Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. *417*, 15–27.
Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.
Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.
Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science (80-.). *360*, 918–921.
Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.
Maniatis, T., Falvo, J., Kim, T., Kim, T., Lin, C., Parekh, B., and Wathelet, M. (1998). Structure and function of the Interferon-B Enhanceosome. Cold Spring Harb. Symp. Quant. Biol.
Manley, J.L. (2002). Nuclear coupling: RNA processing reaches back to transcription. Nat. Struct. Biol. *9*, 790–791.
Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science (80-.). *346*, 1373–1377.
Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat. Cell Biol. *13*, 95–101.

Martin, D.I.K., Fiering, S., and Groudine, M. (1996). Regulation of β-globin gene expression: straightening out the locus. Curr. Opin. Genet. Dev. *6*, 488–495.
Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional Regulatory Elements in the Human Genome. Annu. Rev. Genomics Hum. Genet. *7*, 29–59.
Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. *255*, 11992–11996.
McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature.
McKnight, S.L., and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. Science (80-.). *217*, 316–324.
Mikhaleva, S., and Lemke, E.A. (2018). Beyond the Transport Function of Import Receptors: What's All the FUS about? Cell *173*, 549–553.
Milovanovic, D., Wu, Y., Bian, X., and Camilli, P. De (2018). A liquid phase of synapsin and lipid vesicles. *5671*, 1–8.
Misteli, T., and Phair, R.D. (2000). High mobility of proteins in the mammalian cell nucleus. Nature *404*, 604–609.
Mitchell, P.J., and Tjlan, R. (1989). Transcriptional Regulation in Mammalian Cells by DNA Binding Proteins. Science (80-.). *509*, 371–378.
Moen, P.T., Johnson, C., Bryon, M., Shopland, L., Serna, I., Imbalzano, A.N., and Lawrence, J.B. (2004). Repositioning of Muscle-specific Genes Relative to the Periphery of SC-35 Domains during Skeletal Myogenesis. Mol. Biol. Cell *15*, 197–206.
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.
Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. Cell *136*, 688–700.
Moore, M.J., Query, C.C., and Sharp, P.A. (1993). Splicing of precursors to messenger RNAs by the spliceosome. RNA World.
Müller, H.P., Sogo, J., and Schaffner, W. (1989). An enhancer stimulates transcription in Trans when attached to the promoter via a protein bridge. Cell *58*, 767–777.
Naftelberg, S., Schor, I.E., Ast, G., and Kornblihtt, A.R. (2015). Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. Annu. Rev. Biochem. *84*, 165–198.
Newman, A.J., Lin, R. fang, Cheng, S.C., and Abelson, J. (1985). Molecular consequences of specific intron mutations on yeast mRNA splicing in vivo and in vitro. Cell *42*, 335–344.
Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.
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.
Oesterreich, F.C., Bieberstein, N., and Neugebauer, K.M. (2011). Pause locally, splice globally. Trends Cell Biol. *21*, 328–335.
Osheim, Y.N., O.L. Miller, J., and Beyer, A.L. (1985). RNP particles at splice junction sequences on Drosophila chorion transcripts. Cell *43*, 143–151.
Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F., and Sharp, P.A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science (80-.). *225*, 898–903.
Patel, A., Malinovska, L., Saha, S., Wang, J., Alberti, S., Krishnan, Y., and Hyman, A.A. (2017). ATP as a biological hydrotrope. Science (80-.). *356*, 753–756.
Paul B. Sigler (1988). Acid blobs and Negative Noodles. Nature 210–212.
Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J.A., and Yamamoto, K.R. (1981). Purified glucocorticoid receptors bind selectively in vitro to a cloned DNA fragment whose transcription is regulated by glucocorticoids in vivo. Proc. Natl. Acad. Sci. *78*, 6628–6632.
Pena, V., Lee, C.-T., Dybkov, O., Kumar, V., Kastner, B., Urlaub, H., Luhrmann, R., Rauhut, R., Hartmuth, K., Fabrizio, P., et al. (2016). Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science (80-.). *353*, 1399–1405.
Perales, R., and Bentley, D. (2009). "Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions. Mol. Cell *36*, 178–191.
Plaschka, C., Nozawa, K., and Cramer, P. (2016). Mediator Architecture and RNA Polymerase II Interaction. J. Mol. Biol. *428*, 2569–2574.
Plys, A.J., Davis, C.P., Kim, J., Rizki, G., Keenen, M.M., Marr, S.K., and Kingston, R.E. (2019). Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. Genes Dev. *33*, 799–813.
Ptashne, M. (1967). Isolation of the lambda phage repressor. Proc. Natl. Acad. Sci. 306–313.
Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature *320*, 697–701.
Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683–689.
Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744-757.e24.
Rai, A.K., Chen, J.-X., Selbach, M., and Pelkmans, L. (2018). Kinase-controlled phase transition of membraneless organelles in mitosis. Nature.
Ransone, L.J., Wamsley, P., Morley, K.L., and Verma, I.M. (1990). Domain swapping reveals the modular nature of Fos, Jun, and CREB proteins. Mol. Cell. Biol. *10*, 4565–4573.
Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr. Opin. Genet. Dev. *43*, 73–81.
Ren, B., and Yue, F. (2015). Transcriptional Enhancers: Bridging the Genome and Phenome. Cold Spring Harb. Symp. Quant. Biol. *80*, 17–26.
Riedel, N., Wolin, S., and Guthrie, C. (1987). A subset of yeast snRNA's contains functional binding sites for the highly conserved Sm antigen. Science (80-.). *235*.
Ries, R.J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B.F., Patil, D.P., Kwak, H., Lee, J.H., and Jaffrey, S.R. (2019). m6A enhances the phase separation potential of mRNA. Nature *571*, 424–428.

Roberts, S.G.E. (2000). Mechanisms of action of transcription activation and repression domains. Cell. Mol. Life Sci. *57*, 1149–1160.
Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.-J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell *166*, 1411-1422.e16.
Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234–237.
Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317–331.
Russell, J., and Zomerdijk, J.C.B.M. (2006). Europe PMC Funders Group The RNA polymerase I transcription machinery. Biochem. Soc. *73*, 203–216.
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 (80-.). *361*, eaar3958.
Saha, S., Weber, C.A., Nousch, M., Adame-Arana, O., Hoege, C., Hein, M.Y., Osborne-Nishimura, E., Mahamid, J., Jahnel, M., Jawerth, L., et al. (2016). Polar Positioning of Phase-Separated Liquid Compartments in Cells Regulated by an mRNA Competition Mechanism. Cell *166*, 1572-1584.e16.
Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. J. Mol. Biol. *428*, 2623–2635.
Samuels, M., Fire, A., and Sharp, P.A. (1982). Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. *257*, 14419–14427.
Sawadogo, M. (1990). RNA polymerase B (II) and general transcription factors. Annu. Rev. Biochem. *59*, 711–754.
Sayre, M.H., Tschochner, H., and Kornberg, R.D. (1992). Reconstitution of transcription with five purified initiation factors and RNA polymerase II from Saccharomyces cerevisiae. J. Biol. Chem. *267*, 23376–23382.
Scaife, J., and Beckwith, J.R. (1966). Mutational alteration of the maximal level of Lac operon expression. Cold Spring Harb. Symp. Quant. Biol. *31*, 403–408.
Schuijers, J., Manteiga, J.C., Weintraub, A.S., Day, D.S., Zamudio, A.V., Hnisz, D., Lee, T.I., and Young, R.A. (2018). Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. Cell Rep. *23*, 349–360.
Schwer, B., and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. Nature *349*, 494–499.
Shevtsov, S.P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. Nat. Cell Biol. *13*, 167–173.
Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science (80-.). *357*, eaaf4382.
Shin, Y., Chang, Y.-C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell *175*, 1481-1491.e13.
Shopland, L.S., Johnson, C. V., Byron, M., McNeil, J., and Lawrence, J.B. (2003). Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains. J. Cell Biol. *162*, 981–990.
Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A. V, Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., et al. (2019). Enhancer features that drive formation of transcriptional condensates. Mol. Cell *75*.
Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. *11*, 4047–4057.
Soutourina, J. (2017). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol.
Spector, D.L., and Lamond, A.I. (2011). Nuclear Speckles. Cold Spring Harb. Perspect. Biol. *3*, a000646–a000646.
Spector, D.L., Schrier, W.H., and Busch, H. (1983). Immunoelectron microscopic localization of snRNPs. Biol Cell *49*.
Sperling, R. (2017). The nuts and bolts of the endogenous spliceosome. Wiley Interdiscip. Rev. RNA *8*.
Staby, L., O'Shea, C., Willemoës, M., Theisen, F., Kragelund, B.B., and Skriver, K. (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. Biochem. J. *474*, 2509–2532.
Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a Segmentation Stripe by Overlapping Activators and Repressors in the Drosophila Embryo. Science (80-.). *254*.
Steitz, J.A., Wolin, S.L., Rinke, J., Pettersson, I., Mount, S.M., Lerner, E.A., Hinterberger, M., and Gottlieb, E. (1983). Small ribonucleoproteins from eukaryotes: Structures and roles in RNA biogenesis. Cold Spring Harb. Symp. Quant. Biol. *47*.
Stockmayer, W.H. (1952). Molecular distribution in condensation polymers. J. Polym. Sci. *9*.
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.
Struhl, K. (1988). The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast. Nature *332*.
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 (80-.). *352*, 595–599.
Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E., and Chambon, P. (1989). The human estrogen receptor has two independent nonacidic transcriptional activation functions. Cell *59*, 477–487.
Triezenberg, S.J. (1995). Structure and function of transcriptional activation domains. Curr. Opin. Genet. Dev. *5*, 190–196.
Tuttle, L.M., Pacheco, D., Warfield, L., Luo, J., Ranish, J., Hahn, S., and Klevit, R.E. (2018). Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. Cell Rep. *22*, 3251–3264.
Vannini, A., and Cramer, P. (2012). Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. Mol. Cell *45*, 439–446.
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.
Wang, Z., and Burge, C.B. (2008). Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. RNA *14*, 802–813.
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 *0*, 1–12.
Wang, Y., Ma, M., Xiao, X., and Wang, Z. (2012). Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. Nat. Struct. Mol. Biol. *19*, 1044–1052.
Wang, Y., Zhang, T., Kwiatkowski, N., Abraham, B.J., Lee, T.I., Xie, S., Yuzugullu, H., Von, T., Li, H., Lin, Z., et al. (2015a). CDK7-

Dependent Transcriptional Addiction in Triple-Negative Breast Cancer. Cell *163*, 174–186.
Wang, Y., Liu, J., Huang, B., XU, Y.-M., LI, J., HUANG, L.-F., LIN, J., ZHANG, J., MIN, Q.-H., YANG, W.-M., et al. (2015b). Mechanism of alternative splicing and its regulation. Biomed. Reports *3*, 152–158.
Warfield, L., Tuttle, L.M., Pacheco, D., Klevit, R.E., and Hahn, S. (2014). A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. Proc. Natl. Acad. Sci. *111*, E3506–E3513.
Weil, A.P., Luse, D.S., Segall, J., and Roeder, R.G. (1979). Selective and accurate initiation of transcription at the ad2 major late promotor in a soluble system dependent on purified rna polymerase ii and dna. Cell *18*, 469–484.
Weinberg, R.A., and Penman, S. (1968). Small molecular weight monodisperse nuclear RNA. J. Mol. Biol. *38*.
Weintraub, A.S., Li, C.H., Zamudio, A. V, Sigova, A.A., Hannet, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2018). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell 1–16.
Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.
Will, C.L., and Luhrmann, R. (2011). Spliceosome Structure and Function. Cold Spring Harb. Perspect. Biol. *3*.
Wippich, F., Bodenmiller, B., Trajkovska, M.G., Wanka, S., Aebersold, R., and Pelkmans, L. (2013). Dual Specificity Kinase DYRK3 Couples Stress Granule Condensation/Dissolution to mTORC1 Signaling. Cell *152*, 791–805.
Wise, J.A., Tollervey, D., Maloney, D., Swerdlow, H., Dunn, E.J., and Guthrie, C. (1983). Yeast contains small nuclear RNAs encoded by single copy genes. Cell *35*, 743–751.
Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. Mol. Cell *54*, 601–612.
Workman, J.L., and Buchman, A.R. (1993). Multiple functions of nucleosomes and regulatory factors in transcription. Trends Biochem. Sci. *18*, 90–95.
Wu, Z., Murphy, C., Callan, H.G., and Gall, J.G. (1991). Small nuclear ribonucleoproteins and heterogeneous nuclear ribonucleoproteins in the amphibian germinal vesicle: Loops spheres, and snurposomes. J. Cell Biol. *113*, 465–483.
Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol. Cell. Biol. *14*, 7219–7225.
Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N., and Obradovic, Z. (2007). Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. J. Proteome Res. *6*, 1882–1898.
Xing, Y. (1995). Nonrandom gene organization: structural arrangements of specific pre- mRNA transcription and splicing with SC-35 domains. J. Cell Biol. *131*, 1635–1647.
Yin, J. -w., and Wang, G. (2014). The Mediator complex: a master coordinator of transcription and cell lineage development. Development *141*, 977–987.
Young, R.A. (1991). RNA polymerase II. Annu. Rev. Biochem. *60*, 689–715.
Yuan, W., Condorelli, G., Caruso, M., Felsani, A., and Giordanoi, A. (1996). Human p300 protein is a coactivator for the transcription factor MyoD. J. Biol. Chem. *271*, 9009–9013.
Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors : establishing competence for gene expression Parameters affecting transcription factor access to target sites in chromatin Initiating events in chromatin : pioneer factors bind first. Genes Dev. 2227–2241.

CHAPTER 2: TRANSCRIPTIONAL DYSREGULATION OF *MYC* REVEALS COMMON ENHANCER-DOCKING MECHANISM

Originally published in *Cell Reports* Volume 23, Issue 2. (2018)

Jurian Schuijers[1,3], John Colonnese Manteiga[1,2,3], Abraham Selby Weintraub[1,2], Daniel Sindt Day[1], Alicia Viridiana Zamudio[1,2], Denes Hnisz[1], Tong Ihn Lee[1], Richard Allen Young[1,2,4,*]

[1] Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142, USA.

[2] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.

[3] These authors contributed equally

[4] lead contact

* Correspondence to: young@wi.mit.edu

SUMMARY

Transcriptional dysregulation of the *MYC* oncogene is among the most frequent events in aggressive tumor cells, and this is generally accomplished by acquisition of a super-enhancer somewhere within the 2.8 Mb TAD where *MYC* resides. We find that these diverse cancer-specific super-enhancers, differing in size and location, interact with the *MYC* gene through a common and conserved CTCF binding site located 2 kb upstream of the *MYC* promoter. Genetic perturbation of this enhancer-docking site in tumor cells reduces CTCF binding, super-enhancer interaction, *MYC* gene expression, and cell proliferation. CTCF binding is highly sensitive to DNA methylation, and this enhancer-docking site, which is hypomethylated in diverse cancers, can be inactivated through epigenetic editing with dCas9-DNMT. Similar enhancer-docking sites occur at other genes, including genes with prominent roles in multiple cancers, suggesting a mechanism by which tumor cell oncogenes can generally hijack enhancers. These results provide insights into mechanisms that allow a single target gene to be regulated by diverse enhancer elements in different cell types.

KEYWORDS: Gene regulation, super-enhancers, chromosome structure, enhancer-docking

INTRODUCTION

Elevated expression of the c-MYC transcription factor occurs in a broad spectrum of human cancers and is associated with tumor aggression and poor clinical outcome (Berns et al., 1992; Dang, 2012; Gabay et al., 2014; Grotzer et al., 2001). Activation of the *MYC* gene, which encodes c-MYC, is a hallmark of cancer initiation and maintenance. Dysregulation of *MYC* is often achieved through the formation of large tumor-specific super-enhancers in the region surrounding the *MYC* gene (Chapuy et al., 2013; Fulco et al., 2016; Herranz et al., 2014; Hnisz et al., 2013; Lin et al., 2016; Liu et al., 2015; Lovén et al., 2013; Shi et al., 2013; Whyte et al., 2013; Zhang et al., 2016). These large enhancer clusters differ in size, composition, and distance from the *MYC* promoter, yet all accomplish the same task of stimulating *MYC* overexpression across a broad spectrum of tumors.

Selective gene activation is essential to the gene expression programs that define both normal and cancer cells. During gene activation, transcription factors (TFs) bind enhancer elements and regulate transcription from the promoters of nearby or distant genes through physical contacts that involve looping of DNA between enhancers and promoters (Bonev et al., 2016; Buecker and Wysocka, 2012; Bulger and Groudine, 2011; Fraser et al., 2015; Müller et al., 1989; Spitz, 2016; de Wit et al., 2013). The mechanisms that ensure that specific enhancers interact with specific promoters are not fully understood. Some enhancer-promoter interactions are likely determined by the nature of transcription factors bound at the two sites (Muerdter and Stark, 2016; Weintraub et al., 2017).

Recent studies have revealed that specific chromosome structures play important roles in gene control. Enhancer-promoter interactions generally occur within larger chromosomal loop structures formed by the interaction of CTCF proteins bound to each of the loop anchors (Dekker and Mirny, 2016; Fraser et al., 2015; Gibcus and Dekker, 2013; Gorkin et al., 2014; Hnisz et al., 2016a, 2018; Ji et al., 2016). These loop structures, variously called topologically associated domains (TADs), sub-TADs, loop domains, CTCF contact domains, and insulated neighborhoods, tend to insulate enhancers and genes within the CTCF-CTCF loops from elements outside those loops (Dixon et al., 2012; Dowen et al., 2014; Franke et al., 2016; Hnisz et al., 2016a, 2016b; Ji et al., 2016; Narendra et al., 2015; Nora et al., 2012; Phillips-Cremins et al., 2013; Rao et al., 2014). Constraining DNA interactions within CTCF-CTCF loop structures in this manner may facilitate proper enhancer-promoter contacts.

CTCF does not generally occupy enhancer and promoter elements (Cuddapah et al., 2009; Dixon et al., 2012; Dowen et al., 2014; Handoko et al., 2011; Ji et al., 2016; Kim et al., 2007; Phillips-Cremins et al., 2013; Rao et al., 2014; Rubio et al., 2008; Tang et al., 2015). Another TF, YY1, generally binds to enhancers and promoters and facilitates their interaction through YY1 dimerization (Weintraub et al., 2017). However, when CTCF does bind these regulatory elements, it can also contribute to enhancer-promoter interactions (Banani et al., 2017; Nora et al., 2017; Splinter et al., 2006; Zuin et al., 2014).

Here, we investigate DNA looping structures in the *MYC* locus in multiple cancers and identify a CTCF-occupied site at the *MYC* promoter that facilitates docking with essentially any enhancers that are formed within the 2.8 Mb *MYC* locus. The CTCF-occupied site at the *MYC* promoter, which we call the *MYC* enhancer-docking site, can be abrogated by genetic and epigenetic editing. Similar enhancer-docking sites occur at other oncogenes. This suggests a mechanism by which tumor cells can generally hijack enhancers and, with editing, a potential therapeutic vulnerability.

RESULTS

Cell type-specific *MYC* enhancers loop to a common upstream CTCF site

Previous studies have established that tumor cells acquire tumor-specific super-enhancers at various sites throughout the *MYC* locus (Figure 1A and S1A) (Bradner et al., 2017; Chapuy et al., 2013; Gabay et al., 2014; Gröschel et al., 2014; Herranz et al., 2014; Hnisz et al., 2013; Lin et al., 2016; Lovén et al., 2013; Parker et al., 2013; Zhang et al., 2016; Shi et al., 2013), but the mechanisms by which these diverse enhancer structures control *MYC* are not clear. In one case, for example, a super-enhancer located ~2 Mb downstream of the *MYC* gene has been shown to physically interact with *MYC*, but the mechanisms responsible for this specific interaction are unclear (Shi et al., 2013). To gain insights into the potential role of DNA loop structures in gene control at the *MYC* locus, we generated cohesin HiChIP data for HCT-116 cells and collected published DNA interaction data for three other cancer cell types for comparison (Figure 1B; Table S1 and S5) (Hnisz et al., 2016a; Pope et al., 2014). Among the DNA loop structures observed in these datasets, a large 2.8 Mb DNA loop was evident in all four cell types. This loop connects CTCF sites encompassing the *MYC* gene and qualifies as an insulated neighborhood. The DNA anchor sites of this 2.8 Mb DNA loop occur at the boundaries of a TAD found in all cells (Figure S1B). The *MYC* TAD encompasses a region previously described as a “gene desert” because this large span of DNA contains no other annotated protein-coding genes (Montavon and Duboule, 2012; Ovcharenko et al., 2005).

While all cells examined appear to share the TAD-spanning 2.8 Mb loop encompassing *MYC*, the loop structures within the neighborhood were found to be markedly different among the tumor types. The internal loops were dominated by interactions between a *MYC* promoter-proximal CTCF site and the diverse super-enhancers (Figures 1B and 1C). The major differences between these internal structures in the different tumor cells involved the different positions of the tumor-specific super-enhancer elements. Examination of Hi-C data for a broader spectrum of tumor cell types suggests that tumor cells generally have DNA contacts between the *MYC* promoter-proximal site and other sites within the 2.8 Mb *MYC* TAD (Figure S1B). This looping was not limited to cancer cells, because examination of enhancer and promoter-capture Hi-C data in a variety of normal cell types that express *MYC* (Javierre et al., 2016) revealed that cell-type-specific enhancers do indeed loop to the *MYC* proximal CTCF site (Figures S1C and S1D). This indicated that this CTCF site is also used during normal development by cell-type-specific enhancers to facilitate *MYC* expression and cellular proliferation.

Further examination of the *MYC* promoter-proximal region revealed three constitutive CTCF binding sites (Figure 1C). All three sites were found to be occupied by CTCF in a wide variety of normal cells and tumor cells, and this binding pattern is shared across species (Figure S1C). Previous studies have examined the role of CTCF binding at all three sites (Filippova et al., 1996; Gombert and Krumm, 2009; Gombert et al., 2003; Klenova et al., 1993; Rubio et al., 2008). The two sites located within the *MYC* gene have been shown to play roles in *MYC* transcript start site selection and in promoter-proximal pausing of RNA polymerase II (Filippova et al., 1996). The CTCF binding site located 2 kb upstream of the major transcript start site has been reported to protect the promoter from methylation and to be an insulator element (Gombert and Krumm, 2009; Gombert et al., 2003). The DNA interaction data described here, however, suggests that this upstream site dominates connections with distal enhancer elements, as the majority of reads in the DNA interaction data are associated with this site in all tumor cells examined (Figures 1C and S1E). The -2 kb CTCF binding site contains a number of putative

CTCF binding motifs; one of these most closely matches the canonical CTCF motif in the JASPAR database (Sandelin et al., 2004) and occurs within a highly conserved sequence (Figure 1D). These features, the presence of CTCF sites in tumor super-enhancers, and the ability of two CTCF-bound sites to be brought together through CTCF homodimerization (Saldana-Meyer et al., 2014; Yusufzai et al., 2004) led us to further study the possibility that the -2 kb site has an enhancer-docking function critical to *MYC* expression.

FIGURE 1

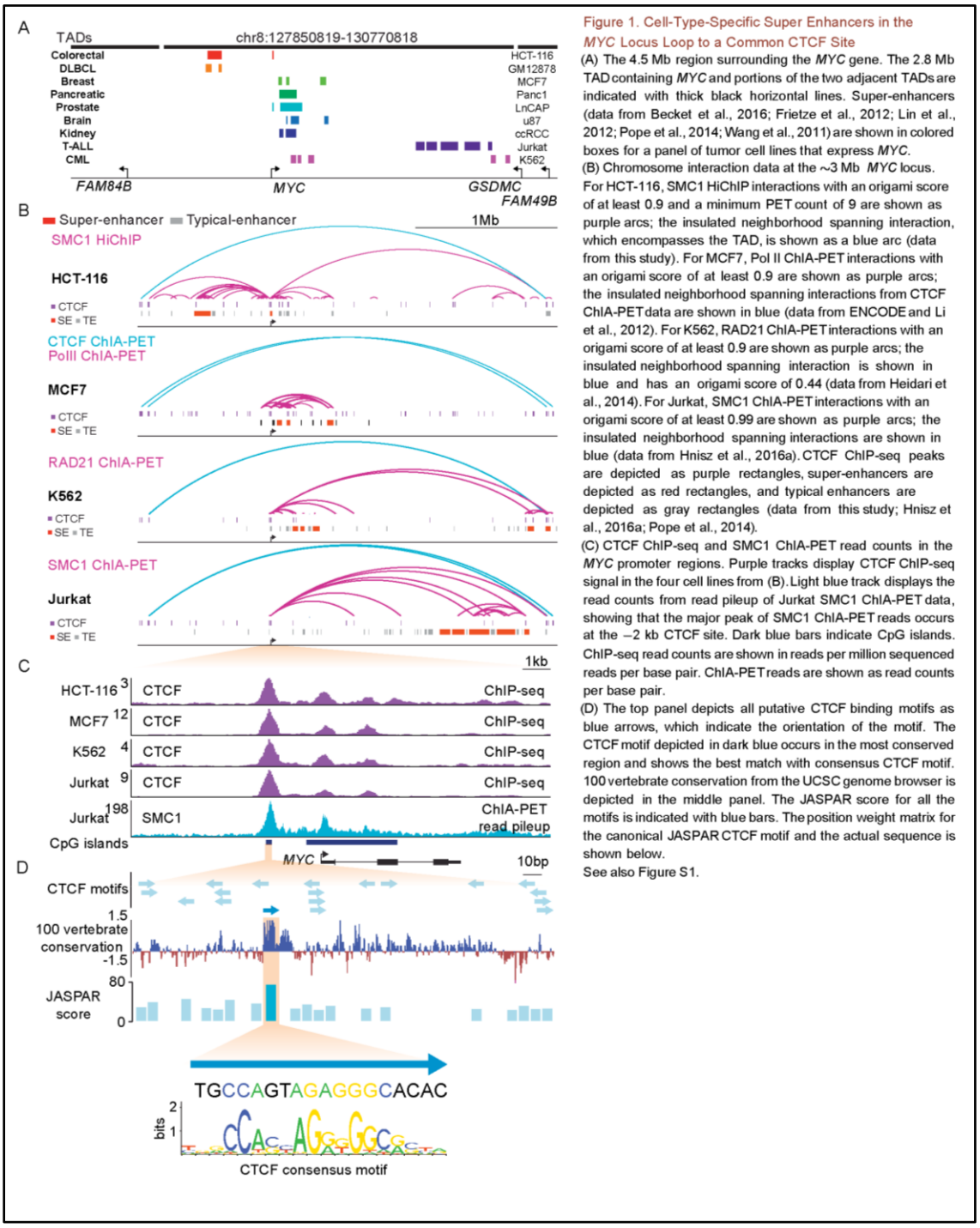


Figure 1. Cell-Type-Specific Super Enhancers in the *MYC* Locus Loop to a Common CTCF Site

(A) The 4.5 Mb region surrounding the *MYC* gene. The 2.8 Mb TAD containing *MYC* and portions of the two adjacent TADs are indicated with thick black horizontal lines. Super-enhancers (data from Becket et al., 2016; Frietze et al., 2012; Lin et al., 2012; Pope et al., 2014; Wang et al., 2011) are shown in colored boxes for a panel of tumor cell lines that express *MYC*.

(B) Chromosome interaction data at the ~3 Mb *MYC* locus. For HCT-116, SMC1 HiChIP interactions with an origami score of at least 0.9 and a minimum PET count of 9 are shown as purple arcs; the insulated neighborhood spanning interaction, which encompasses the TAD, is shown as a blue arc (data from this study). For MCF7, Pol II ChIA-PET interactions with an origami score of at least 0.9 are shown as purple arcs; the insulated neighborhood spanning interactions from CTCF ChIA-PET data are shown in blue (data from ENCODE and Li et al., 2012). For K562, RAD21 ChIA-PET interactions with an origami score of at least 0.9 are shown as purple arcs; the insulated neighborhood spanning interaction is shown in blue and has an origami score of 0.44 (data from Heidari et al., 2014). For Jurkat, SMC1 ChIA-PET interactions with an origami score of at least 0.99 are shown as purple arcs; the insulated neighborhood spanning interactions are shown in blue (data from Hnisz et al., 2016a). CTCF ChIP-seq peaks are depicted as purple rectangles, super-enhancers are depicted as red rectangles, and typical enhancers are depicted as gray rectangles (data from this study; Hnisz et al., 2016a; Pope et al., 2014).

(C) CTCF ChIP-seq and SMC1 ChIA-PET read counts in the *MYC* promoter regions. Purple tracks display CTCF ChIP-seq signal in the four cell lines from (B). Light blue track displays the read counts from read pileup of Jurkat SMC1 ChIA-PET data, showing that the major peak of SMC1 ChIA-PET reads occurs at the −2 kb CTCF site. Dark blue bars indicate CpG islands. ChIP-seq read counts are shown in reads per million sequenced reads per base pair. ChIA-PET reads are shown as read counts per base pair.

(D) The top panel depicts all putative CTCF binding motifs as blue arrows, which indicate the orientation of the motif. The CTCF motif depicted in dark blue occurs in the most conserved region and shows the best match with consensus CTCF motif. 100 vertebrate conservation from the UCSC genome browser is depicted in the middle panel. The JASPAR score for all the motifs is indicated with blue bars. The position weight matrix for the canonical JASPAR CTCF motif and the actual sequence is shown below.

See also Figure S1.

MYC promoter proximal CTCF site is necessary for enhancer-promoter looping and high *MYC* expression

To determine whether the putative enhancer-docking site plays a functional role in *MYC* expression through DNA loop formation, small perturbations of the CTCF binding site were generated in both alleles of the tumor cell lines K562, HCT-116, Jurkat, and MCF7 using CRISPR/Cas9 (Figure 2A, B). Attempts at genetic perturbation by transfection with constructs carrying CRISPR/Cas9 with a guide RNA specifically targeting the CTCF motif upstream of the *MYC* gene did not yield viable clones. To allow cells to continue to proliferate if the CTCF motif deletion was lethal, cells were virally transduced with an exogenous *MYC* gene driven by a phosphoglycerate kinase (PGK) promoter (Figure S2A). This construct contained sequence differences in the 3' UTR that allowed discrimination between the endogenous and exogenous *MYC* mRNAs. Cells expressing this exogenous *MYC* gene were then subjected to CRISPR/Cas9 perturbation. Clones were selected with small deletions or insertions disrupting the canonical CTCF motif (Figures 2B and S2B) and these cells were further characterized. CTCF ChIP-seq showed complete loss of CTCF binding to this site in K562 and HCT-116 cells and a 60-70% reduction in CTCF binding at this site in Jurkat and MCF7 cells (Figure 2C). RNA analysis revealed a 70-80% reduction of endogenous *MYC* mRNA in the absence of the enhancer-docking site in all of these cell types (Figure 2D). Furthermore, an inducible CRISPR/Cas9 perturbation model showed reduced proliferation for these four cell types upon induction of CTCF-site deletions (Figures S2C-S2G). These results indicate that the CTCF motif in the *MYC* enhancer-docking site is necessary for CTCF binding, for high levels of *MYC* expression and for cellular proliferation.

If the putative *MYC* enhancer-docking site contributes to looping interactions with distal enhancers, then the loss of this site should cause a decrease in DNA interactions between the *MYC* promoter and the distal super-enhancers. We used chromosome conformation capture combined with high-throughput sequencing (4C-seq) to compare super-enhancer interactions in K562 and HCT-116 cells with normal or perturbed CTCF binding motifs. The 4C-seq data in K562 cells indicated that the *MYC* enhancer-docking site interacts predominantly with distal super-enhancers located ~0.3 Mb and ~2 Mb downstream of the *MYC* gene, and that the majority of these interactions were significantly reduced when the putative CTCF motif was perturbed (Figures 3A and S3A). In order to control for any direct effects of a genetic alteration near the viewpoint, 4C-seq was performed with a viewpoint placed in the downstream super-enhancer. This showed clear interactions with the *MYC* enhancer-docking site as well as with the nearby super-enhancer, and these interactions were significantly reduced upon perturbation of the CTCF motif (Figures 3B and S3B). Similar results were obtained in HCT-116 cells, where the viewpoint was centered on the super-enhancer located ~0.4 Mb upstream of the *MYC* gene (Figure S3C). These results showed that the CTCF site in the promoter-proximal region of *MYC* is important for optimal interaction with distal enhancers and supports the idea that this CTCF site functions as an enhancer-docking site.

FIGURE 2

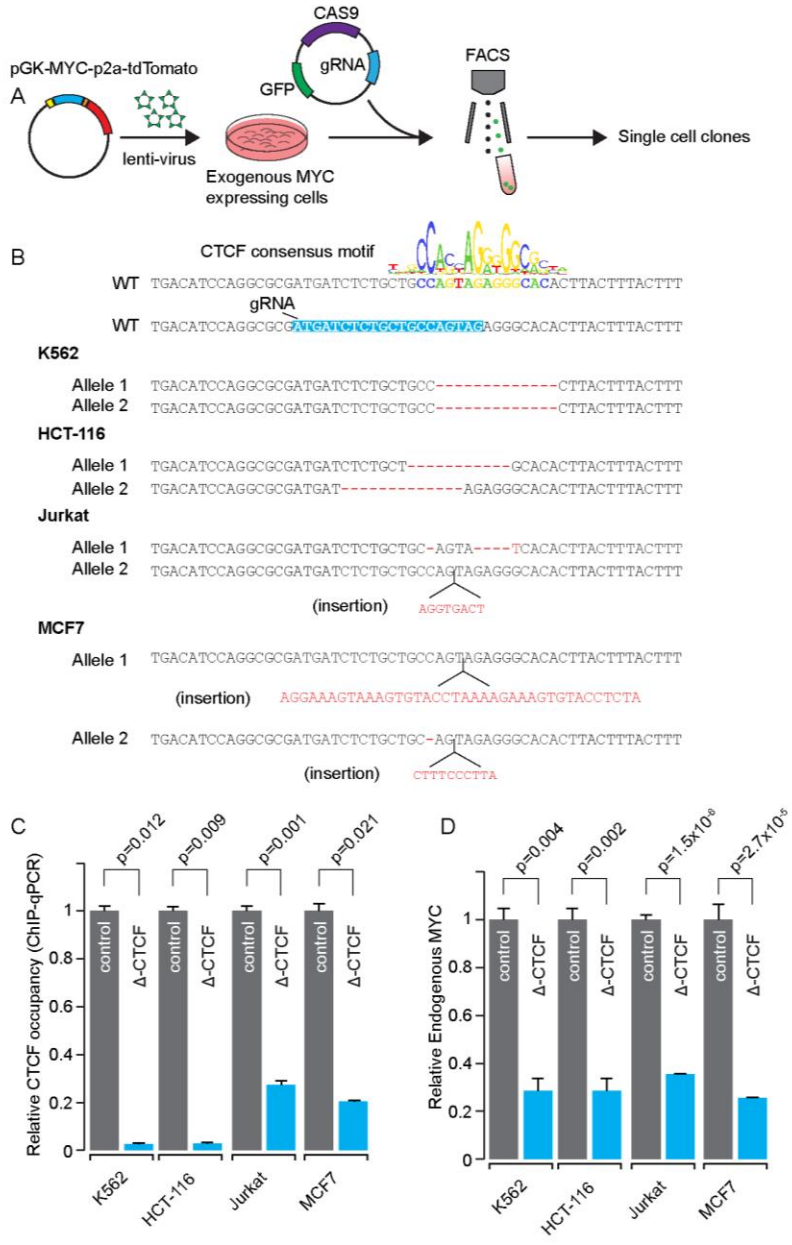


Figure 2. Perturbation of the Core CTCF Motif in the *MYC* CTCF Loop-Anchor Reduces CTCF Occupancy and *MYC* Expression

(A) Schematic representation of the experiment. HCT-116, K562, Jurkat, and MCF7 cells were transduced with a construct expressing *MYC* under a PGK promoter and selected for successful integration. These cells were then transiently transfected with plasmid carrying Cas9 and a gRNA targeting the CTCF binding motif. Positive cells were identified and selected using fluorescence-activated cell sorting (FACS). These cells were multiplied, and clonal populations were characterized.

(B) The DNA sequences in the vicinity of the core CTCF motif and the mutations generated in clonal populations of K562, HCT-116, Jurkat, and MCF7 cell lines. The reference (WT, wild-type) sequence highlighted in blue is complementary to the gRNA sequence targeting the most prominent CTCF motif (shown here in bold colored sequence). For the aneuploid MCF7 cell line, the two most common mutations are depicted.

(C) ChIP-qPCR showing reduction of the CTCF occupancy in D-CTCF K562, HCT-116, Jurkat, and MCF7 cells. p values were generated with a Student's t test. Error bars represent the SD of the mean from three technical replicates.

(D) qPCR showing reduction of endogenous *MYC* mRNA levels in D-CTCF K562, HCT-116, Jurkat, and MCF7 cells. p values were generated with a Student's t test. Error bars represent the SD of the mean from three biological replicates.

See also Figure S2.

FIGURE 3

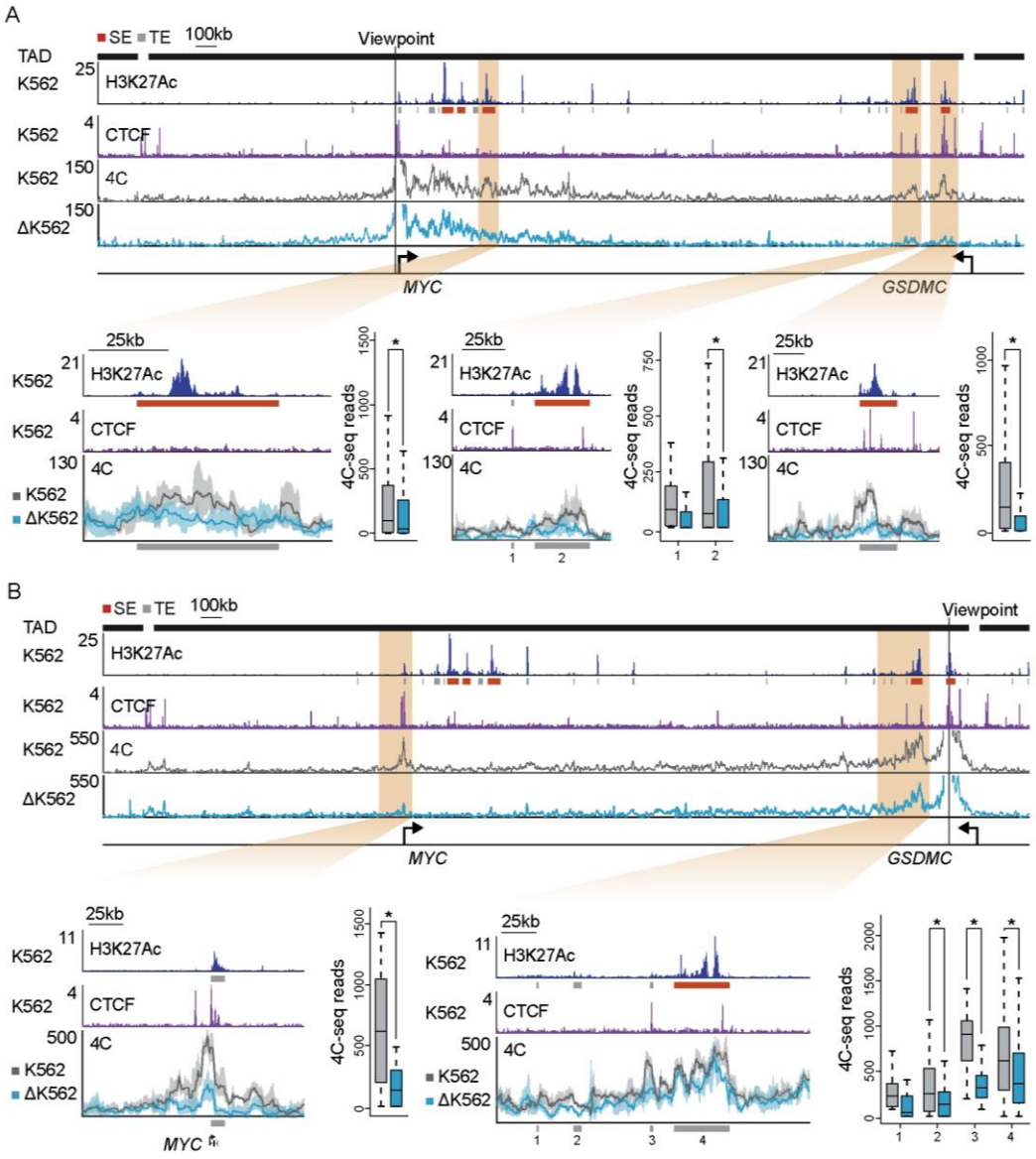


Figure 3. Perturbation of the *MYC* Enhancer-Docking Site Reduces Looping to Super-Enhancers

(A) 4C analysis showing reduced looping of the *MYC* promoter proximal CTCF site to super-enhancers in CTCF motif deletion cells (ΔK562) versus unmodified cells (K562). H3K27Ac ChIP-seq and CTCF ChIP-seq are shown in blue and purple colors respectively. Blowups show the 4C interactions for three K562 specific super-enhancers. The 4C viewpoint is situated 112 base pairs upstream of the deleted loop-anchor region.

(B) 4C analysis showing reduced looping of the *MYC* promoter proximal CTCF site to super-enhancers in CTCF motif deletion cells (ΔK562) versus unmodified cells (K562) using a viewpoint centered on the most distant super-enhancer downstream of the *MYC* gene. H3K27Ac ChIP-seq and CTCF ChIP-seq are shown in blue and purple, respectively. Blowups show the 4C interactions at the *MYC* promoter and distant super-enhancer near the viewpoint. Shading represents the 90% confidence interval based on three biological replicates. Peak calls from the H3K27Ac ChIP-seq were used to define the regions to be quantified and are indicated in gray boxes at the bottom of the panels. Boxplots show quantification of the reads per fragment for the indicated regions. p values were generated using Student's t test, and data pairs with a p value < 0.05 are indicated with an asterisk. Reads are shown in reads per million sequenced reads per base pair. Typical enhancers and super-enhancers are shown as gray boxes and red boxes, respectively.

See also Figure S3.

Loss of *MYC* expression with methylation of enhancer-docking site

CTCF binding is abrogated when its sequence motif is methylated (Bell and Felsenfeld, 2000; Maurano et al., 2015), and the *MYC* enhancer-docking site occurs within a CpG island that is consistently hypomethylated in different tumor types as well as in different normal tissues (Figures S4A and S4B). The recent development of tools that permit site-specific DNA methylation (Liu et al., 2016; Siddique et al., 2013) suggested a means to disrupt *MYC* expression by methylation of the enhancer-docking site. To achieve targeted methylation, we created a construct to express a dCas9 fusion protein consisting of the catalytic domain of DNMT3A and the interacting domain of DMNT3L. This dCas9-DNMT3A-3L protein was targeted to the *MYC* enhancer-docking site in K562 and HCT-116 cells using multiple guide RNAs that span the region (Figures 4A, and 4B). The targeting of dCas9-DNMT3A-3L resulted in robust local DNA methylation (Figure 4C) and a 40-50 % reduction in mRNA levels in both cell types (Figure 4D). The methylated region likely contains binding sites for additional transcription factors that may be sensitive to DNA methylation, so it is possible that the reduced mRNA levels are due to multiple factors. In order to test whether disruption of transcription factors other than CTCF contribute to the reduction in *MYC* mRNA levels, the dCas9-DNMT3A-3L was targeted to the *MYC* enhancer-docking site in CTCF-site deleted K562 cells. No further reduction of *MYC* mRNA levels was observed under these conditions (Figures S4C, and S4D) indicating that loss of CTCF was a major contributor to the observed reduction of *MYC* expression upon targeted methylation of the *MYC* enhancer-docking site. These results demonstrate that epigenetic editing of the enhancer-docking site can reduce *MYC* expression.

FIGURE 4

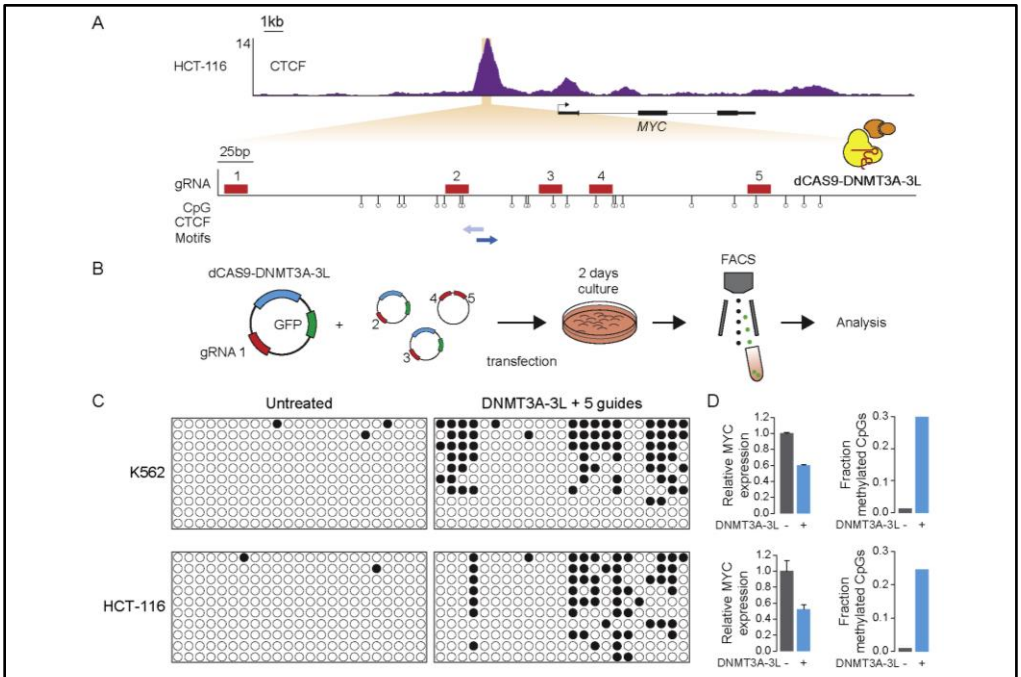


Figure 4. dCas9-Mediated Methylation of the CTCF Loop-Anchor Site Reduces *MYC* Expression in Tumor Cells

(A) Top panel shows CTCF ChIP-seq at the *MYC* gene region in HCT-116 cells. ChIP-seq reads are shown in reads per million sequenced reads per base pair. Bottom panels shows a blowup of the ~700-bp region underneath the CTCF peak depicting the CTCF motifs (blue arrows) and the gRNAs (red rectangles) used to target dCas9-DNMT3A-3L to the enhancer anchor. Lollipop symbols indicate the location of CpGs that are assayed for methylation levels in (C).

(B) Schematic representation of the experiment. HCT-116 or K562 cells were transfected with plasmids encoding the dCAS9-DNMT3A-3L, GFP, and a gRNA together with a plasmid encoding 2 additional gRNAs. HCT-116 or K562 cells were isolated by FACS after 2 days, and DNA and RNA were isolated.

(C) Methylation at *MYC* promoter loop-anchor site in untreated cells and cells transfected with dCas9-DNMT3A-3L in conjunction with the 5 indicated gRNAs.

(D) qPCR analysis of *MYC* mRNA levels and fraction of methylated CpGs for untreated and dCas9-DNMT3A-3L + 5 gRNA transfected cells. Error bars represent the SD of the mean for three biological replicates.

See also Figure S4.

CTCF enhancer-docking sites at additional genes

Previous genomic studies have noted that CTCF might engender enhancer-promoter interactions at a minority of genes (Banani et al., 2017; Nora et al., 2017; Splinter et al., 2006; Zuin et al., 2014). We therefore identified the set of genes whose promoter-proximal regions contain CTCF-bound sites and that show evidence of enhancer interactions in K562, Jurkat, and HCT-116 cells. We identified all active transcription start sites (TSSs) that have at least one CTCF-bound site within 2.5 kb of the TSS that interacts with at least one enhancer. This yielded between 555 and 1108 TSSs with a nearby CTCF site that loops to an active enhancer (Figure 5A; Table S2). We define these TSSs as having a putative CTCF enhancer-docking site. The majority of TSSs identified in this analysis were identified in only one cell type, with only 52 TSSs identified in all three cell types (Figure 5B). Nonetheless, these putative enhancer-docking sites tended to be constitutively bound by CTCF in all three cell types, and the CTCF motifs in these sites showed high sequence conservation (Figures 5C and 5D). This suggests that these putative enhancer-docking sites are occupied by CTCF regardless of interaction with active enhancers, and that differences in cell-type-specific enhancers are largely responsible for differential use of enhancer-docking site genes in these cells.

Gene ontology analysis of the genes with putative enhancer-docking sites found different processes to be significantly enriched in each cell type, and these processes were dominated by the cellular identity of the cell lines (Figure S5A; Table S3). Common processes among the three cell types include cell cycle and other cancer-related processes such as gene expression and response to signaling (Figure S5A). A number of cancer associated genes were found, including *TGIF1*, *VEGFA*, *RUNX1*, and *PIM1* (Figure 5E), as well as others (Figure S5B). We conclude that genes other than *MYC* are likely regulated by CTCF-bound enhancer-docking sites and that these include multiple cancer-associated genes.

DISCUSSION

Aberrant transcriptional activation of the *MYC* oncogene occurs frequently in tumor cells and is associated with tumor aggression. *MYC* resides within a 2.8 Mb TAD and its aberrant activation is generally accomplished by acquisition of a super-enhancer somewhere within that TAD. How these diverse cancer-specific super-enhancers loop long distances to specifically interact with *MYC* has not been clear. We find that the diverse super-enhancers commonly interact with, and depend on, a conserved CTCF binding site located 2 kb upstream of the *MYC* promoter. Because tumor super-enhancers can encompass genomic regions as large as 200 kb, and CTCF occupies sites that occur on average every 10 kb, there is considerable opportunity for super-enhancers to adventitiously contain a CTCF-bound site, which in turn could serve to interact with the *MYC* CTCF site (Table S6). Thus, different tumor super-enhancers have the opportunity to form through diverse mechanisms throughout this large TAD and can exploit the *MYC* CTCF site to interact with and activate *MYC* expression.

The concept that enhancer-promoter interactions generally occur within larger chromosomal loop structures such as TADs, which are themselves often formed by the interaction of CTCF proteins bound to each of the TAD loop anchors (Dekker and Mirny, 2016; Fraser et al., 2015; Gorkin et al., 2014; Hnisz et al., 2016a), is supported by the observations described here. These larger loop structures tend to insulate enhancers and genes within the CTCF-CTCF loops from elements outside those loops. Constraining DNA interactions within CTCF-CTCF loop structures in this manner may facilitate proper enhancer-promoter contacts.

FIGURE 5

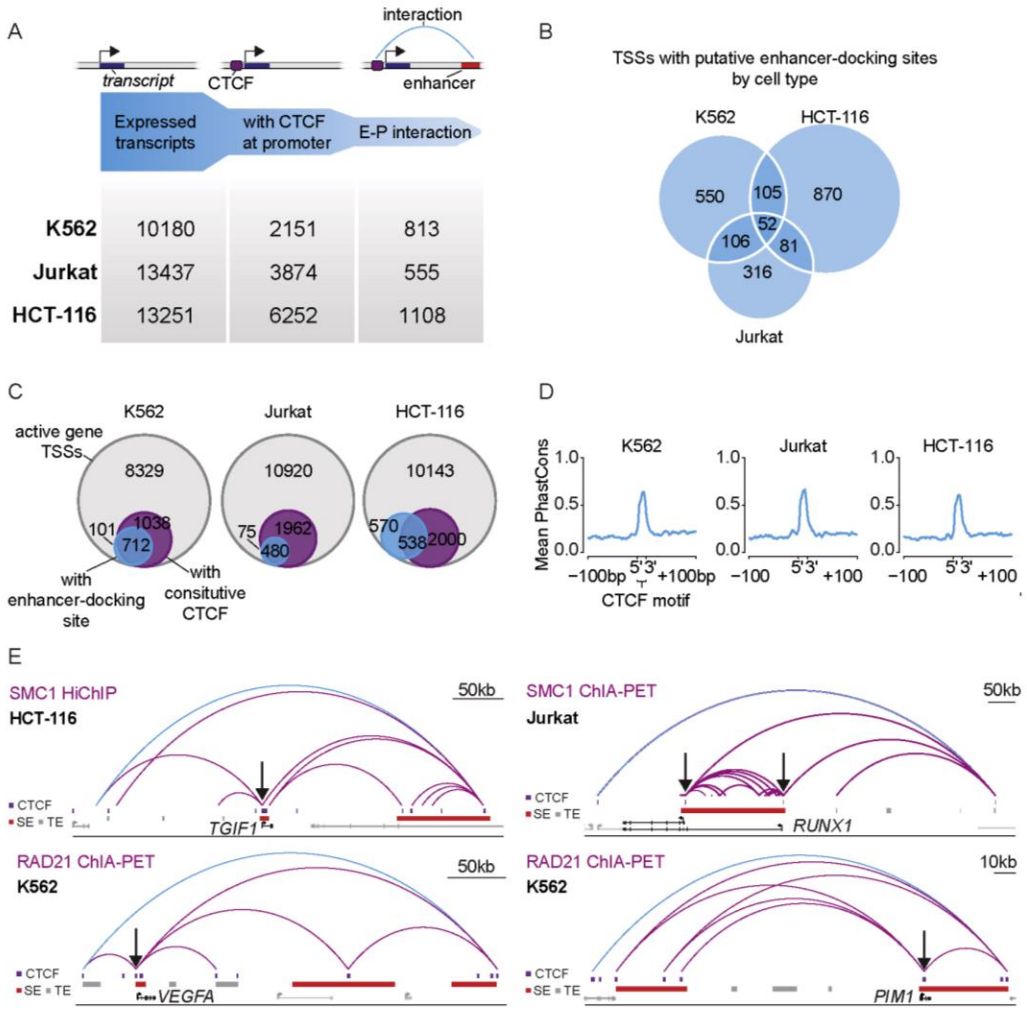


Figure 5. Putative Enhancer-Docking Sites Occur at Additional Genes with Prominent Roles in Cancer

(A) Identification of genes with putative enhancer-docking sites. Genes were filtered for their expression status, presence of a CTCF binding site within 2.5 kb of the TSS and evidence of looping to an active enhancer, defined by H3K27Ac ChIP-seq.

(B) Venn-diagram showing the overlap of TSSs with putative CTCF enhancer docking in K562, HCT-116, and Jurkat cells.

(C) Venn-diagrams showing the number of TSSs from active genes, the number of these that exhibit putative CTCF enhancer-docking and how many of these have a constitutive CTCF site within 2.5 kb of the TSS.

(D) Conservation analysis of the CTCF motifs in the CTCF-bound elements in putative enhancer-docking sites. The mean 46-way PhastCons score of the highest JASPAR scoring motifs in CTCF peaks within putative CTCF enhancer docking and their flanking regions are shown.

(E) Examples of genes with putative CTCF enhancer-docking sites from the different cell types analyzed. CTCF ChIP-seq peaks are shown as purple rectangles, typical enhancers are shown as gray rectangles, and super-enhancers are shown as red rectangles. Black arrows indicate the CTCF sites that may facilitate enhancer docking. The insulated neighborhood loop is shown in blue and loops internal to it are shown in purple. HCT-116 HiChIP interactions internal to the neighborhood with an origami score of at least 0.9 and a minimum PET count of 15 are shown for the *TGIF1* locus. Jurkat SMC1 ChIA-PET interactions internal to the neighborhood with an origami score of at least 0.97 are shown for the *RUNX1* locus. K562 RAD21 ChIA-PET interactions internal to the neighborhood with an origami score of at least 0.9 and a minimum PET count of 30 are shown for the *VEGFA* locus. K562 RAD21 ChIA-PET interactions internal to the neighborhood with an origami score of at least 0.9 and a minimum PET count of 30 are shown for the *PIM1* locus. Data are from this study and two others (Hnisz et al., 2016a; Heidari et al., 2014).

See also Figure S5.

The evidence described here argues that diverse human tumor cell super-enhancers depend on the *MYC* CTCF site for optimal levels of enhancer-promoter looping and mRNA expression. A recent independent study in K562 cells used a tiling CRISPR screen to systematically perturb the *MYC* locus and also found that full *MYC* expression and cell proliferation is dependent on this region (Fulco et al., 2016). However, deletion of the -2 kb

CTCF site has limited effects on *MYC* expression in mice (Dave et al., 2017; Gombert and Krumm, 2009), and some translocated enhancers can drive *MYC* expression in the absence of this CTCF site (Shiramizu et al., 1991). There are several potential explanations for these diverse results. It is possible that the -2 kb CTCF site is important for optimal *MYC* expression levels in human cells, but not in mice. It is conceivable that the deletion of a region containing the CTCF site can be compensated by features of the new enhancer landscape in the deletion mutations. Furthermore, additional mechanisms normally involved in enhancer-promoter interactions, such as YY1-YY1 interactions, may mask the loss of the CTCF site *in vivo*; YY1 is present in the *MYC* promoter region and is thus likely to contribute to DNA looping and expression (Weintraub et al., 2017).

Our studies suggest that an additional set of human genes, beyond *MYC*, may utilize promoter-proximal enhancer-docking sites to mediate cell-type-specific enhancer-promoter interactions. Such CTCF-mediated enhancer-promoter interactions are generally nested within larger CTCF-mediated loops that would function as insulated neighborhoods. At these genes with CTCF-mediated enhancer docking, the promoter-proximal enhancer-docking sites tend to be constitutively bound by CTCF and these binding sites tend to be highly conserved. Indeed, two studies have reported that these genes tend to lose expression upon perturbation of CTCF (Nora et al., 2017; Zuin et al., 2014), consistent with a role for CTCF in enhancer-promoter looping. Among these genes are cancer-associated genes that likely employ this mechanism to engender interactions with tumor-specific enhancers. For example, at *CSNK1A1*, a drug target in acute myeloid leukemia (AML) tumor cells (Järås et al., 2014), *VEGFA*, which is upregulated in many cancers (Goel and Mercurio, 2013), and *RUNX1*, a well-defined oncogene in AML (Deltcheva and Nimmo, 2017), the evidence suggests that super-enhancers in these cancer cells use a CTCF enhancer-docking mechanism to interact with the oncogene. Thus, a CTCF-dependent enhancer-docking mechanism, which presumably facilitates interaction with different cell-specific enhancers during development, is exploited by cancer cells to dysregulate expression of prominent oncogenes.

MYC dysregulation is a hallmark of cancer (Bradner et al., 2017). The c-MYC TF is an attractive target for cancer therapy because of the role that excessive c-MYC levels play in a broad spectrum of aggressive cancers (Felsher and Bishop, 1999; Jain et al., 2002; Soucek et al., 2008), but direct pharmacologic inhibition of c-MYC remains an elusive challenge in drug discovery (Bradner et al., 2017). The *MYC* enhancer-docking site, and presumably those of other oncogenes, can be repressed by dCas9-DNMT-mediated DNA methylation. Oncogene enhancer-docking sites may thus represent a vulnerability in multiple human cancers.

Acknowledgements
We thank X. Shawn Liu and Rudolf Jaenisch for their help with the targeted methylation of the CTCF bound element. We thank the Whitehead Institute Genome Technology Core and the FACS facility for their contribution to this work. This work is supported by NIH grant HG002668 (R.A.Y.), Rubicon fellowship by NWO (J.S.), Ludwig Graduate Fellowship funds (A.S.W.), American Cancer Society fellowship PF-16-146-01-DMC (D.S.D.), Margaret and Herman Sokol Postdoctoral Award (D.H.).

Author Contributions
JS, JM and RAY wrote the manuscript. JS and JM performed methylation, perturbation and other experiments. DH performed HiChIP. DSD analyzed HiChIP and ChIA-PET data and wrote ORIGAMI, JS, JM, and AZ performed 4C experiments. ASW made inducible CTCF-site deletion lines. TIL and RAY supervised.

Declaration of interests
The Whitehead Institute filed a patent application based on this study. R.A.Y. is a founder of Syros Pharmaceuticals, of Marauder Therapeutics and of Omega Therapeutics.

References
Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol.

Becket, E., Chopra, S., Duymich, C.E., Lin, J.J., You, J.S., Pandiyan, K., Nichols, P.W., Siegmund, K.D., Charlet, J., Weisenberger, D.J., et al. (2016). Identification of DNA methylation-independent epigenetic events underlying clear cell renal cell carcinoma. Cancer Res. *76*, 1954–1964.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Berns, E.M.J.J., Klijn, J.G.M., van Staveren, I.L., Portengen, H., and Foekens, J.A. (1992). c-myc Amplification Is a Better Prognostic Factor than HER2/neu Amplification in Primary Breast Cancer. Cancer Res. *52*, 1107–1113.

Bonev, B., Cavalli, G., Bonev Boyan, and Cavalli Giacomo (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. Cell *168*, 629–643.

Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: Lessons from genomics. Trends Genet. *28*, 276–284.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.

Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G.M., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell *24*, 777–790.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active

and repressive domains. Genome Res. *19*, 24–32.

Dang, C. V. (2012). MYC on the path to cancer. Cell *149*, 22–35.

Dave, K., Sur, I., Yan, J., Zhang, J., Kaasinen, E., Zhong, F., Blaas, L., Li, X., Kharazi, S., Gustafsson, C., et al. (2017). Mice deficient of Myc super-enhancer region reveal differential control mechanism between normal and pathological growth. Elife *6*, 1–25.

Dekker, J., and Mirny, L. (2016). The 3D genome as moderator of chromosomal communication. Cell *164*, 1110–1121.

Deltcheva, E., and Nimmo, R. (2017). RUNX transcription factors at the interface of stem cells and cancer. Biochem. J. *474*, 1755–1768.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Felsher, D.W., and Bishop, J.M. (1999). Reversible tumorigenesis by MYC in hematopoietic lineages. Mol. Cell *4*, 199–207.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V. V (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell. Biol. *16*, 2802–2813.

Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.

Fraser, J., Ferrai, C., Chiariello, A.M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B.L., Kraemer, D.C., Aitken, S., et al. (2015). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol Syst Biol *11*, 1–14.

Frietze, S., Wang, R., Yao, L., Tak, Y.G., Ye, Z., Gaddis, M., Witt, H., Farnham, P.J., and Jin, V.X. (2012). Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. Genome Biol. *13*, R52.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science *354*, 769–773.

Gabay, M., Li, Y., and Felsher, D.W. (2014). MYC activation is a hallmark of cancer initiation and maintenance. Cold Spring Harb. Perspect. Med. *4*, 1–14.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol. Cell *49*, 773–782.

Goel, H.L., and Mercurio, A.M. (2013). VEGF targets the tumour cell. Nat. Rev. Cancer *13*, 871–882.

Gombert, W.M., and Krumm, A. (2009). Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression. PLoS One *4*, 1–

8.

Gombert, W.M., Farris, S.D., Rubio, E.D., Morey-Rosler, K.M., Schubach, W.H., and Krumm, A. (2003). The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain. Mol. Cell. Biol. *23*, 9338–9348.

Gorkin, D.U., Leung, D., and Ren, B. (2014a). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell *14*, 762–775.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., De Wit, E., Bouwman, B.A.M., Erpelinck, C., Van Der Velden, V.H.J., Havermans, M., Avellino, R., Van Lom, K., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. Cell *157*, 369–381.

Grotzer, M.A., Hogarty, M.D., Janss, A.J., Liu, X., Zhao, H., Eggert, A., Sutton, L.N., Rorke, L.B., Brodeur, G.M., and Phillips, P.C. (2001). MYC messenger RNA expression predicts survival outcome in childhood primitive neuroectodermal tumor/medulloblastoma. Clin. Cancer Res. *7*, 2425–2433.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. *43*, 630–638.

Heidari, N., Phanstiel, D.H., He, C.C.C., Grubert, F., Jahanbani, F., Kasowski, M., Zhang, M.Q., Snyder, M.P., Jahanbanian, F., Kasowski, M., et al. (2014). Genome-wide map of regulatory interactions in the human genome. Genome Res. *24*, 1905–1917.

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S.A., Belver, L., Wendorff, A.A., Xu, L., Castillo-Martin, M., Llobet-Navás, D., Cordon-Cardo, C., et al. (2014). A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat. Med. *20*, 1130–1137.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016a). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454–1458.

Hnisz, D., Day, D.S., and Young, R.A. (2016b). Insulated Neighborhoods: structural and functional units of mammalian gene control. Cell *167*, 1188–1200.

Hnisz, D., Schuijers, J., Li, C.H., and Young, R.A. (2018). Regulation and Dysregulation of Chromosome Structure in Cancer. Annu. Rev. Cancer Biol. *2*, annurev-cancerbio-030617-050134.

Jain, M., Arvanitis, C., Chu, K., Dewey, W., Leonhardt, E., Trinh, M., Sundberg, C.D., Bishop, J.M., and Felsher, D.W. (2002). Sustained loss of a neoplastic phenotype by brief inactivation of MYC. Science *297*, 102–104.

Järås, M., Miller, P.G., Chu, L.P., Puram, R. V., Fink, E.C., Schneider, R.K., Al-Shahrour, F., Peña, P., Breyfogle, L.J., Hartwell, K.A., et al. (2014). Csnk1a1 inhibition has p53-dependent therapeutic efficacy in acute myeloid leukemia. J. Exp. Med. *211*, 605–612.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J.,

Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369–1384.e19.

Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., et al. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell *18*, 262–275.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D.D., Zhang, M.Q., Lobanenkov, V. V., and Ren, B. (2007). Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. Cell *128*, 1231–1245.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V. V (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol. Cell. Biol. *13*, 7612–7624.

Lee, T., Johnston, S., and Young, R. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

Lin, C.Y., Lovén, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. Cell *151*, 56–67.

Lin, C.Y., Erkek, S., Tong, Y., Yin, L., Federation, A.J., Zapatka, M., Haldipur, P., Kawauchi, D., Risch, T., Warnatz, H.-J., et al. (2016). Active medulloblastoma enhancers reveal subgroup-specific cellular origins. Nature *530*, 57–62.

Liu, M., Maurano, M.T., Wang, H., Qi, H., Song, C.-Z., Navas, P.A., Emery, D.W., Stamatoyannopoulos, J.A., and Stamatoyannopoulos, G. (2015). Genomic discovery of potent chromatin insulators for human gene therapy. Nat. Biotechnol. *33*, 198–203.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233–247.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. *12*, 1184–1195.

Montavon, T., and Duboule, D. (2012). Landscapes and archipelagos: Spatial organization of gene regulation in vertebrates. Trends Cell Biol. *22*, 347–354.

Muerdter, F., and Stark, A. (2016). Gene Regulation: Activation through Space. Curr. Biol. *26*, R895–R898.

Müller, H.P., Sogo, J., and Schaffner, W. (1989). An enhancer stimulates transcription in Trans when attached to the promoter via a protein bridge. Cell *58*, 767–777.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.

Ovcharenko, I., Loots, G.G., Nobrega, M.A., Hardison, R.C., Miller, W., and Stubbs, L. (2005). Evolution and functional classification of vertebrate gene deserts. Genome Res. *15*, 137–145.

Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. U. S. A. *110*, 17921–17926.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.-T.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc. Natl. Acad. Sci. U. S. A. *105*, 8309–8314.

Saldana-Meyer, R., Gonzalez-Buendia, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Sandelin, A. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, 91D–94.

Shi, J., Whyte, W. a, Zepeda-mendoza, C.J., Milazzo, J.P., Shen, C., Roe, J., Minder, J.L., Mercan, F., Wang, E., Eckersley-maslin, M. a, et al. (2013). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes Dev. *27*, 2648–2662.

Shiramizu, B., Barriga, F., Neequaye, J., Jafri, A., Dalla-Favera, R., Neri, A., Guttierez, M.,

Levine, P., and Magrath, I. (1991). Patterns of chromosomal breakpoint locations in Burkitt's lymphoma: relevance to geography and Epstein-Barr virus association. Blood *77*, 1516–1526.

Siddique, A.N., Nunna, S., Rajavelu, A., Zhang, Y., Jurkowska, R.Z., Reinhardt, R., Rots, M.G., Ragozin, S., Jurkowski, T.P., and Jeltsch, A. (2013). Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. J. Mol. Biol. *425*, 479–491.

Soucek, L., Whitfield, J., Martins, C.P., Finch, A.J., Murphy, D.J., Sodir, N.M., Karnezis, A.N., Swigart, L.B., Nasi, S., and Evan, G.I. (2008). Modelling Myc inhibition as a cancer therapy. Nature *455*, 679–683.

Spitz, F. (2016). Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles. Semin. Cell Dev. Biol. *57*, 57–67.

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. *20*, 2349–2354.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell *163*, 1611–1627.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K. a, et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–394.

Weintraub, A.S., Li, C.H., Zamudio, A. V, Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.

van de Werken, H.J.G., Landan, G., Holwerda, S.J.B., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Öz, Y., Bouwman, B.A.M., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods *9*, 969–972.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227–231.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell *13*, 291–298.

Zhang, X., Choi, P.S., Francis, J.M., Imielinski, M., Watanabe, H., Cherniack, A.D., and Meyerson, M. (2016). Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. Nat. Genet. *48*, 176–182.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. *111*, 996–1001.

Experimental Procedures

Further details and an outline of resources used in this work can be found in Supplemental Experimental Procedures online at https://doi.org/10.1016/j.celrep.2018.03.056.

CRISPR/Cas9 genome editing

Genome editing was performed using CRISPR/Cas9 essentially as described (Ran et al., 2013). The genomic sequences complementary to all guide RNAs are listed in Table S4

ChIP-seq

ChIP was performed as described in (Lee et al., 2006). ~30 million cells were crosslinked for 10 min at room temperature by the addition of one-tenth of the volume of 11% formaldehyde solution to the growth media followed by 5 min quenching with 125 mM glycine. Cells were washed twice with PBS, then the supernatant was aspirated and the cell pellet was flash frozen at -80C. 100µl of Protein G Dynabeads (Thermo 10003D) were blocked with 0.5%BSA (w/v) in PBS. Magnetic beads were bound with 40 µl of anti-CTCF antibody (Millipore 07-729). Nuclei were isolated as previously described (Lee et al., 2006), and sonicated in lysis buffer on a Misonix 3000 sonicator for 5 cycles at 30s each on ice (18-21 W) with 60 s on ice between cycles. Sonicated lysates were cleared once by centrifugation and incubated overnight at 4°C with magnetic beads bound with antibody to enrich for DNA fragments bound by the indicated factor. Beads were washed with wash buffers A , B , C and D sequentially. DNA was eluted, Cross-links were reversed and DNA was purified with phenol chloroform extraction and ethanol precipitation. Libraries for Illumina sequencing were prepared following the Illumina TruSeq DNA Sample Preparation v2 kit and sequenced on the Illumina HiSeq 2500 for 40 bases in single read mode.

4C-seq

A modified version of 4C-seq (van de Werken et al., 2012) was developed (Suppelentary Experimental Procdures). The major change was the ligation is performed in intact nuclei (in situ). This change was incorporated because previous work has noted that in situ ligation dramatically decreases the rate of chimeric ligations and background interactions (Rao et al., 2014).

HiChIP

HiChIP was performed essentially as described (Mumbach et al., 2016). 10 million HCT116 cells were crosslinked for 10 min at room temperature by the addition of one-tenth of the volume of 11% formaldehyde solution to the growth media followed by 5 min quenching with glycine. Cells were washed twice with PBS, then the cell pellet was flash frozen in liquid nitrogen. Frozen were processed according to protocol (Supplementary Experimental Procedures).

Targeted methylation and bisulfite sequencing.

To perform targeted methylation cells were transfected with a dCas9-DNMT3A-3L construct with or without guides. To generate the dCas9-DNMT3A-3L construct, dCas9 was isolated from pSQL1658 (Addgene: 51023) by PCR. Cas9 was removed from pX330-Cas9 (Addgene: 42230) and replaced by DNMT3A-3L (Siddique et al., 2013). Guide RNAs used for targeting can be found in table S4.

Statistical Methods

ChIP-seq data analysis

ChIP-Seq datasets were generated for this study as well as collated from previous studies (Table S5), and were aligned using Bowtie (version 0.12.2) to the human genome (build hg19, GRCh37) with parameter -k 1 -m 1 -n 2. We used the MACS version 1.4.2 with the parameter "--no-model --keep-dup=auto". A p-value threshold of enrichment of 1e-09 was used.

4C analysis

The 4C-seq reads were trimmed and mapped using bowtie with options –k 1 –m 1 against the hg19 genome assembly. We only used the reads from non-blind fragments for further analysis. The normalized profile of each sample was smoothened using a 6 kb running mean at 500 bp steps across the genome. Quantification of the 4C signal counted the reads per fragment per million sequenced reads in the super-enhancers or the CTCF MACS peak calls.

HiChIP and ChIA-PET data analysis

We developted a new software pipeline and analytical method called *origami* to process HiChIP and ChIA-PET. The software and releases can be found at https://github.com/younglab/origami using version alpha20160828. The ChIA-PET data sets analyzed along with their corresponding linker sequence and called interactions in and around the *MYC* TAD can be found in Table S4. Each ChIA-PET data sets was processed as follows: the reads were first trimmed and aligned using *origami-alignment*. Each end of a PET with a linker sequence were separately mapped to the hg19 genome assembly using bowtie with the following options: -v 1 –k 1 –m 1. After alignment, the separated PETs were re-paired in the final BAM output. After repairing, all duplicated PETs within the data were removed. Peaks were called on the re-paired ChIA-PET reads using MACS1 v1.4.2 with the following parameters: --nolambda –nomodel –p 1e-9.

Data and Software Availability

The accession number for the sequencing data reported in this paper is GEO: GSE92881. The ORIGAMI algorithm is accessible at: https://github.com/younglab/origami, version alpha20160828.

Additional Supplemental Items

Supplemental tables 1-6 can be found online at https://doi.org/10.1016/j.celrep.2018.03.056.

FIGURE S1

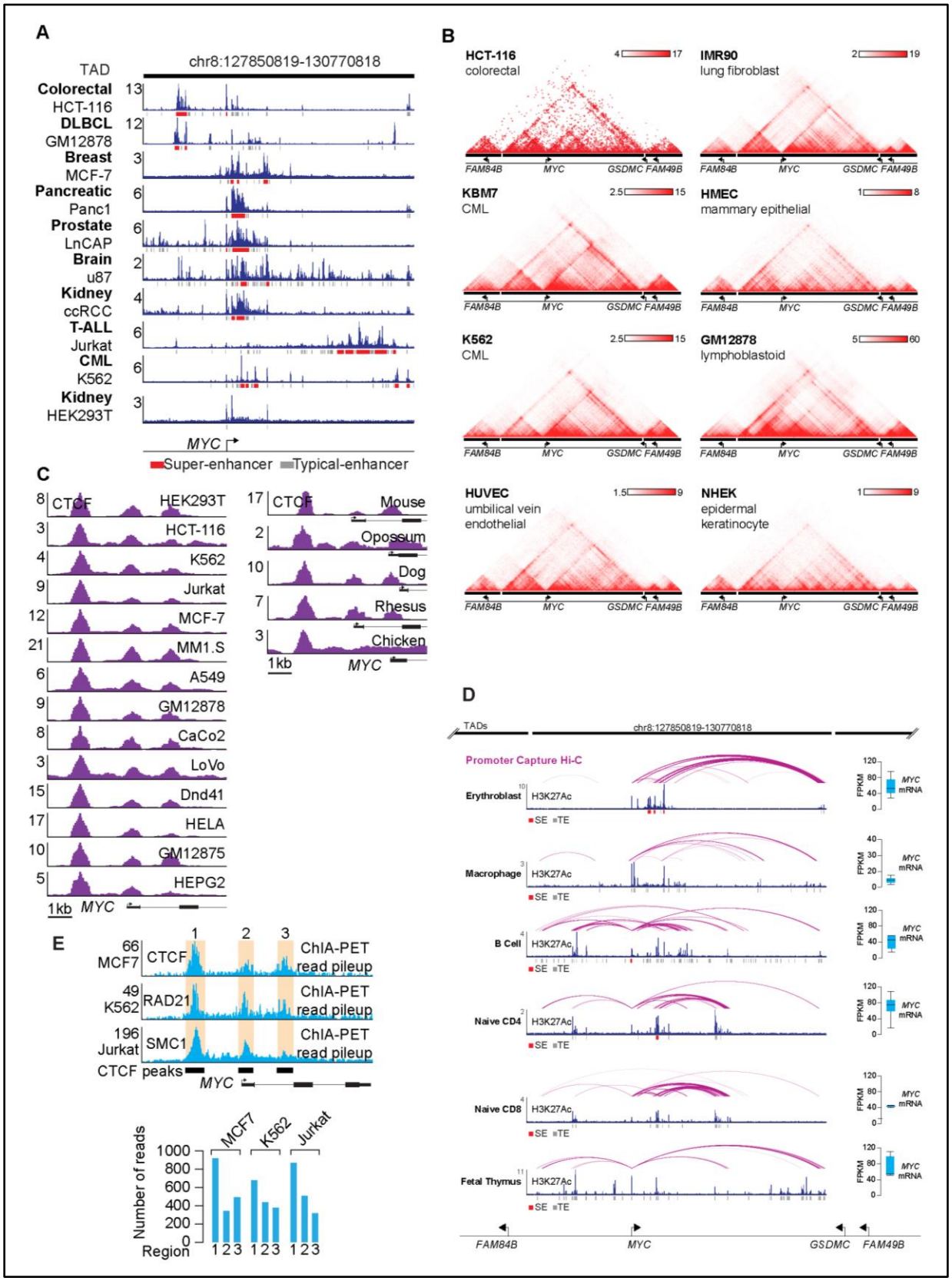

A
TAD
chr8:127850819-130770818
Colorectal
HCT-116
DLBCL
GM12878
Breast
MCF-7
Pancreatic
Panc1
Prostate
LnCAP
Brain
u87
Kidney
ccRCC
T-ALL
Jurkat
CML
K562
Kidney
HEK293T
MYC
Super-enhancer
Typical-enhancer
B
HCT-116
colorectal
IMR90
lung fibroblast
KBM7
CML
HMEC
mammary epithelial
K562
CML
GM12878
lymphoblastoid
HUVEC
umbilical vein endothelial
NHEK
epidermal keratinocyte
FAM84B
MYC
GSDMC
FAM49B
C
CTCF
HEK293T
HCT-116
K562
Jurkat
MCF-7
MM1.S
A549
GM12878
CaCo2
LoVo
Dnd41
HELA
GM12875
HEPG2
Mouse
Opossum
Dog
Rhesus
Chicken
1kb
MYC
D
TADs
chr8:127850819-130770818
Promoter Capture Hi-C
Erythroblast
Macrophage
B Cell
Naive CD4
Naive CD8
Fetal Thymus
H3K27Ac
SE
TE
FPKM
MYC mRNA
E
MCF7
CTCF
K562
RAD21
Jurkat
SMC1
ChIA-PET read pileup
CTCF peaks
MYC
Number of reads
Region

Figure S1. Cell type-specific super-enhancers in the *MYC* locus loop to a common CTCF site. Related to Figure 1.

(**A**) The 2.8 Mb TAD containing the *MYC* gene is indicated with thick black horizontal lines. H3K27Ac ChIP-seq signal (reads per million sequenced reads per base pair, data from this manuscript, (Becket et al., 2016; Frietze et al., 2012; Lin et al., 2012; Pope et al., 2014; Wang et al., 2011)) is shown in dark blue for a panel of tumor cell lines that express MYC. Tumor super-enhancers in the *MYC* TAD are depicted as red rectangles and typical enhancers are depicted as grey rectangles. (**B**) Heatmap of the ORIGAMI processed HiChIP, unfiltered data showing the *MYC* TAD with flanking regions (chr8:127100000-131525000) and Heatmaps of Hi-C interaction data showing the *MYC* TAD with flanking regions (chr8:127100000-131525000) across seven different cell types (data from (Rao et al., 2014)). The HiChIP was not smoothened as opposed to the smoothened HiC data. Scale bars represent the contrast settings used, numbers indicate the maximum intensity cutoff. The color intensity represents the PET count and the cutoff is represented in PET numbers for the HiChIP data. (**C**) CTCF ChIP-seq across a panel of tumor cell lines (data from (Anders et al., 2014; Encode Consortium, 2012; Hnisz et al., 2016; Pope et al., 2014; Wang et al., 2012; Yan et al., 2013)), and from mouse T-helper cells, Opossum, Dog, and Rhesus macaque liver (data from (Schmidt et al., 2012; Stadler et al., 2011)). Read counts are shown in reads per million sequenced reads per base pair.

(**D**) Promoter Hi-C interaction data and H3K27Ac ChIP-seq at the *MYC* TAD for cell types that represent different stages in hematopoietic development. The 2.8 Mb TAD containing *MYC* and part of the two adjacent TADs are indicated with thick black horizontal lines. Promoter Hi-C interactions are shown as purple colored arcs; the intensity of purple color reflects the confidence score from (Javierre et al., 2016). H3K27Ac ChIP-seq signal is shown, measured in reads per million sequenced reads per base pair (data from(Bernstein et al., 2010; Encode Consortium, 2012; Schmidt et al., 2016; Xu et al., 2012)). Super-enhancers are depicted as red rectangles and typical enhancers as grey rectangles. The relative level of *MYC* transcripts in the corresponding cell types are shown as boxplots in fragments per kilobase of exon per million sequenced reads (FPKM), expression data from the BLUEPRINT consortium, fetal thymus expression data from the ENCODE consortium. (**E**) ChIA-PET read pileups at the *MYC* promoter and quantification of the reads in the three CTCF peaks indicated. Light blue tracks display the read counts from read pileups of MCF7 CTCF, K562 RAD21 and Jurkat SMC1 ChIA-PET data showing that the majority of the ChIA-PET reads are found at the enhancer-docking site. Reads are shown as read counts per base pair.

FIGURE S2

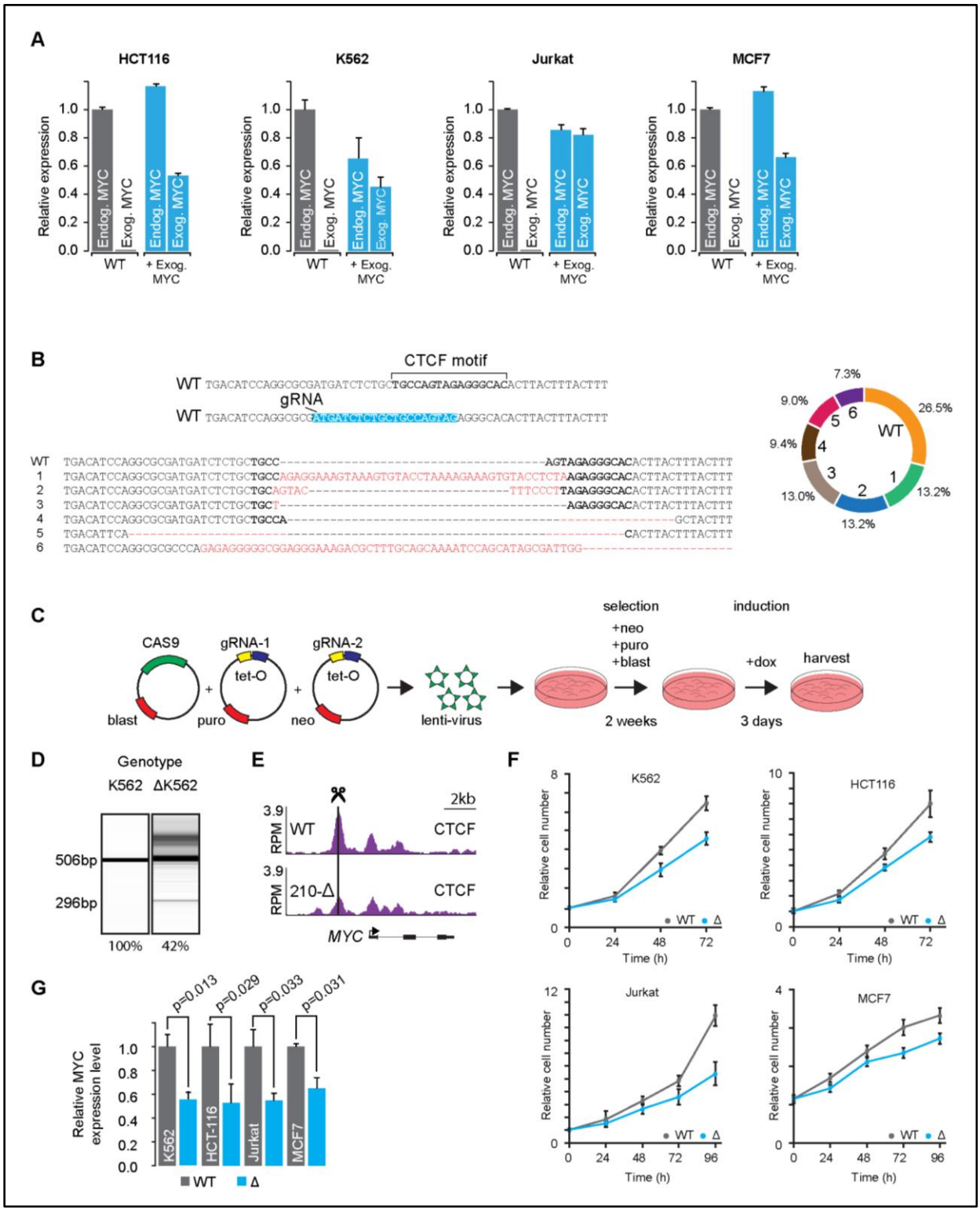

A
HCT116
K562
Jurkat
MCF7
Relative expression
Endog. MYC
Exog. MYC
WT
+ Exog. MYC
B
CTCF motif
gRNA
WT
7.3%
9.0%
26.5%
9.4%
13.0%
13.2%
13.2%
C
CAS9
gRNA-1
gRNA-2
tet-O
blast
puro
neo
lenti-virus
selection
+neo
+puro
+blast
2 weeks
induction
+dox
3 days
harvest
D
Genotype
K562 ΔK562
506bp
296bp
100%
42%
E
2kb
RPM
WT
210-Δ
CTCF
MYC
F
K562
HCT116
Jurkat
MCF7
Relative cell number
Time (h)
WT
Δ
G
Relative MYC expression level
p=0.013
p=0.029
p=0.033
p=0.031
K562
HCT-116
Jurkat
MCF7
WT
Δ

Figure S2. Perturbation of the core CTCF motif in the *MYC* CTCF loop-anchor reduces CTCF occupancy and MYC expression. Related to Figure 2.

(**A**) qPCR measuring the mRNA levels of endogenous (endog.) and exogenous (exog.) *MYC* in parental (wild type) and exogenous *MYC* expressing HCT-116, K562, MCF7 and Jurkat cells. Endogenous and exogenous *MYC* were detected using primers directed against the 3' UTR of the *MYC* mRNA and the *MYC*-tdTomato junction respectively. (**B**) Sequencing of mutant alleles in the selected MCF7 clone with mutated enhancer-docking site. The CRISPR targeted region was amplified, fragmented and sequenced to identify the composition and frequency of mutant alleles. The 6 most common mutant alleles are displayed. (**C**) Perturbation of *MYC* promoter proximal CTCF site reduces *MYC* expression and proliferation rate across cancers. Schematic representation of the experiment. Cells were transduced with one virus carrying Cas9 and two viruses each carrying one guide RNA (gRNA) under a doxycycline inducible promoter. After selection for all three components, cells were induced with doxycycline for 3 days prior to harvest and testing. (**D**) Heat map of fragment lengths after genotyping PCR of wild type K562 and ΔK562 cells. PCR product was analyzed with a Fragment Analyzer and fragments of different lengths were quantified. The percentage of fragments with a length expected from wild type cells (506) is indicated under the lanes. The expected length of the PCR product for deleted and recombined alleles is 296bp. (**E**) ChIP-seq of CTCF in induced or uninduced K562 cells targeting the 210bp fragment containing the -2kb CTCF site. The -2kb CTCF binding is reduced, while the other CTCF binding sites in the *MYC* locus are unaffected. (F) Proliferation of parental (grey) and CTCF site deleted (blue) K562, HCT-116, Jurkat and MCF7 cells. Error bars represent the standard deviation of the mean from six biological replicates. (G) qPCR showing the *MYC* mRNA levels after deletion of the CTCF site in K562, HCT-116, Jurkat and MCF7 cells. p-values were generated with a Students T-test. Error bars represent the standard deviation of the mean from three biological replicates.

FIGURE S3

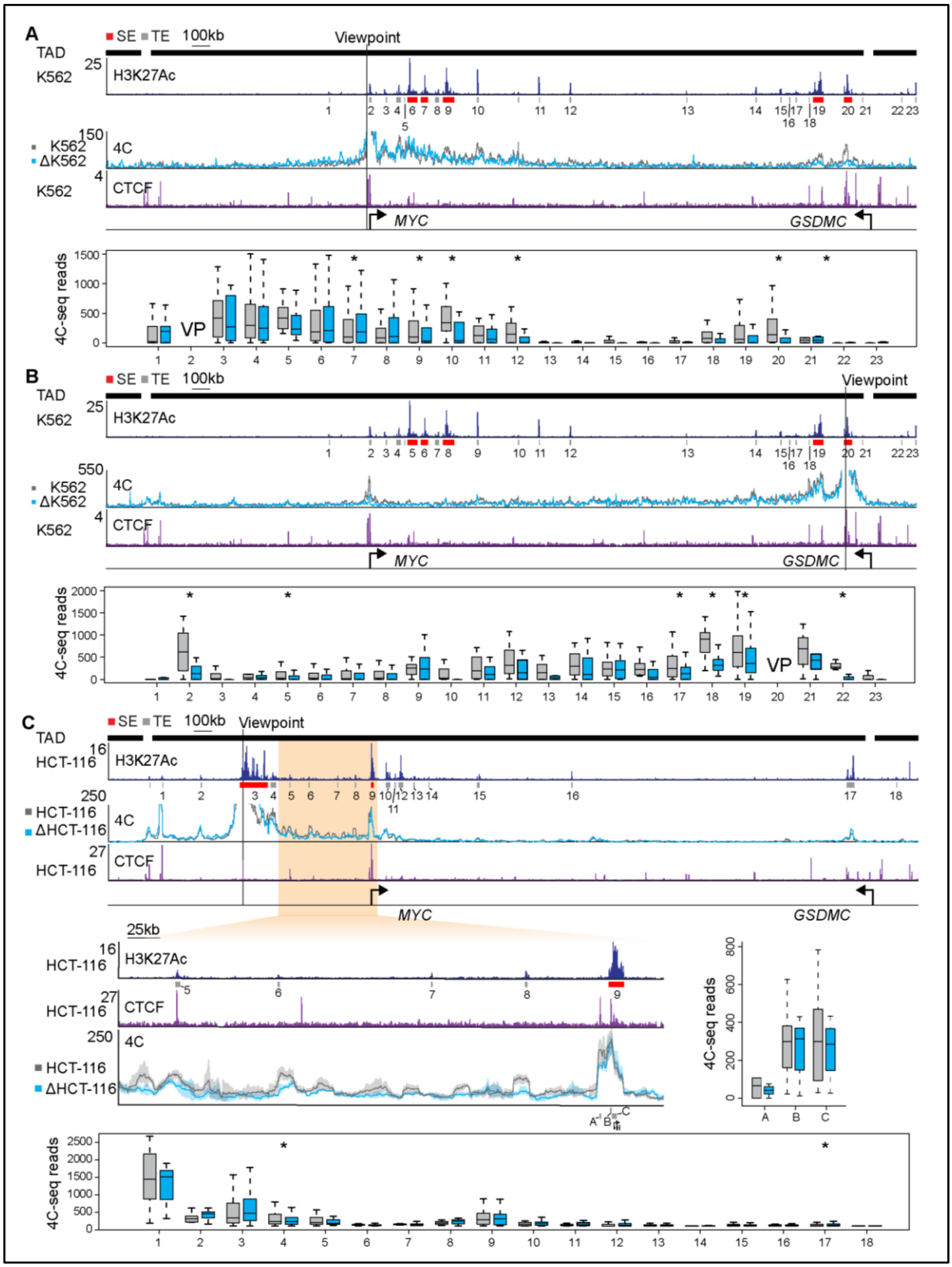

A
SE TE 100kb
Viewpoint
TAD
K562
H3K27Ac
K562
ΔK562
4C
K562
CTCF
MYC
GSDMC
4C-seq reads
VP
B
SE TE 100kb
Viewpoint
TAD
K562
H3K27Ac
K562
ΔK562
4C
K562
CTCF
MYC
GSDMC
4C-seq reads
VP
C
SE TE 100kb
Viewpoint
TAD
HCT-116
H3K27Ac
HCT-116
ΔHCT-116
4C
HCT-116
CTCF
MYC
GSDMC
25kb
HCT-116
H3K27Ac
HCT-116
CTCF
HCT-116
ΔHCT-116
4C
4C-seq reads
4C-seq reads

Figure S3. Perturbation of the *MYC* enhancer-docking site reduces looping to super-enhancers. Related to figure 3.

(**A**) 4C analysis showing quantification of regions based on H3K27Ac ChIP-seq peak calls for the 4C experiment with the viewpoint (VP) at the -2kb upstream CTCF site in K562 cells. (**B**) 4C analysis showing quantification of regions based on H3K27Ac ChIP-seq peak calls for the 4C experiment with the viewpoint (VP) at downstream super-enhancer in K562 cells. (**C**) 4C analysis showing quantification of regions based on H3K27Ac or CTCF ChIP-seq peak calls for the 4C experiment with the viewpoint (VP) at upstream super-enhancer in HCT-116 cells. Blowup shows the 4C interactions for the intervening region between the super-enhancer and the *MYC* gene. H3K27Ac ChIP-seq and CTCF ChIP-seq are shown in blue and purple, respectively. Shading represents the 90% confidence interval based on three biological replicates. Grey bars indicate the regions that are quantified. Box plots show quantification of the reads per fragment for the indicated regions. p-values were generated using Student's t-test and data pairs with a p-value of < 0.05 are indicated with an asterisk. Reads are shown in reads per million sequenced reads per base pair. Typical-enhancers and super-enhancers are shown as grey boxes and red boxes respectively.

FIGURE S4

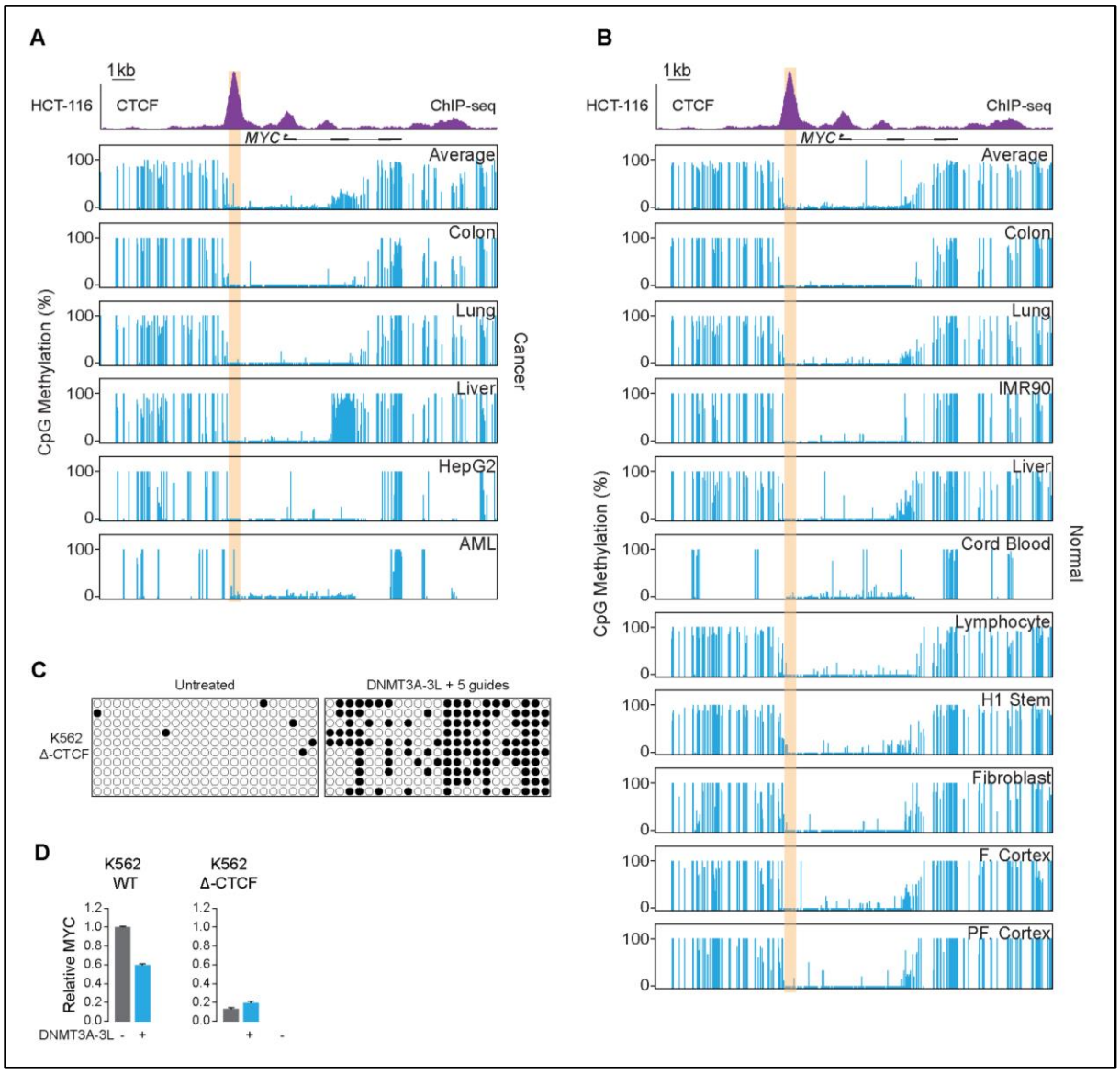

A
1kb
HCT-116
CTCF
ChIP-seq
MYC
Average
Colon
Lung
Liver
HepG2
AML
CpG Methylation (%)
Cancer
B
1kb
HCT-116
CTCF
ChIP-seq
MYC
Average
Colon
Lung
IMR90
Liver
Cord Blood
Lymphocyte
H1 Stem
Fibroblast
F. Cortex
PF. Cortex
CpG Methylation (%)
Normal
C
Untreated
DNMT3A-3L + 5 guides
K562
Δ-CTCF
D
K562
WT
K562
Δ-CTCF
Relative MYC
DNMT3A-3L

Figure S4. The *MYC* CTCF loop-anchor site is hypomethylated in a swathe of cancer and normal cells. Loss of *MYC* expression upon docking site methylation is dependent on presence of CTCF. Related to figure 4.

(**A**) Percentage of methylation of CpG's at the *MYC* locus in cancer cells. Percent methylation of each CpG in the region for which data was available for is represented as a blue line. HCT-116 CTCF ChIP-seq signal is shown in purple with the *MYC* promoter proximal CTCF site highlighted in yellow for reference. ChIP-seq read counts are shown in reads per million sequenced reads per base pair. (**B**) Percentage of methylation of CpG's at the *MYC* locus in normal cells. Percent methylation of each CpG in the region for which data was available for is represented as a blue line. Whole genome bisulfite sequencing data from ENCODE, (Barabé et al., 2016; Ziller et al., 2013).

(**C**) Methylation at *MYC* promoter loop-anchor site in untreated and dCas9-DNMT3A-3L + 5 gRNA transfected K562 ∆-CTCF cells. (**D**) qPCR analysis of *MYC* mRNA levels for untreated and dCas9-DNMT3A-3L + 5 gRNA transfected K562 WT cells (data from Figure 4D) and K562 ∆-CTCF cells. Error bars represent the standard deviation of the mean for three biological replicates.

FIGURE S5

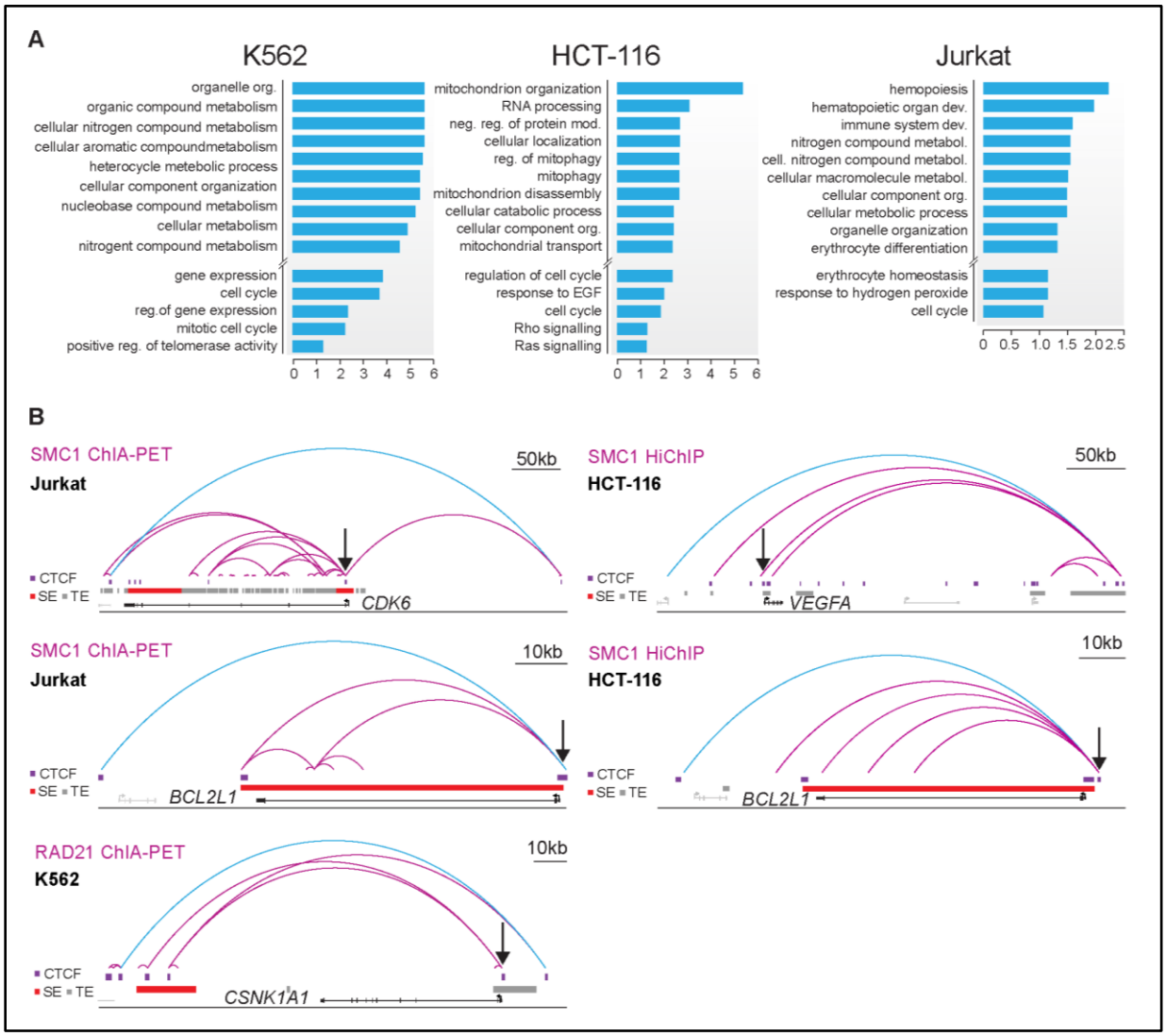


Figure S5. Putative enhancer-docking sites occur at additional genes with prominent roles in cancer. Related to figure 5.

(**A**) Gene ontology analysis of the genes with putative CTCF enhancer-docking in K562, HCT-116 and Jurkat cells. Blue bars indicate the –log10 of the q-value associated with the enrichment of the GO term indicated. The top ten most enriched GO terms are shown in addition to selected cancer associated GO terms that are significantly enriched (q-value < 0.1). (**B**) Displays of additional cancer associated genes with putative CTCF enhancer-docking. Black arrows indicate the CTCF sites that may facilitate enhancer-docking. CTCF ChIP-seq peaks are shown as purple rectangles, typical enhancers are shown as grey rectangles and super-enhancers are shown as red rectangles. The insulated neighborhood loop is shown in blue and loops internal to it are shown in purple. ChIA-PET data used is indicated in purple lettering.

CHAPTER 3: POL II PHOSPHORYLATION REGULATES A SWITCH BETWEEN TRANSCRIPTIONAL AND SPLICING CONDENSATES

Originally published in *Nature* Volume 572, Issue 7770. (2019)
Reprinted with permission from Nature Publishing Group.

Yang Eric Guo[1,10], John C. Manteiga[1,2,10], Jonathan E. Henninger[1], Benjamin R. Sabari[1], Alessandra Dall'Agnese[1], Nancy M. Hannett[1], Jan-Hendrik Spille[3,8], Lena K. Afeyan[1,2], Alicia V. Zamudio[1,2], Krishna Shrinivas[4,5], Brian J. Abraham[1,9], Ann Boija[1], Tim-Michael Decker[6], Jenna K. Rimel[6], Charli B. Fant[6], Tong Ihn Lee[1], Ibrahim I. Cisse[3], Phillip A. Sharp[2,7], Dylan J. Taatjes[6], and Richard A. Young[1,2]*

[1]Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142, USA
[2]Department of Biology
[3]Department of Physics
[4]Department of Chemical Engineering
[5]Institute of Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[6]Department of Biochemistry, University of Colorado, Boulder, CO 80303, USA
[7]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[8]Present address: Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA
[9]Present address: Computational Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
[10]These authors contributed equally to this work
*Correspondence to: Richard A. Young (young@wi.mit.edu)

The synthesis of pre-mRNA by RNA polymerase II (Pol II) involves the formation of a transcription initiation complex and a transition to an elongation complex[1-4]. The large subunit of Pol II contains an intrinsically disordered C-terminal domain (CTD), which is phosphorylated by cyclin-dependent kinases (CDKs) during the initiation-to-elongation transition, thus influencing the CTD's interaction with different components of the initiation or the RNA splicing apparatus (Fig. 1a)[5,6]. Recent observations suggest that this model provides only a partial picture of the effects of CTD phosphorylation. Both the transcription initiation machinery and the splicing machinery can form phase-separated condensates containing large numbers of component molecules; hundreds of Pol II and Mediator molecules are concentrated in condensates at super-enhancers[7,8] and large numbers of splicing factors are concentrated in nuclear speckles, some of which occur at highly active transcription sites[9-12]. Here we investigate whether phosphorylation of the CTD regulates its incorporation into phase-separated condensates associated with transcription initiation and splicing. We find that the hypophosphorylated Pol II CTD is incorporated into Mediator condensates and that phosphorylation by regulatory CDKs reduces this incorporation. We also find that the hyperphosphorylated CTD is preferentially incorporated into condensates formed by splicing factors. These results suggest that Pol II CTD phosphorylation drives an exchange from condensates involved in transcription initiation to those involved in RNA processing and implicates phosphorylation as a mechanism to regulate condensate preference.

Previous studies have shown that the hypophosphorylated Pol II CTD can interact with Mediator[5,6] and that Pol II and Mediator occur in condensates at super-enhancers[7,8] (Fig. 1b, Extended Data Fig. 1a), but have not established whether the CTD contributes to Pol II interactions with Mediator condensates. To investigate whether the Pol II CTD is incorporated into Mediator condensates, we purified the human Mediator complex and recombinant CTD fused to GFP (full length GFP-CTD52 and truncated forms GFP-CTD26 and GFP-CTD10) and measured condensate formation in an in vitro droplet assay. Mediator droplets incorporated and concentrated GFP-CTD52 to a much greater extent than the truncated forms or control GFP (Fig. 1c). We further investigated the interaction of the CTD with Mediator using MED1, the largest subunit of the Mediator complex[13]. MED1 has proven to be a useful surrogate for the Mediator condensate in previous studies[8,14], has an exceptionally large intrinsically disordered region (IDR) that contributes to condensate formation[8], and has been shown to associate with Pol II in human cells[15]. Droplet assays revealed that mCherry-MED1-IDR condensates incorporated and concentrated GFP-CTD52 to a greater extent than its truncated forms or GFP alone (Fig. 1d). The GFP-CTD52/MED1-IDR condensates exhibited liquid-like fusion behavior (Extended Data Fig. 1b) and showed evidence of dynamic internal rearrangement and internal-external exchange of molecules by fluorescence recovery after photobleaching (FRAP; Extended Data Fig. 1c), consistent with liquid-liquid phase-separated condensates[16-18]. These results are consistent with the idea that the Pol II CTD contributes to Pol II incorporation into Mediator condensates.

FIGURE 1

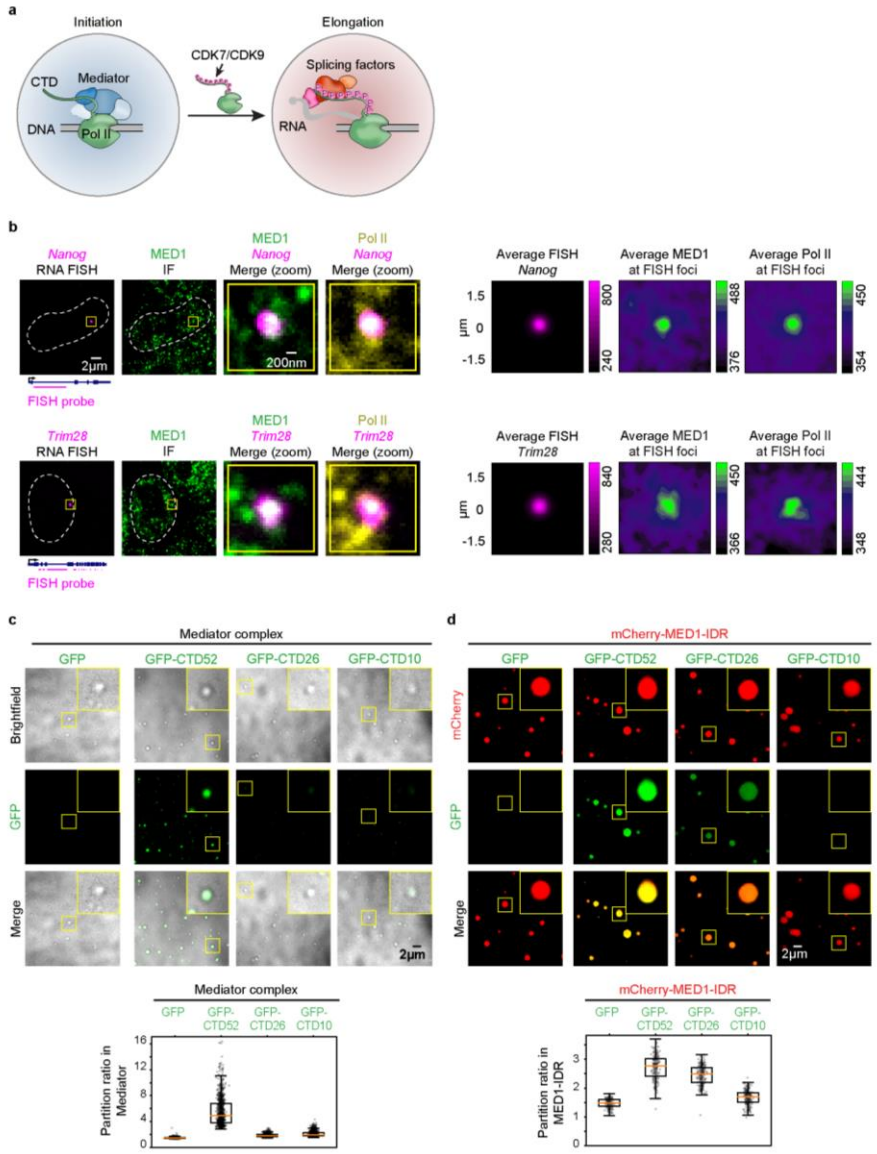


Figure 1. The CTD of Pol II is integrated and concentrated in Mediator condensates

a. A model depicting the role of Pol II CTD phosphorylation in the transition from transcription initiation to elongation. This multistep process[1] is depicted in simplified form here, with CDK7/CDK9 phosphorylating the CTD, leading to elongation. During elongation, Pol II with a hyperphosphorylated CTD interacts with various splicing factors[5,6]. The colored compartments surrounding the initiating and elongating polymerases represent initiation and splicing factor condensates, respectively.

b. Representative images exhibiting overlap between immunofluorescence (IF) of MED1 and Pol II with nascent RNA FISH of *Nanog* and *Trim28* in fixed mouse embryonic stem cells (mESCs). The three columns on the right show average RNA FISH signal and average MED1 or Pol II IF signal centered on RNA FISH foci (see methods).

c. Representative images and quantification of partition ratios of droplet experiments measuring full length or truncated CTD incorporation into human Mediator complex droplets. Purified human Mediator complex (~200 nM; see methods) was mixed with 10 μM GFP, GFP-CTD52, or truncated forms of GFP-CTD in droplet formation buffers with 140 mM monovalent salt and 16% Ficoll-400 and visualized on a fluorescence microscope with the indicated filters.

d. Representative images and quantification of partition ratios of droplet experiments measuring full length or truncated CTD incorporation into MED1-IDR droplets. Purified human MED1-IDR fused to mCherry (mCherry-MED1-IDR) at 10 μM was mixed with 3.3 μM GFP, GFP-CTD52, or truncated forms of GFP-CTD in droplet formation buffers with 125 mM NaCl and 16% Ficoll-400.

We next sought to determine whether splice factor condensates occur at super-enhancer-associated genes because these genes are transcribed at especially high rates[19], RNA splicing can occur co-transcriptionally[20-22], and some nuclear speckles have been reported to occur in the vicinity of highly transcribed genes[9-12]. We selected eight different components of the splicing apparatus and used immunofluorescence (IF) microscopy with concurrent nascent RNA fluorescence in situ hybridization (FISH) for *Nanog* and *Trim28* to determine whether the splicing apparatus occurs in puncta in the vicinity of these super-enhancer-associated genes. The results showed that all eight splicing factors occur in puncta at these two genes (Fig. 2a, Extended Data Fig. 2a). To gain additional insights into splice factor puncta that colocalize with

Pol II, mESCs engineered to express endogenously tagged proteins were studied using lattice light sheet imaging in live cells. We previously showed that large numbers of Mediator and Pol II molecules can occur in puncta and that these sometimes colocalize[7,8]; using a similar approach, we found here that large numbers of SRSF2 molecules occur in puncta and some of these (~15%) overlap with Pol II puncta (Fig. 2b, Extended Data Fig. 2b). Treatment of cells with an inhibitor of splicing, Pladienolide B, which reduced splicing as determined by a splicing reporter (Extended Data Fig. 3a), also reduced the levels of splicing factors, but not RNA polymerase II, in puncta at *Trim28* DNA FISH foci (Fig. 2c, Extended Data Fig. 3b, c). This treatment also led to splicing factor incorporation into “mega-speckles” at some distance from the gene (Fig. 2c), a phenomenon observed previously when splicing is inhibited[23]. These results suggest that functional RNA splicing apparatus is present in condensates at active super-enhancer-associated genes.

Actively transcribed genes may become associated with nuclear speckles or obtain splicing apparatus stored in speckles[9-12], which are thought to be phase-separated[16]. In live cell imaging, we found that the SRSF2 puncta exhibited features of liquid-like condensates: they all showed evidence of dynamic internal rearrangement and internal-external exchange of molecules by FRAP (Extended Data Fig. 4a-c), were sensitive to treatment with 1,6-hexanediol (Extended Data Fig. 4d), and some would occasionally fuse (Extended Data Fig. 4e). These results are consistent with previous reports regarding speckle behavior[9] and suggest that the SRSF2-containing puncta that come into contact with active super-enhancer-associated genes are liquid-like condensates. At highly transcribed genes such as those driven by super-enhancers, large numbers of Pol II molecules may be engaged in transcription elongation[7,8,17], and these might serve to recruit into condensates some portion of the apparatus otherwise located in speckles.

We next investigated whether hypophosphorylated Pol II tends to be associated with MED1 condensates whereas hyperphosphorylated Pol II tends to be associated with SRSF2 condensates. Using immunofluorescence and antibodies against hypophosphorylated or serine 2 phosphorylated CTD, we confirmed this prediction: MED1 puncta more frequently overlapped with signals for the hypophosphorylated CTD whereas SRSF2 puncta more frequently overlapped with signals for the serine 2 phosphorylated CTD (Fig. 3a). A control experiment showed that there was essentially no overlap between SRSF2 puncta and the heterochromatin protein HP1a, and strong overlap of SRSF2 puncta visualized using independent methods (Fig. 3a). An independent experimental approach using ChIP-seq with antibodies against MED1, SRSF2, and the two phosphoforms of Pol II also confirmed that MED1 tends to occupy super-enhancers and promoters together with Pol II containing hypophosphorylated CTD, whereas SRSF2 is observed across the transcription unit and is prominent at the ends of genes together with Pol II containing serine 2 phosphorylated CTD (Fig. 3b).

FIGURE 2

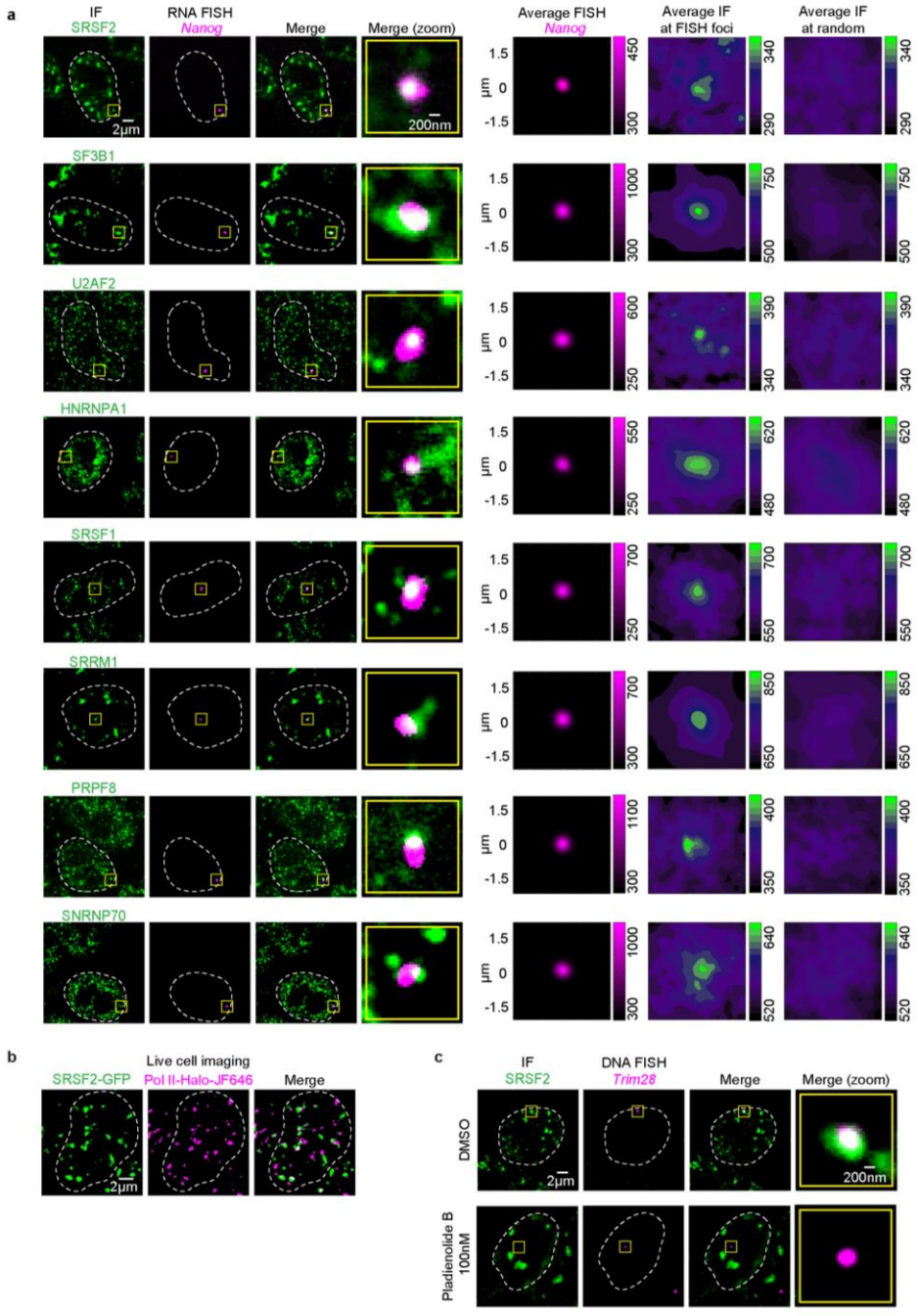


Figure 2. Splicing factor condensates occur at active super-enhancer-driven genes

a. Representative images exhibiting overlap between IF of splicing factors SRSF2, SF3B1, U2AF2, HNRNPA1, SRSF1, SRRM1, PRPF8, or SNRNP70 with nascent RNA FISH of *Nanog* in fixed mESCs. The rightmost column shows average IF signal for splicing factors centered on randomly selected nuclear positions (see methods).

b. Representative lattice light sheet images of live mESCs engineered to express GFP tagged SRSF2 and Halo-JF646 tagged Pol II. Maximum intensity projection after background subtraction.

c. Representative images exhibiting overlap or absence of overlap between IF of SRSF2 and DNA FISH of *Trim28* in mESCs treated with DMSO or splicing inhibitor Pladienolide B for 4 hrs.

FIGURE 3

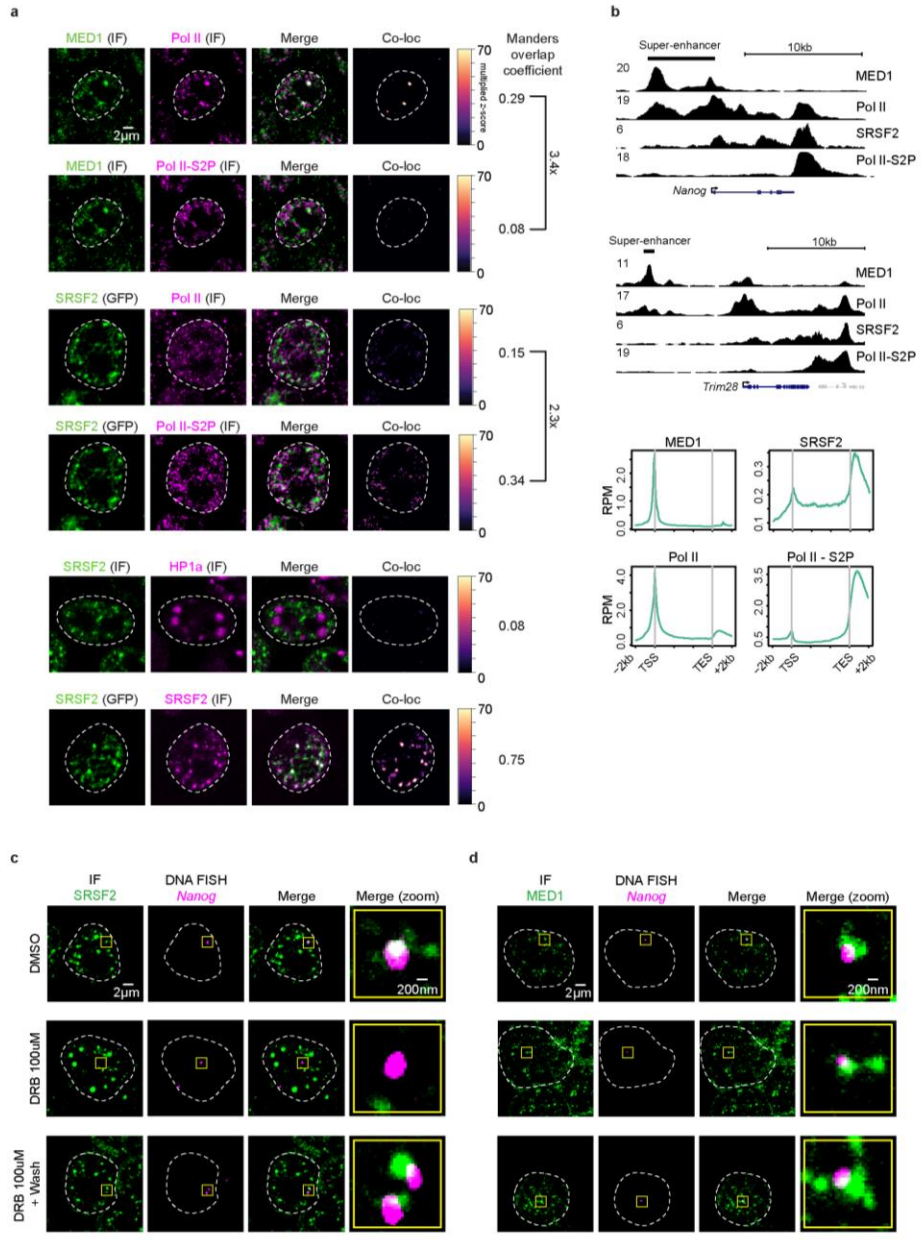


Figure 3. Pol II partitioning in transcriptional and splicing factor condensates

a. First four rows: IF imaging using antibodies for the hypophosphorylated and serine 2 phosphorylated (S2P) Pol II CTD, coupled with IF for MED1 or direct visualization of SRSF2 in the GFP-SRSF2 mESCs. Last two rows: IF for SRSF2 coupled to IF for HP1a or direct visualization of SRSF2 in the GFP-SRSF2 mESCs. The "Co-loc" column highlights overlapped pixels for each factor in an example z-slice, and the Manders' overlap coefficient gives a relative score for the degree of overlap from multiple cells and images (see methods). For each experimental comparison, one coverslip of cells was stained for the indicated factors and 5 independent fields were imaged and analyzed.

b. Top: Representative ChIP-seq tracks of MED1, SRSF2 and the hypophosphorylated or serine 2 phosphorylated forms of Pol II in mESCs. Y-axis in reads per million (RPM). Bottom: Metagene plots of average ChIP-seq RPM for the same factors across gene bodies (see methods). ChIP-seq was performed once for each factor with approximately 100 million cells.

c. Representative images exhibiting overlap or lack of overlap between IF of SRSF2 and DNA FISH of *Nanog* in mESCs treated with DMSO for 2 hrs, DRB for 2 hrs, or DRB for 2 hrs followed by a 2 hr washout.

d. Representative images exhibiting overlap between IF of MED1 and DNA FISH of *Nanog* in mESCs treated as in panel c.

If the formation or maintenance of splicing factor condensates is dependent on CTD phosphorylation, we would expect that inhibition of CTD phosphorylation in cells would prevent the formation of splicing factor condensates at super-enhancer-driven genes. Indeed, inhibition of CTD phosphorylation by the CDK inhibitor D-ribofuranosylbenzimidazole (DRB) caused a marked reduction in the occupancy of multiple splicing factor condensate components (SRSF2, SRSF1, SF3B1, U2AF2, PRPF8) at *Nanog* or *Trim28* DNA FISH foci, and a washout of the drug led to a partial reestablishment of most of these splicing factors within 2 hours (Fig. 3c, Extended Data Fig. 5a, d). In contrast, DRB treatment had minor effects on Mediator and Pol II condensates (Fig. 3d, Extended Data Fig. 5b-d). These DRB treatment and washout results suggest that CTD phosphorylation is necessary for formation of splicing factor condensates at these genes in vivo, although it is possible that altered phosphorylation of other CDK substrates may contribute to these observations.

The transition of Pol II from initiation to elongation is accompanied by phosphorylation of the CTD heptapeptide repeat by CDK7 and CDK9[24,25]. Phosphorylation of the CTD has been shown to affect its interaction with hydrogels formed by the low-complexity domains of FET (FUS/EWS/TAF15) proteins[26], suggesting that phosphorylation may affect the condensate interacting properties of the CTD. We investigated whether phosphorylation of the CTD by CDK7 or CDK9 would affect its incorporation into Mediator condensates. We found that CTD phosphorylation by either CDK7 or CDK9 (Extended Data Fig. 6a, b) caused a reduction in CTD incorporation into Mediator droplets (Fig. 4a). Similarly, CTD phosphorylation caused a reduction in CTD incorporation into MED1-IDR droplets (Extended Data Fig. 6c, d). These results are consistent with the model that Pol II CTD phosphorylation causes eviction from a Mediator condensate.

The observation that CTD phosphorylation is necessary for splicing factor condensate formation at highly transcribed genes (Fig. 3c) suggests that CTD phosphorylation might enhance its partitioning into condensates formed by splicing components. To investigate this idea, we first selected four human splicing factors (SRSF2, SRSF1, U2AF2 and hnRNPA1) as surrogates for the more complex splicing factor condensates and explored their condensate forming properties. Each of the four purified human proteins, fused to mCherry, formed phase-separated droplets (Extended Data Fig. 7a, b). SRSF2 is among the multiple proteins involved in pre-mRNA splicing that contain serine:arginine (SR) dipeptide repeats and has an especially large SR-rich domain[27], so we used SRSF2 as a core component to study whether it could concentrate the other three factors into heterotypic droplets. Indeed, all of these factors could form binary heterotypic droplets with SRSF2 (Extended Data Fig. 7c). We then asked whether CTD phosphorylation influences its incorporation into splicing factor condensates in vitro using recombinant SRSF1 and SRSF2. The results showed that unphosphorylated CTD was not efficiently incorporated into SRSF1 or SRSF2 droplets, whereas CDK7 or CDK9-phosphorylated CTD was incorporated and concentrated in both SRSF1 and SRSF2 droplets (Fig. 4b, c, Extended Data Fig. 8a-c). The ability of SRSF2 to incorporate phosphorylated CTD was dependent on CTD length (Extended Data Fig. 8d), as expected for a high valency condensate interaction[16-18] and consistent with models where CTD truncation leads to splicing defects[28]. We conclude that phosphorylation of the Pol II CTD leads to a switch in its preference for interactions between Mediator and SR-protein-containing condensates.

FIGURE 4

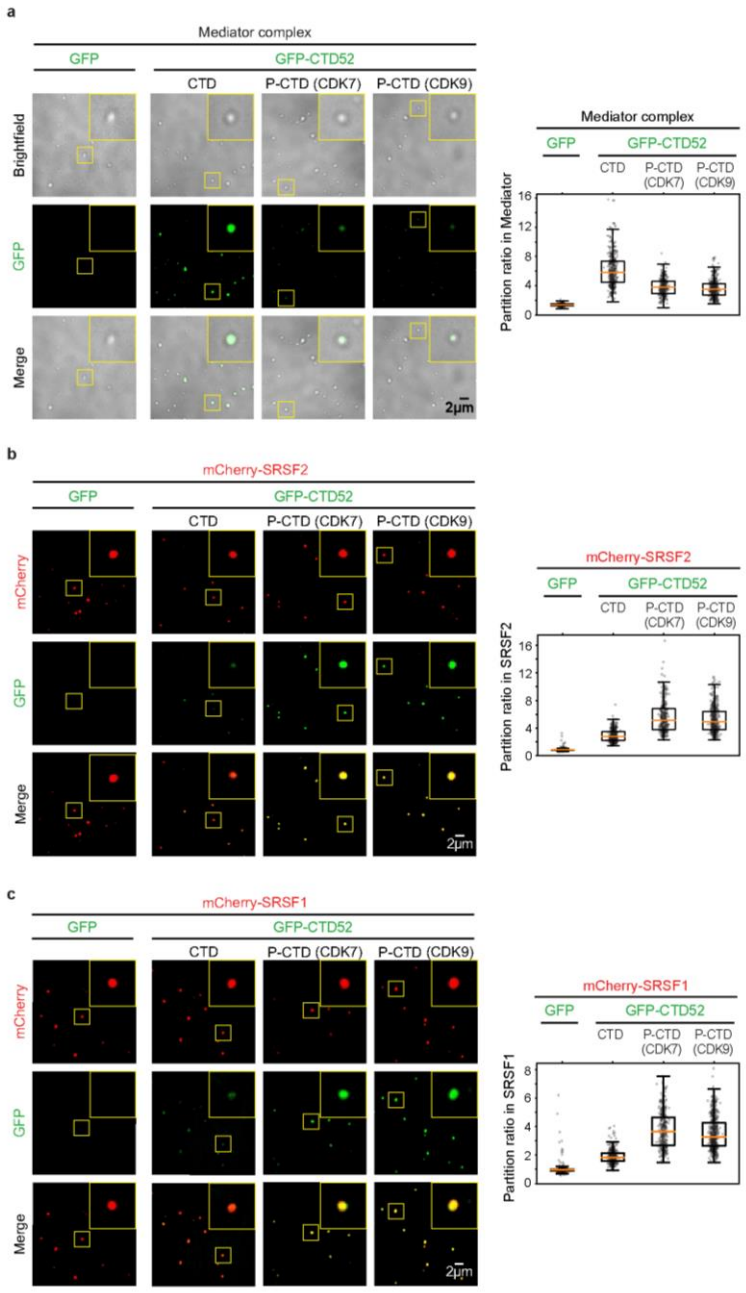


Figure 4. CTD phosphorylation promotes a condensate preference switch between Mediator and splicing factor condensates

a. Representative images and quantification of partition ratios of droplet experiments measuring CTD incorporation into Mediator droplets. Purified human Mediator complex (~200 nM; see methods) was mixed with 10 μM GFP, GFP-CTD52 or GFP-CTD52 phosphorylated with CDK7 or CDK9 in droplet formation buffers with 140 mM monovalent salt and 16% Ficoll-400.

b. Representative images and quantification of partition ratios of droplet experiments measuring CTD incorporation into SRSF2 droplets. Purified human SRSF2 fused to mCherry (mCherry-SRSF2) at 2.5 μM was mixed with 3.3 μM GFP, GFP-CTD52 or GFP-CTD52 phosphorylated with CDK7 or CDK9 in droplet formation buffers with 120 mM NaCl and 10% PEG-8000.

c. Representative images and quantification of partition ratios of droplet experiments measuring CTD incorporation into SRSF1 droplets. Same conditions as in panel b.

Our results indicate that Pol II CTD phosphorylation alters its condensate partitioning behavior and may thus drive an exchange of Pol II from condensates involved in transcription initiation to those involved in RNA splicing at super-enhancer-associated genes. This model is consistent with evidence from previous studies that large clusters of Pol II can fuse with Mediator condensates in cells[7], that phosphorylation dissolves CTD-mediated Pol II clusters[29], that CDK9/Cyclin T can interact with the CTD through a phase separation mechanism[30], that Pol II is no longer associated with Mediator during transcription elongation[13], and that nuclear speckles containing splicing factors can be observed at loci with high transcriptional activity[9-12]. Previous studies have shown that the CTD can interact with components of the transcription initiation apparatus and RNA processing machinery in a phosphoform-specific manner[5,6], but did not explore the possibility that these components occur in condensates or that phosphorylation of the Pol II CTD alters its partitioning behavior between these condensates. Our results reveal that Mediator and splicing factor condensates occur at the same super-enhancer driven genes and suggest that the transition of Pol II from interactions with components involved in initiation to those involved in splicing can be mediated in part through a CTD phosphorylation regulated change in condensate partitioning. These results also suggest that phosphorylation may be among the mechanisms that regulate condensate partitioning of proteins in processes where protein function involves eviction from one condensate and migration to another.

Acknowledgements
We thank Isaac A. Klein, Jurian Schuijers, Charles H. Li, Eliot L. Coffey and other members in the Young lab for helpful discussions, Wendy Salmon of the W.M Keck Microscopy Facility, Tom Volkert, Jennifer Love and Sumeet Gupta of the Whitehead Genomics Core facility for technical assistance, Michael Stubna for help with droplet and biochemical assays, and Steven L. McKnight, Ilmin Kwon, and Masato Kato for CTD constructs. The work was supported by NIH grant GM123511 (R.A.Y.), NSF grant PHY1743900 (R.A.Y. and P.A.S.), NIH GM117370 and GM110064 (D.J.T.), NIH grant R01-GM034277 (P.A.S.), Cancer Research Institute Irvington Fellowship (Y.E.G.), Damon Runyon Cancer Research Foundation Fellowship (2309-17) (B.R.S.), Hope Funds for Cancer Research fellowship (B.J.A.), Swedish Research Council Postdoctoral Fellowship (VR 2017-00372) (A.B.), DFG Postdoctoral Fellowship (SP 1680/1-1) (J.H.S.), the German Research Foundation DFG DE 3069/1-1 (T-M.D.), NIH T32 GM008759 (C.B.F.).

Author contributions
Y.E.G. and R.A.Y. conceived the project. Y.E.G., J.C.M., and R.A.Y. organized the studies and wrote the manuscript. Y.E.G. and B.R.S. performed in vitro droplet formation assays. J.C.M., A.D., and A.B. performed immunofluorescence experiments. J.E.H. and K.S. developed and performed computational analyses. B.R.S., Y.E.G, B.J.A. and J.C.M. performed ChIP/analyzed data. N.M.H. purified recombinant proteins. Y.E.G., J.E.H. and L.K.A. generated cell lines. J.H.S. and A.V.Z. performed lattice light-sheet microscopy and analysis. T-M.D., J.R., C.F. and D.T. purified human Mediator. J.C.M., Y.E.G., and J.E.H. generated constructs. J.E.H. performed live cell imaging. J.E.H. and T.I.L. contributed to writing the manuscript. P.A.S. and I.I.C. provided input into experimental design and interpretation. R.A.Y. supervised the project with the help from T.I.L. All authors contributed to editing the manuscript.

Author information
R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics and Dewpoint Therapeutics. P.A.S. is a member of the Board and shareholder in Syros and a member of the Scientific Advisory Board of Dewpoint. B.J.A. and T.I.L. are shareholders of Syros Pharmaceuticals. T.I.L. is a consultant to Camp4 Therapeutics and I.I.C. is a consultant to Dewpoint Therapeutics. All other authors declare no competing interests. Datasets generated in this study have been deposited in the Gene Expression Omnibus under accession number GSE120656.
Correspondence to: Richard A. Young (young@wi.mit.edu)

References
1 Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-731, doi:10.1038/nrg3293 (2012).
2 Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* **18**, 263-273, doi:10.1038/nrm.2017.10 (2017).
3 Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13-25, doi:10.1016/j.cell.2014.02.009 (2014).
4 Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129-143, doi:10.1038/nrm3952 (2015).
5 Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* **113**, 8456-8490, doi:10.1021/cr400071f (2013).
6 Ebmeier, C. C. *et al.* Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. *Cell Rep* **20**, 1173-1186, doi:10.1016/j.celrep.2017.07.021 (2017).

7 Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415, doi:10.1126/science.aar4199 (2018).
8 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, doi:10.1126/science.aar3958 (2018).
9 Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb Perspect Biol* **3**, doi:10.1101/cshperspect.a000646 (2011).
10 Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J Cell Biol*, doi:10.1083/jcb.201807108 (2018).
11 Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757 e724, doi:10.1016/j.cell.2018.05.024 (2018).
12 Hall, L. L., Smith, K. P., Byron, M. & Lawrence, J. B. Molecular anatomy of a speckle. *Anat Rec A Discov Mol Cell Evol Biol* **288**, 664-675, doi:10.1002/ar.a.20336 (2006).
13 Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**, 155-166, doi:10.1038/nrm3951 (2015).
14 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816, doi:10.1016/j.cell.2018.10.042 (2018).
15 Zhang, X. *et al.* MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Mol Cell* **19**, 89-100, doi:10.1016/j.molcel.2005.05.015 (2005).
16 Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298, doi:10.1038/nrm.2017.7 (2017).
17 Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23, doi:10.1016/j.cell.2017.02.007 (2017).
18 Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, doi:10.1126/science.aaf4382 (2017).
19 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
20 Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252-1269, doi:10.1016/j.cell.2013.02.034 (2013).
21 Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**, 637-650, doi:10.1038/nrm.2017.63 (2017).
22 Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**, 2119-2137, doi:10.1101/gad.200303.112 (2012).
23 Kotake, Y. *et al.* Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat Chem Biol* **3**, 570-575, doi:10.1038/nchembio.2007.16 (2007).
24 Akhtar, M. S. *et al.* TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34**, 387-393, doi:10.1016/j.molcel.2009.04.016 (2009).
25 Czudnochowski, N., Bosken, C. A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun* **3**, 842, doi:10.1038/ncomms1846 (2012).
26 Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049-1060, doi:10.1016/j.cell.2013.10.033 (2013).
27 Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* **417**, 15-27, doi:10.1042/BJ20081501 (2009).

28 McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-361, doi:10.1038/385357a0 (1997).
29 Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol*, doi:10.1038/s41594-018-0112-y (2018).
30 Lu, H. *et al.* Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318-323, doi:10.1038/s41586-018-0174-3 (2018).

METHODS

Cell culture

V6.5 murine embryonic stem cells (mESCs) were a gift from the Jaenisch lab. Cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in 2i media, DMEM-F12 (Life Technologies, 11320082), 0.5X B27 supplement (Life Technologies, 17504044), 0.5X N2 supplement (Life Technologies, 17502048), an extra 0.5 mM L-glutamine (Gibco, 25030-081), 0.1 mM beta-mercaptoethanol (Sigma, M7522), 1% Penicillin Streptomycin (Life Technologies, 15140163), 1X nonessential amino acids (Gibco, 11140-050), 1000 U/ml LIF (Chemico, ESG1107), 1 μM PD0325901 (Stemgent, 04-0006-10), 3 μM CHIR99021 (Stemgent, 04-0004-10). Cells were grown at 37°C with 5% CO2 in a humidified incubator. For confocal imaging, cells were grown on glass coverslips (Carolina Biological Supply, 633029), coated with 5 μg/mL of poly-L-ornithine (Sigma Aldrich, P4957) for at least 30 min at 37°C and with 5 μg/ml of Laminin (Corning, 354232) for 2-16 hrs at 37 °C. For passaging, cells were washed in PBS (Life Technologies, AM9625), 1000 U/mL LIF. 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, 2 mM L-Glutamine, 0.1 mM beta-mercaptoethanol and 15% Fetal Bovine Serum, FBS, (Sigma Aldrich, F4135).

Western blot

Purified phosphorylated CTD was mixed in 1X XT buffer (Bio-Rad) and run on 3-8% Criterion™ XT Tris-acetate Precast Gels (Bio-Rad) at 100 V until the 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 250 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, with shaking. The membrane was then incubated with a 1:5,000 dilution of anti-GFP (Abcam #ab290) antibodies in 5% non-fat milk in TBST overnight at 4°C, with shaking. The membrane was washed three times with TBST for 10 min at room temperature with shaking. The membrane was incubated with 1:10,000 secondary antibodies (GE health) for 1 hr at RT and washed three times in TBST for 5 mins. Membranes were developed with Femto ECL substrate (Thermo Scientific, 34095) and imaged using a CCD camera.

Immunofluorescence with RNA FISH

Coverslips were coated at 37°C with 5 μg/mL poly-L-ornithine (Sigma-Aldrich, P4957) for 30 minutes and 5 μg/mL of Laminin (Corning, 354232) for 2 hrs. Cells were plated on the pre-coated cover slips 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 was performed using 0.5% triton X100 (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. Cells were then incubated with the indicated primary antibody at a concentration of 1:500 in PBS for 4-16 hrs. Antibodies used for IF in this study include SRSF2 (Abcam ab11826), MED1 (Abcam ab64965), Pol II-CTD (Abcam

ab817), Pol II-CTD-S2 (Millipore 04-1571), SF3B1 (Sigma HPA050275), U2AF2 (Abcam ab37530), HNRNPA1 (Abcam ab5832), SRSF1 (Santa Cruz 33652), SRRM1 (Abcam ab221061), PRPF8 (Santa Cruz 55533), SNRNP70 (Sigma HPA043516), and HP1a (Abcam ab203432). SRSF2, MED1, Pol II-CTD, U2AF2, HNRNPA1, SRSF1, and SRRM1 antibodies were validated in house by siRNA knockdown. Pol II-CTD-S2 antibody was validated in house by treatment of cells with DRB. SF3B1 and SNRNP70 antibodies were validated by The Cell Atlas and meet the “enhanced” validation criteria. HP1a antibody was knockout validated by Abcam. Cells were washed with PBS three times followed by incubation with secondary antibody at a concentration of 1:500 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 µM RNA probe 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 in 20 µm/mL Hoechst 33258 (Life Technologies, H3569) for 5 minutes, followed by a 5 minute wash in Wash buffer B (Biosearch Technologies, SMFWB1-20). Cells were washed once in water followed by mounting the coverslip onto glass slides with Vectashield (VWR, 101098-042) and finally sealing the cover slip with nail polish (Electron Microscopy Science Nm, 72180). Images were acquired on the RPI Spinning Disk confocal microscope with 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 probes were custom designed and generated by Agilent to target Nanog and Trim28 intronic regions to visualize nascent RNA.

Cell line generation

V6.5 murine embryonic stem cells were a gift from the Jaenisch lab, and have been verified using short tandom repeat (STR) analysis and tested negative for mycoplasma contamination. Other cell lines in this manuscript were generated from these V6.5 cells. CRISPR/Cas9 was used to generate endogenously-mEGFP-tagged SRSF2 mESC, and endogenously-Halo-tagged RPB1/mEGPF-tagged SRSF2 mESC. Oligos coding for guide RNAs targeting the N-terminus were cloned into a px330 vector expressing Cas9 and mCherry (gift from R. Jaenisch). The sequence that was targeted for SRSF2 was 5’-CGTAGCTCATGGCTGCGAAG-3’. The sequence that was targeted for RPB1 was 5’-TGCCTCGCCATGCACGGGGG-3’. Repair templates were cloned into a pUC19 vector (NEB) containing mEGFP, a GS linker, and 800 bp homology arms flanking the insert. 500K mouse ESCs were transfected with 1.25 µg px330 vector and 1.25 µg repair templates using Lipofectamine-3000 (Thermofisher). Cells were sorted 2 days after transfection for mCherry and 1 week after first sort for mEGFP. 50K cell were serially diluted in a 6 well plate and colonies were picked 4 days after seeding into a 96 well plate. 2-4 days after colony picking, cells were passaged into 3 plates. 1 plate was used for genotyping and the other 2 were frozen down at -80°C in 10% DMSO, 10% FBS and 80% 1x DMEM. The primer pairs that were used for genotyping were the following:

mEGFP-SRSF2
FW - 5’ TTTGGCGGGCTTTCTAACTGC 3’
RV - 5’ CGGTAGGTCAGGTTGTCCAC 3’

Halo-RPB1
FW - 5’ GAGCCCTAGCGTCAACAACT 3’
RV - 5’ CCTCTGGTATCAGCTCCCCT 3’

A clone with heterozygous mEGFP-SRSF2 was subsequently passaged for all assays. A clone of homozygous Halo-RPB1 and heterozygous mEGFP-SRSF2 were used for live cell lattice light sheet imaging.

Live cell imaging of GFP-SRSF2 cell line
Cells were grown on glass dishes (Mattek Corporation P35G-1.5-20-C) coated with 5 µg/mL of poly-L-ornithine (Sigma-Aldrich, P4957) for 30 min at 37°C and with 5 µg/mL of Laminin (Corning, 354232) for 2-16 hrs at 37°C. Before imaging cells, culture media was replaced with phenol red-free 2i media and imaged using the Andor Revolution Spinning Disk Confocal. Raw Andor images were processed using FIJI/ImageJ.

Fluorescence recovery after photobleaching in live cells
FRAP was performed on an Andor Revolution Spinning Disk Confocal with 488 nm laser. Bleaching was performed using 100% laser power with 30 microseconds dwell time for 5 cycles and images were collected every 500 milliseconds. Fluorescence intensity at the bleached spot, a control unbleached spot, and background was measured using the FIJI plugin FRAP Profiler. Background intensity was subtracted and values are reported relative to the unbleached spot to control for photobleaching during image acquisition.

Immunofluorescence with DNA FISH
Immunofluorescence was performed as previously described in the “Immunofluorescence with RNA FISH” section. 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 by mixing 7 mL of FISH Hybridization Buffer (Agilent G9400A), 1 mL of FISH probe (see below for region) and 2 mL of water. 5 mL of mixture was added on a slide and the coverslip was placed on top (cell-side toward the hybridization mixture). Coverslips were 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 37°C in the dark overnight. The coverslip was removed from the 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 room temperature. Slides were air dried and nuclei were stained with Hoechst in PBS for 5 minutes at RT. Coverslips were washed three times in PBS, mounted on slides using Vectashield and sealed with nail polish. Images were acquired on 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. DNA FISH probes were custom designed and generated by Agilent to target Nanog and Trim28 super-enhancers.

Nanog
Design Input Region – mm9
chr6 122605249 – 122705248
Design Region – mm9
chr6: 122605985-122705394

Trim28
Design Input Region – mm9
Chr7:13551990-13651989

Design Region – mm9
chr7:13552132-13651971

Drug treatments
V6.5 mESCs were grown in 24 well plates on coated glass coverslips as described previously. Cells were treated with drugs or vehicle (Pladienolide B at 100nM, DRB at 100uM, or DMSO at 0.1%), followed by fixation with 4% PFA in PBS. For DRB washout experiments, cells were treated with DRB for 2 hours, gently washed twice with fresh 2i media, and left to recover in 2i media containing 0.1% DMSO for 2 hrs before fixation.

Co-immunofluorescence co-localization analysis
For analysis of co-localization data (Figure 3a), custom Python scripts were written to process and analyze 3D image data gathered in IF and DAPI channels. Nuclei were detected by Otsu thresholding, and a mask of nuclei were applied to the IF channels. Manual minimal thresholds were called for the IF channels. Manders coefficients were then calculated for masked IF channels A and B in 3D with the following formulas: $M1 = I_A[I_B > 0] / \sum I_A$, $M2 = I_B[I_A > 0] / \sum I_B$ (see Bolte and Cordelieres. 2006)[31]. Manders coefficients were calculated for at least 3 images per sample and then averaged. To generate the heatmaps of co-localization in Figure 3a, a representative z-slice was selected for each dataset. Each channel was standardized by subtracting the mean and dividing by the standard deviation of the slice to generate a per-pixel z-score. The z-scores between channels were then multiplied and a heatmap was generated from multiplied z-scores using the Python package *matplotlib* with the *magma* heatmap. All heatmaps are displayed on the same scale (vmax = 70). This method highlights areas where both channels have overlapping pixels.

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 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. For MED1-IDR, the inserted sequence encodes residues 948 to 1574 of the full length MED1 protein. Vector expressing mEGFP alone contains the linker sequence followed by a STOP codon. All expression constructs were sequenced to ensure sequence identity.

For protein expression, plasmids were transformed into LOBSTR cells (gift of Chessman Lab). A fresh bacterial colony was inoculated into LB media containing kanamycin and chloramphenicol and grown overnight at 37°C. Cells containing CTD constructs 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 1 mM and growth continued for 20 hours. Cells were collected and stored frozen at -80°C. Cells containing all other constructs were treated in a similar manner except they were grown for 5 hours at 37°C after IPTG induction.

For wildtype (GFP-CTD52, full-length CTD with 52 heptapeptide repeats) and mutant (GFP-CTD26, C-terminal 26 repeats; GFP-CTD10, C-terminal 10 repeats) CTD and hnRNPA1, pellets of 500 ml of cells were resuspended in 15 ml of Buffer A (50 mM Tris pH7.5, 500 mM NaCl) with cOmplete protease inhibitors, (Roche,11873580001) and sonicated (ten cycles of 15 seconds on, 60 seconds off). The lysates were cleared by centrifugation at 12,000 xg for 30 minutes and added to 1 ml of Ni-NTA agarose (Invitrogen, R901-15) that had been pre-equilibrated with 10 volumes of the same buffer. and rotated at 4°C for 1.5 hours. The slurry was centrifuged at 3,000 rpm for 10 minutes in a Thermo Legend XTR swinging bucket rotor. The resin pellets

were washed 2 X with 5 ml of Buffer A followed by centrifugation as above. Protein was eluted 3 X with 2 ml of buffer A plus 250 mM imidazole. For each cycle the elution buffer was added and rotated at least 10 minutes at 4C and centrifuged as above. Elutes were analyzed on a 12% acrylamide gel stained with Coomassie. Fractions containing protein of the expected size were pooled, diluted 1:1 with the 250 mM imidazole buffer and dialyzed against two changes of buffer containing 50 mM Tris 7.5, 12 5mM NaCl, 10% glycerol and 1 mM DTT at 4°C. Protein concentration was measured by Thermo BCA Protein Assay Kit – Reducing Agent Compatible.

For SRSF2, SRSF1 and U2AF2, pellets of 500ml of cells were resuspended in 15 ml of denaturing buffer (50 mM Tris 7.5, 300 mM NaCl, 10 mM imidazole, 8 M Urea) with cOmplete protease inhibitors, sonicated and the cleared as above. The lysates were added to 1 ml of Ni-NTA agarose that had been pre-equilibrated with 10 volumes of the same denaturing buffer. Tubes containing this agarose lysate slurry were rotated for 1.5 hours at room temperature. The slurry was centrifuged and washed 2 X as above. Protein was eluted with 3 X 2ml of denaturing buffer containing 250 mM imidazole. Fractions containing protein of the expected size were diluted 1:1 and dialyzed vs 50 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT with 4 M Urea, followed by the same buffer containing 2M Urea and lastly 2 changes of buffer with 10% Glycerol and no Urea. Any precipitate after dialysis was removed by centrifugation at 3,000 rpm for 10 minutes and concentration determined assayed by BCA.

Purification of Mediator

The Mediator samples were purified as previously described[32] 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 20 mM KCl, 20 mM HEPES, 0.1 mM EDTA, 2 mM $MgCl_2$, 20% glycerol and then dialyzed against pH 7.9 buffer containing 0.15 M KCl, 20 mM HEPES, 0.1 mM EDTA, 20% glycerol and 0.02% NP-40 prior to the affinity purification step. Affinity purification was carried out as described[32], eluted material was loaded onto a 2.2 mL centrifuge tube containing 2 mL 0.15M KCl HEMG (20 mM HEPES, 0.1 mM EDTA, 2 mM $MgCl_2$, 10% glycerol) and centrifuged at 50K RPM for 4h at 4ºC. This served to remove excess free GST-VP16 and to concentrate the Mediator in the final fraction. Prior to droplet assays, purified Mediator was further concentrated using Microcon-30kDa Centrifugal Filter Unit with Ultracel-30 membrane (Millipore MRCF0R030) to reach ~650 nM of Mediator complex. Concentrated Mediator was added to the droplet assay to a final concentration of ~200 nM with 10 μM indicated GFP-tagged protein. Droplet reactions contained 16% Ficoll-400 and 140 mM salt.

Chromatin immunoprecipitation sequencing (ChIP-seq)

mES were grown to 80% confluence in 2i media. About 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). For ChIPs using phospho-specific antibodies, all buffers contained freshly prepared PhosSTOP phosphatase inhibitor cocktail (Roche, 4906837001). Frozen crosslinked cells were thawed on ice and then resuspended in LB1 (50 mM Hepes-KOH, pH7.9, 140 mM NaCl, 1 mM EDTA 0.5 mL 0.5 M, 10% glycerol, 0.5% NP-40, 1% Triton X-100, 1x protease inhibitor) and incubated for 20 minutes rotating at 4°C. Cells were pelleted for 5 minutes at 1350 g, resuspended in LB2 (10 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1x protease inhibitor) and incubated for 5 minutes rotating at 4°C. Pellets were resuspended in LB3 (10 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium-deoxycholate, 0.5% sodium lauroyl sarcosinate, 1% Triton X-100, 1x protease inhibitor)

at a concentration of 30-50 million cells/ml. Cells were sonicated using Covaris S220 for 12 minutes (Duty cycle: 5%, intensity: 4, cycles per burst: 200). Sonicated material was clarified by spinning at 20000 xg for 30 minutes at 4°C. The supernatant is the soluble chromatin used for the ChIP. Dynabeads pre-blocked with 0.5% BSA were incubated with indicated antibodies for 2 hours. Chromatin was added to antibody-bead complex and incubated rotating overnight at 4°C. Beads were washed three times each with Wash buffer 1 (50 mM Hepes pH7.5, 500 mM NaCl, 1 mM EDTA, 1 mM EGTA,1% Triton, 0.1% NaDoc, 0.1% SDS) and Wash Buffer 2 (20 mM Tris pH 8, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% NaDoc) at 4°C, followed by washing one time with TE at room temperature. Chromatin was eluted by adding Elution buffer (50 mM, Tris pH 8.0, 10 mM EDTA, 1% sodium dodecyl sulfate) to the beads and incubated shaking at 60°C for 30 minutes. Reversal of crosslinking was performed overnight at 58°C. RNaseA was added and incubated for 1 hour at 50°C for RNA removal. Proteinase K was added and incubated for 1 hour at 60°C for protein removal. DNA was purified using Qiagen PCR purification kit, as per manufacturer's instructions, and eluted in 50 µL 10 mM Tris-HCl, pH 8.5, which was used for quantitation and ChIP library preparation. ChIP Libraries were prepared with the Swift Biosciences Accel-NGS® 2S Plus DNA Library Kit according to kit instructions with an additional size selection step on the PippinHT system from Sage Science. Following library prep, 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.

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. Metagene plots were made using ngs.plot[33] (v2.61) using default parameters. Top 20% of expressed genes were calculated from a published RNA-seq dataset (GSE112807)[8].

SRSF2 and Ser2-P Pol II ChIP-seqs were generated in this study using antibodies against SRSF2 (Abcam ab11826) and Pol II Ser2 phospho CTD (Millipore 04-1571), whereas MED1 and total Pol II ChIP-seqs were previously published (GSE112808)[8].

RNA FISH average 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 IF channels. Nuclear stains were blurred with a Gaussian filter (sigma = 2.0), maximally projected in the z plane, and clustered into 2 clusters (nuclei and background) by K-means. FISH foci were either manually called with ImageJ or automatically called using the *scipy ndimage* package. For automatic detection, an intensity threshold (mean + 3*standard deviation) was applied to the FISH channel. 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 (minimum 100 voxels), circularity of a max z-projection ($circularity = 4\pi * \frac{area}{perimeter\text{^}2}$; 0.7), and being present in a nucleus (determined by nuclear mask described above). For manual calling, FISH foci were identified in maximum z-projections of the FISH channel, and the x and y coordinates were used as reference points to guide the automatic detection described above. The FISH foci were then centered in a 3D-box (length size l = 3.0 µm). The IF signal centered at FISH foci for each FISH

and IF pair are then combined and an average intensity projection is calculated, providing averaged data for IF signal intensity within a l x l square centered at FISH foci. As a control, this same process was carried out for IF signal 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 are 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 IF 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, lime. This was done to ensure that the reader's eye could more readily detect the contrast in signal. The generated colormap was employed to 15 evenly spaced intensity bins for all IF plots. The averaged IF centered at FISH or at randomly selected nuclear locations are plotted using the same color scale, set to include the minimum and maximum signal from each plot.

Lattice light sheet microscopy

For lattice light sheet microscopy cells were plated on coated coverslip 24 hrs before imaging. Before imaging cells were incubated for 20 min with 250 nM Halo-JF646 ligand[34] and washed in growth medium for 20 min. Dual-color stacks were acquired with 100 ms exposure time and 340 nm effective z-spacing. Light sheet data was processed (deskewed, deconvolved, chromatic aberration correction) using LLSpy[35] and analyzed using custom MATLAB scripts. Foci were localized in 3D following a two-step procedure. First, background was subtracted by subtracting a median filtered image from each slice in a z-stack, and intensity peaks were detected using the MTT algorithm[36]. Foci were then identified as peaks that were found in at least 4 subsequent z-slices within a 100nm radius in x-y. A 3D Gaussian peak function was fitted to the intensity distribution to obtain 3D center coordinates for foci of SRSF2 and Pol II. Detection of foci was performed on deconvolved, background-subtracted data, whereas subsequent quantitative analysis of foci brightness was performed on deskewed, background-subtracted data. To estimate the number of fluorescently tagged proteins in foci cells were fixed in 4% PFA for 10 mins, washed 3x in PBS, and imaged on the lattice light sheet microscope until bleached almost entirely. A single plane was then imaged continuously for 1000 frames to detect single emitter signals. The apparent brightness of Halo-JF646 (n = 204) or GFP (n = 236) single emitters was determined by calculating the integrated intensity above background[37] (Extended Data Fig. 2b). We note that signals are close to the noise floor of the camera, and the integrated intensity measure can sometimes (~10% of the cases) yield negative values when nearby emitters lead to overestimation of the local background intensity. For the subsequent analysis we excluded emitters with negative integrated intensity measures. We used the mean estimates obtained for the remaining single emitters to normalize values obtained by the same metric for foci of SRSF2-GFP, and Halo-JF646-Pol II in the z slice closest to the 3D center position, taking into account higher laser power densities that were used to detect single emitters. To assess co-localization we paired Pol II-Halo foci with their nearest neighbor in 3D space. We note that axial resolution of the imaging method is considerably lower than xy-resolution. We therefore call only those foci co-localized with center coordinates separated by less than the optical resolution of 300 nm laterally (xy), and 900nm axially (z).

Splicing reporter assay

The splicing reporter assay was performed as described in Younis et al. (2010)[38]. Briefly, mESCs were transfected with a plasmid encoding luciferase with an intervening intron (Addgene 62858) or a plasmid encoding luciferase with no intervening intron (Addgene 62857). Cells were then treated with DMSO or 100 nM Pladienolide B for 4 hrs, at which point they were

lysed and assayed for luciferase activity. Relative splicing levels in each condition were determined by normalizing the luciferase activity detected in cells transfected with the intron containing plasmid to the luciferase activity detected in cells transfected with the intronless plasmid.

FISH-IF overlap analysis

DNA FISH spots were identified as described in "RNA FISH average image analysis". Images of the spots with the corresponding IF channel for all conditions were randomized and blindly scored for FISH-IF overlap (at least 25% of the FISH spot overlapping with an IF puncta) or no overlap. Overlap scores for each condition were then tallied and compared. For presentation, the FISH foci overlap with IF was indexed with the DMSO condition set to 1.

In vitro droplet assay

Recombinant GFP or mCherry fusion proteins were concentrated and desalted to an appropriate protein concentration and 125-500 mM NaCl using Amicon Ultra centrifugal filters (30K MWCO, Millipore). Recombinant proteins were added to solutions at varying concentrations with 120-125 mM final salt and 16% Ficoll-400 or 10% PEG-8000 as crowding agent in droplet formation buffer (50 mM Tris-HCl pH 7.5, 10% glycerol, 1 mM DTT) as described in figure legends. The protein solution was incubated for 1 hour and loaded onto a homemade chamber comprising a glass slide with a coverslip attached by two parallel strips of double-sided tape. Slides were then imaged with the Andor confocal microscope with a 150x objective. Unless indicated, images presented are of droplets settled on the glass coverslip. For FRAP of in vitro droplets, 2 pulses of laser (20% power) at a 20 us dwell time were applied to the droplet, and recovery was imaged on the Andor microscope every 1s for the indicated time periods. For CDK7 or CDK9 mediated CTD phosphorylation, commercially available active CDK7/MAT1/CCNH (CAK complex; Millipore 14-476) or CDK9/Cyclin T1 (Millipore 14-685) was used to phosphorylate GFP-CTD52 in kinase reaction buffer (20 mM MOPs-NaOH pH 7.0, 1 mM EDTA, 0.001% NP-40, 2.5 % glycerol, 0.05% beta-mercaptoethanol, 10 mM MgAc, 10 μM ATP) at room temperature for 3 hours. The CTD to enzyme ratio is ~1 μM CTD to ~5 ng/ul CDK7 or CDK9.

In vitro droplet quantification

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 three standard deviations above the mean of the image, (2) size thresholds (20 pixel minimum droplet size), (3) and a minimum circularity ($circularity = 4\pi * \frac{area}{perimeter\text{^}2}$) 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 (see Banani et al Cell 2016)[39]. Hundreds of droplets identified in typically 10 independent fields of view were quantified. The mean intensity within the droplets (C-in) and in the bulk (C-out) were calculated for each channel. The partition ratio was computed as (C-in)/(C-out). The box plots show the distributions of all droplets. Each dot represents an individual droplet. The measured datasets for partition ratio versus the protein concentration in Extended Data Figure 7b were fitted by the logistic equation (see Wang et al Cell 2018)[40]:

$$f = \frac{a}{1 + e^{-\frac{(x-x0)}{b}}}$$

Where *f* is the partition ratio and *x* is the corresponding protein concentration.

Statistics and reproducibility

For all immunofluorescence and fluorescence in situ hybridization experiments, one coverslip of cells was stained for the indicated factors and at least eight independent imaging fields were acquired, which typically contained 50-200 FISH foci. The exact number of FISH foci analyzed, and the fraction of those foci that overlap with IF puncta and relevant comparative statistics for experiments where these comparisons were made are as follows:

Fig. 1b/Extended Data Fig. 1a: 86 *Nanog* foci, 131 *Trim28* foci.

Fig. 2a: *Nanog* FISH foci counts for each IF experiment- SRSF2: 97, SF3B1: 122, U2AF2: 74, HNRNPA1: 88, SRSF1: 109, SRRM1: 137, PRPF8: 103, SNRNP70: 119.

Fig. 2c/Extended Data Fig. 3b, c: The numbers of overlapped foci for each factor in the DMSO and Pladienolide B condition, and the p-values associated with each DMSO vs. Pladienolide B comparison are as follows: (SRSF2, 31/61, 19/125, <0.0001; SF3B1, 29/61, 30/126, 0.0014; Pol II, 16/71, 15/65, >0.9999).

Fig. 3c, d/Extended Data Fig. 5a-d: The numbers of overlapped foci for each factor in the DMSO, DRB, and washout condition, and the p-values associated with the DMSO vs. DRB and DRB vs. washout comparison are as follows: (SRSF2 Nanog, 40/91, 11/146, 19/78, <0.0001, 0.0008; Pol II Nanog, 33/114, 23/92, 36/160, 0.6368, 0.6467; MED1 Nanog, 28/89, 32/133, 27/84, 0.2804, 0.2122; SRSF2 Trim28, 26/71, 5/111, 12/92, <0.0001, 0.0403; SRSF1 Trim28, 19/36, 12/55, 22/73, 0.0347, 0.3189; SF3B1 Trim28, 48/91, 14/63, 27/99, 0.0002, 0.5788; U2AF2 Trim28, 21/42, 17/50, 27/78, 0.1406, >0.9999; PRPF8 Trim28, 15/80, 6/69, 17/102, 0.0996, 0.1721; MED1 Trim28, 19/73, 30/96, 21/91, 0.4971, 0.251; Pol II Trim28, 25/102, 22/93, 21/78, >0.9999, 0.7238).

Extended Data Fig. 2a: *Trim28* FISH foci counts for each IF experiment - SRSF2: 115, SF3B1: 151, U2AF2: 104, HNRNPA1: 90, SRSF1: 145, SRRM1: 127, PRPF8: 175, SNRNP70: 157.

For lattice light sheet imaging, the number of images acquired and data points plotted are as follows:
Fig. 2b: full image dataset comprises 102 cells from 10 independent fields of view.
Extended Data Fig. 2b: 288 colocalized condensates were plotted in the scatter plot.

For all in vitro droplet experiments, one slide of droplet mix was imaged and at least seven independent fields of view were acquired, which typically contained ~100-1000 droplets. In all cases where the droplet data is quantified and displayed as a box plot, the box in the boxplot extends from the 25th to 75th percentiles, the line in the middle of the box is plotted at the median, the whiskers represent the range within 1.5x interquartile, and each dot represents an individual droplet. The exact number of fields and droplets analyzed are as follows:

Fig. 1c (fields/droplets): GFP, 10/540; GFP-CTD52, 7/842; GFP-CTD26, 10/879; GFP-CTD10 10/1293.

Fig. 1d (fields/droplets): GFP 11/159, GFP-CTD52 10/175, GFP-CTD26 11/207, GFP-CTD10 10/206.

Fig. 4a (fields/droplets): GFP 10/114, CTD 7/420, P-CTD (CDK7) 8/342, P-CTD (CDK9) 8/464.

Fig. 4b (fields/droplets): GFP 10/385, CTD 10/328, P-CTD (CDK7) 10/294, P-CTD (CDK9) 12/361.

Fig. 4c (fields/droplets): GFP 10/400, CTD 10/369, P-CTD (CDK7) 10/314, P-CTD (CDK9) 15/513.

Extended Data Fig. 1b: One fusion event was captured from one field.

Extended Data Fig. 6c, d (fields/droplets): 16% Ficoll, GFP 21/231, CTD 21/289, P-CTD (CDK7) 10/134, P-CTD (CDK9) 11/144; 10% PEG, GFP 24/147, CTD 21/227, P-CTD (CDK7) 10/106, P-CTD (CDK9) 10/83.

Extended Data Fig. 7a, b: The number of fields acquired are as follows, SRSF2: 5μM 10, 2.5μM 10, 1.25μM 10, 0.625μM 10, 0.313μM 10, 0.156μM 11, 0.078μM 12, 0.039μM 12, 0.0195μM 10, 0.0098μM 10, 0.0049μM 10, 0.0024μM 10, 0.0012μM 10, 0.0006μM 10; SRSF1: 5μM 10, 2.5μM 10, 1.25μM 10, 0.625μM 10, 0.313μM 10, 0.156μM 10, 0.078μM 10, 0.039μM 10, 0.0195μM 10, 0.0098μM 10, 0.0049μM 10, 0.0024μM 11, 0.0012μM 11, 0.0006μM 10; U2AF2: 10 for each sample; hnRNPA1: 10 for each sample; mCherry: 10 for each sample.

Extended Data Fig. 7c: 10 independent fields of view were acquired for each sample.

Extended Data Fig. 8a, b: same as Fig. 4b, c.

Extended Data Fig. 8c (fields/droplets): GFP 10/365, CTD 10/321, P-CTD (CDK7) 10/325, P-CTD (CDK9) 10/313.

Extended Data Fig. 8d: (fields/droplets): GFP 10/423, CTD 11/437, P-CTD (CDK7) 10/412, P-CTD (CDK9) 10/381.

Code availability
All custom code used in this study is available upon request.

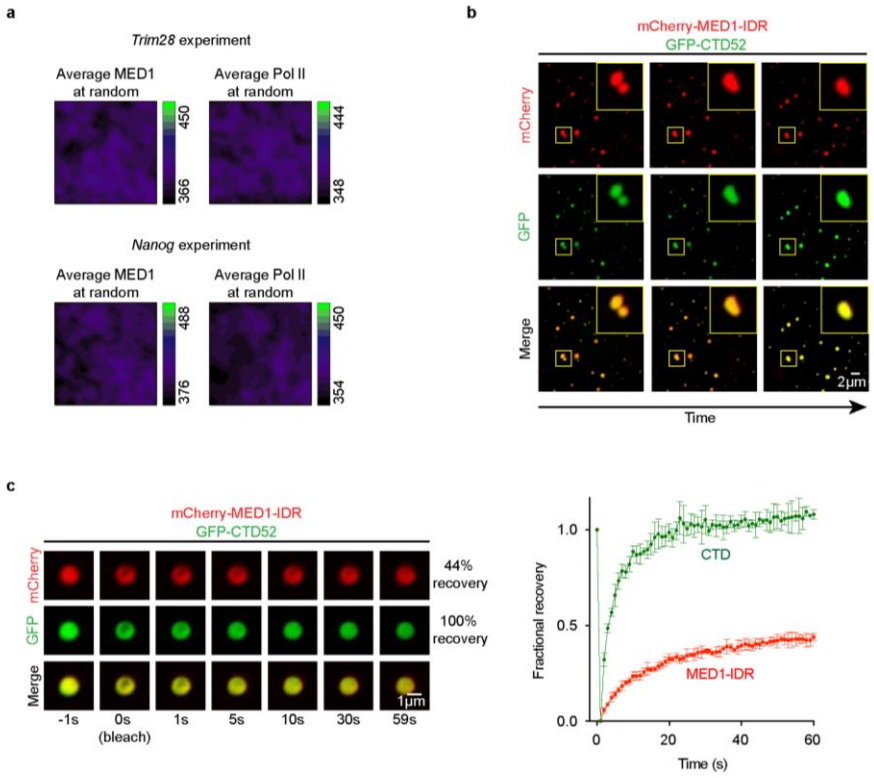
Data availability
Datasets generated in this study have been deposited in the Gene Expression Omnibus under accession number GSE120656.

Methods references

31 Bolte, S. & Cordelieres, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* **224**, 213-232, doi:10.1111/j.1365-2818.2006.01706.x (2006).
32 Meyer, K. D. *et al.* Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *EMBO J* **27**, 1447-1457, doi:10.1038/emboj.2008.78 (2008).
33 Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284, doi:10.1186/1471-2164-15-284 (2014).
34 Grimm, J. B. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* **12**, 244-250, 243 p following 250, doi:10.1038/nmeth.3256 (2015).
35 tlambert03/LLSpy: v0.3.8. (Zenodo, 2018).

36 Serge, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat Methods* **5**, 687-694, doi:10.1038/nmeth.1233 (2008).
37 Verdaasdonk, J. S., Lawrimore, J. & Bloom, K. Determining absolute protein numbers by quantitative fluorescence microscopy. *Methods Cell Biol* **123**, 347-365, doi:10.1016/B978-0-12-420138-5.00019-7 (2014).
38 Younis, I. *et al.* Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol Cell Biol* **30**, 1718-1728, doi:10.1128/MCB.01301-09 (2010).
39 Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663, doi:10.1016/j.cell.2016.06.010 (2016).
40 Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 e616, doi:10.1016/j.cell.2018.06.006 (2018).

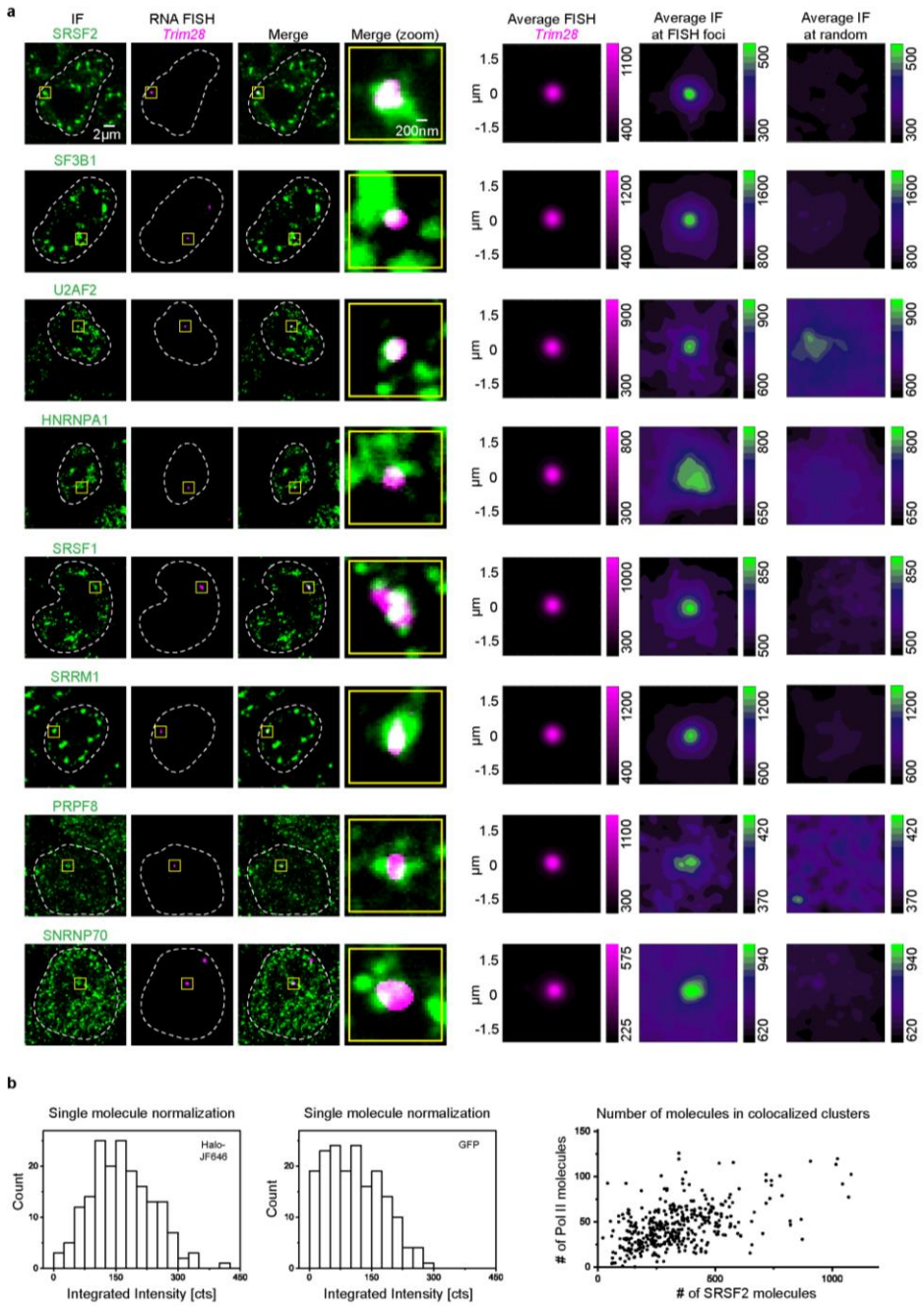
EXTENDED DATA FIGURE 1



Extended Data Figure 1. CTD partitioning in Mediator condensates

a. Average IF signal for MED1 and Pol II centered on randomly selected nuclear positions.
b. Images of a fusion event between two full-length CTD/MED1-IDR droplets. GFP-CTD52 at 10 μM was mixed with 10 μM mCherry-MED1-IDR in droplet formation buffer with 125 mM NaCl and 16% Ficoll-400.
c. Top: representative images of FRAP of heterotypic droplets of mCherry-MED1-IDR and GFP-CTD52. Droplet formation conditions are the same as in b. Bottom: Quantification of the fraction of fluorescence recovery following photobleaching of mCherry-MED1-IDR and GFP-CTD52. Data represent mean +/- SD (n = 3).

EXTENDED DATA FIGURE 2

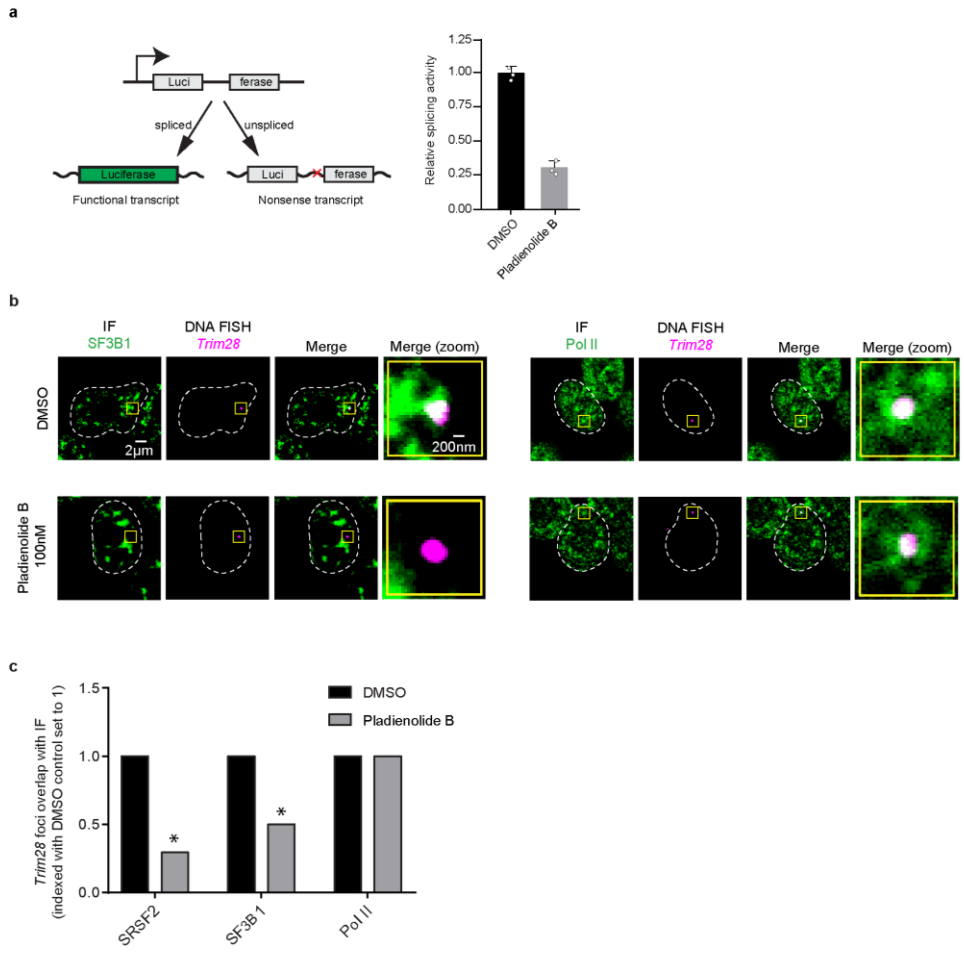


Extended Data Figure 2. Splicing factors at *Trim28*, SRSF2/Pol II molecule quantification.

a. Representative images exhibiting overlap between IF of splicing factors SRSF2, SF3B1, U2AF2, HNRNPA1, SRSF1, SRRM1, PRPF8, or SNRNP70 with nascent RNA FISH of *Trim28* in fixed mESCs.

b. Left: Histogram of the integrated intensity of single Halo-JF646 (n=178) and single GFP emitters (n=177). Mean values of 164.8 +/- 5.6 cts (mean +/- s.e.m.) and 108.6 +/- 5.1 (mean +/- s.e.m.) were used to normalize the integrated intensity of Pol II-Halo-JF646 and SRSF2-GFP, respectively. Right: Scatter plot depicting estimated numbers of Pol II and SRSF2 molecules in colocalizing Pol II and SRSF2 puncta (see methods).

EXTENDED DATA FIGURE 3



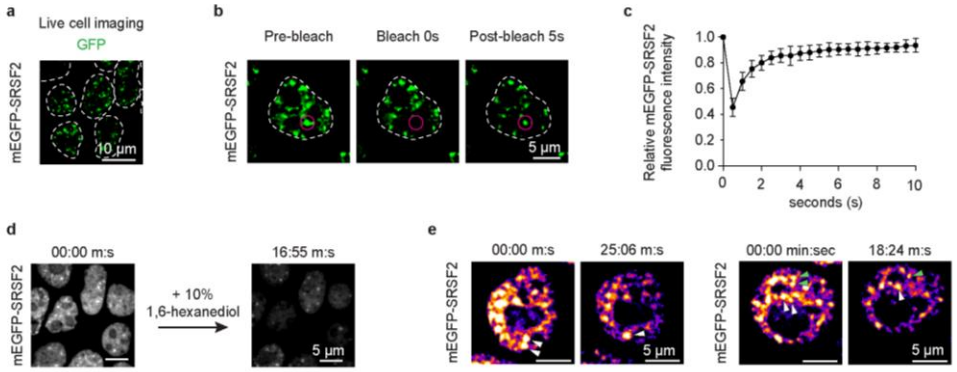
Extended Data Figure 3. Splicing inhibition and splicing factor condensates

a. Left: diagram depicting the splicing reporter used to measure splicing inhibition following treatment with Pladienolide B. Right: relative levels of splicing in cells treated with DMSO vs. cells treated with Pladienolide B for 4 hrs. The mean of 3 biological replicates (each replicate shown as dot) with standard deviations is plotted. See methods for details.

b. Representative images exhibiting overlap or absence of overlap between IF of SF3B1 or Pol II and DNA FISH of *Trim28* in mESCs treated with either DMSO or splicing inhibitor Pladienolide B for 4 hours.

c. Fraction of overlap of FISH foci with IF puncta in cells treated with Pladienolide B relative to cells treated with DMSO. A star above the drug treated bar indicates a p-value of <0.05 from a two-tailed Chi-squared test comparing the number of overlapping and non-overlapping FISH spots in the DMSO vs. drug condition. See methods for details.

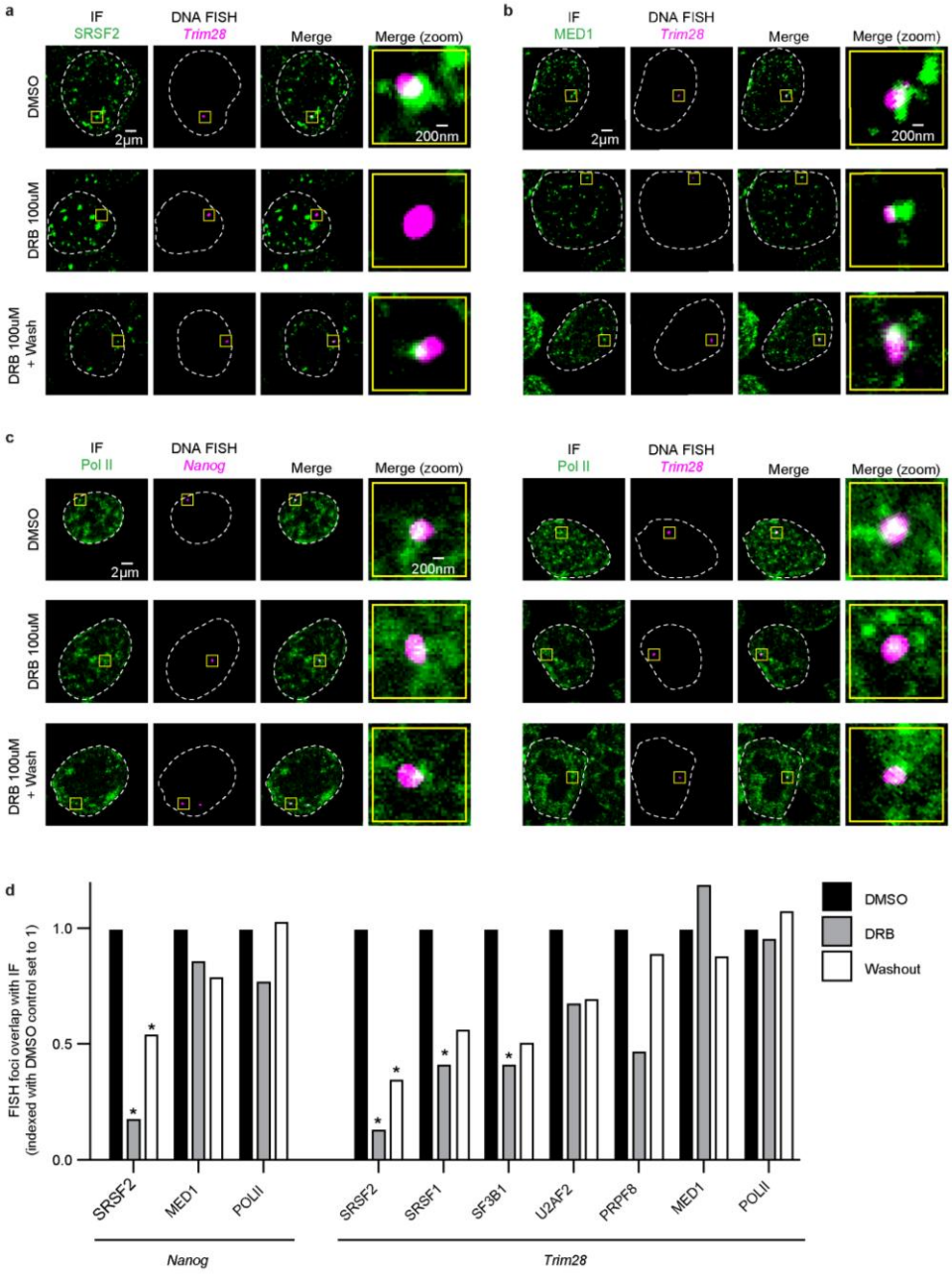
EXTENDED DATA FIGURE 4



Extended Data Figure 4. Liquid-like properties of SRSF2 condensates

a. Live cell imaging of GFP-SRSF2 mESCs. Ten independent fields from one plate of cells were imaged.
b. Representative images of fluorescence recovery after photobleaching (FRAP) experiments performed on the GFP-SRSF2 mESC cell line.
c. Quantification of experiment depicted in b. Data represent means +/- SD (n = 9).
d. Representative images of live cells before and after treatment with 10% 1,6-hexanediol for ~17 minutes. Five independent fields from one plate of cells were imaged before and after treatment.
e. Examples of fusion events occurring between SRSF2 puncta in the GFP-SRSF2 cell line. Two fields from two independent plates of cells were imaged over a two hour time course, and each showed at least one fusion event.

EXTENDED DATA FIGURE 5



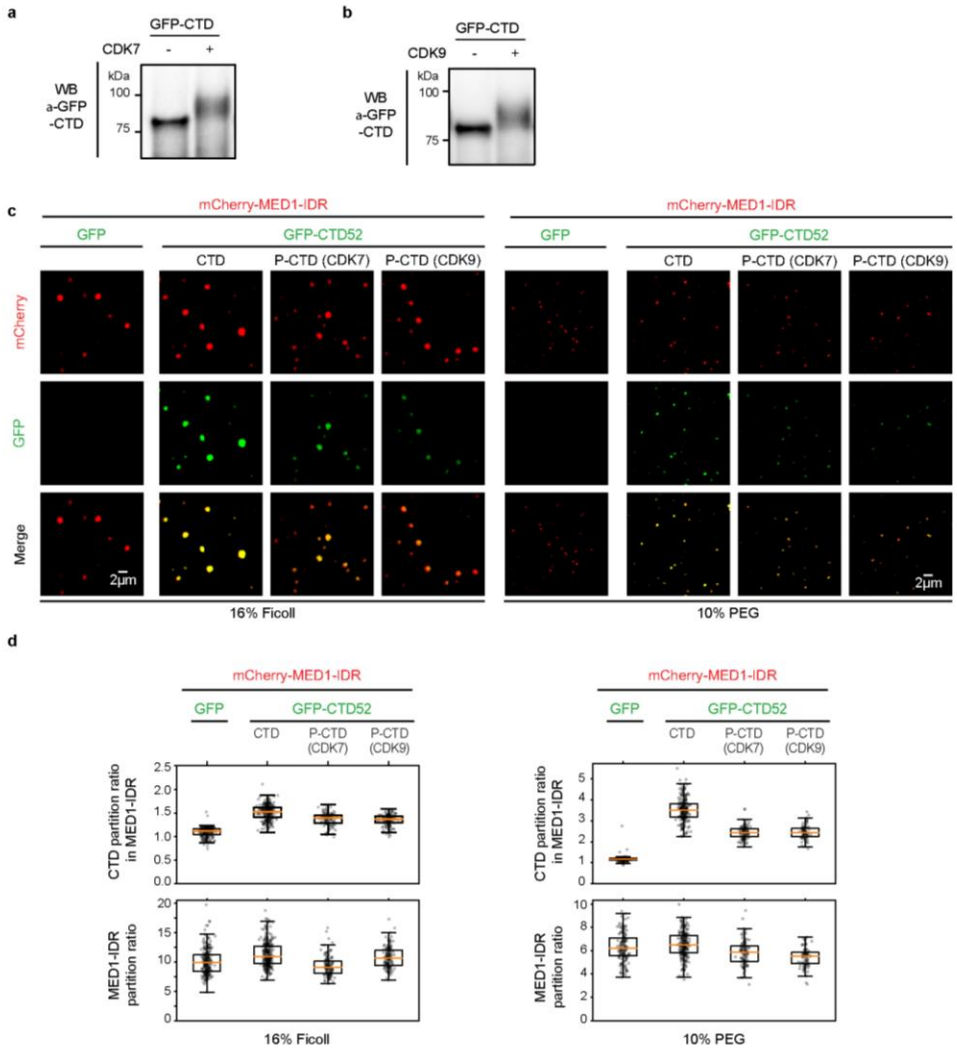
Extended Data Figure 5. DRB treatment effects on splicing factor and transcriptional condensates

a. Representative images exhibiting overlap or lack of overlap between IF of SRSF2 and DNA FISH of *Trim28* in mESCs treated with DMSO for 2 hrs, DRB for 2 hrs, or DRB for 2 hrs followed by a 2 hr DRB washout.

b. Representative images exhibiting overlap between IF of MED1 and DNA FISH of *Trim28* in mESCs treated with DMSO for 2 hrs, DRB for 2 hrs, or DRB for 2 hrs followed by a 2 hr DRB washout.

c. Representative images exhibiting overlap between IF of Pol II and DNA FISH of *Nanog* or *Trim28* in mESCs treated with DMSO for 2 hrs, DRB for 2 hrs, or DRB for 2 hrs followed by a 2 hr DRB washout.
d. Fraction of overlap of FISH foci with IF puncta in cells treated with DRB or DRB followed by washout relative to cells treated with DMSO. A star above the DRB treated bar indicates a p-value of <0.05 from a two-tailed Chi-squared test comparing the number of overlapping and non-overlapping FISH spots in the DMSO vs. DRB condition. A star over the washout bar indicates a p-value of <0.05 from a two-tailed Chi-squared test comparing the DRB vs. washout condition. Splicing factors tested showed significantly decreased or trended towards decreased overlap with FISH spots in the DRB condition compared to DMSO, while MED1 and Pol II exhibited limited changes. See methods for details.

EXTENDED DATA FIGURE 6



Extended Data Figure 6. CTD in vitro phosphorylation and partitioning in MED1-IDR droplets

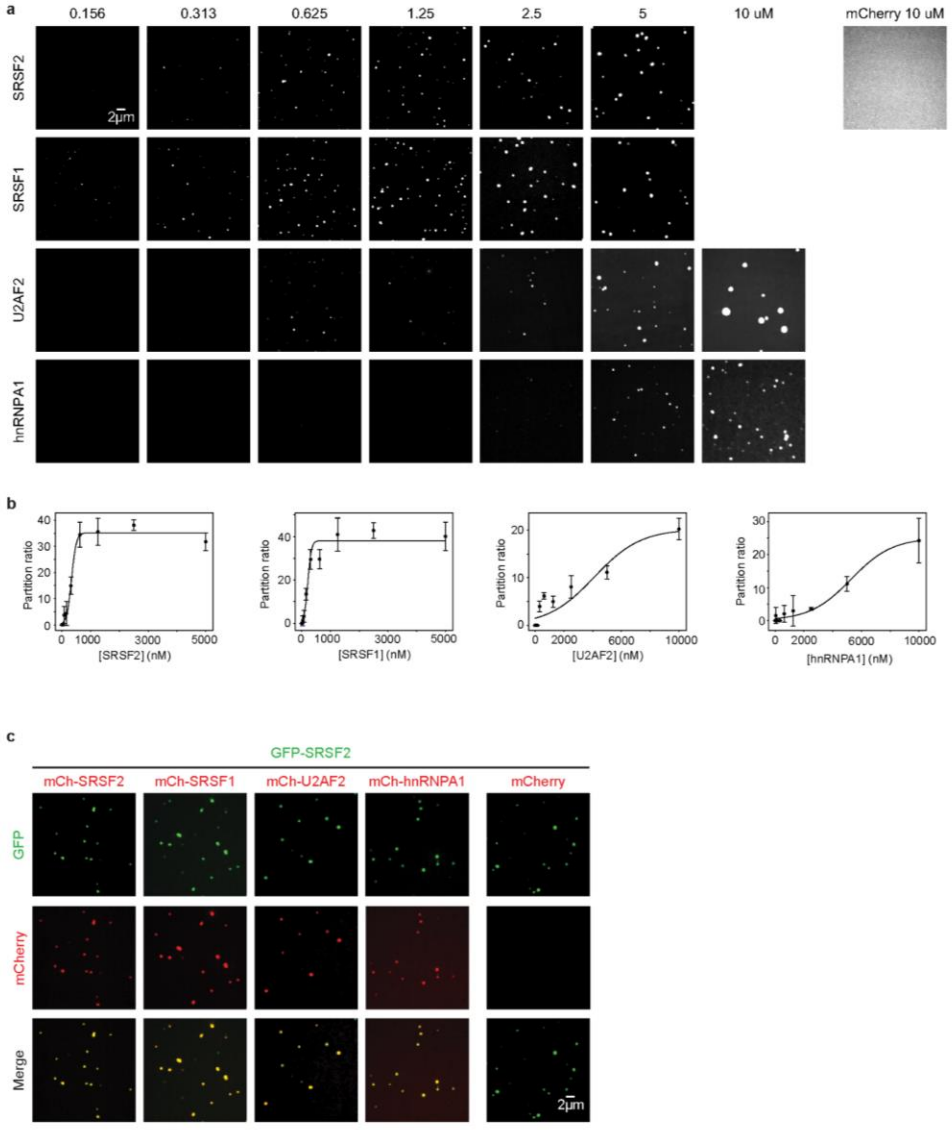
a. Western blot showing phosphorylation of CTD by CDK7. Equal amounts of unphosphorylated and phosphorylated GFP-CTD52 (see methods) were run on a 3-8% SDS PAGE gel and analyzed by western blot using anti-GFP antibodies. Similar results were obtained in two biological replicates. For gel source data, see Supplementary Figure 1.

b. Western blot showing phosphorylation of CTD by CDK9. Similar results were obtained in two biological replicates. For gel source data, see Supplementary Figure 1.

c. Representative images of droplet experiments measuring CTD incorporation into MED1-IDR droplets. Purified human MED1-IDR fused to mCherry (mCherry-MED1-IDR) at 10 μM was mixed with 3.3 μM GFP, GFP-CTD52, GFP-CTD52 phosphorylated with CDK7 or CDK9 in droplet formation buffers with 125 mM NaCl and 16% Ficoll-400 or 10% PEG-8000.

d. Partition ratios of GFP and GFP-CTD in MED1-IDR droplets, and partition ratios of mCherry-MED1-IDR from experiments depicted in c.

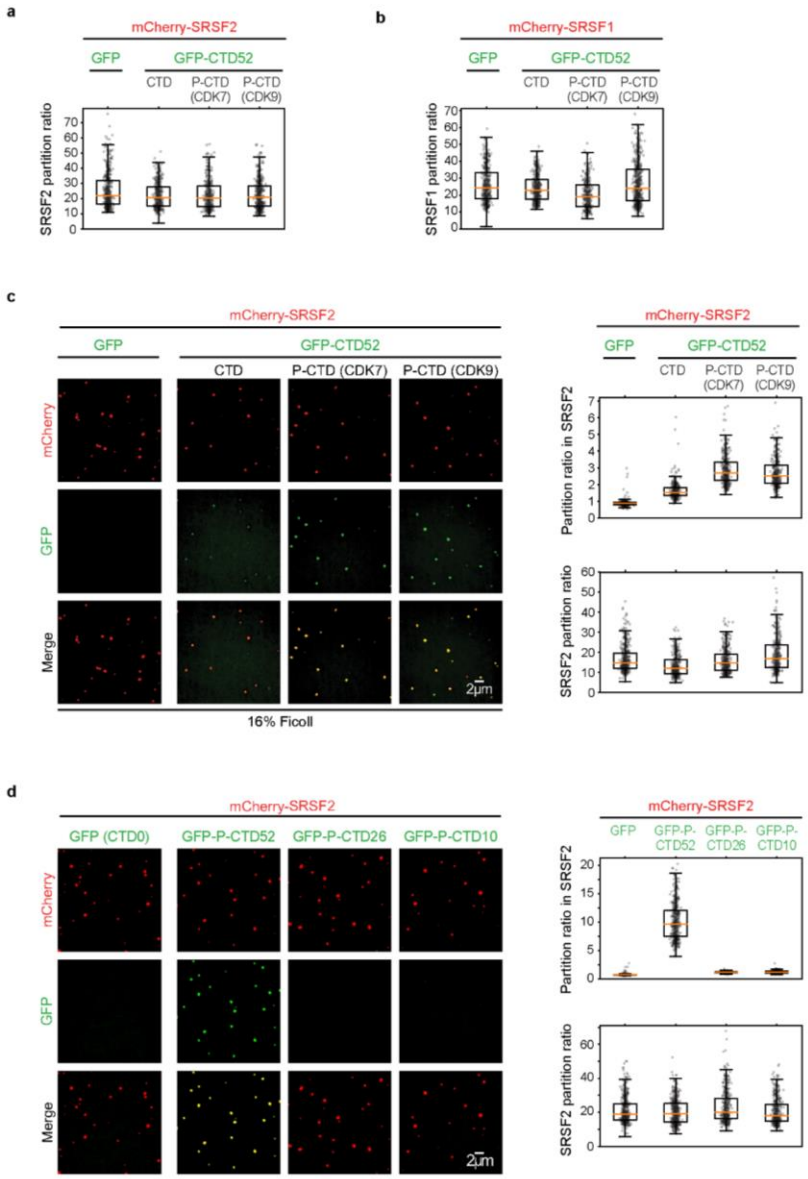
EXTENDED DATA FIGURE 7



Extended Data Figure 7. Splicing factors form droplets in vitro

a. Representative images of droplet formation by mCherry-SRSF2, SRSF1, U2AF2 and hnRNPA1 with increasing protein concentrations. All assays were performed in the presence of 120 mM NaCl and 10% PEG-8000.
b. Quantification of the partition ratios from the experiments depicted in a.
c. Representative images of heterotypic droplets formed between SRSF2 and other splicing factors, including SRSF1, U2AF2 and hnRNPA1. Purified human SRSF2 fused to GFP (GFP-SRSF2) at 2.5 μM was mixed with 2.5 μM purified mCherry-fused to human SRSF2, SRSF1, U2AF2 or hnRNPA1 in droplet formation buffers with 120mM NaCl and 10% PEG-8000.

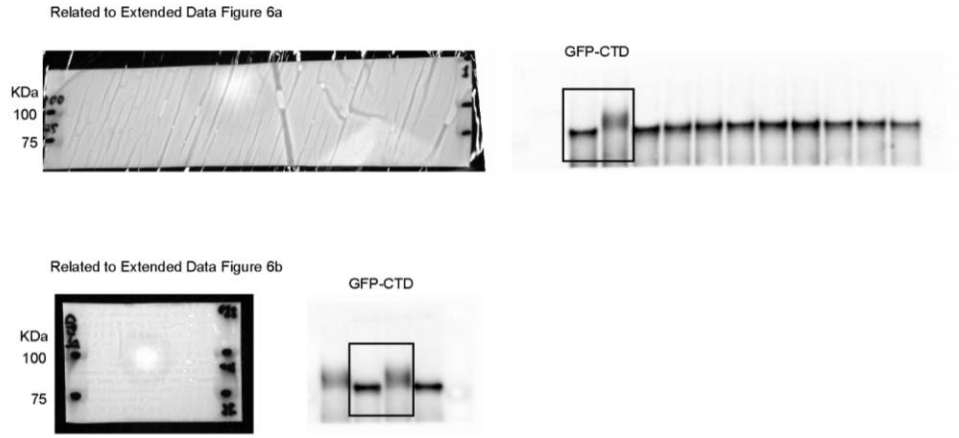
EXTENDED DATA FIGURE 8



Extended Data Figure 8. CTD partitioning in SR-protein droplets

a. Quantification of the partition ratios of SRSF2 from experiments depicted in Fig. 4b.
b. Quantification of the partition ratios of SRSF1 from experiments depicted in Fig. 4c.
c. Representative images of droplet experiments measuring CTD incorporation into SRSF2 droplets. Same reagents and conditions were used as in Fig. 4b except that 16% Ficoll-400 was used as a crowding agent.
d. Representative images and quantification of partition ratios of droplet experiments measuring phosphorylated full length or truncated CTD incorporation into SRSF2 droplets. Droplet formation conditions are the same as described in Fig. 4b.

SUPPLEMENTARY FIGURE 1



Supplementary Figure 1. Uncropped western blots with size markers

Uncropped blots related to extended data figures 6a and 6b. The regions shown in extended data figures 6a and 6b are boxed.

CHAPTER 4: FUTURE DIRECTIONS AND DISCUSSION

While our understanding of condensate biology has advanced rapidly over the last decade, the field is still in its infancy and there are many open questions and research areas to explore. Some of these include identifying additional processes that involve condensation, gaining a more complete understanding of the roles IDRs and weak multivalent interactions have played over the course of evolution (Gao et al., 2018), identifying the molecular rules that govern biological phase separation, and developing new therapeutic approaches informed by condensate biology. In this Chapter I will discuss the latter two.

The molecular grammar of phase separation

While many processes have been found to involve phase separation, the interactions between the phase-separated components that dictate the formation, dissolution, physical properties, and the ability to partition specific clients of condensates remains poorly understood. For example, in Chapter 3, we showed empirically that CTD phosphorylation results in exclusion from Mediator condensates, but the molecular interactions that govern this exclusion remain to be determined. This “molecular grammar” is likely defined by the physicochemical properties of specific amino acids in specific conformations and the multivalent interaction capacity of proteins and nucleic acids. A few studies have begun to determine some of these rules, but much work remains to determine a more general code that can be applied to any system. Here, I will detail some of these emerging rules before describing the potential benefits of understanding the molecular grammar of phase separation and proposing strategies to rapidly advance this understanding.

Pioneering studies have dissected the determinants of phase separation for the protein FUS (Wang et al., 2018). FUS is an RNA binding protein that forms compartments in the nucleus and cytoplasm and has been reported to play roles in transcription, DNA repair, rRNA biogenesis, and various neurodegenerative diseases (Mikhaleva and Lemke, 2018; Polymenidou et al., 2012; Wang et al., 2013, 2008). The protein consists of an intrinsically disordered prion-like domain and an RNA binding domain. Through precise mutational analysis, the authors determined that the multivalent interactions that drive FUS phase separation are primarily mediated by transient pi-cation interactions between tyrosine residues in the IDR and arginine residues in the RNA binding domain. These tyrosine-arginine interactions were enhanced by neighboring aspartic acid residues. Repulsion between aspartic acid residues in the IDR further prevented non-productive intramolecular interactions between tyrosines in the IDR. The identity of amino acids found in the spacer regions between tyrosine residues affected the physical properties of FUS condensates. Spacers rich in glycine allowed for greater flexibility and more liquid-like dynamic condensates, whereas the presence of glutamine in the spacers appeared to contribute to droplet hardening, perhaps by allowing for the formation of cross-beta sheets (Murray et al., 2017). It remains to be determined whether these rules for FUS phase separation can be generally applied to other families of proteins, but the methodology of precise mutational dissection used in this study could be applied to other proteins as well.

Additional physical properties of spacer regions appear to govern the formation of condensates (Harmon et al., 2017). Model based approaches show that an intermediate spacer length is optimal for condensation. If the spacers are too short, multiple interactions on a single factor may be sterically hindered and the effective valency of the factor is decreased. If the spacers are too long, interaction domains can interact independently without the formation of a dense phase. The solubility of spacers in the surrounding environment also affects condensate formation, and again

an intermediate level is optimal for phase separation. If the spacer solubility is too low, the propensity for self-interaction and clustering of spacers leads to a more ordered conformation not optimal for condensation. If the spacer solubility is too high, the spacers will prefer to be solvated and will resist confinement and concentration within droplets.

Nucleic acids are key components of condensates, and some general rules that govern the inclusion of specific RNAs in condensates have been determined (Langdon and Gladfelter, 2018; Langdon et al., 2018; Zhang et al., 2015). As described in Chapter 1 for P granules, RNA can form multivalent interactions with proteins and contribute to condensate formation. In the fungus *Ashbya gossypii*, multiple condensates with distinct cellular localization containing distinct sets of mRNAs co-exist and appear to not mix with one another. The specificity of RNA inclusion in these condensates stems from different RNA secondary structures. RNAs adopt secondary structures that allow for intermolecular base pairing interactions with other RNAs in the same condensate, while hiding sequences that would be complementary to RNAs found in other condensates. These secondary structures can be modified by specific RNA binding proteins to change the RNAs that occur in a given condensate in response to stimulus. The inclusion of different RNAs in condensates may in turn lead to the inclusion of specific proteins that interact with those RNAs. While this phenomena has yet to be observed in higher eukaryotes, it seems that the general principles discovered in these studies could function in any system where multiple condensates with distinct constituents co-exist.

A full understanding of the rules that govern phase separation and client partitioning could be useful for a number of reasons. On a basic science level, we could better understand the regulatory consequences of alterations to proteins. For example, in addition to thinking about the creation or destruction of structured binding sites upon post translational modification, we could also consider potential changes in condensate partitioning preference. Understanding molecular grammar could also provide insights into human pathology. The altered phase behavior of a mutated factor could lead to the formation of a new condensate, dissolution of a condensate, altered physical properties of a condensate, or partitioning of proteins into the incorrect condensates, and predicting these changes could inspire new treatment options. Finally, the ability to design proteins such that they form or partition into condensates could prove useful in the industrial production of biological compounds. As discussed in Chapter 1, one function of phase separation is to concentrate reaction components to improve the efficiency of reactions, and even subtle increases in rates or yields provided by condensation could have major impacts at the industrial scale.

The future may hold a “phase separation code” that predicts the phase behavior of any protein or nucleic acid based on its amino acid residues, domain structures, multivalent interaction capacity, and other relevant features. Generation of this code will require the synthesis of currently published (Das and Pappu, 2013; Harmon et al., 2017; Langdon and Gladfelter, 2018; Langdon et al., 2018; Nott et al., 2015; Pak et al., 2016; Quiroz and Chilkoti, 2015; Wang et al., 2018; Zhang et al., 2015) and future studies dissecting the contributions of specific amino acids, protein conformations, and nucleic acid sequences and structures. This process could be expedited by mining existing knowledge of the components of condensates. Multiple nuclear bodies can be purified (Hacot et al., 2010; Hubstenberger et al., 2017; Lam et al., 2002; Saitoh et al., 2004) and thus the sequence features of their known constituents could be analyzed in an attempt to extract information about the key driving forces of formation for different types of condensates. Further, many biomolecular condensates can be reconstituted in vitro, so the rules dictating client partitioning into different condensates could be examined using screening based approaches looking at the preferential incorporation of specific sequences from peptide or nucleic acid

libraries. No matter what strategies are used to uncover the code, determining the molecular grammar of phase separation may lead to advances in basic science, industry, and medicine.

Utilizing condensate biology in therapeutics

Condensates may represent novel drug targets in disease. In many maladies, and especially in cancer, the overexpression of a few key proteins can drive the disease phenotype. As discussed in chapter 1, concentration is a key feature that determines whether a factor will condense. If the factor in question reaches the concentration necessary for phase separation only in the diseased cells, there may be condensates that exist in diseased cells that do not exist in healthy cells, thus providing an opportunity for selective targeting. Additionally, alterations that change the multivalent interaction capacity of proteins or nucleic acids can lead to the formation of pathogenic condensates. In this section I will first provide examples of known pathogenic condensates, and then discuss how condensates could be targeted by drugs and how small molecules could be modified to better inhibit targets within condensates.

The misregulation of condensation may contribute to multiple diseases. Thus far, condensate-related diseases-associated alterations have been characterized as gain of function; in all cases alterations cause an increased propensity to form condensates which may then transition into aggregates. The well-studied phase-separating protein FUS exhibits this behavior. In some versions of the neurodegenerative diseases amyotrophic lateral sclerosis and frontotemporal lobar degeneration, FUS forms aggregate structures in the cytoplasm that may contribute to disease. In these diseases, FUS has characteristic point mutations or loss of arginine methylation on certain residues. Through slightly different mechanisms, all of these alterations effectively increase the valency of FUS interaction. Point mutations prevent binding to other proteins that normally shield interaction, or directly increase the strength of FUS interactions (Guo et al., 2018; Hofweber et al., 2018; Patel et al., 2015; Qamar et al., 2018; Yoshizawa et al., 2018), while loss of arginine methylation reveals additional sites with interaction capability (arginine methylation prevents the tyrosine-arginine interactions that drive phase separation discussed in the previous section) (Hofweber et al., 2018; Qamar et al., 2018). Each of these changes leads to enhanced ability to form condensates that are more likely to mature into aggregates. Another class of neurodegenerative diseases that involve aggregation are repeat expansion disorders. Recent work suggests aberrant phase transitions may be at the heart of these diseases as well (Jain and Vale, 2017). In vitro and in vivo experiments show that RNA transcripts containing high numbers of repeats can form condensates, and further increases to the number of repeats causes aggregation in vitro. These RNA condensates or aggregates may contribute to disease by sequestering RNA binding proteins, causing nucleolar stress, or disrupting nucleocytoplasmic transport (Freibaum et al., 2015; Miller, 2000; Tsoi et al., 2012). Finally, inappropriate condensate formation can also contribute to disease in the context of misregulated gene expression. A number of cancers are caused by fusion events between DNA binding domains and the low complexity IDRs of proteins from the FET (FUS/EWS/TAF15) family (Denny and Arvand, 2001; Guipaud et al., 2006; Lessnick and Ladanyi, 2012). These mutant fusion proteins may bind to DNA near oncogenes and form aberrant condensate-like structures that recruit RNAPII and promote transcription (Kwon et al., 2013; Wei et al., 2019).

There are multiple features of factors that form or partition into condensates that could be targeted to disrupt pathogenic condensation. Most of the phase-separating proteins discussed in this thesis contain an intrinsically disordered region as well as a structured, high affinity interaction domain that binds to nucleic acids or other proteins. The intrinsically disordered region typically mediates

the multivalent interactions that drive phase separation while the structured domain typically targets the protein to a specific location to increase effective concentration or further increase valency. For example, in addition to their IDRs, FUS, P granule proteins, and SR proteins all have structured RNA binding domains, PRC1, HP1a, and BRD4 all have structured histone binding domains, TFs have a structured DNA binding domain, and Mediator subunits have structured domains that enable interaction with other subunits. RNA that participates in phase separation may rely on multiple specific base pairing interactions. Theoretically, preventing any of these interaction types could disrupt pathogenic condensates.

Direct therapeutic targeting of the weak multivalent interactions that drive phase separation may require the development of new classes of small molecules. Historically, successful small molecule drugs have been designed to bind tightly and specifically to binding pockets on proteins leading to the disruption of structured protein interactions or the inhibition of enzymatic activity (Gao and Skolnick, 2013). Given the transient and dynamic nature of the interactions that drive condensate formation and the lack of structure in IDRs, this type of small molecule is poorly suited to disrupt condensates directly. Instead, molecules that affect the physical chemistry of phase separating proteins perhaps by competing for transient interactions are required. A recent preprint has provided proof of principle for these types of small molecules and shown that they can act with specificity towards different types of condensates (Wheeler et al., 2019). An especially exciting prospect is that these “physicochemical” drugs could be used to target infamously undruggable proteins (Dang et al., 2017) such as the Myc transcription factor that lack binding pockets but participate in condensation.

Because phase-separating factors typically participate in specific structured interactions in addition to IDR-mediated interactions, more conventional drugs may also disrupt condensates. Preliminary data from the Young lab shows that inhibition of the structured binding of BRD4 to acetylated histones by the small molecule JQ1 leads to a near complete loss of detectable BRD4 containing condensates. This may occur because BRD4 is no longer locally concentrated on the genome and thus its effective concentration falls below that which is required for condensation. It is likely that many drugs have similar effects on condensates in addition to their known effects, and we are simply unaware of these because we have not yet viewed therapeutics through the condensate lens. Because RNA appears to be a key factor in many condensates, the use of antisense oligos to disrupt certain secondary structures or hide protein interaction sites could also prove beneficial. Thus disrupting specific structured interactions of key pathogenic condensate components could lead to condensate dissolution and therapeutic benefit.

A better understanding of the rules that govern client partitioning in condensates could lead to the development of small molecules with improved pharmacodynamic properties. Condensates selectively concentrate or exclude specific proteins based on their physicochemical properties, so it is likely that this concept extends to small molecules as well. Indeed, preliminary data from the Young lab has shown that a panel of clinically successful drugs are preferentially concentrated in the condensates that contain their targets as compared to unrelated condensates. In some cases, this elevated concentration in specific condensates appears to occur even in the absence of the target, indicating that the presence of the target is not required for the concentrating property. Thus the ability to partition into the right condensate may be a general feature of successful small molecules, and it is possible that some drugs fail simply because they cannot gain access to their targets in condensates. Understanding the molecular grammar of phase separation could allow the development of an array of chemical moieties that specifically partition into different condensates. These moieties could then be attached to small molecule warheads to guide drugs to the proper condensate that contains their target. This type of strategy would likely

increase the efficacy and reduce the required dosage of the drug. Thus future therapeutics may be able to take advantage of multiple aspects of condensate biology.

Concluding thoughts

The past half century of research has provided great insight into the control of gene expression, and in the last several years the importance of phase separation in this control has been realized. Transcription factors bind to regulatory regions to form enhancers that regulate gene transcription by forming a DNA loop with promoters, and enhancer activity is constrained by genome structure. At some genes, enhancers loop to promoters via interactions mediated by the protein CTCF (Chapter 2). Exceptionally strong enhancers known as super-enhancers can form phase-separated condensates that concentrate and compartmentalize the transcriptional machinery (Appendix I and II). RNA Polymerase II partitions in these condensates through interactions mediated by its C-terminal domain, and phosphorylation of the C-terminal domain leads to eviction from transcriptional condensates and partitioning into condensates associated with co-transcriptional splicing (Chapter 3). In the coming years, we may gain a more complete understanding of the specific interactions that drive biological condensation. This understanding is likely to drive advances in basic science, industry, and medicine. Discoveries made in the next decade of research will continue to uncover the utility of phase separation in biology and force us to constantly update our views on the control of gene expression.

References

Dang, C. V., Reddy, E.P., Shokat, K.M., and Soucek, L. (2017). Drugging the “undruggable” cancer targets. Nat. Rev. Cancer *17*, 502–508.

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

Denny, C.T., and Arvand, A. (2001). Biology of EWS/ETS fusions in Ewing’s family tumors. Oncogene *20*, 5747–5754.

Freibaum, B.D., Lu, Y., Lopez-Gonzalez, R., Kim, N.C., Almeida, S., Lee, K.-H., Badders, N., Valentine, M., Miller, B.L., Wong, P.C., et al. (2015). GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature *525*, 129–133.

Gao, M., and Skolnick, J. (2013). A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. PLoS Comput. Biol. *9*.

Gao, A., Shrinivas, K., Lepeudry, P., Suzuki, H.I., Sharp, P.A., and Chakraborty, A.K. (2018). Evolution of weak cooperative interactions for biological specificity. Proc. Natl. Acad. Sci. 201815912.

Guipaud, O., Guillonneau, F., Labas, V., Praseuth, D., Rossier, J., Lopez, B., and Bertrand, P. (2006). An in vitro enzymatic assay coupled to proteomics analysis reveals a new DNA processing activity for Ewing sarcoma and TAF(II)68 proteins. Proteomics *6*, 5962–5972.

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.

Hacot, S., Coute, Y., Belin, S., Alexandra, M., Mertani, H.C., Sanchez, J.-C., Rosa-Calatrava, M., and Diaz, J.J. (2010). Isolation of nucleoli. Curr. Protoc. Cell Biol.

Harmon, T.S., Holehouse, A.S., Rosen, M.K., and Pappu, R. V. (2017). Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. Elife *6*, 1–31.

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.

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

Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. Nature *546*, 243–247.

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.

Lam, Y., Lyon, C.E., and Lamond, A.I. (2002). Large-scale isolation of Cajal bodies from HeLa cells. Mol. Biol. Cell *13*.

Langdon, E.M., and Gladfelter, A.S. (2018). A New Lens for RNA Localization: Liquid-Liquid Phase Separation. Annu. Rev. Microbiol. *72*, 255–271.

Langdon, E.M., Qiu, Y., Ghanbari Niaki, A., McLaughlin, G.A., Weidmann, C.A., Gerbich, T.M., Smith, J.A., Crutchley, J.M., Termini, C.M., Weeks, K.M., et al. (2018). mRNA structure determines specificity of a polyQ-driven phase separation. Science (80-.). *360*,

922–927.
Lessnick, S.L., and Ladanyi, M. (2012). Molecular Pathogenesis of Ewing Sarcoma: New Therapeutic and Transcriptional Targets. Annu. Rev. Pathol. Mech. Dis. *7*, 145–159.
Mikhaleva, S., and Lemke, E.A. (2018). Beyond the Transport Function of Import Receptors: What's All the FUS about? Cell *173*, 549–553.
Miller, J.W. (2000). Recruitment of human muscleblind proteins to (CUG)n expansions associated with myotonic dystrophy. EMBO J. *19*, 4439–4448.
Murray, D.T., Kato, M., Lin, Y., Thurber, K.R., Hung, I., McKnight, S.L., and Tycko, R. (2017). Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. Cell *171*, 615-627.e16.
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.
Pak, C.W., Kosno, M., Holehouse, A.S., Padrick, S.B., Mittal, A., Ali, R., Yunus, A.A., Liu, D.R., Pappu, R. V., and Rosen, M.K. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. Mol. Cell *63*, 72–85.
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.
Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Bennett, C.F., Cleveland, D.W., and Yeo, G.W. (2012). Misregulated RNA processing in amyotrophic lateral sclerosis. Brain Res. *1462*, 3–15.
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.
Quiroz, F.G., and Chilkoti, A. (2015). Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. Nat. Mater. *14*, 1164–1171.
Saitoh, N., Spahr, C.S., Patterson, S.D., Bubulya, P., Neuwald, A.F., and Spector, D.L. (2004). Proteomic Analysis of Interchromatin Granule Clusters. Mol. Biol. Cell *15*.
Tsoi, H., Lau, T.C.-K., Tsang, S.-Y., Lau, K.-F., and Chan, H.Y.E. (2012). CAG expansion induces nucleolar stress in polyglutamine diseases. Proc. Natl. Acad. Sci. *109*, 13428–13433.
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 *0*, 1–12.
Wang, W.Y., Pan, L., Su, S.C., Quinn, E.J., Sasaki, M., Jimenez, J.C., MacKenzie, I.R.A., Huang, E.J., and Tsai, L.H. (2013). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. Nat. Neurosci. *16*, 1383–1391.
Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126–130.
Wei, M., Chang, Y., Shimobayashi, S.F., Shin, Y., and Brangwynne, C.P. (2019). Nucleated transcriptional condensates amplify gene expression. BioRxiv.
Wheeler, R.J., Lee, H.O., Boczek, E., Bickle, M., Alberti, S., and Hyman, A.A. (2019). Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. BioRxiv.
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.
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.

APPENDIX I: COACTIVATOR CONDENSATION AT SUPER-ENHANCERS LINKS PHASE SEPARATION AND GENE CONTROL

Originally published in *Science* Volume 361, Issue 6400. (2018)

RESEARCH ARTICLE

MOLECULAR BIOLOGY

Coactivator condensation at super-enhancers links phase separation and gene control

Benjamin R. Sabari[1]*, Alessandra Dall'Agnese[1]*, Ann Boija[1], Isaac A. Klein[1,2], Eliot L. Coffey[1,3], Krishna Shrinivas[4,5], Brian J. Abraham[1], Nancy M. Hannett[1], Alicia V. Zamudio[1,3], John C. Manteiga[1,3], Charles H. Li[1,3], Yang E. Guo[1], Daniel S. Day[1], Jurian Schuijers[1], Eliza Vasile[6], Sohail Malik[7], Denes Hnisz[1], Tong Ihn Lee[1], Ibrahim I. Cisse[8], Robert G. Roeder[7], Phillip A. Sharp[3,6], Arup K. Chakraborty[4,5,8,9,10,11], Richard A. Young[1,3]†

Super-enhancers (SEs) are clusters of enhancers that cooperatively assemble a high density of the transcriptional apparatus to drive robust expression of genes with prominent roles in cell identity. Here we demonstrate that the SE-enriched transcriptional coactivators BRD4 and MED1 form nuclear puncta at SEs that exhibit properties of liquid-like condensates and are disrupted by chemicals that perturb condensates. The intrinsically disordered regions (IDRs) of BRD4 and MED1 can form phase-separated droplets, and MED1-IDR droplets can compartmentalize and concentrate the transcription apparatus from nuclear extracts. These results support the idea that coactivators form phase-separated condensates at SEs that compartmentalize and concentrate the transcription apparatus, suggest a role for coactivator IDRs in this process, and offer insights into mechanisms involved in the control of key cell-identity genes.

Phase separation of fluids is a physicochemical process by which molecules separate into a dense phase and a dilute phase. Phase-separated biomolecular condensates, which include the nucleolus, nuclear speckles, stress granules, and others, provide a mechanism to compartmentalize and concentrate biochemical reactions within cells (*1*–*3*). Biomolecular condensates produced by liquid-liquid phase separation allow rapid movement of components into and within the dense phase and exhibit properties of liquid droplets such as fusion and fission (*4*). Dynamic and cooperative multivalent interactions among molecules, such as those produced by certain intrinsically disordered regions (IDRs) of proteins, have been implicated in liquid-liquid phase separation (*5*–*7*).

Enhancers are gene regulatory elements bound by transcription factors (TFs) and other components of the transcription apparatus that function to regulate expression of cell type–specific genes (*8*–*13*). Super-enhancers (SEs)—clusters of enhancers that are occupied by exceptionally high densities of transcriptional machinery—regulate genes with especially important roles in cell identity (*14*, *15*). DNA interaction data show that enhancer elements in the clusters are in close spatial proximity with each other and the promoters of the genes that they regulate (*16*–*18*), consistent with the notion of a dense assembly of transcriptional machinery at these sites. This high-density assembly at SEs has been shown to exhibit sharp transitions of formation and dissolution, forming as the consequence of a single nucleation event (*19*, *20*) and collapsing when concentrated factors are depleted from chromatin (*21*–*25*) or when nucleation sites are deleted (*26*–*29*). These properties of SEs led to the proposal that the high-density assembly of biomolecules at active SEs is due to phase separation of enriched factors at these genetic elements (*30*). Here we provide experimental evidence that the transcriptional coactivators BRD4 and MED1 (a subunit of the Mediator complex) form condensates at SEs. This establishes a new framework to account for the diverse properties described for these regulatory elements and expands the known biochemical processes regulated by phase separation to include the control of cell-identity genes.

[1]Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142, USA. [2]Department of Medical Oncology, 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 Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, NY 10065, USA. [8]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [9]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [10]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [11]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard, Cambridge, MA 02139, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: young@wi.mit.edu

BRD4 and MED1 coactivators form nuclear puncta

The enhancer clusters that make up SEs are occupied by master TFs and unusually high densities of factors, including BRD4 and MED1, that are coactivators (*31*–*35*) whose presence can be used to define SEs (*14*, *15*, *21*). We reasoned that if BRD4 and MED1 are components of nuclear condensates, then they might be visualized as discrete puncta in the nuclei of cells, and the properties of these puncta could be investigated. Fixed cell immunofluorescence (IF) with antibodies against BRD4 and MED1 in murine embryonic stem cells (mESCs) revealed nuclear puncta for both factors (Fig. 1A). To determine whether such puncta occur in live cells, we engineered mESCs by using CRISPR-Cas9 to tag endogenous BRD4 and MED1 with monomeric enhanced green fluorescent protein (mEGFP) (fig. S1). Live-cell fluorescence microscopy of the engineered mESC lines also revealed discrete nuclear puncta (Fig. 1B). Analysis of these images showed that there were 1034 ± 130 BRD4 and 983 ± 102 MED1 puncta per nucleus (means ± SEM) (table S1). These results demonstrate that BRD4 and MED1 are components of puncta within the nuclei of mESCs.

SEs are associated with coactivator puncta

Several lines of evidence suggest that SEs are likely to be associated with some of the BRD4 and MED1 puncta in mESCs. ChIP-seq (chromatin immunoprecipitation followed by sequencing) data for BRD4 and MED1 show that SEs are especially enriched in these coactivators (*14*, *15*). DNA interaction data suggest that SE constituents occupied by BRD4 and MED1 are in close spatial proximity to one another (Fig. 1C and fig. S2A). Co-occupancy of the genome by BRD4 and MED1 is most evident at SEs (fig. S2B) (*14*, *15*). To determine whether SEs are associated with some of the BRD4 and MED1 puncta, we performed IF for BRD4 or MED1 together with DNA-FISH or nascent RNA-FISH for the genomic region containing the *Nanog* gene and its SEs (FISH, fluorescence in situ hybridization) (Fig. 1, D to G). We found that BRD4 and MED1 puncta consistently overlapped the DNA-FISH foci (Fig. 1D) or RNA-FISH foci (Fig. 1F). An average image analysis (details are given in the methods) of the BRD4 and MED1 IF signals centered at DNA-FISH foci (n = 137 for BRD4 and 125 for MED1) and RNA-FISH foci (n = 121 for BRD4 and 181 for MED1) revealed that, on average, BRD4 and MED1 fluorescence intensities are most enriched at the center of FISH foci (Fig. 1, E and G); this trend was not observed for average images centered at randomly selected nuclear positions (Fig. 1, E and G). Radial distribution functions of the averaged images for FISH and IF pairs show a significant correlation (Spearman correlation

coefficients > 0.6; P values < 1×10^{-16}), with both BRD4 and MED1 having their highest signal intensities at the center of the FISH focus; signals decay with distance from this center (fig. S3). The radial distributions of FISH and IF at randomly selected nuclear positions are not correlated (Spearman correlation coefficients < 0.2) (fig. S3). Similar results were obtained when we performed IF for BRD4 or MED1 together with nascent RNA-FISH for the SE-regulated genes *Klf4*, *Mir290*, and *Trim28* (figs. S3 and S4, A to F). When a similar experiment was conducted for two genes expressed in mESCs but not associated with a SE (*Fam168b* and *Zfp606*), there was no evident overlap between FISH foci

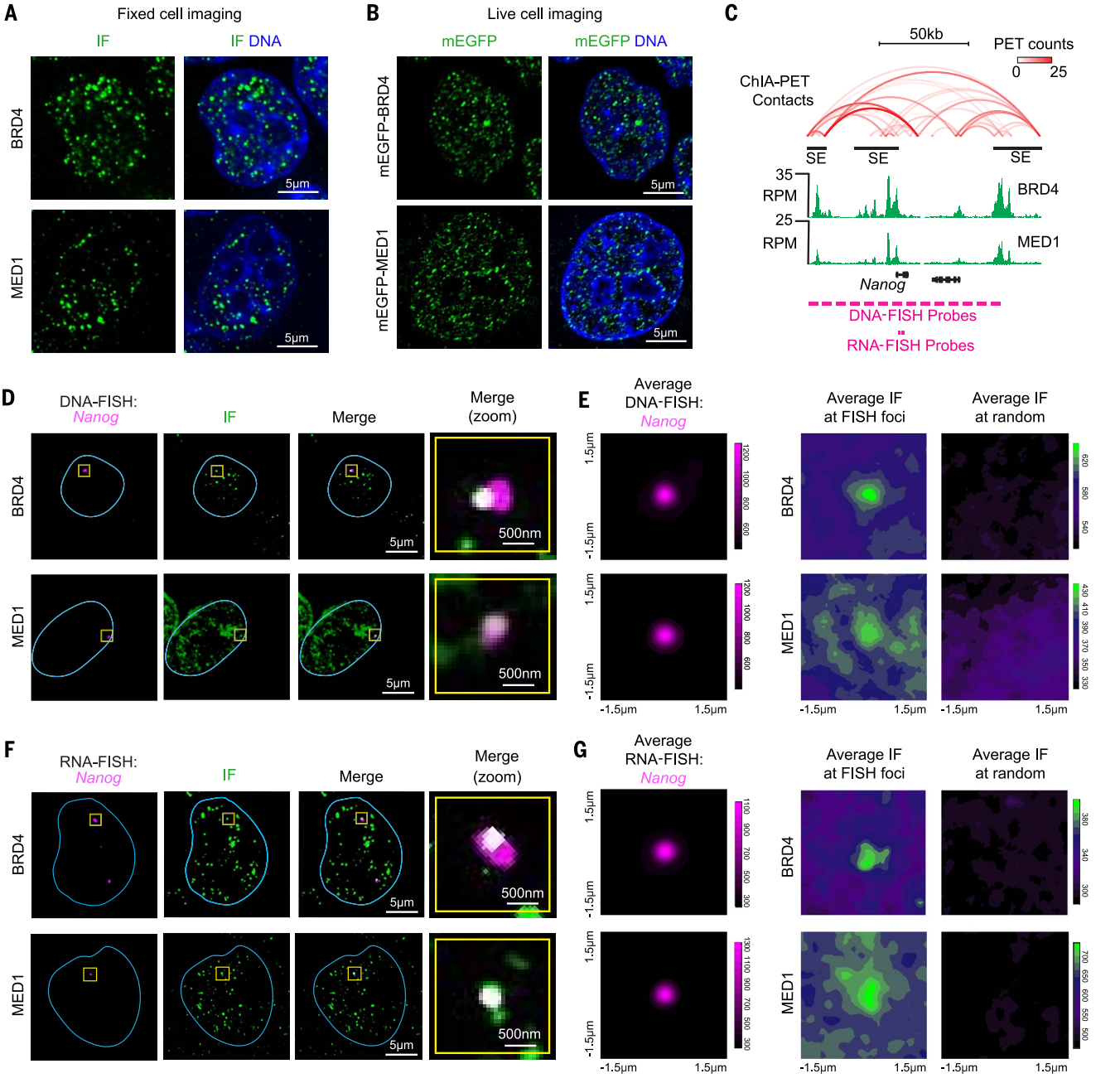


Fig. 1. BRD4 and MED1 form puncta at super-enhancers (SEs). (**A**) Immunofluorescence (IF) imaging of BRD4 and MED1 in mouse embryonic stem cells (mESCs). Fluorescence signal is shown alone (left) and merged with Hoechst stain (right). (**B**) Live imaging of endogenously tagged mEGFP-BRD4 and mEGFP-MED1 in mESCs. (**C**) Depiction of *Nanog* locus, associated SEs (black bars), DNA contacts (red arcs), BRD4 and MED1 ChIP-seq (green histograms), and location of FISH probes. ChIA-PET, chromatin interaction analysis with paired-end tag; RPM, reads per million. (**D**) Colocalization between BRD4 or MED1 and the *Nanog* locus by IF and DNA-FISH in fixed mESCs. Separate images of the indicated IF and FISH are shown, along with an image showing the merged channels (overlapping signal in white). The blue line highlights the nuclear periphery, determined by Hoechst staining (not shown). The rightmost column shows the area in the yellow box in greater detail. (**E**) Averaged signal of (left) DNA-FISH for *Nanog* and (right) IF for BRD4 or MED1 centered at *Nanog* DNA-FISH foci or randomly selected nuclear positions. (**F**) Colocalization between BRD4 or MED1 and the nascent RNA of *Nanog*, determined by IF and RNA-FISH in fixed mESCs. Data are shown as in (D). (**G**) Averaged signal of (left) RNA-FISH for *Nanog* and (right) IF for BRD4 or MED1 centered at *Nanog* RNA-FISH foci or randomly selected nuclear positions.

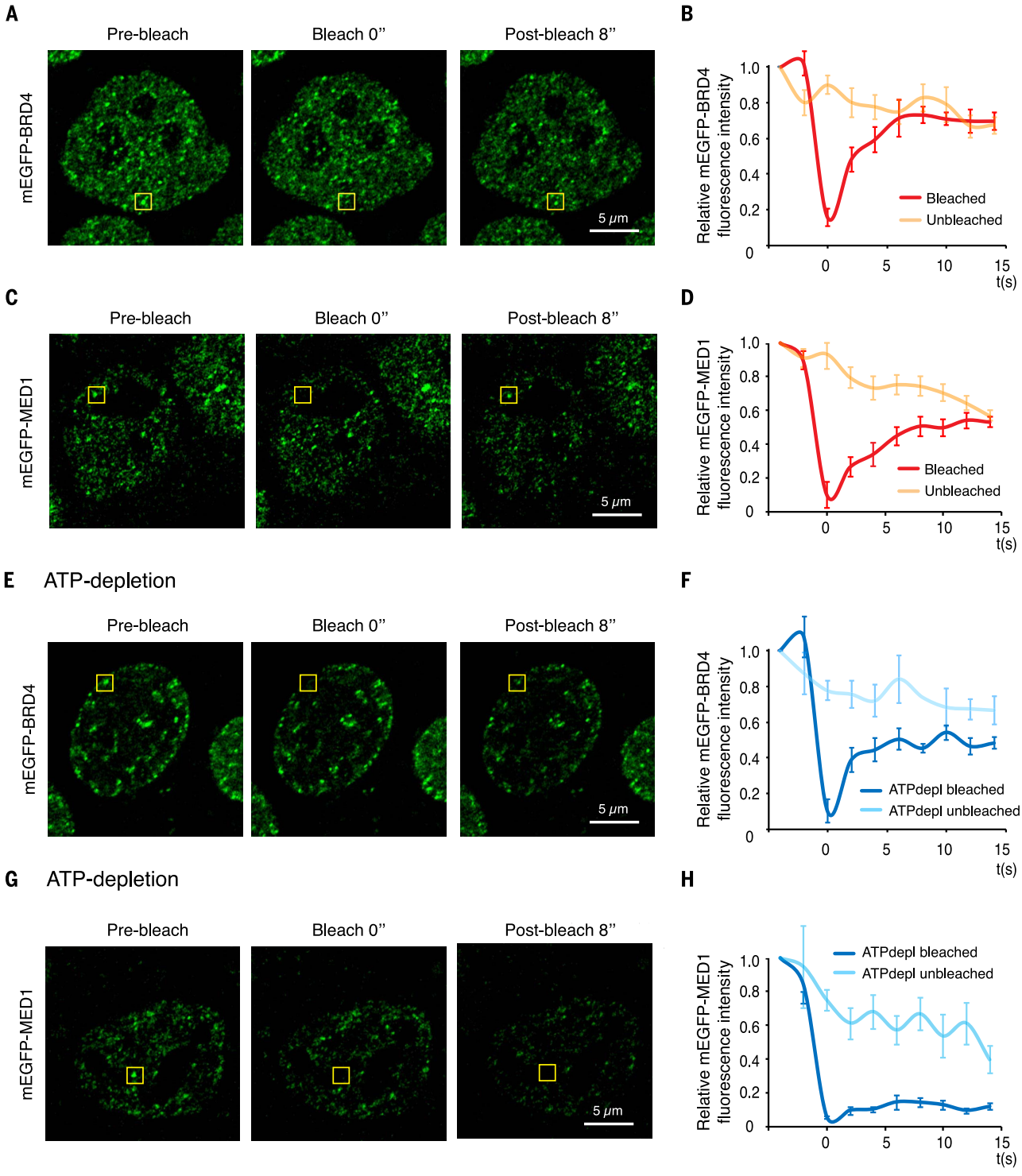


Fig. 2. BRD4 and MED1 nuclear puncta exhibit properties expected for biomolecular condensates. (**A**) Representative images of the FRAP experiment with mEGFP-BRD4–engineered mESCs (where " indicates time in seconds). The yellow box highlights the punctum undergoing targeted bleaching. (**B**) Quantification of FRAP data for mEGFP-BRD4 puncta. The bleaching event occurs at t = 0 s. For the bleached area and the unbleached control, background-subtracted fluorescence intensities are plotted relative to a prebleach time point (t = −4 s). Data are plotted as means ± SEM (n = 9). (**C**) Same as (A), but with mEGFP-MED1–engineered mESCs. (**D**) Same as (B), but for mEGFP-MED1 puncta (n = 9). (**E**) Representative images of the FRAP experiment with mEGFP-BRD4–engineered mESCs upon ATP depletion. (**F**) Quantification of FRAP data for mEGFP-BRD4 puncta upon ATP depletion (n = 8), as in (B). (**G**) Representative images of the FRAP experiment with mEGFP-MED1–engineered mESCs upon ATP depletion. (**H**) Quantification of FRAP data for mEGFP-MED1 puncta upon ATP depletion (n = 8), as in (B). Images were taken using the Zeiss LSM 880 confocal microscope with an Airyscan detector and a 63× objective at 37°C.

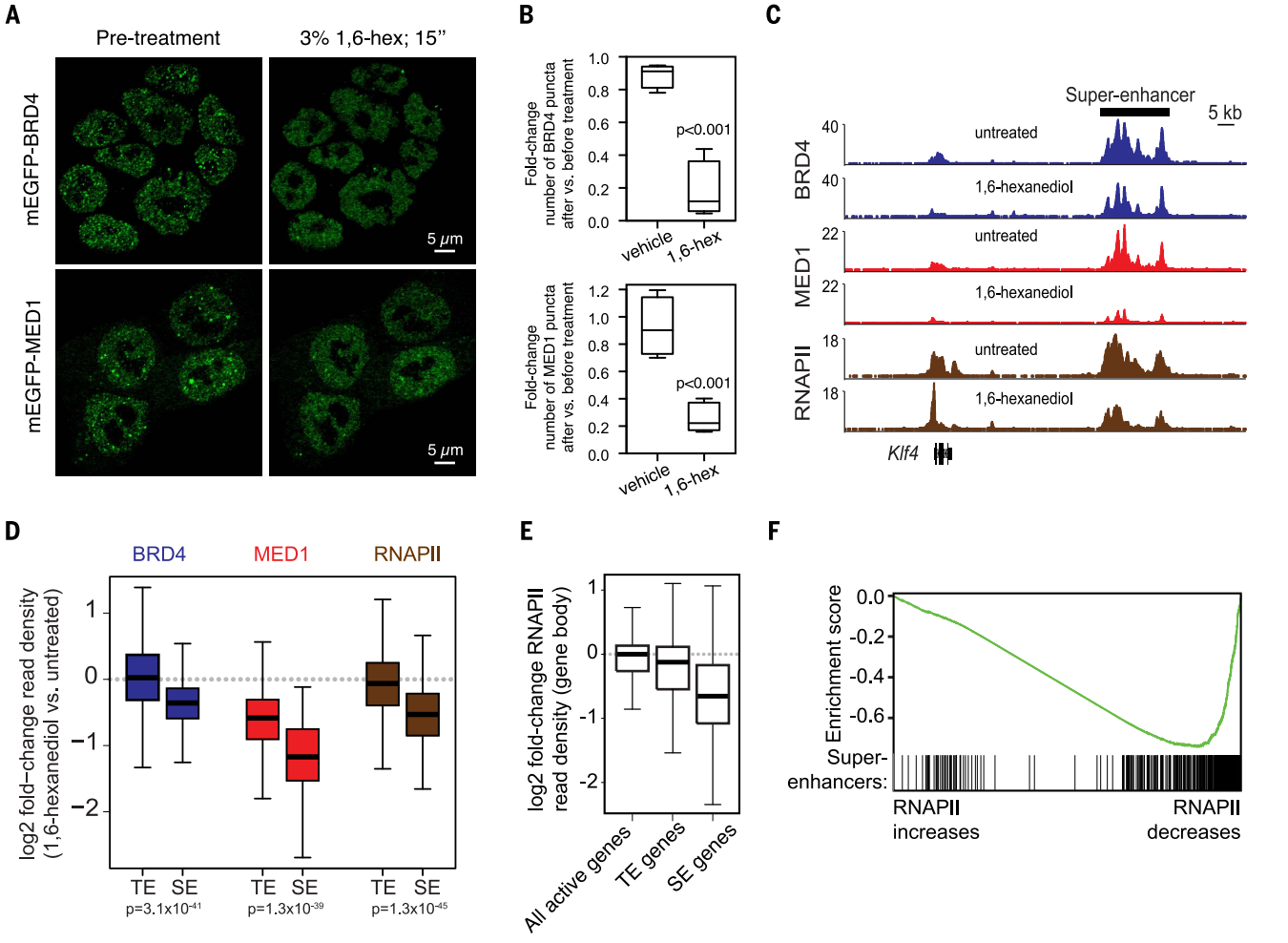


Fig. 3. 1,6-hexanediol disrupts BRD4 and MED1 puncta and disrupts BRD4, MED1, and RNA Pol II occupancy at SEs and SE-driven genes. (**A**) Representative images of mEGFP-BRD4– or mEGP-MED1–engineered mESCs before and after treatment with 3% hexanediol for 15 s. (**B**) Box-plot representation of the fold change in the number of mEGFP-BRD4 or mEGFP-MED1 puncta observed after addition of vehicle or 1,6-hexanediol to a final concentration of 3%. (**C**) Genome browser view of BRD4 (blue), MED1 (red), and RNA Pol II (RNAPII, brown) ChIP-seq data from untreated or 1,6-hexanediol–treated (1.5% for 30 min) mESCs at the *Klf4* locus. The *y* axis shows reads per million. (**D**) Box-plot representation of the log2 fold change in BRD4 (blue), MED1 (red), and RNA Pol II (brown) ChIP-seq read density (1,6-hexanediol–treated versus untreated) for regions defined as SEs or typical enhancers (TEs) (methods and table S2). (**E**) Box-plot representation of the log2 fold change in RNA Pol II ChIP-seq density (1,6-hexanediol–treated versus untreated) within the gene body (transcription start site to transcription end site) of all active genes (reads per kilobase per million > 1), TE-associated genes, or SE-associated genes. (**F**) Gene set enrichment analysis, with genes ranked by their log2 fold change in RNA Pol II ChIP-seq density within the gene body and annotated against the set of SE-associated genes. Enrichment score profile and the position of SE-associated genes are shown.

and BRD4 puncta (fig. S4G). These results indicate that both BRD4 and MED1 puncta are present at SEs.

Coactivator puncta exhibit liquid-like rates of fluorescence recovery after photobleaching

We next sought to examine whether BRD4 and MED1 puncta exhibit features characteristic of liquid-like condensates. A hallmark of liquid-like condensates is internal dynamical reorganization and rapid exchange kinetics (*1*–*3*), which can be interrogated by measuring the rate of fluorescence recovery after photobleaching (FRAP). To study the dynamics of BRD4 and MED1 foci in live cells, we performed FRAP experiments on endogenously tagged mEGFP-BRD4 or mEGFP-MED1 cell lines. After photobleaching, mEGFP-BRD4 and mEGFP-MED1 puncta recovered fluorescence on a time scale of seconds (Fig. 2, A to D), with apparent diffusion coefficients of ~0.37 ± 0.13 and ~0.14 ± 0.04 $\mu m^2/s$, respectively. These values are similar to those previously reported for components of liquid-like condensates (*36*, *37*). Adenosine triphosphate (ATP) has been implicated in promoting condensate fluidity by driving energy-dependent processes and/or through its intrinsic hydrotrope activity (*38*, *39*). Depletion of cellular ATP by glucose deprivation and oligomycin treatment altered fluorescence recovery after photobleaching for both mEGFP-BRD4 and mEGFP-MED1 foci; the rate of recovery for MED1 was reduced, and the extent of recovery for BRD4 was diminished (Fig. 2, E to H). These results indicate that puncta containing BRD4 and MED1 have liquid-like properties in cells, consistent with previously described phase-separated condensates.

Coactivator puncta and SE occupancy are sensitive to condensate perturbation

To further investigate the biophysical properties of BRD4 and MED1 puncta, we investigated their sensitivity to 1,6-hexanediol, a compound known to disrupt liquid-like condensates, possibly by disruption of hydrophobic interactions (*40*). We found that treatment of mESCs expressing endogenously tagged mEGFP-BRD4 or mEGFP-MED1

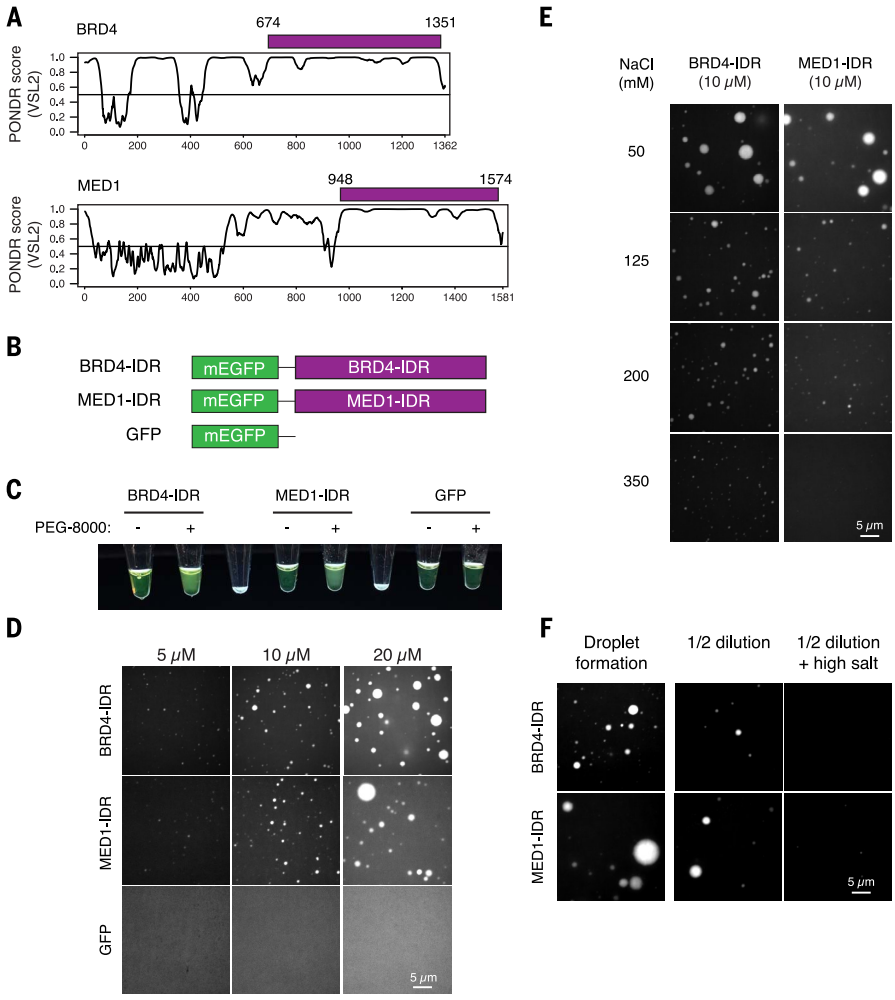


Fig. 4. Intrinsically disordered regions (IDRs) of BRD4 and MED1 phase-separate in vitro. (**A**) Graphs plotting intrinsic disorder for BRD4 and MED1. PONDR (Predictor of Natural Disordered Regions) VSL2 scores are shown on the *y* axis, and amino acid positions are shown on the *x* axis. The purple bar designates the IDR under investigation. (**B**) Schematic of recombinant mEGFP fusion proteins used in this study. Purple boxes indicate the IDRs of BRD4 and MED1 shown in (A). (**C**) Visualization of turbidity associated with droplet formation. Tubes containing BRD4-IDR (left pair), MED1-IDR (middle pair), or GFP (right pair) in the presence (+) or absence (−) of PEG-8000 are shown. Blank tubes are included between pairs for contrast. (**D**) Representative images of droplet formation at different protein concentrations. BRD4-IDR, MED1-IDR, or mEGFP were added to the droplet formation buffer to the final concentrations indicated. (**E**) Representative images of droplet formation at different salt concentrations. BRD4-IDR or MED1-IDR was added to droplet formation buffer to achieve 10 μM protein concentration with a final NaCl concentration as indicated. (**F**) Representative images of the droplet reversibility experiment with BRD4-IDR (top row) or MED1-IDR (bottom row) [20 μM protein and 75 mM NaCl (initial), followed by a 1:1 dilution (1/2 dilution) or a 1:1 dilution with an increase to 425 mM NaCl (1/2 dilution + NaCl)].

with 1,6-hexanediol caused a reduction in the number of BRD4 and MED1 puncta (Fig. 3, A and B).

To determine the effect of 1,6-hexanediol on BRD4, MED1, and RNA polymerase II (RNA Pol II) occupancy at enhancers and genes, ChIP-seq was performed with antibodies against these proteins in untreated or 1,6-hexanediol–treated mESCs. Treatment with 1,6-hexanediol caused a reduction in all three proteins at enhancers, with the most profound effects occurring at SEs (Fig. 3, C and D, and fig. S5A). For example, at the *Klf4* SE, the levels of BRD4 were reduced by 44%, those of MED1 by 80%, and those of RNA Pol II by 56% (Fig. 3C). Similar effects were observed genome-wide, where reductions in BRD4, MED1, and RNA Pol II were substantially larger at SEs than at typical enhancers (Fig. 3D), and the degrees to which BRD4 and MED1 were lost from SEs were positively correlated (fig. S5B). These results are consistent with the notion that BRD4 and MED1 form condensates at SEs that are sensitive to 1,6-hexanediol.

The level of RNA Pol II occupancy across gene bodies can be used as a measure of transcriptional output (*41*). The ChIP-seq data revealed that the reduction in BRD4 and MED1 occupancy at SEs was associated with a loss of RNA Pol II occupancy across SE-associated gene bodies (Fig. 3, C and E, and fig. S5A). When genes were ranked by the extent to which RNA Pol II was lost upon 1,6-hexanediol treatment, SE-associated genes were highly enriched among those that lost the most RNA Pol II (Fig. 3F). These results are consistent with the idea that BRD4 and MED1 condensates are associated with SEs and that loss of condensate integrity adversely affects transcription.

IDRs of BRD4 and MED1 phase-separate in vitro

BRD4 and MED1 contain large IDRs (Fig. 4A) and share features with the IDRs of several proteins known to facilitate condensate formation (*2*, *3*), including high proline and glutamine content (BRD4), high serine content (MED1), and acidic and basic regions (BRD4 and MED1). The purified IDRs of several proteins involved in

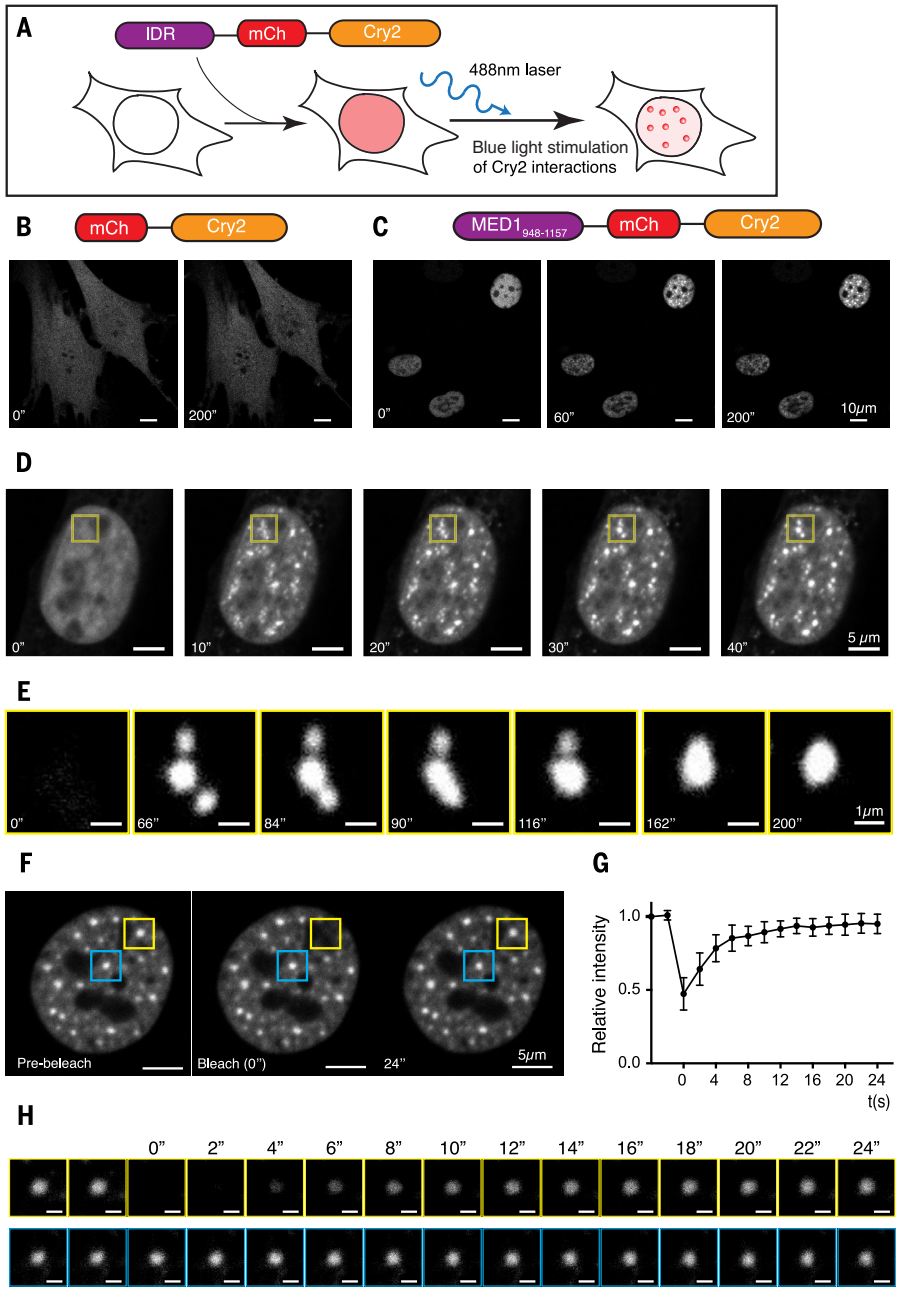


Fig. 5. The IDR of MED1 participates in phase separation in cells. (**A**) Schematic of the optoIDR assay, depicting recombinant protein with an IDR (purple), mCherry (red), and Cry2 (orange) expressed in cells exposed to blue light. (**B** and **C**) Images of NIH3T3 cells expressing either (B) mCherry-Cry2 or (C) a portion of the MED1-IDR (amino acids 948 to 1157) fused to mCherry-Cry2 (MED1-optoIDR). Cells were subjected to laser excitation every 2 s for the indicated times. (**D**) Time-lapse images of the nucleus of an NIH3T3 cell expressing MED1-optoIDR subjected to laser excitation every 2 s for the times indicated. A droplet fusion event occurs in the region highlighted by the yellow box. (**E**) The droplet fusion event highlighted in (D) at higher resolution and extended times as indicated. (**F**) Image of a MED1-optoIDR optoDroplet (yellow box) before (left), during (middle), and after (right) photobleaching. The blue box highlights an unbleached region for comparison. Time relative to photobleaching (0 s) is indicated. (**G**) Signal intensity relative to the prebleaching signal (*y* axis) versus time relative to photobleaching (*x* axis). Data are shown as average relative intensity ± SD (n = 15). (**H**) Time-lapse and close-up view of droplet recovery for regions highlighted in (F). Times relative to photobleaching are indicated. Scale bars, 1 μm.

condensate formation form phase-separated droplets in vitro (*7*, *36*, *37*, *42*), so we investigated whether the IDRs of BRD4 or MED1 form such droplets in vitro. Purified recombinant mEGFP-IDR fusion proteins (BRD4-IDR and MED1-IDR) (Fig. 4B) were added to buffers containing 10% PEG-8000 (polyethylene glycol, molecular weight 8000; materials and methods), turning the solution opaque, whereas equivalent solutions with mEGFP alone remained clear (Fig. 4C). Fluorescence microscopy of the opaque MED1-IDR and BRD4-IDR solutions revealed GFP-positive, micron-sized spherical droplets freely moving in solution (Movies 1 and 2) and falling onto and wetting the surface of the glass coverslip, where they remained stationary (Movie 3). As determined by aspect ratio analysis, the MED1-IDR and BRD4-IDR droplets were highly spherical (fig. S6A), a property expected for liquid-like droplets (*1*–*3*).

Phase-separated droplets typically scale in size according to the concentration of components in the system (*43*). We performed the droplet formation assay with varying concentrations of BRD4-IDR, MED1-IDR, and mEGFP, ranging from 0.625 to 20 μM. BRD4-IDR and MED1-IDR formed droplets with concentration-dependent size distributions, whereas mEGFP remained diffuse in all conditions tested (Fig. 4D and fig. S6B). Although these droplets were smaller at lower concentrations, we observed BRD4-IDR and MED1-IDR droplets at the lowest concentration tested (0.625 μM) (fig. S6C).

To investigate the biophysical properties of these droplets, we tested their ability to form under varying salt concentrations (to probe the contribution of electrostatic interactions) or upon 1,6-hexanediol treatment (to probe the contribution of hydrophobic interactions). The size distributions of both BRD4-IDR and MED1-IDR droplets shifted toward smaller droplets with increasing NaCl concentration (from 50 to 350 mM) (Fig. 4E and fig. S6D), and opacity was reduced with 10% 1,6-hexanediol treatment (fig. S7A). These results demonstrate that a variety of molecular interactions contribute to BRD4-IDR and MED1-IDR droplet formation.

We next sought to test whether the droplets are irreversible aggregates or reversible phase-separated condensates. To do this, BRD4-IDR and MED1-IDR were allowed to form droplets in an initial solution. The protein concentration was then diluted by half in equimolar salt or in

a higher salt solution (Fig. 4F). The preformed BRD4-IDR and MED1-IDR droplets were reduced in size and number with dilution and even further reduced with elevated salt concentration (Fig. 4F and fig. S7B). These results show that the BRD4-IDR and MED1-IDR droplets form a distribution of sizes that is dependent on the conditions of the system and, once formed, respond to changes in the system, with rapid adjustments in size. These features are characteristic of phase-separated condensates formed by networks of weak protein-protein interactions (*1*–*3*).

MED1-IDR participates in liquid-liquid phase separation in cells

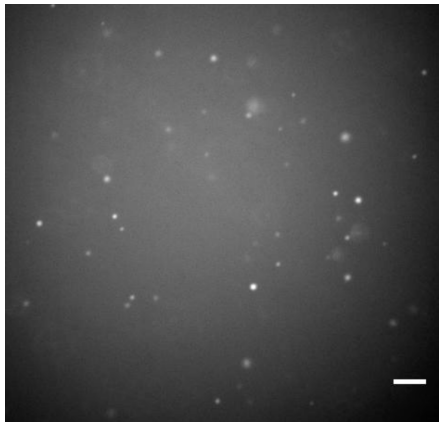
To investigate whether the coactivator IDRs facilitate phase separation in cells, we used a previously developed assay to manipulate local protein concentrations within the cell; this optoIDR assay tests IDR-dependent, light-inducible droplet formation in vivo (*44*). Briefly, the photoactivatable, self-associating Cry2 protein was labeled with mCherry and fused to an IDR of interest. This fusion mediates a blue light–inducible increase in the local concentration of selected IDRs within the cell (Fig. 5A) (*44*). In this assay, IDRs known to promote phase separation enhance the photoresponsive clustering properties of Cry2 (*45*, *46*), causing rapid formation of liquid-like spherical droplets under stimulation by blue light. Fusion of a portion of the MED1-IDR to Cry2-mCherry facilitated the rapid formation of micron-sized spherical droplets upon blue light stimulation (optoDroplets) (Fig. 5, B and C, and fig. S8). During stimulation, proximal droplets were observed to fuse (Fig. 5, D and E, and Movie 4). The fusions exhibited characteristic liquid-like fusion properties of necking and relaxation to spherical shape (Fig. 5E). The MED1-IDR droplets persisted after blue light stimulation and exhibited liquid-like FRAP recovery rates in the absence of blue light stimulation (Fig. 5, F to H). The rapid FRAP kinetics in the absence of light-activated Cry2 interactions suggests that the MED1-IDR optoDroplets established by blue light are dynamic assemblies exchanging with the dilute phase.

Conserved serine bias in the MED1-IDR is necessary for phase separation

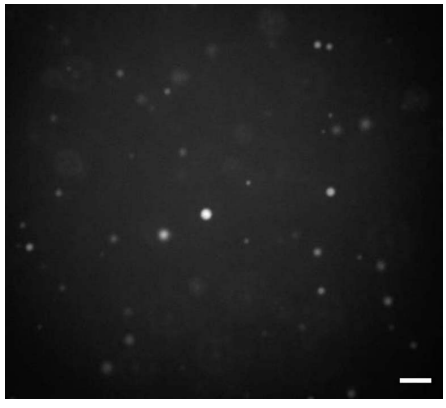
Previous studies have implicated low-complexity IDRs of proteins in liquid-liquid phase separation (*7*, *36*, *37*, *42*). An examination of the amino acid content of MED1 revealed that the IDR contains a compositional bias for serine (Fig. 6A). This serine compositional bias is conserved among vertebrates (Fig. 6B). To investigate whether the serine bias is necessary for the MED1-IDR's capacity to phase-separate, we mutated all the serine (S) residues to alanine (A) and investigated the ability of this mutated IDR to form phase-separated droplets in vitro. The MED1-IDR S-to-A mutant was incapable of forming phase-separated droplets under conditions in which the wild-type IDR readily formed droplets (Fig. 6C), indicating that the conserved serine bias in the MED1-IDR is necessary for droplet formation.

MED1-IDR droplets can incorporate proteins necessary for transcription

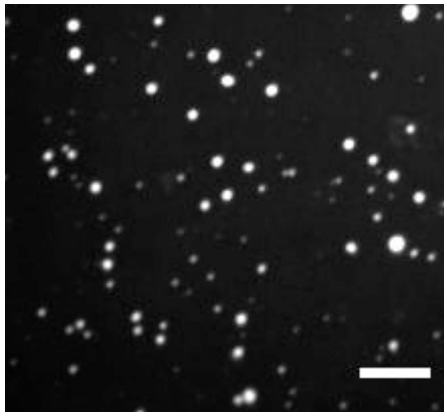
A proposed function of phase separation at SEs is the ability to compartmentalize and concentrate factors within a biomolecular condensate, so we sought to test whether MED1-IDR droplets could recapitulate this compartmentalization function in vitro. We identified conditions under which the MED1-IDR could form droplets but the BRD4-IDR could not (fig. S9). We then investigated whether the MED1-IDR droplets could compartmentalize BRD4-IDR protein under these conditions (Fig. 7A). Using differentially labeled proteins (mCherry–MED1-IDR and mEGFP–BRD4-IDR), we found that the MED1-IDR droplets could incorporate, and thus concentrate, the BRD4-IDR protein (Fig. 7A). The MED1-IDR droplets did not incorporate mEGFP (Fig. 7A). To probe the approximate mesh size of the MED1-IDR droplets (*47*), we incubated them with fluorescently labeled dextrans with average molecular weights of 4, 10, and 40 kDa. We found that the 4-kDa dextrans were incorporated into the MED1-IDR droplets, the 10-kDa dextrans were incorporated with less efficiency, and the 40-kDa dextrans were excluded (fig. S10). These results suggest that the incorporation of mEGFP–BRD4-IDR (105 kDa) into the MED1-IDR droplet is due to attractive molecular interactions, as opposed to passive diffusion through the droplet mesh.



Movie 1. BRD4-IDR droplets in solution. Each frame represents 1 s. The movie is rendered at 12 frames per second. 20 μM protein, 125 mM NaCl. Scale bar, 5 μm.



Movie 2. MED1-IDR droplets in solution. Each frame represents 1 s. The movie is rendered at 12 frames per second. 20 μM protein, 125 mM NaCl. Scale bar, 5 μm.



Movie 3. MED1-IDR droplets settling onto a glass coverslip. Each frame represents 1 s. The movie is rendered at 12 frames per second. 10 μM protein, 125 mM NaCl. Scale bar, 5 μm.

We next investigated whether the MED1-IDR, introduced into a transcription-competent nuclear extract, would form droplets that might incorporate BRD4 or other transcriptional components. We found that the wild-type MED1-IDR, but not the MED1-IDR S-to-A mutant, formed droplets in these extracts (Fig. 7B). The MED1-IDR phase-separated droplets were denser than the surrounding extract and thus could be purified from solution by centrifugation. Immunoblot analysis revealed that BRD4 and the largest subunit of RNA Pol II (RPB1) were enriched in pelleted droplets in a MED1-IDR dose–dependent manner (Fig. 7C). These results indicate that the MED1-IDR droplets can incorporate BRD4 and RNA Pol II.

The ability of the MED1-IDR protein to incorporate BRD4 and RNA Pol II into an artificial phase-separated compartment suggests that it sequesters key components of the transcription apparatus and might thus "squelch" transcription in the nuclear extract. We carried out an in vitro transcription assay with these extracts and found that the wild-type MED1-IDR protein does squelch transcription, correlating with the amount of material separated from solution by the MED1-IDR droplets (Fig. 7D). We did not observe these effects with equivalent concentrations of mEGFP or with the MED1-IDR S-to-A mutant (Fig. 7D). These results demonstrate that the MED1-IDR has the capacity to compartmentalize and concentrate transcriptional machinery from a complex nuclear extract.

Discussion

SEs regulate genes with prominent roles in healthy and diseased cellular states (*14*, *15*, *19*–*25*, *48*, *49*). SEs and their components have been proposed to form phase-separated condensates (*30*), but with no direct evidence. Here we demonstrate that two key components of SEs, BRD4 and MED1, form nuclear condensates at sites of SE-driven

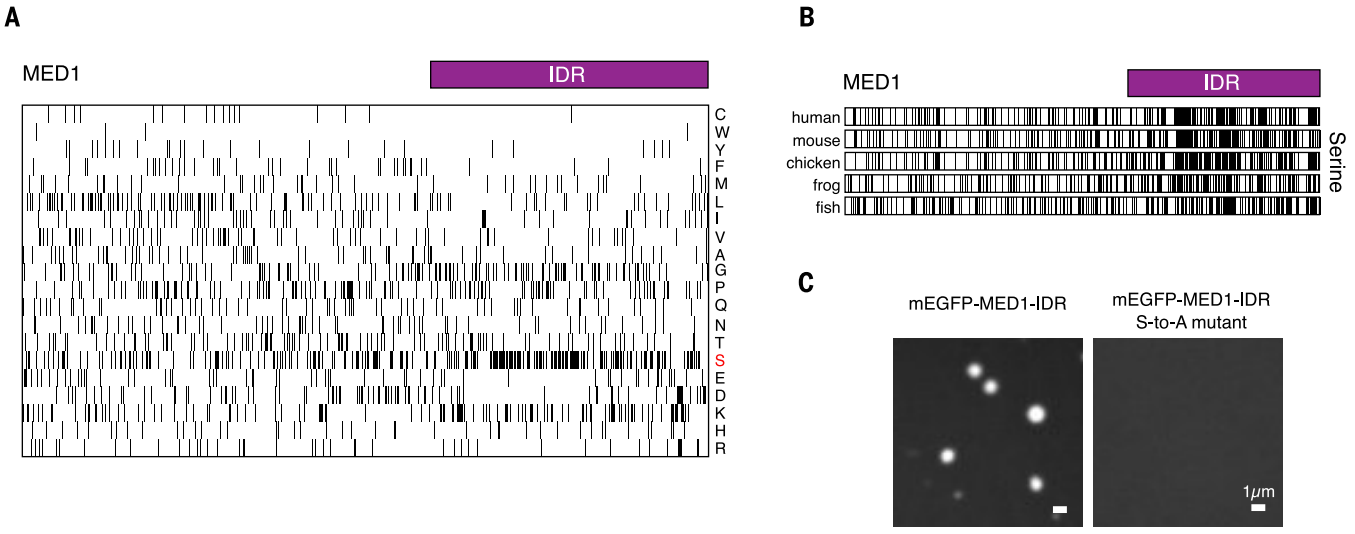
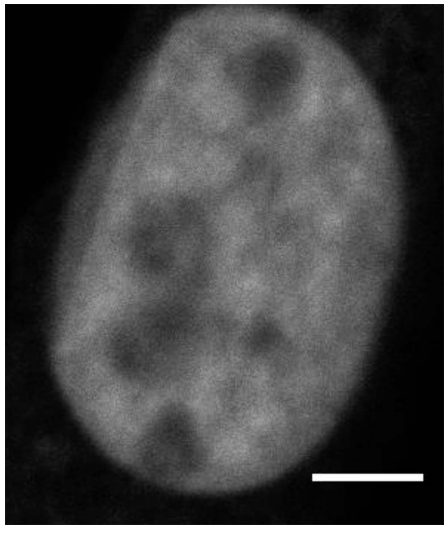


Fig. 6. Conserved serine bias is necessary for MED1-IDR phase separation. (**A**) Amino acid composition of the MED1 protein. Each row represents information for a single amino acid. Single-letter amino code abbreviations (right) are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. The length of the row corresponds to the length of the MED1 protein. Black bars represent the occurrence of the indicated amino acid at that position in MED1. The purple bar represents the IDR of MED1 under investigation. (**B**) Serine composition of MED1 protein from indicated organisms, presented as in (A). (**C**) Mutating all serines to alanine (S to A) disrupts phase separation. Representative images of wild-type MED1-IDR or MED1-IDR S-to-A mutant fused to mEGFP in the droplet formation assay (10 μM protein, 125 mM NaCl, and 10% Ficoll-400).

transcription. Within these condensates, BRD4 and MED1 exhibit apparent diffusion coefficients similar to those previously reported for other proteins in phase-separated condensates in vivo (*36*, *37*). The IDRs of both BRD4 and MED1 are sufficient to form phase-separated droplets in vitro, and the MED1-IDR facilitates phase separation in living cells. Droplets formed by MED1-IDR are capable of concentrating transcriptional machinery in a transcriptionally competent nuclear extract. These results support a model in which transcriptional coactivators form phase-separated condensates that compartmentalize and concentrate the transcription apparatus at SE-regulated genes and identify SE components that likely play a role in phase separation.

SEs are established by the binding of master TFs to enhancer clusters (*14*, *15*). These TFs typically consist of a structured DNA-binding domain and an intrinsically disordered transcriptional activation domain (*50*–*52*). The activation domains of these TFs recruit high densities of many transcription proteins, which, as a class, are enriched for IDRs (*53*). Although the exact client-scaffold relationship (*54*) between these components remains unknown, it is likely that these protein sequences mediate weak multivalent interactions, thereby facilitating condensation. We propose that condensation of such high-valency factors at SEs creates a reaction crucible within the separated dense phase, where high local concentrations of the transcriptional machinery ensure robust gene expression.

The nuclear organization of chromosomes is likely influenced by condensates at SEs. DNA interaction technologies indicate that the individual enhancers within the SEs have exceptionally high interaction frequencies with one another (*16*–*18*), consistent with the idea that condensates draw these elements into close proximity in the dense phase. Several recent studies suggest that SEs can interact with one another and may also contribute in this fashion to chromosome organization (*55*, *56*). Cohesin, an SMC (structural maintenance of chromosomes) protein complex, has been implicated in constraining SE-SE interactions because its loss causes extensive fusion of SEs within the nucleus (*56*). These SE-SE interactions may be due to a tendency of liquid-phase condensates to undergo fusion (*1*–*3*).



Movie 4. Formation of MED1-IDR optoDroplets upon stimulation with blue light. NIH3T3 cells expressing the MED1-optoIDR construct were subjected to 488-nm laser light in 2-s intervals. Each frame represents 2 s. The movie is rendered at 12 frames per second. Scale bar, 5 μm.

The model whereby phase separation of coactivators compartmentalizes and concentrates the transcription apparatus at SEs and their regulated genes, described here and corroborated by (*57*), raises many questions. How does condensation contribute to regulation of transcriptional output? A study of RNA Pol II clusters, which may be phase-separated condensates, suggests a positive correlation between condensate lifetime and transcriptional output (*58*). What components drive formation and dissolution of transcriptional condensates? Our studies indicate that BRD4 and MED1 likely participate, but the roles of DNA-binding TFs, RNA Pol II, and regulatory RNAs require further study. Why do some proteins, such as HP1a, contribute to phase-separated heterochromatin condensates (*59*, *60*) and others contribute to euchromatic condensates? The rules that govern partitioning into specific types of condensates have begun to be studied (*61*–*65*) and will need to be defined for proteins involved in transcriptional condensates. Does condensate misregulation contribute to pathological processes in disease, and will new insights into condensate behaviors present new opportunities for therapy? Mutations within IDRs and misregulation of phase separation have already been implicated in a number of neurodegenerative diseases (*66*–*68*). Tumor cells have exceptionally large SEs at driver oncogenes that are not found in their cell of origin, and some of these are exceptionally sensitive to drugs that target SE components (*22*–*25*). How do we take advantage of phase separation

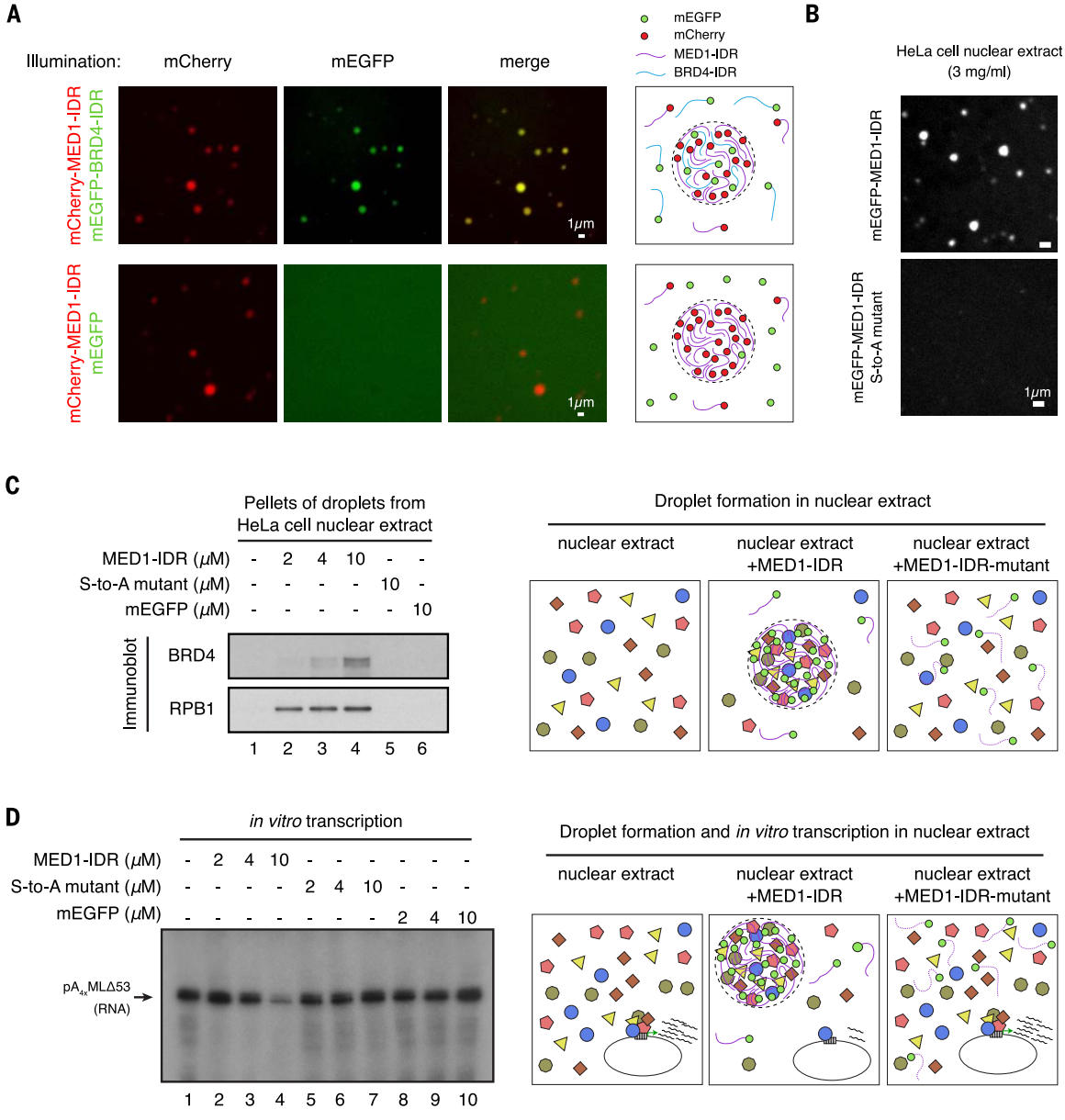


Fig. 7. MED1-IDR droplets compartmentalize and concentrate proteins necessary for transcription. (**A**) MED1-IDR droplets incorporate BRD4-IDR protein in vitro. The indicated mEGFP or mCherry fusion proteins were mixed at 10 μM each in buffer D containing 10% Ficoll-400 and 125 mM NaCl. Indicated fluorescence channels are presented for each mixture. Illustrations summarizing results are shown on the right. (**B**) MED1-IDR forms droplets in an in vitro transcription reaction containing HeLa cell nuclear extract, whereas the MED1-IDR S-to-A mutant does not. Shown are representative images of the indicated mEGFP-fusion protein when added to an in vitro transcription reaction containing HeLa cell nuclear extract at a final concentration of 3 mg/ml (a complete list of components is given in the methods). (**C**) MED1-IDR droplets compartmentalize transcriptional machinery from a nuclear extract. Shown are immunoblots of the pellet fraction of the indicated protein added to in vitro transcription reactions [as described in (B)]. A proposed model of molecular interactions taking place within MED1-IDR droplets in the nuclear extract is illustrated on the right. (**D**) MED1-IDR droplets compartmentalize machinery necessary for the in vitro transcription reaction. An autoradiograph of radiolabeled RNA products of in vitro transcription reactions under indicated conditions is shown on the left. The arrow indicates the expected RNA product. Reactions were conducted as in (*68*) with minor modifications (full details are given in the methods). A proposed model of molecular interactions taking place within MED1-IDR droplets in nuclear extract and the impact on the in vitro transcription reaction is illustrated on the right.

principles established in physics and chemistry to more effectively improve our understanding of this form of regulatory biology? Addressing these questions at the crossroads of physics, chemistry, and biology will require collaboration across these diverse sciences.

Methods summary

Immunofluorescence against BRD4 and MED1, coupled with DNA-FISH or RNA-FISH against SEs or SE-driven nascent transcripts, was performed in mESCs to visualize the colocalization between BRD4 or MED1 puncta and SEs. BRD4 and MED1 were endogenously tagged with mEGFP in mESCs to visualize the organization of BRD4 and MED1 and to study their dynamics by FRAP and drug treatments in live cells. ChIP-seq was performed to investigate the effect of 1,6-hexanediol treatment on the chromatin occupancy of BRD4, MED1, and RNA Pol II. Recombinant BRD4-IDR and MED1-IDR were purified to test their capacity to phase-separate in vitro. The optoIDR assay (*45*) was implemented to test the capacity of a section of MED1-IDR to phase-separate in live cells. Mutations were introduced into MED1-IDR to study the sequence determinants of MED1-IDR phase separation. BRD4-IDR and MED1-IDR fused to different fluorescent tags were used to demonstrate the capacity of MED1-IDR droplets to compartmentalize and concentrate BRD4-IDR. Formation of MED1-IDR droplets in a transcriptionally competent nuclear extract was used to study the ability of MED1-IDR droplets to compartmentalize and concentrate BRD4 and RNA Pol II from a complex extract. In vitro transcription assays were used to measure the effect of synthetic droplet formation on transcription. All procedures are described in detail in the supplementary materials.

REFERENCES AND NOTES

1. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014). doi: 10.1146/annurev-cellbio-100913-013325; pmid: 25288112
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). doi: 10.1038/nrm.2017.7; pmid: 28225081
3. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017). doi: 10.1126/science.aaf4382; pmid: 28935776
4. C. P. Brangwynne *et al*., Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009). doi: 10.1126/science.1172046; pmid: 19460965
5. M. Kato *et al*., Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012). doi: 10.1016/j.cell.2012.04.017; pmid: 22579281
6. P. Li *et al*., Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012). doi: 10.1038/nature10879; pmid: 22398450
7. Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208–219 (2015). doi: 10.1016/j.molcel.2015.08.018; pmid: 26412307
8. K. Adelman, J. T. Lis, Promoter-proximal pausing of RNA polymerase II: Emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012). doi: 10.1038/nrg3293; pmid: 22986266
9. M. Bulger, M. Groudine, Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011). doi: 10.1016/j.cell.2011.01.024; pmid: 21295696
10. E. Calo, J. Wysocka, Modification of enhancer chromatin: What, how, and why? *Mol. Cell* **49**, 825–837 (2013). doi: 10.1016/j.molcel.2013.01.038; pmid: 23473601
11. F. Spitz, E. E. M. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012). doi: 10.1038/nrg3207; pmid: 22868264
12. W. Xie, B. Ren, Developmental biology. Enhancing pluripotency and lineage specification. *Science* **341**, 245–247 (2013). doi: 10.1126/science.1236254; pmid: 23869010
13. M. Levine, C. Cattoglio, R. Tjian, Looping back to leap forward: Transcription enters a new era. *Cell* **157**, 13–25 (2014). doi: 10.1016/j.cell.2014.02.009; pmid: 24679523
14. W. A. Whyte *et al*., Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013). doi: 10.1016/j.cell.2013.03.035; pmid: 23582322
15. D. Hnisz *et al*., Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013). doi: 10.1016/j.cell.2013.09.053; pmid: 24119843
16. J. M. Dowen *et al*., Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014). doi: 10.1016/j.cell.2014.09.030; pmid: 25303531
17. D. Hnisz *et al*., Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016). doi: 10.1126/science.aad9024; pmid: 26940867
18. X. Ji *et al*., 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18**, 262–275 (2016). doi: 10.1016/j.stem.2015.11.007; pmid: 26686465
19. M. R. Mansour *et al*., An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014). doi: 10.1126/science.1259037; pmid: 25394790
20. J. D. Brown *et al*., NF-κB directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol. Cell* **56**, 219–231 (2014). doi: 10.1016/j.molcel.2014.08.024; pmid: 25263595
21. B. Chapuy *et al*., Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013). doi: 10.1016/j.ccr.2013.11.003; pmid: 24332044
22. J. Lovén *et al*., Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013). doi: 10.1016/j.cell.2013.03.036; pmid: 23582323
23. E. Chipumuro *et al*., CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126–1139 (2014). doi: 10.1016/j.cell.2014.10.024; pmid: 25416950
24. N. Kwiatkowski *et al*., Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014). doi: 10.1038/nature13393; pmid: 25043025
25. Y. Wang *et al*., CDK7-dependent transcriptional addiction in triple-negative breast cancer. *Cell* **163**, 174–186 (2015). doi: 10.1016/j.cell.2015.08.063; pmid: 26406377
26. D. Hnisz *et al*., Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015). doi: 10.1016/j.molcel.2015.02.014; pmid: 25801169
27. T. Jiang *et al*., Identification of multi-loci hubs from 4C-seq demonstrates the functional importance of simultaneous interactions. *Nucleic Acids Res.* **44**, 8714–8725 (2016). doi: 10.1093/nar/gkw568; pmid: 27439714
28. C. Proudhon *et al*., Active and Inactive Enhancers Cooperate to Exert Localized and Long-Range Control of Gene Regulation. *Cell Rep.* **15**, 2159–2169 (2016). pmid: 27239026
29. H. Y. Shin *et al*., Hierarchy within the mammary STAT5-driven *Wap* super-enhancer. *Nat. Genet.* **48**, 904–911 (2016). doi: 10.1038/ng.3606; pmid: 27376239
30. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017). doi: 10.1016/j.cell.2017.02.007; pmid: 28340338
31. Z. Yang *et al*., Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19**, 535–545 (2005). doi: 10.1016/j.molcel.2005.06.029; pmid: 16109377
32. M. K. Jang *et al*., The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19**, 523–534 (2005). doi: 10.1016/j.molcel.2005.06.027; pmid: 16109376
33. R. Di Micco *et al*., Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.* **9**, 234–247 (2014). pmid: 25263550
34. J. Soutourina, S. Wydau, Y. Ambroise, C. Boschiero, M. Werner, Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **331**, 1451–1454 (2011). doi: 10.1126/science.1200188; pmid: 21415355
35. J. Soutourina, Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **19**, 262–274 (2018). pmid: 29209056
36. T. J. Nott *et al*., Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015). doi: 10.1016/j.molcel.2015.01.013; pmid: 25747659
37. C. W. Pak *et al*., Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell* **63**, 72–85 (2016). doi: 10.1016/j.molcel.2016.05.042; pmid: 27392146
38. C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, 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). doi: 10.1073/pnas.1017150108; pmid: 21368180
39. A. Patel *et al*., ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017). doi: 10.1126/science.aaf6846; pmid: 28522535
40. S. Kroschwald, S. Maharana, A. Simon, Hexanediol: A chemical probe to investigate the material properties of membrane-less compartments. *Matters* 10.19185/matters.201702000010 (2017).
41. C. Y. Lin *et al*., Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012). doi: 10.1016/j.cell.2012.08.026; pmid: 23021215
42. S. Elbaum-Garfinkle *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). doi: 10.1073/pnas.1504822112; pmid: 26015579
43. C. P. Brangwynne, Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.* **203**, 875–881 (2013). doi: 10.1083/jcb.201308087; pmid: 24368804
44. Y. Shin *et al*., Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159–171.e14 (2017). doi: 10.1016/j.cell.2016.11.054; pmid: 28041848
45. I. Ozkan-Dagliyan *et al*., Formation of Arabidopsis Cryptochrome 2 photobodies in mammalian nuclei: Application as an optogenetic DNA damage checkpoint switch. *J. Biol. Chem.* **288**, 23244–23251 (2013). doi: 10.1074/jbc.M113.493361; pmid: 23833191
46. X. Yu *et al*., Formation of nuclear bodies of Arabidopsis CRY2 in response to blue light is associated with its blue light-dependent degradation. *Plant Cell* **21**, 118–130 (2009). doi: 10.1105/tpc.108.061663; pmid: 19141709
47. M.-T. Wei *et al*., Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* **9**, 1118–1125 (2017). doi: 10.1038/nchem.2803; pmid: 29064502
48. H. I. Suzuki, R. A. Young, P. A. Sharp, Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis. *Cell* **168**, 1000–1014.e15 (2017). doi: 10.1016/j.cell.2017.02.015; pmid: 28283057
49. J. E. Bradner, D. Hnisz, R. A. Young, Transcriptional Addiction in Cancer. *Cell* **168**, 629–643 (2017). doi: 10.1016/j.cell.2016.12.013; pmid: 28187285
50. M. Ptashne, How eukaryotic transcriptional activators work. *Nature* **335**, 683–689 (1988). doi: 10.1038/335683a0; pmid: 3050531
51. P. J. Mitchell, R. Tjian, Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371–378 (1989). doi: 10.1126/science.2667136; pmid: 2667136
52. J. Liu *et al*., Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888 (2006). doi: 10.1021/bi0602718; pmid: 16734424
53. H. Xie *et al*., Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* **6**, 1882–1898 (2007). doi: 10.1021/pr060392u; pmid: 17391014
54. S. F. Banani *et al*., Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016). doi: 10.1016/j.cell.2016.06.010; pmid: 27374333
55. R. A. Beagrie *et al*., Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017). pmid: 28273065
56. S. S. P. Rao *et al*., Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24 (2017). doi: 10.1016/j.cell.2017.09.026; pmid: 28985562
57. W.-K. Cho *et al*., Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
58. W.-K. Cho *et al*., RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5**, 1123 (2016). doi: 10.7554/eLife.13617; pmid: 27138339
59. A. G. Larson *et al*., Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017). doi: 10.1038/nature22822; pmid: 28636604
60. A. R. Strom *et al*., Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017). doi: 10.1038/nature22989; pmid: 28636597
61. M. Feric *et al*., Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016). doi: 10.1016/j.cell.2016.04.047; pmid: 27212236
62. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017). doi: 10.7554/eLife.30294; pmid: 29091028
63. J. A. Riback *et al*., Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028–1040.e19 (2017). doi: 10.1016/j.cell.2017.02.027; pmid: 28283059
64. S. Boeynaems *et al*., Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol. Cell* **65**, 1044–1055.e5 (2017). doi: 10.1016/j.molcel.2017.02.013; pmid: 28306503
65. J. P. Brady *et al*., Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8194–E8203 (2017). doi: 10.1073/pnas.1706197114; pmid: 28894006
66. A. Patel *et al*., A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015). doi: 10.1016/j.cell.2015.07.047; pmid: 26317470
67. A. Molliex *et al*., Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015). doi: 10.1016/j.cell.2015.09.015; pmid: 26406374
68. A. Jain, R. D. Vale, RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017). pmid: 28562589

ACKNOWLEDGMENTS

We thank W. Salmon of the W. M. Keck Microscopy Facility; D. Richardson and S. Terclavers of the Harvard Center for Biological Imaging; and T. Volkert, D. Reynolds, S. Mraz, and S. Gupta of the Whitehead Genome Technologies Core for technical

assistance. We thank the Imaging Platform at the Broad Institute for assistance with CellProfiler. **Funding:** The work was supported by NIH grants GM123511 (R.A.Y.) and P01-CA042063 (P.A.S.), NSF grant PHY-1743900 (A.K.C., R.A.Y., and P.A.S.), Koch Institute Support (core) grant P30-CA14051 from the NCI (P.A.S.), Damon Runyon Cancer Research Foundation Fellowship 2309-17 (B.R.S.), Swedish Research Council Postdoctoral Fellowship VR 2017-00372 (A.B.), a Hope Funds for Cancer Research fellowship (B.J.A.), an NSF Graduate Research Fellowship (A.V.Z.), a Cancer Research Institute Irvington Fellowship (Y.E.G.), American Cancer Society New England Division Postdoctoral Fellowship PF-16-146-01-DMC (D.S.D.), and a NWO Rubicon Fellowship (J.S.). **Author contributions:** B.R.S., A.D., and R.A.Y. conceptualized and organized the project and wrote the manuscript. A.D., A.B., J.C.M., and Y.E.G. performed cell-imaging experiments and image analysis. I.A.K. and A.V.Z. generated endogenously tagged cell lines. B.R.S. and A.B. performed ChIP-seq. B.R.S. and E.L.C. performed in vitro droplet assays and optoIDR experiments. K.S. and B.J.A. developed and performed image analysis and produced visualizations. B.J.A. performed ChIP-seq analysis and produced visualizations. N.M.H. produced and purified recombinant proteins. A.V.Z. helped with biochemical experiments. C.H.L. performed protein amino acid analysis. D.S.D. performed ChIA-PET analysis and visualization. B.R.S., I.A.K., E.L.C., J.S., and A.V.Z. generated constructs. S.M. performed in vitro transcription assays. D.H., E.V., T.I.L., I.I.C., R.G.R., P.A.S., A.K.C., and R.A.Y. provided input into experimental design and interpretation. P.A.S., A.K.C., and R.A.Y. acquired funding for this study. R.A.Y. supervised the project with help from T.I.L. and A.K.C. All authors contributed to editing the manuscript. **Competing interests:** The Whitehead Institute filed a patent application based on this paper. R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, and Omega Therapeutics. B.J.A. and T.I.L. are shareholders of Syros Pharmaceuticals, and T.I.L. is a consultant to Camp4 Therapeutics. All other authors declare no competing interests. **Data and materials availability:** Datasets generated in this study have been deposited in the Gene Expression Omnibus under accession number GSE112808.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6400/eaar3958/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S3
References (*69*–*82*)
Data S1

4 November 2017; resubmitted 9 April 2018
Accepted 6 June 2018
Published online 21 June 2018
10.1126/science.aar3958

APPENDIX II: TRANSCRIPTION FACTORS ACTIVATE GENES THROUGH THE PHASE SEPARATION CAPACITY OF THEIR ACTIVATION DOMAINS

Originally published in *Cell* Volume 175, Issue 7. (2018)

Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains

Ann Boija,[1,7] Isaac A. Klein,[1,2,7] Benjamin R. Sabari,[1] Alessandra Dall'Agnese,[1] Eliot L. Coffey,[1,3] Alicia V. Zamudio,[1,3] Charles H. Li,[1,3] Krishna Shrinivas,[4,5] John C. Manteiga,[1,3] Nancy M. Hannett,[1] Brian J. Abraham,[1] Lena K. Afeyan,[1,3] Yang E. Guo,[1] Jenna K. Rimel,[6] Charli B. Fant,[6] Jurian Schuijers,[1] Tong Ihn Lee,[1] Dylan J. Taatjes,[6,8] and Richard A. Young[1,3,8,9,*]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Medical Oncology, 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 Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Institute of Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Department of Biochemistry, University of Colorado, Boulder, CO 80303, USA
[7]These authors contributed equally
[8]Senior author
[9]Lead Contact
*Correspondence: young@wi.mit.edu
https://doi.org/10.1016/j.cell.2018.10.042

SUMMARY

Gene expression is controlled by transcription factors (TFs) that consist of DNA-binding domains (DBDs) and activation domains (ADs). The DBDs have been well characterized, but little is known about the mechanisms by which ADs effect gene activation. Here, we report that diverse ADs form phase-separated condensates with the Mediator coactivator. For the OCT4 and GCN4 TFs, we show that the ability to form phase-separated droplets with Mediator *in vitro* and the ability to activate genes *in vivo* are dependent on the same amino acid residues. For the estrogen receptor (ER), a ligand-dependent activator, we show that estrogen enhances phase separation with Mediator, again linking phase separation with gene activation. These results suggest that diverse TFs can interact with Mediator through the phase-separating capacity of their ADs and that formation of condensates with Mediator is involved in gene activation.

INTRODUCTION

Regulation of gene expression requires that the transcription apparatus be efficiently assembled at specific genomic sites. DNA-binding transcription factors (TFs) ensure this specificity by occupying specific DNA sequences at enhancers and promoter-proximal elements. TFs typically consist of one or more DNA-binding domains (DBDs) and one or more separate activation domains (ADs) (Brent and Ptashne, 1985; Keegan et al., 1986). While the structure and function of TF DBDs are well documented, comparatively little is understood about the structure of ADs and how these interact with coactivators to drive gene expression.

The structure of TF DBDs and their interaction with cognate DNA sequences has been described at atomic resolution for many TFs, and TFs are generally classified according to the structural features of their DBDs (Fulton et al., 2009; Vaquerizas et al., 2009). For example, DBDs can be composed of zinc-coordinating, basic helix-loop-helix, basic-leucine zipper, or helix-turn-helix DNA-binding structures. These DBDs selectively bind specific DNA sequences that range from 4 to 12 bp, and the DNA binding sequences favored by hundreds of TFs have been described (Hume et al., 2015; Jolma et al., 2013; Khan et al., 2018). Multiple TF molecules typically bind together at any one enhancer or promoter-proximal element. For example, at least eight different TF molecules bind a 50-bp core component of the interferon (IFN)-β enhancer (Panne et al., 2007).

Anchored in place by the DBD, the AD interacts with coactivators, which integrate signals from multiple TFs to regulate transcriptional output (Allen and Taatjes, 2015; Juven-Gershon and Kadonaga, 2010; Malik and Roeder, 2010; Plaschka et al., 2016; Reiter et al., 2017; Soutourina, 2018; Taatjes, 2010). In contrast to the structured DBD, the ADs of most TFs are low-complexity amino acid sequences not amenable to crystallography. These intrinsically disordered regions (IDRs) have therefore been classified by their amino acid profile as acidic, proline, serine/threonine, or glutamine rich or by their hypothetical shape as acid blobs, negative noodles, or peptide lassos (Mitchell and Tjian, 1989; Roberts, 2000; Sigler, 1988; Staby et al., 2017; Triezenberg, 1995). Remarkably, hundreds of TFs are thought to interact with the same small set of coactivator complexes, which include Mediator and p300 (Allen and Taatjes, 2015; Avantaggiati et al., 1996; Dai and Markham, 2001; Eckner et al., 1996; Green, 2005; Merika et al., 1998; Oliner et al., 1996; Yin and Wang, 2014; Yuan et al., 1996). ADs that share little

sequence homology are functionally interchangeable among TFs (Godowski et al., 1988; Hope and Struhl, 1986; Jin et al., 2016; Lech et al., 1988; Ransone et al., 1990; Sadowski et al., 1988; Struhl, 1988; Tora et al., 1989); this interchangeability is not readily explained by traditional lock-and-key models of protein-protein interaction. Thus, how the diverse ADs of hundreds of different TFs interact with a similar small set of coactivators remains a conundrum.

Recent studies have shown that the AD of the yeast TF GCN4 binds to the Mediator subunit MED15 at multiple sites and in multiple orientations and conformations (Brzovic et al., 2011; Jedidi et al., 2010; Tuttle et al., 2018; Warfield et al., 2014). The products of this type of protein-protein interaction, where the interaction interface cannot be described by a single conformation, have been termed ''fuzzy complexes'' (Tompa and Fuxreiter, 2008). These dynamic interactions are also typical of the IDR-IDR interactions that facilitate formation of phase-separated biomolecular condensates (Alberti, 2017; Banani et al., 2017; Hyman et al., 2014; Shin and Brangwynne, 2017; Wheeler and Hyman, 2018).

We recently proposed that transcriptional control may be driven by the formation of phase-separated condensates (Hnisz et al., 2017) and demonstrated that the coactivator proteins MED1 and BRD4 form phase-separated condensates at super-enhancers (SEs) (Sabari et al., 2018). Here, we report that diverse TF ADs phase separate with the Mediator coactivator. We show that the embryonic stem cell (ESC) pluripotency TF OCT4, the estrogen receptor (ER), and the yeast TF GCN4 form phase-separated condensates with Mediator and require the same amino acids or ligands for both activation and phase separation. We propose that IDR-mediated phase separation with coactivators is a mechanism by which TF ADs activate genes.

RESULTS

Mediator Condensates at ESC SEs Depend on OCT4

OCT4 is a master TF essential for the pluripotent state of ESCs and is a defining TF at ESC SEs (Whyte et al., 2013). The Mediator coactivator, which forms condensates at ESC SEs (Sabari et al., 2018), is thought to interact with OCT4 via the MED1 subunit (Table S1) (Apostolou et al., 2013). If OCT4 contributes to the formation of Mediator condensates, then OCT4 puncta should be present at the SEs where MED1 puncta have been observed. Indeed, immunofluorescence (IF) microscopy with concurrent nascent RNA fluorescence in situ hybridization (FISH) revealed discrete OCT4 puncta at the SEs of the key pluripotency genes *Esrrb*, *Nanog*, *Trim28*, and *Mir290* (Figure 1). Average image analysis confirmed that OCT4 IF was enriched at center of RNA FISH foci. This enrichment was not seen using a randomly selected nuclear position (Figure S1). These results confirm that OCT4 occurs in puncta at the same SEs where Mediator forms condensates (Sabari et al., 2018) and where chromatin immunoprecipitation sequencing (ChIP-seq) shows co-occupancy of OCT4 and MED1 (Figure 1).

We investigated whether the Mediator condensates present at SEs are dependent on OCT4 using a degradation strategy (Nabet et al., 2018). Degradation of OCT4 in an ESC line bearing endogenous knockin of DNA encoding the FKBP protein fused to OCT4 was induced by addition of dTag for 24 hr (Weintraub et al., 2017) (Figures 2A and S2A). Induction of OCT4 degradation reduced OCT4 protein levels but did not affect MED1 levels (Figure S2B). ChIP-seq analysis showed a reduction of OCT4 and MED1 occupancy at enhancers, with the most profound effects occurring at SEs, as compared to typical enhancers (TEs) (Figure 2B). RNA sequencing (RNA-seq) revealed that expression of SE-driven genes was concomitantly decreased (Figure 2B). For example, OCT4 and MED1 occupancy was reduced by approximately 90% at the *Nanog* SE (Figure 2C), associated with a 60% reduction in *Nanog* mRNA levels (Figure 2D). IF microscopy with concurrent DNA FISH showed that OCT4 degradation caused a reduction in MED1 condensates at *Nanog* (Figures 2E and S2C). These results indicate that the presence of Mediator condensates at an ESC SE is dependent on OCT4.

ESC differentiation causes a loss of OCT4 binding at certain ESC SEs, which leads to a loss of these OCT4-dependent SEs, and thus should cause a loss of Mediator condensates at these sites. To test this idea, we differentiated ESCs by leukemia inhibitory factor (LIF) withdrawal. In the differentiated cell population, we observed reduced OCT4 and MED1 occupancy at the *Mir290* SE (Figures 2F, 2G, and S2D) and reduced levels of *Mir290* miRNA (Figure 2H), despite continued expression of MED1 protein (Figure S2E). Correspondingly, MED1 condensates were reduced at *Mir290* (Figures 2I and S2F) in the differentiated cell population. These results are consistent with those obtained with the OCT4 degron experiment and support the idea that Mediator condensates at these ESC SEs are dependent on occupancy of the enhancer elements by OCT4.

OCT4 Is Incorporated into MED1 Liquid Droplets

OCT4 has two intrinsically disordered ADs responsible for gene activation, which flank a structured DBD (Figure 3A) (Brehm et al., 1997). Since IDRs are capable of forming dynamic networks of weak interactions, and the purified IDRs of proteins involved in condensate formation can form phase-separated droplets (Burke et al., 2015; Lin et al., 2015; Nott et al., 2015), we next investigated whether OCT4 is capable of forming droplets *in vitro*, with and without the IDR of the MED1 subunit of Mediator.

Recombinant OCT4-GFP fusion protein was purified and added to droplet formation buffers containing a crowding agent (10% PEG-8000) to simulate the densely crowded environment of the nucleus. Fluorescent microscopy of the droplet mixture revealed that OCT4 alone did not form droplets throughout the range of concentrations tested (Figure 3B). In contrast, purified recombinant MED1-IDR-GFP fusion protein exhibited concentration-dependent liquid-liquid phase separation (Figure 3B), as described previously (Sabari et al., 2018).

We then mixed the two proteins and found that droplets of MED1-IDR incorporate and concentrate purified OCT4-GFP to form heterotypic droplets (Figure 3C). In contrast, purified GFP was not concentrated into MED1-IDR droplets (Figures 3C and S3A). OCT4-MED1-IDR droplets were near micrometer sized (Figure S3B), exhibited fast recovery after photobleaching (Figure 3D), had a spherical shape (Figure S3C), and were salt

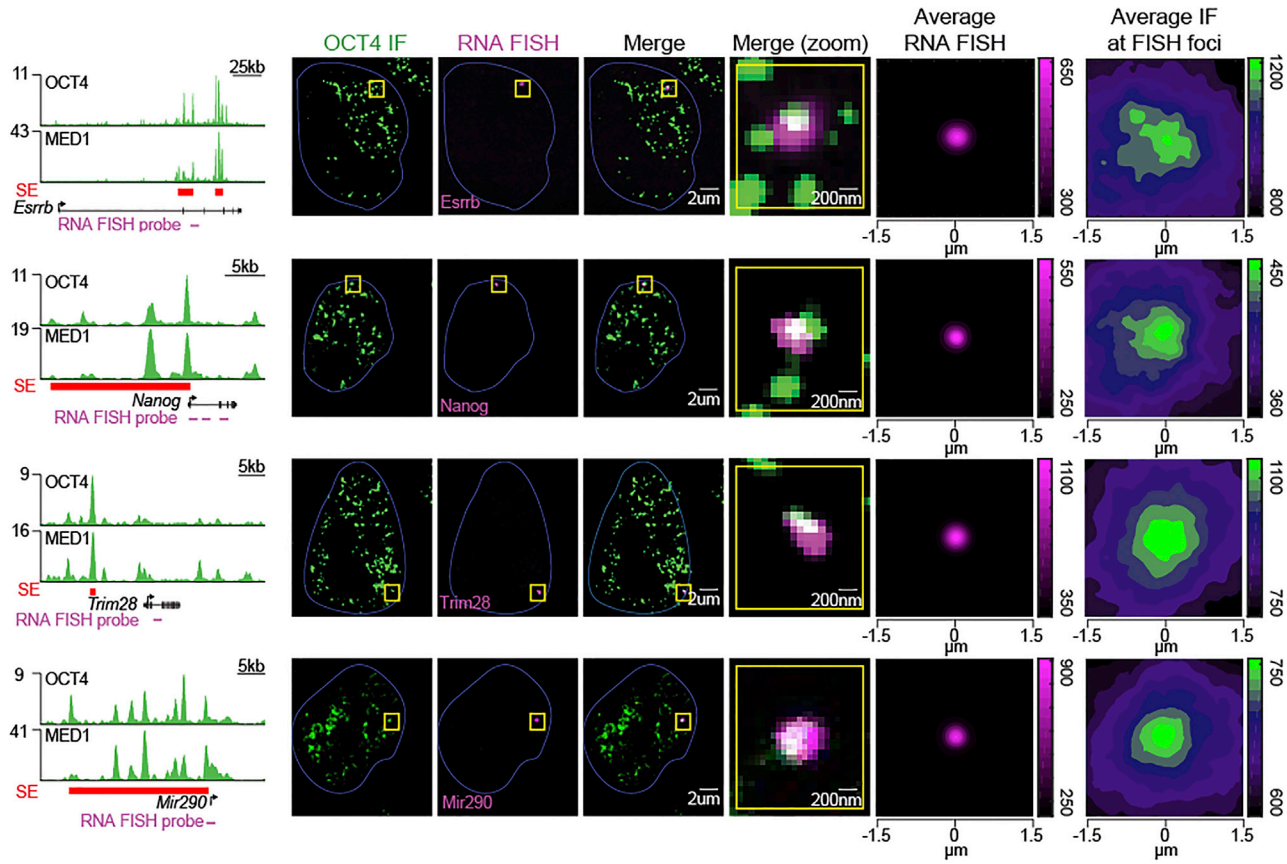


Figure 1. OCT4 and Mediator Occupy SEs *In Vivo*

ChIP-seq tracks of OCT4 and MED1 in ESCs at SEs (left column) and OCT4 IF with concurrent RNA-FISH demonstrating occupancy of OCT4 at *Esrrb*, *Nanog*, *Trim28*, and *Mir290*. Hoechst staining was used to determine the nuclear periphery, highlighted with a blue line. The two rightmost columns show average RNA FISH signal and average OCT4 IF signal centered on the RNA-FISH focus from at least 11 images. Average OCT4 IF signal at a randomly selected nuclear position is displayed in Figure S1.

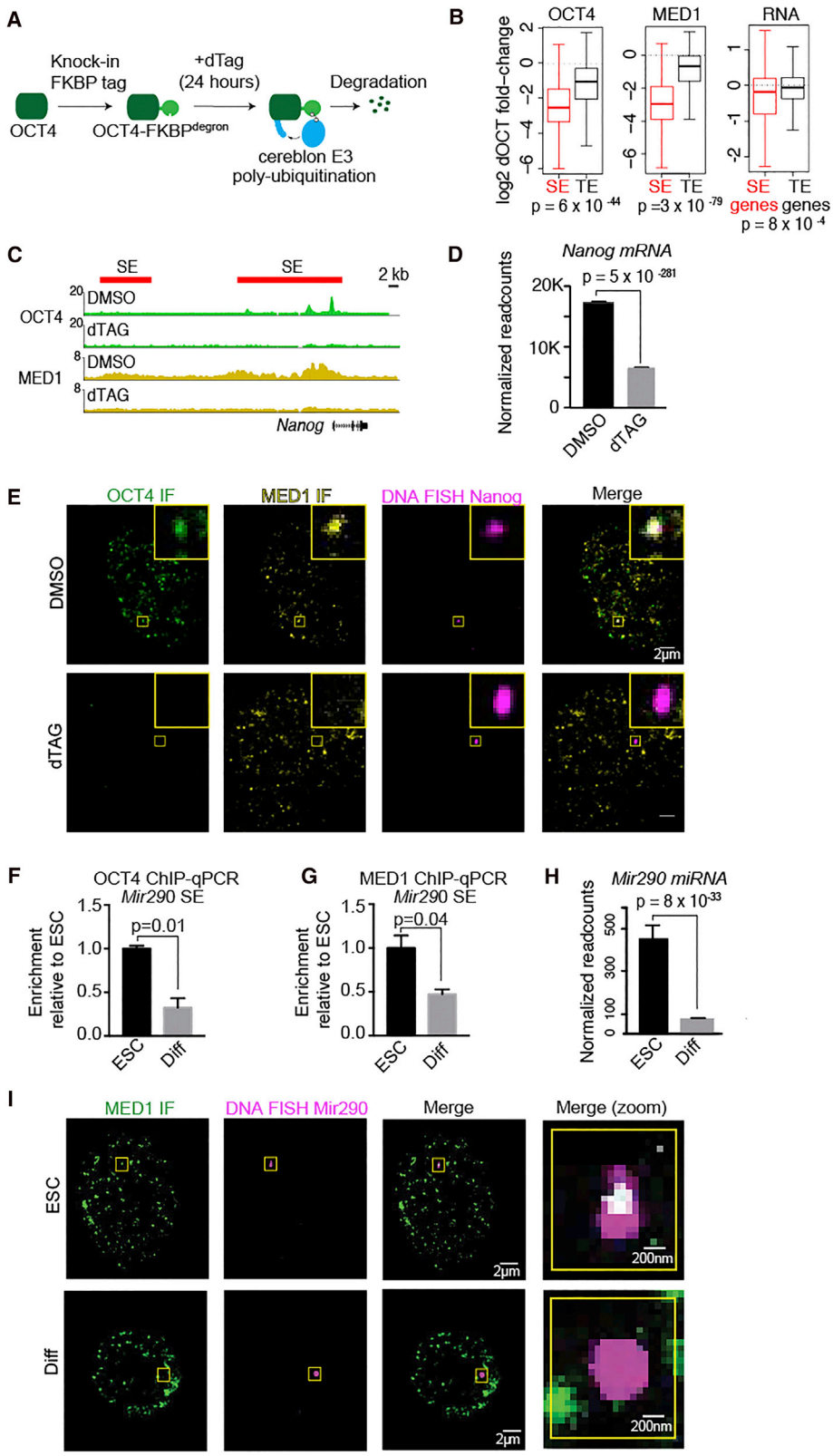
See also Figure S1 and Tables S1, S2, and S4.

sensitive (Figures 3E and S3D). Thus, they exhibited characteristics associated with phase-separated liquid condensates (Banani et al., 2017; Shin and Brangwynne, 2017). Furthermore, we found that OCT4-MED1-IDR droplets could form in the absence of any crowding agent (Figures S3E and S3F).

Residues Required for OCT4-MED1-IDR Droplet Formation and Gene Activation

We next investigated whether specific OCT4 amino acid residues are required for the formation of OCT4-MED1-IDR phase-separated droplets, as multiple categories of amino acid interaction have been implicated in forming condensates. For example, serine residues are required for MED1 phase separation (Sabari et al., 2018). We asked whether amino acid enrichments in the OCT4 ADs might point to a mechanism for interaction. An analysis of amino acid frequency and charge bias showed that the OCT4 IDRs are enriched in proline and glycine and have an overall acidic charge (Figure 4A). ADs are known to be enriched in acidic amino acids and proline and have historically been classified on this basis (Frietze and Farnham, 2011), but the mechanism by which these enrichments might cause gene activation is not known. We hypothesized that proline or acidic amino acids in the ADs might facilitate interaction with the phase-separated MED1-IDR droplet. To test this, we designed fluorescently labeled proline and glutamic acid decapeptides and investigated whether these peptides can be concentrated in MED1-IDR droplets. When added to droplet formation buffer alone, these peptides remained in solution (Figure S4A). When mixed with MED1-IDR-GFP, however, proline peptides were not incorporated into MED1-IDR droplets, while the glutamic acid peptides were concentrated within (Figures 4B and S4B). These results show that peptides with acidic residues are amenable to incorporation within MED1 phase-separated droplets.

Based on these results, we deduced that an OCT4 protein lacking acidic amino acids in its ADs might be defective in its ability to phase separate with MED1-IDR. Such a dependence on acidic residues would be consistent with our observation that OCT4-MED1-IDR droplets are highly salt sensitive. To test this idea, we generated a mutant OCT4 in which all acidic residues in the ADs were replaced with alanine (thus changing 17 AAs in the N-terminal AD and 6 in the C-terminal AD) (Figure 4C). When this GFP-fused OCT4 mutant was mixed with purified



(legend on next page)

MED1-IDR, entry into droplets was highly attenuated (Figures 4C and S4C). To test whether this effect was specific for acidic residues, we generated a mutant of OCT4 in which all the aromatic amino acids within the ADs were changed to alanine. We found that this mutant was still incorporated into MED1-IDR droplets (Figures S4C and S4D). These results indicate that the ability of OCT4 to phase separate with MED1-IDR is dependent on acidic residues in the OCT4 IDRs.

To ensure that these results were not specific to the MED1-IDR, we explored whether purified Mediator complexes would form droplets *in vitro* and incorporate OCT4. The human Mediator complex was purified as previously described (Meyer et al., 2008) and then concentrated for use in the droplet formation assay (Figure S4E). Because purified endogenous Mediator does not contain a fluorescent tag, we monitored droplet formation by differential interference contrast (DIC) microscopy and found it to form droplets alone at ~200–400 nM (Figure 4D). Consistent with the results for MED1-IDR droplets, OCT4 was incorporated within human Mediator complex droplets, but incorporation of the OCT4 acidic mutant was attenuated. These results indicate that the MED1-IDR and the complete Mediator complex each exhibit phase-separating behaviors and suggest that they both incorporate OCT4 in a manner that is dependent on electrostatic interactions provided by acidic amino acids.

To test whether the OCT4 AD acidic mutations affect the ability of the factor to activate transcription *in vivo*, we utilized a GAL4 transactivation assay (Figure 4E). In this system, ADs or their mutant counterparts are fused to the GAL4 DBD and expressed in cells carrying a luciferase reporter plasmid. We found that the wild-type OCT4-AD fused to the GAL4-DBD was able to activate transcription, while the acidic mutant lost this function (Figure 4E). These results indicate that the acidic residues of the OCT4 ADs are necessary for both incorporation into MED1 phase-separated droplets *in vitro* and for gene activation *in vivo*.

Multiple TFs Phase Separate with Mediator Subunit Droplets

TFs with diverse types of ADs have been shown to interact with Mediator subunits, and MED1 is among the subunits that is most targeted by TFs (Table S1). An analysis of mammalian TFs confirmed that TFs and their putative ADs are enriched in IDRs, as previous analyses have shown (Liu et al., 2006; Staby et al., 2017) (Figure 5A). We reasoned that many different TFs might interact with the MED1-IDR to generate liquid droplets and therefore be incorporated into MED1 condensates. To assess whether diverse MED1-interacting TFs can phase separate with MED1, we prepared purified recombinant, mEGFP-tagged, full-length MYC, p53, NANOG, SOX2, RARa, GATA2, and ER (Table S3). When added to droplet formation buffers, most TFs formed droplets alone (Figure 5B). When added to droplet formation buffers with MED1-IDR, all 7 of these TFs concentrated into MED1-IDR droplets (Figures 5C and S5A). We selected p53 droplets for fluorescence recovery after photobleaching (FRAP) analysis; they exhibited rapid and dynamic internal reorganization (Figure S5B), supporting the notion that they are liquid condensates. These results indicate that TFs previously shown to interact with the MED1 subunit of Mediator can do so by forming phase-separated condensates with MED1.

Estrogen Stimulates Phase Separation of ER with MED1

ER is a well-studied example of a ligand-dependent TF. ER consists of an N-terminal ligand-independent AD, a central DBD, and a C-terminal ligand-dependent AD (also called the ligand binding domain [LBD]) (Figure 6A). Estrogen facilitates the interaction of ER with MED1 by binding the LBD of ER, which exposes a binding pocket for LXXLL motifs within the MED1-IDR (Figures 6A and 6B) (Manavathi et al., 2014). We noted that ER can form heterotypic droplets with the MED1-IDR recombinant protein used thus far in these studies (Figure 5C), which lacks the LXXLL motifs. This led us to investigate whether ER-MED1 droplet formation is responsive to estrogen and whether this involves the MED1 LXXLL motifs.

We performed droplet formation assays using a MED1-IDR recombinant protein containing LXXLL motifs (MED1-IDRXL-mCherry) and found that, similar to MED1-IDR and complete Mediator, it had the ability to form droplets alone (Figure 6C). We then tested the ability of ER to phase separate with MED1-IDRXL-mCherry and MED1-IDR-mCherry droplets.

Figure 2. MED1 Condensates Are Dependent on OCT4 Binding *In Vivo*

(A) Schematic of OCT4 degradation. The C terminus of OCT4 is endogenously biallelically tagged with the FKBP protein; when exposed to the small molecule dTag, OCT4 is ubiquitylated and rapidly degraded.

(B) Boxplot representation of log2 fold change in OCT4 and MED1 ChIP-seq reads and RNA-seq reads of super-enhancer (SE)- or typical enhancer (TE)-driven genes, in ESCs carrying the OCT4 FKBP tag, treated with DMSO, or dTAG for 24 hr. The box represents the interquartile; whiskers are 1.5× the interquartile range.

(C) Genome browser view of OCT4 (green) and MED1 (yellow) ChIP-seq data at the *Nanog* locus. The *Nanog* SE (red) shows a 90% reduction of OCT4 and MED1 binding after OCT4 degradation.

(D) Normalized RNA-seq read counts of *Nanog* mRNA shows a 60% reduction upon OCT4 degradation. Error bars represent SEM.

(E) Confocal microscopy images OCT4 and MED1 IF with DNA FISH to the *Nanog* locus in ESCs carrying the OCT4 FKBP tag, treated with DMSO or dTAG. Inset represents a zoomed in view of the yellow box. The merge view displays all three channels (OCT4 IF, MED1 IF, and *Nanog* DNA FISH) together.

(F) OCT4 ChIP-qPCR to the *Mir290* SE in ESCs and differentiated cells (Diff) presented as enrichment over control, relative to signal in ESCs. Error bars represent the SEM from two biological replicates.

(G) MED1 ChIP-qPCR to the *Mir290* SE in ESCs and differentiated cells (Diff) presented as enrichment over control, relative to signal in ESCs. Error bars represent the SEM from two biological replicates.

(H) Normalized RNA-seq read counts of *Mir290* miRNA in ESCs or differentiated cells (Diff). Error bars represents the SEM from two biological replicates.

(I) Confocal microscopy images of MED1 IF and DNA FISH to the *Mir290* genomic locus in ESCs and differentiated cells. Merge (zoom) represents a zoomed-in view of the yellow box in the merged channel.

See also Figure S2 and Tables S2 and S4.

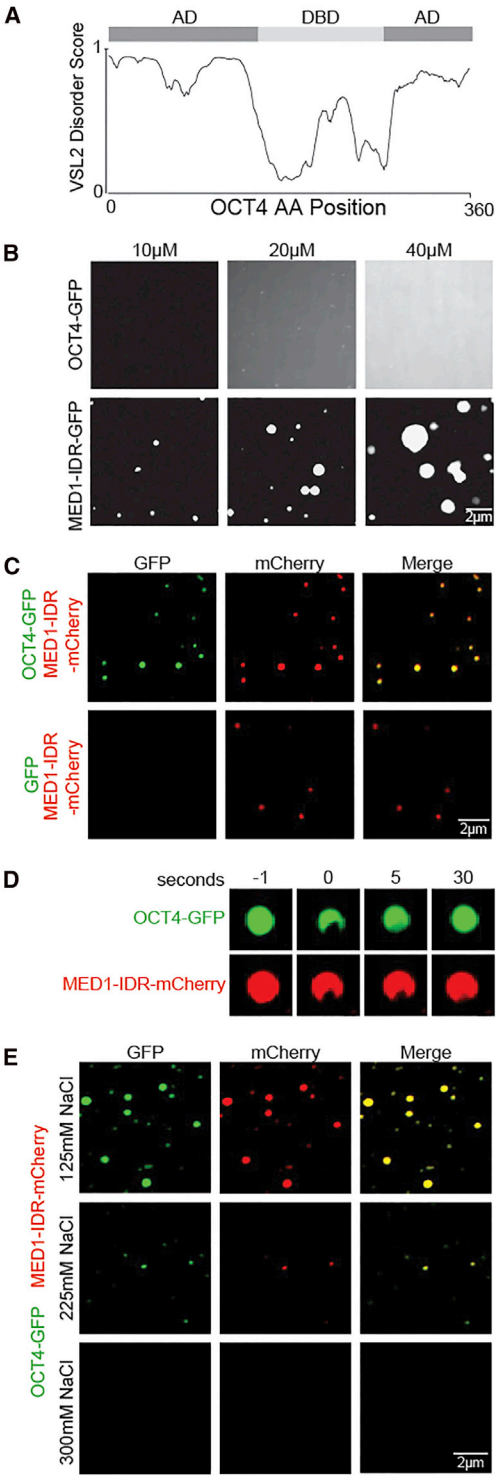


Figure 3. OCT4 Forms Liquid Droplets with MED1 *In Vitro*

(A) Graph of intrinsic disorder of OCT4 as calculated by the VSL2 algorithm (http://www.pondr.com/). The DNA binding domain (DBD) and activation domains (ADs) are indicated above the disorder score graph (Brehm et al., 1997).

(B) Representative images of droplet formation of OCT4-GFP (top row) and MED1-IDR-GFP (bottom row) at the indicated concentration in droplet formation buffer with 125 mM NaCl and 10% PEG-8000.

(C) Representative images of droplet formation of MED1-IDR-mCherry mixed with GFP or OCT4-GFP at 10 μM each in droplet formation buffer with 125 mM NaCl and 10% PEG-8000.

(D) FRAP of heterotypic droplets of OCT4-GFP and MED1-IDR-mCherry. Confocal images were taken at indicated time points relative to photobleaching (0).

(E) Representative images of droplet formation of 10 μM MED1-IDR-mCherry and OCT4-GFP in droplet formation buffer with varying concentrations of salt and 10% PEG-8000.

See also Figure S3 and Table S3.

Some recombinant ER was incorporated and concentrated into MED1-IDRXL-mCherry droplets, but the addition of estrogen considerably enhanced heterotypic droplet formation (Figures 6D and 6E). In contrast, the addition of estrogen had little effect on droplet formation when the experiment was conducted with MED1-IDR-mCherry, which lacks the LXXLL motifs (Figure S6). These results show that estrogen, which stimulates ER-mediated transcription *in vivo*, also stimulates incorporation of ER into MED1-IDR droplets *in vitro*. Thus, OCT4 and ER both require the same amino acids/ligands for both phase separation and activation. Furthermore, since the LBD is a structured domain that undergoes a conformation shift upon estrogen binding to interact with MED1, it appears that structured interactions may contribute to transcriptional condensate formation.

GCN4 and MED15 Phase Separation Is Dependent on Residues Required for Activation

Among the best-studied TF-coactivator systems is the yeast TF GCN4 and its interaction with the MED15 subunit of Mediator (Brzovic et al., 2011; Herbig et al., 2010; Jedidi et al., 2010). The GCN4 AD has been dissected genetically, the amino acids that contribute to activation have been identified (Drysdale et al., 1995; Staller et al., 2018), and recent studies have shown that the GCN4 AD interacts with MED15 in multiple orientations and conformations to form a "fuzzy complex" (Tuttle et al., 2018). Weak interactions that form fuzzy complexes have features of the IDR-IDR interactions that are thought to produce phase-separated condensates.

To test whether GCN4 and MED15 can form phase-separated droplets, we purified recombinant yeast GCN4-GFP and the N-terminal portion of yeast MED15-mCherry containing residues 6–651 (hereafter called MED15), which are responsible for the interaction with GCN4. When added separately to droplet formation buffer, GCN4 formed micrometer-sized droplets only at quite high concentrations (40 μM), and MED15 formed only small droplets at this high concentration (Figure 7A). When mixed together, however, the GCN4 and MED15 recombinant proteins formed double-positive, micrometer-sized, spherical droplets at lower concentrations (Figures 7B and S7A). These GCN4-MED15 droplets exhibited rapid FRAP kinetics (Figure S7B), consistent with liquid-like behavior. We generated a phase diagram of these two proteins and found that they formed droplets together at low concentration (Figures S7C and S7D). This suggests that interaction between the two is required for phase separation at low concentration.

The ability of GCN4 to interact with MED15 and activate gene expression has been attributed to specific hydrophobic patches

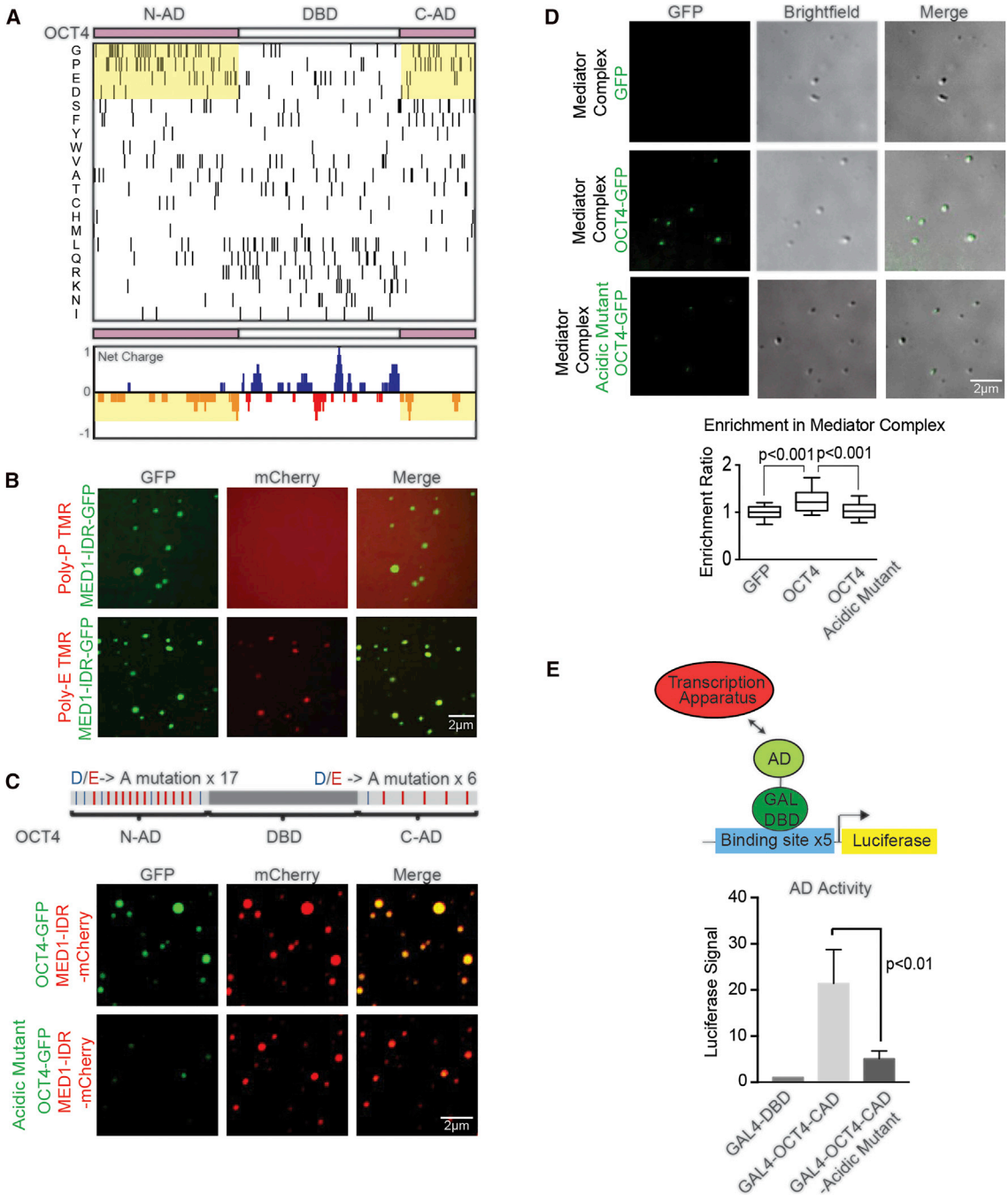


Figure 4. OCT4 Phase Separation with MED1 Is Dependent on Specific Interactions

(A) Amino acid enrichment analysis ordered by frequency of amino acid in the ADs (top). Net charge per amino acid residue analysis of OCT4 (bottom).

(B) Representative images of droplet formation showing that Poly-E peptides are incorporated into MED1-IDR droplets. MED1-GFP and a tetramethylrhodamine (TMR)-labeled proline or glutamic acid decapeptide (Poly-P and Poly-E, respectively) were added to droplet formation buffers at 10 μM each with 125 mM NaCl and 10% PEG-8000.

(C) Top: schematic of OCT4 protein, horizontal lines in the AD mark acidic D residues (blue) and acidic E residues (red). All 17 acidic residues in the N-AD and 6 acidic residues in the C-AD were mutated to alanine to generate an OCT4-acidic mutant. Bottom: representative confocal images of droplet formation showing that the OCT4 acidic mutant has an attenuated ability to concentrate into MED1-IDR droplets. 10 μM MED1-IDR-mCherry and OCT4-GFP or OCT4-acidic mutant-GFP was added to droplet formation buffers with 125 mM NaCl and 10% PEG-8000.

(D) Top: representative images of droplet formation showing that OCT4 but not the OCT4 acidic mutant is incorporated into Mediator complex droplets. Purified Mediator complex was mixed with 10 μM GFP, OCT4-GFP, or OCT4-acidic mutant-GFP in droplet formation buffers with 140 mM NaCl and 10% PEG-8000.

(legend continued on next page)

and aromatic residues in the GCN4 AD (Drysdale et al., 1995; Staller et al., 2018; Tuttle et al., 2018). We created a mutant of GCN4 in which the 11 aromatic residues contained in these hydrophobic patches were changed to alanine (Figure 7C). When added to droplet formation buffers, the ability of the mutant protein to form droplets alone was attenuated (Figure S7E). Next, we tested whether droplet formation with MED15 was affected; indeed, the mutated protein has a compromised ability to form droplets with MED15 (Figures 7C and S7F). Similar results were obtained when GCN4 and the aromatic mutant of GCN4 was added to droplet formation buffers with the complete Mediator complex; while GCN4 was incorporated into Mediator droplets, the incorporation of the GCN4 mutant into Mediator droplets was attenuated (Figures 7D and S7G). These results demonstrate that multivalent, weak interactions between the AD of GCN4 and MED15 promote phase separation into liquid-like droplets.

The ADs of yeast TFs can function in mammalian cells and can do so by interacting with human Mediator (Oliviero et al., 1992). To investigate whether the aromatic mutant of GCN4 AD is impaired in its ability to recruit Mediator *in vivo*, the GCN4 AD and the GCN4 mutant AD were tethered to a Lac array in U2OS cells (Figure 7E) (Janicki et al., 2004). While the tethered GCN4 AD caused robust Mediator recruitment, the GCN4 aromatic mutant did not (Figure 7E). We used the GAL4 transactivation assay described previously to confirm that the GCN4 AD was capable of transcriptional activation *in vivo*, whereas the GCN4 aromatic mutant had lost that property (Figure 7F). These results provide further support for the idea that TF AD amino acids that are essential for phase separation with Mediator are required for gene activation.

DISCUSSION

The results described here support a model whereby TFs interact with Mediator and activate genes by the capacity of their ADs to form phase-separated condensates with this coactivator. For both the mammalian ESC pluripotency TF OCT4 and the yeast TF GCN4, we found that the AD amino acids required for phase separation with Mediator condensates were also required for gene activation *in vivo*. For ER, we found that estrogen stimulates the formation of phase-separated ER-MED1 droplets. ADs and coactivators generally consist of low-complexity amino acid sequences that have been classified as IDRs, and IDR-IDR interactions have been implicated in facilitating the formation of phase-separated condensates. We propose that IDR-mediated phase separation with Mediator is a general mechanism by which TF ADs effect gene expression and provide evidence that this occurs *in vivo* at SEs. We suggest that the ability to phase separate with Mediator, which would employ the features of high valency and low-affinity characteristic of liquid-liquid phase-separated condensates, operates alongside an ability of some TFs to form high-affinity interactions with Mediator (Figure 7G) (Taatjes, 2017).

The model that TF ADs function by forming phase-separated condensates with coactivators explains several observations that are difficult to reconcile with classical lock-and-key models of protein-protein interaction. The mammalian genome encodes many hundreds of TFs with diverse ADs that must interact with a small number of coactivators (Allen and Taatjes, 2015; Arany et al., 1995; Avantaggiati et al., 1996; Dai and Markham, 2001; Eckner et al., 1996; Gelman et al., 1999; Green, 2005; Liu et al., 2009; Merika et al., 1998; Oliner et al., 1996; Yin and Wang, 2014; Yuan et al., 1996), and ADs that share little sequence homology are functionally interchangeable among TFs (Godowski et al., 1988; Hope and Struhl, 1986; Jin et al., 2016; Lech et al., 1988; Ransone et al., 1990; Sadowski et al., 1988; Struhl, 1988; Tora et al., 1989). The common feature of ADs—the possession of low-complexity IDRs—is also a feature that is pronounced in coactivators. The model of coactivator interaction and gene activation by phase-separated condensate formation thus more readily explains how many hundreds of mammalian TFs interact with these coactivators.

Previous studies have provided important insights that prompted us to investigate the possibility that TF ADs function by forming phase-separated condensates. TF ADs have been classified by their amino acid profile as acidic, proline rich, serine/threonine rich, glutamine rich, or by their hypothetical shape as acid blobs, negative noodles, or peptide lassos (Sigler, 1988). Many of these features have been described for IDRs that are capable of forming phase-separated condensates (Babu, 2016; Darling et al., 2018; Das et al., 2015; Dunker et al., 2015; Habchi et al., 2014; van der Lee et al., 2014; Oldfield and Dunker, 2014; Uversky, 2017; Wright and Dyson, 2015). Evidence that the GCN4 AD interacts with MED15 in multiple orientations and conformations to form a "fuzzy complex" (Tuttle et al., 2018) is consistent with the notion of dynamic low-affinity interactions characteristic of phase-separated condensates. Likewise, the low complexity domains of the FET (FUS/EWS/TAF15) RNA-binding proteins (Andersson et al., 2008) can form phase-separated hydrogels and interact with the RNA polymerase II C-terminal domain (CTD) in a CTD phosphorylation-dependent manner (Kwon et al., 2013); this may explain the mechanism by which RNA polymerase II is recruited to active genes in its unphosphorylated state and released for elongation following phosphorylation of the CTD.

The model we describe here for TF AD function may explain the function of a class of heretofore poorly understood fusion oncoproteins. Many malignancies bear fusion-protein translocations involving portions of TFs (Bradner et al., 2017; Kim et al., 2017; Latysheva et al., 2016). These abnormal gene

Bottom: enrichment ratio of GFP, OCT4-GFP, or OCT4-acidic mutant-GFP in Mediator complex droplets. n > 20; error bars represent the distribution between the 10th and 90th percentiles.

(E) Top: GAL4 activation assay schematic. The GAL4 luciferase reporter plasmid was transfected into mouse ESCs with an expression vector for the GAL4-DBD fusion protein. Bottom: the AD activity was measured by luciferase activity of mouse ESCs transfected with GAL4-DBD, GAL4-OCT4-CAD, or GAL4-OCT4-CAD-acidic mutant. Error bars represent SEM.

See also Figure S4 and Table S3.

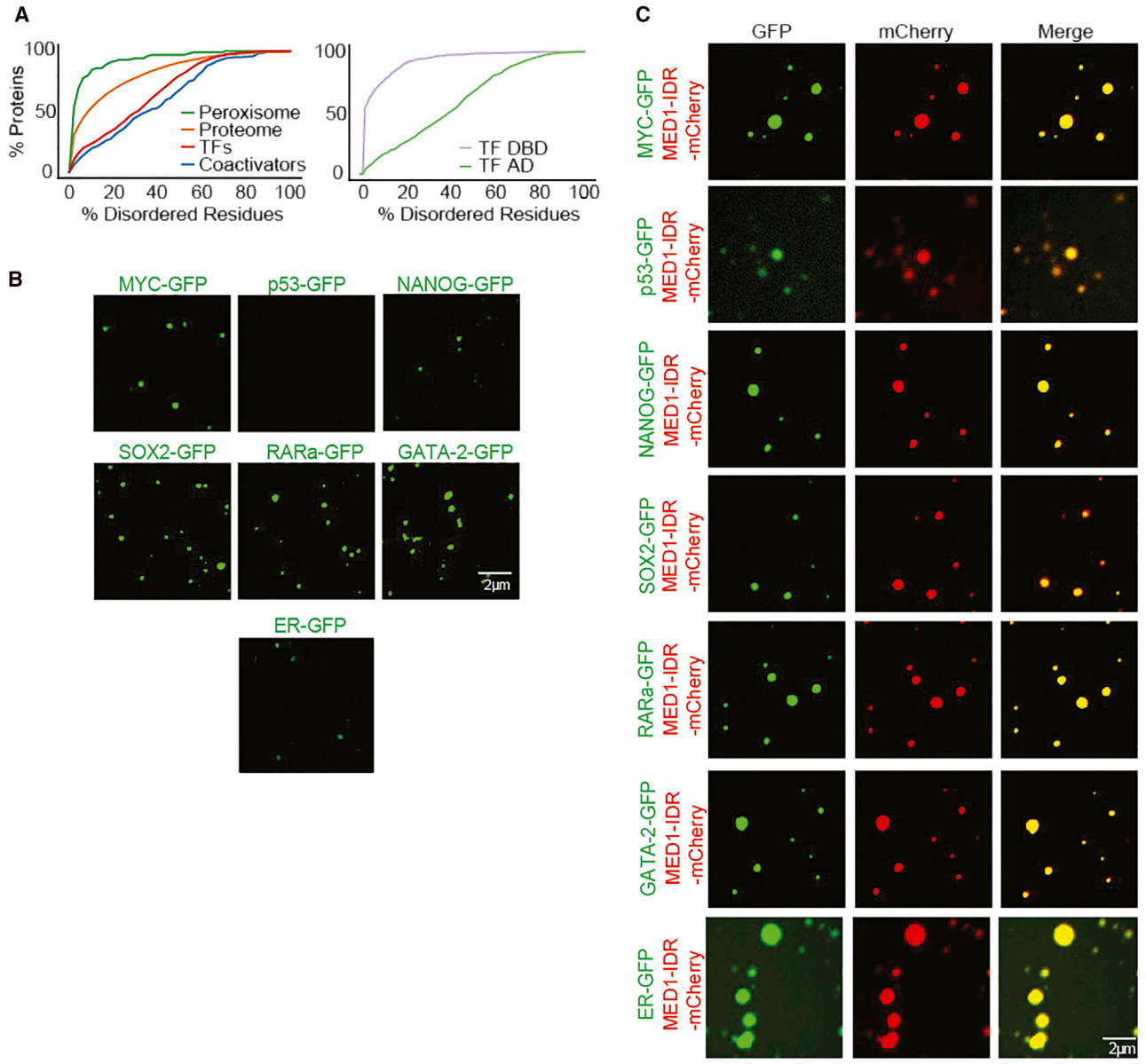


Figure 5. Multiple TFs Phase Separate with Mediator Droplets

(A) Left: percentage of disorder of various protein classes (x axis) plotted against the cumulative fraction of disordered proteins of that class (y axis). Right: disorder content of transcription factor (TF) DNA-binding domains (DBD) and putative activation domains (ADs).

(B) Representative images of droplet formation assaying homotypic droplet formation of indicated TFs. Recombinant MYC-GFP (12 μM), p53-GFP (40 μM), NANOG-GFP (10 μM), SOX2-GFP (40 μM), RARa-GFP (40 μM), GATA-2-GFP (40 μM), and ER-GFP (40 μM) was added to droplet formation buffers with 125 mM NaCl and 10% PEG-8000.

(C) Representative images of droplet formation showing that all tested TFs were incorporated into MED1-IDR droplets. 10 μM of MED1-IDRmCherry and 10 μM of MYC-GFP, p53-GFP, NANOG-GFP, SOX2-GFP, RARa-GFP, GATA-2-GFP, or ER-GFP were added to droplet formation buffers with 125 mM NaCl and 10% PEG-8000.

See also Figure S5 and Table S3.

products often fuse a DNA- or chromatin-binding domain to a wide array of partners, many of which are IDRs. For example, MLL may be fused to 80 different partner genes in AML (Winters and Bernt, 2017), the EWS-FLI rearrangement in Ewing's sarcoma causes malignant transformation by recruitment of a disordered domain to oncogenes (Boulay et al., 2017; Chong et al., 2017), and the disordered phase-separating protein FUS is found fused to a DBD in certain sarcomas (Crozat et al., 1993; Patel et al., 2015). Phase separation provides a mechanism by which such gene products result in aberrant

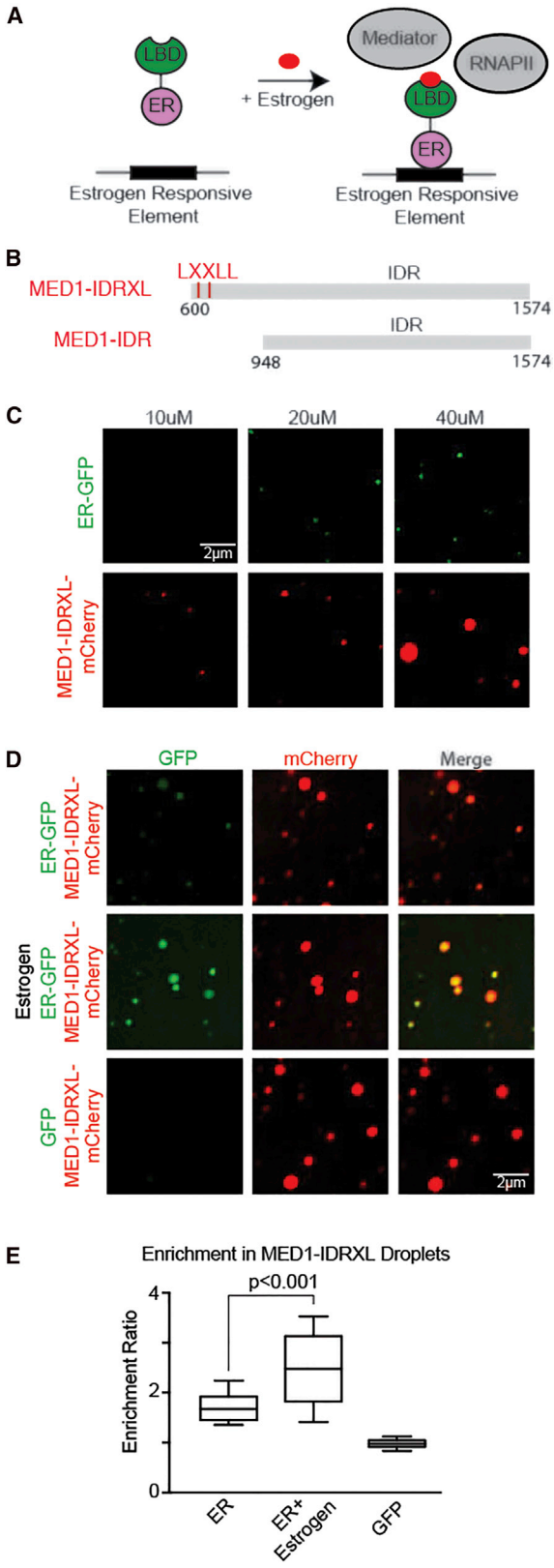


Figure 6. Estrogen Stimulates Phase Separation of ER with MED1

(A) Schematic of estrogen stimulated gene activation. Estrogen facilitates the interaction of ER with Mediator and RNAPII by binding the ligand binding domain (LBD) of ER, which exposes a binding pocket for LXXLL motifs within the MED1-IDR.

(B) Schematic view of MED1-IDRXL and MED1-IDR used for recombinant protein production.

(C) Representative images of droplet formation, assaying homotypic droplet formation of ER-GFP, and MED1-IDRXL-mCherry performed with the indicated protein concentration in droplet formation buffers with 125 mM NaCl and 10% PEG-8000.

(D) Representative confocal images of droplet formation showing that ER is incorporated into MED1-IDRXL droplets and the addition of estrogen enhanced heterotypic droplet formation. ER-GFP, ER-GFP in the presence of estrogen, or GFP is mixed with MED1-IDRXL. 10 μM of each indicated protein was added to droplet formation buffers with 125 mM NaCl and 10% PEG-8000.

(E) Enrichment ratio in MED1-IDRXL droplets of ER-GFP, ER-GFP in the presence of estrogen, or GFP. n > 20; error bars represent the distribution between the 10th and 90th percentiles.

See also Figure S6 and Table S3.

gene expression programs; by recruiting a disordered protein to the chromatin, diverse coactivators may form phase-separated condensates to drive oncogene expression. Understanding the interactions that compose these aberrant transcriptional condensates, their structures, and behaviors may open new therapeutic avenues.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cells
  - Cell culture conditions
- METHOD DETAILS
  - Immunofluorescence with RNA FISH
  - Immunofluorescence with DNA FISH
  - Tissue culture
  - Western blot
  - Chromatin immunoprecipitation (ChIP) qPCR and sequencing
  - RNA-Seq
  - Protein purification
  - *In vitro* droplet assay
  - Genome editing and protein degradation
  - GAL4 transcription assay
  - Lac binding assay
  - Purification of CDK8-Mediator
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Experimental design
  - Average image and radial distribution analysis
  - Chromatin immunoprecipitation PCR and sequencing (ChIP) analysis
  - Super-enhancer identification
  - RNA-Seq analysis
  - Enrichment and charge analysis of OCT4

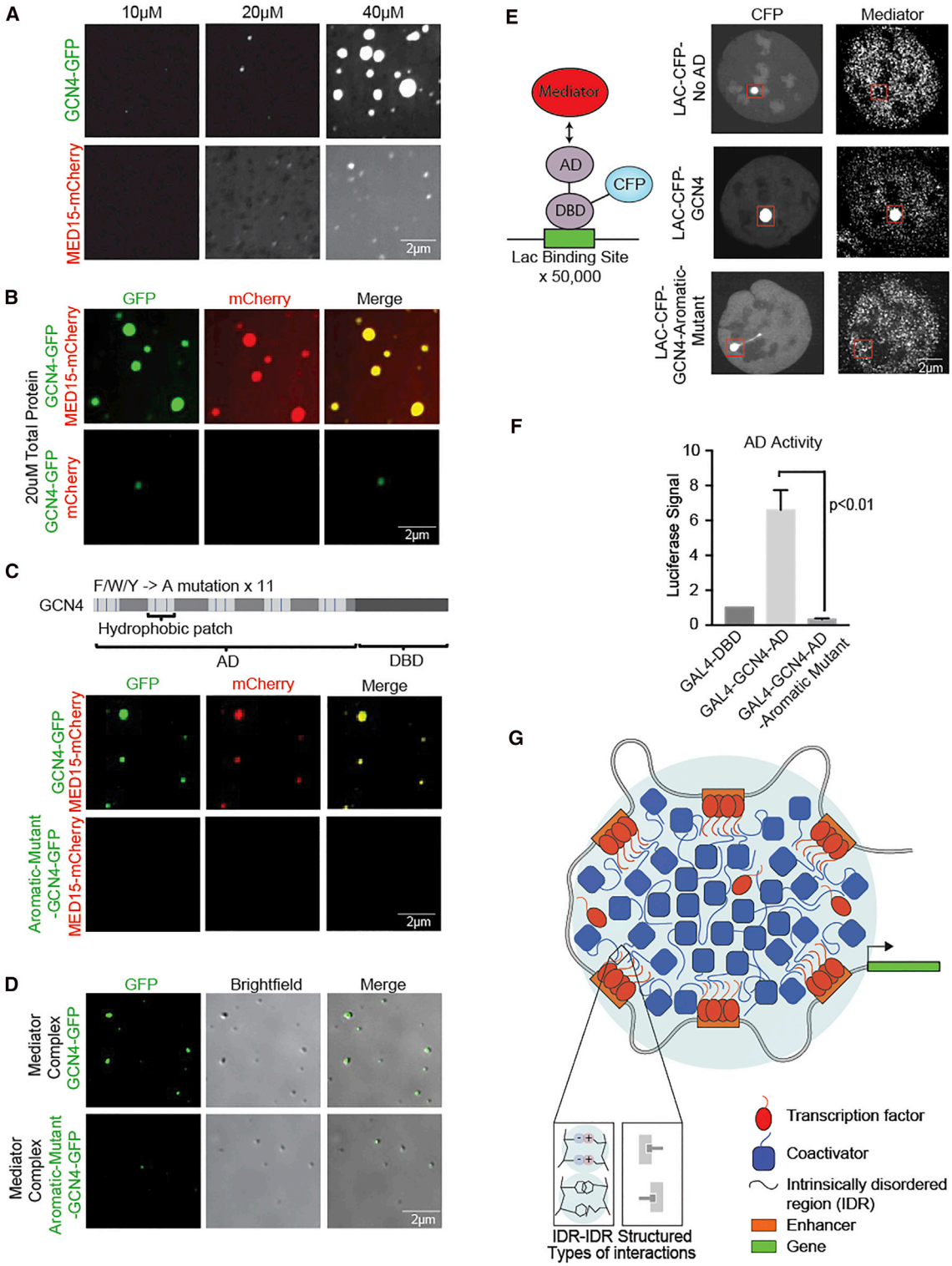


Figure 7. TF-Coactivator Phase Separation Is Dependent on Residues Required for Transactivation

(A) Representative confocal images of droplet formation of GCN4-GFP or MED15-mCherry were added to droplet formation buffers with 125 mM NaCl and 10% PEG-8000.

(B) Representative images of droplet formation showing that GCN4 forms droplets with MED15. GCN4-GFP and mCherry or GCN4-GFP and MED15-mCherry were added to droplet formation buffers at 10 μM with 125 mM NaCl and 10% PEG-8000 and imaged on a fluorescent microscope with the indicated filters.

(legend continued on next page)

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and four tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.10.042.

ACKNOWLEDGMENTS

We thank Wendy Salmon of the W.M. Keck Microscopy Facility, the Whitehead Genomics Core, Arup Chakraborty, and Philip A. Sharp. The work was supported by NIH grant GM123511 and NSF grant PHY1743900 (R.A.Y.), NIH grant GM117370 (D.J.T.), Swedish Research Council Postdoctoral Fellowship (VR 2017-00372) (A.B.), NIH grant T32CA009172 (I.A.K.), a Damon Runyon Cancer Research Foundation Fellowship (2309-17) (B.R.S.), a NSF Graduate Research Fellowship (A.V.Z.), a Hope Funds for Cancer Research fellowship (B.J.A.), a Cancer Research Institute Irvington Fellowship (Y.E.G.), NIH grant T32GM08759 (C.B.F.), and a NWO Rubicon Fellowship (J.S.).

AUTHOR CONTRIBUTIONS

Conceptualization, A.B., I.A.K., and R.A.Y.; Methodology, A.B., I.A.K., B.R.S., A.D., and E.L.C.; Software, K.S. and B.J.A.; Formal Analysis, A.B., I.A.K., C.H.L., K.S., and B.J.A.; Investigation, A.B., I.A.K., B.R.S., A.D., E.L.C., A.V.Z., C.H.L., J.C.M., N.M.H., L.K.A., Y.E.G., J.K.R., C.B.F., and J.S.; Resources, D.J.T. and R.A.Y.; Data Curation, B.J.A.; Writing – Original Draft, A.B., I.A.K., and R.A.Y.; Writing – Reviewing and Editing, all authors; Visualization, A.B., I.A.K., B.R.S., A.D., C.H.L., K.S., B.J.A., and J.S.; Supervision, T.I.L., D.J.T., and R.A.Y.; Project Administration, A.B., I.A.K., and R.A.Y.; Funding Acquisition, D.J.T. and R.A.Y.

DECLARATION OF INTERESTS

The Whitehead Institute filed a patent application based in part on this paper. R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, and Omega Therapeutics. I.A.K. is a consultant to InfiniteMD, Best Doctors, and Foundation Medicine and is a shareholder of InfiniteMD. B.J.A. and T.I.L. are shareholders of Syros Pharmaceuticals. T.I.L. is a consultant to Camp4 Therapeutics.

Received: June 12, 2018
Revised: August 20, 2018
Accepted: October 16, 2018
Published: November 15, 2018

REFERENCES

Alberti, S. (2017). The wisdom of crowds: Regulating cell function through condensed states of living matter. J. Cell Sci. *130*, 2789–2796.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: A central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Andersson, M.K., Ståhlberg, A., Arvidsson, Y., Olofsson, A., Semb, H., Stenman, G., Nilsson, O., and Åman, P. (2008). The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response. BMC Cell Biol. *9*, 37.

Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L., et al. (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell *12*, 699–712.

Arany, Z., Newsome, D., Oldread, E., Livingston, D.M., and Eckner, R. (1995). A family of transcriptional adaptor proteins targeted by the E1A oncoprotein. Nature *374*, 81–84.

Avantaggiati, M.L., Carbone, M., Graessmann, A., Nakatani, Y., Howard, B., and Levine, A.S. (1996). The SV40 large T antigen and adenovirus E1a oncoproteins interact with distinct isoforms of the transcriptional co-activator, p300. EMBO J. *15*, 2236–2248.

Babu, M.M. (2016). The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem. Soc. Trans. *44*, 1185–1200.

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

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.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional addiction in cancer. Cell *9*, 629–643.

Brehm, A., Ohbo, K., and Schöler, H. (1997). The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. Mol. Cell. Biol. *17*, 154–162.

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729–736.

Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., and Hahn, S. (2011). The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. Mol. Cell *44*, 942–953.

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.

(C) Top: schematic of GCN4 protein composed of an activation domain (AD) and DNA-binding domain (DBD). Aromatic residues in the hydrophobic patches of the AD are marked by blue lines. All 11 aromatic residues in the hydrophobic patches were mutated to alanine to generate a GCN4-aromatic mutant. Bottom: representative images of droplet formation showing that the ability of GCN4 aromatic mutant to form droplets with MED15 is attenuated. GCN4-GFP or GCN4-Aromatic mutant-GFP and MED15-mCherry were added to droplet formation at 10 μM each with 125 mM NaCl and 10% PEG-8000.
(D) Representative images of droplet formation showing that GCN4 wild-type but not GCN4 aromatic mutant are incorporated into Mediator complex droplets. 10 μM of GCN4-GFP or GCN4-Aromatic mutant-GFP was mixed with purified Mediator complex in droplet formation buffer with 125 mM NaCl and 10% PEG-8000.
(E) Left: schematic of the Lac assay. A U2OS cell bearing 50,000 repeats of the Lac operon is transfected with a Lac binding domain-CFP-AD fusion protein. Right: IF of MED1 in Lac-U2OS cells transfected with the indicated Lac binding protein construct.
(F) GAL4 activation assay. Transcriptional output as measured by luciferase activity in 293T cells, of the indicated activation domain fused to the GAL4 DBD. Error bars represent SEM.
(G) Model showing TFs and coactivators forming phase-separated condensates at SEs to drive gene activation. In this model, transcriptional condensates incorporate both dynamic and structured interactions.
See also Figure S7 and Table S3.

Chong, S., Dugast-darzacq, C., Liu, Z., Dong, P., and Dailey, G.M. (2017). Dynamic and selective low-complexity domain interactions revealed by live-cell single-molecule imaging. bioRxiv. https://doi.org/10.1101/208710.

Crozat, A., Åman, P., Mandahl, N., and Ron, D. (1993). Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature *363*, 640–644.

Dai, Y.S., and Markham, B.E. (2001). p300 Functions as a coactivator of transcription factor GATA-4. J. Biol. Chem. *276*, 37178–37185.

Darling, A.L., Liu, Y., Oldfield, C.J., and Uversky, V.N. (2018). Intrinsically disordered proteome of human membrane-less organelles. Proteomics *18*, e1700193.

Das, R.K., Ruff, K.M., and Pappu, R.V. (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr. Opin. Struct. Biol. *32*, 102–112.

Drysdale, C.M., Dueñas, E., Jackson, B.M., Reusser, U., Braus, G.H., and Hinnebusch, A.G. (1995). The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. Mol. Cell. Biol. *15*, 1220–1233.

Dunker, A.K., Bondos, S.E., Huang, F., and Oldfield, C.J. (2015). Intrinsically disordered proteins and multicellular organisms. Semin. Cell Dev. Biol. *37*, 44–55.

Eckner, R., Yao, T.P., Oldread, E., and Livingston, D.M. (1996). Interaction and functional collaboration of p300/CBP and bHLH proteins in muscle and B-cell differentiation. Genes Dev. *10*, 2478–2490.

Frietze, S., and Farnham, P.J. (2011). Transcription factor effector domains. Subcell. Biochem. *52*, 261–277.

Fulton, D.L., Sundararajan, S., Badis, G., Hughes, T.R., Wasserman, W.W., Roach, J.C., and Sladek, R. (2009). TFCat: The curated catalog of mouse and human transcription factors. Genome Biol. *10*, R29.

Gelman, L., Zhou, G., Fajas, L., Raspé, E., Fruchart, J.C., and Auwerx, J. (1999). p300 interacts with the N- and C-terminal part of PPARgamma2 in a ligand-independent and -dependent manner, respectively. J. Biol. Chem. *274*, 7681–7688.

Godowski, P.J., Picard, D., and Yamamoto, K.R. (1988). Signal transduction and transcriptional regulation by glucocorticoid receptor-LexA fusion proteins. Science *241*, 812–816.

Green, M.R. (2005). Eukaryotic transcription activation: Right on target. Mol. Cell *18*, 399–402.

Habchi, J., Tompa, P., Longhi, S., and Uversky, V.N. (2014). Introducing protein intrinsic disorder. Chem. Rev. *114*, 6561–6588.

Herbig, E., Warfield, L., Fish, L., Fishburn, J., Knutson, B.A., Moorefield, B., Pacheco, D., and Hahn, S. (2010). Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. Mol. Cell. Biol. *30*, 2376–2390.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model for transcriptional control. Cell *169*, 13–23.

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.

Hope, I.A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell *46*, 885–894.

Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: New tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. *43*, D117–D122.

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

Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H., and Spector, D.L. (2004). From silencing to gene expression: Real-time analysis in single cells. Cell *116*, 683–698.

Jedidi, I., Zhang, F., Qiu, H., Stahl, S.J., Palmer, I., Kaufman, J.D., Nadaud, P.S., Mukherjee, S., Wingfield, P.T., Jaroniec, C.P., and Hinnebusch, A.G. (2010). Activator Gcn4 employs multiple segments of Med15/Gal11, including the KIX domain, to recruit mediator to target genes in vivo. J. Biol. Chem. *285*, 2438–2455.

Jin, W., Wang, L., Zhu, F., Tan, W., Lin, W., Chen, D., Sun, Q., and Xia, Z. (2016). Critical POU domain residues confer Oct4 uniqueness in somatic cell reprogramming. Sci. Rep. *6*, 20818.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. *339*, 225–229.

Keegan, L., Gill, G., and Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science *231*, 699–704.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46* (D1), D260–D266.

Kim, P., Ballester, L.Y., and Zhao, Z. (2017). Domain retention in transcription factor fusion genes and its biological and clinical implications: A pan-cancer study. Oncotarget *8*, 110103–110117.

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.

Latysheva, N.S., Oates, M.E., Maddox, L., Flock, T., Gough, J., Buljan, M., Weatheritt, R.J., and Babu, M.M. (2016). Molecular principles of gene fusion mediated rewiring of protein interaction networks in cancer. Mol. Cell *63*, 579–592.

Lech, K., Anderson, K., and Brent, R. (1988). DNA-bound Fos proteins activate transcription in yeast. Cell *52*, 179–184.

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.

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. Biochemistry *45*, 6873–6888.

Liu, W.-L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R. (2009). Structures of three distinct activator-TFIID complexes. Genes Dev. *23*, 1510–1521.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.

Manavathi, B., Samanthapudi, V.S.K., and Gajulapalli, V.N.R. (2014). Estrogen receptor coregulators and pioneer factors: The orchestrators of mammary gland cell fate and development. Front. Cell Dev. Biol. *2*, 34.

Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol. Cell *1*, 277–287.

Meyer, K.D., Donner, A.J., Knuesel, M.T., York, A.G., Espinosa, J.M., and Taatjes, D.J. (2008). Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. EMBO J. *27*, 1447–1457.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371–378.

Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol. *14*, 431–441.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., and Baldwin,

A.J. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol. Cell *57*, 936–947.

Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B., Dosztányi, Z., Uversky, V.N., Obradovic, Z., Kurgan, L., et al. (2013). D^2P^2: Database of disordered protein predictions. Nucleic Acids Res. *41*, D508–D516.

Oldfield, C.J., and Dunker, A.K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. Annu. Rev. Biochem. *83*, 553–584.

Oliner, J.D., Andresen, J.M., Hansen, S.K., Zhou, S., and Tjian, R. (1996). SREBP transcriptional activity is mediated through an interaction with the CREB-binding protein. Genes Dev. *10*, 2903–2911.

Oliviero, S., Robinson, G.S., Struhl, K., and Spiegelman, B.M. (1992). Yeast GCN4 as a probe for oncogenesis by AP-1 transcription factors: Transcriptional activation through AP-1 sites is not sufficient for cellular transformation. Genes Dev. *6*, 1799–1809.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-β enhanceosome. Cell *129*, 1111–1123.

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.

Plaschka, C., Nozawa, K., and Cramer, P. (2016). Mediator architecture and RNA polymerase II interaction. J. Mol. Biol. *428*, 2569–2574.

Ransone, L.J., Wamsley, P., Morley, K.L., and Verma, I.M. (1990). Domain swapping reveals the modular nature of Fos, Jun, and CREB proteins. Mol. Cell. Biol. *10*, 4565–4573.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr. Opin. Genet. Dev. *43*, 73–81.

Roberts, S.G. (2000). Mechanisms of action of transcription activation and repression domains. Cell. Mol. Life Sci. *57*, 1149–1160.

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*, 80.

Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. Nature *335*, 563–564.

Saint-andré, V., Federation, A.J., Lin, C.Y., Abraham, B.J., Reddy, J., Lee, T.I., Bradner, J.E., and Young, R.A. (2016). Models of human core transcriptional regulatory circuitries. Genome Res. *26*, 385–396.

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.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science. Published online September 22, 2017. https://doi.org/10.1126/science.aaf4382.

Sigler, P.B. (1988). Transcriptional activation. Acid blobs and negative noodles. Nature *333*, 210–212.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. *19*, 262–274.

Staby, L., O'Shea, C., Willemoës, M., Theisen, F., Kragelund, B.B., and Skriver, K. (2017). Eukaryotic transcription factors: Paradigms of protein intrinsic disorder. Biochem. J. *474*, 2509–2532.

Staller, M.V., Holehouse, A.S., Swain-Lenz, D., Das, R.K., Pappu, R.V., and Cohen, B.A. (2018). A high-throughput mutational scan of an intrinsically disordered acidic transcriptional activation domain. Cell Syst. *6*, 444–455.

Struhl, K. (1988). The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast. Nature *332*, 649–650.

Taatjes, D.J. (2010). The human Mediator complex: A versatile, genome-wide regulator of transcription. Trends Biochem. Sci. *35*, 315–322.

Taatjes, D.J. (2017). Transcription factor-mediator interfaces: Multiple and multi-valent. J. Mol. Biol. *429*, 2996–2998.

Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: Polymorphism and structural disorder in protein-protein interactions. Trends Biochem. Sci. *33*, 2–8.

Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E., and Chambon, P. (1989). The human estrogen receptor has two independent nonacidic transcriptional activation functions. Cell *59*, 477–487.

Triezenberg, S.J. (1995). Structure and function of transcriptional activation domains. Curr. Opin. Genet. Dev. *5*, 190–196.

Tuttle, L.M., Pacheco, D., Warfield, L., Luo, J., Ranish, J., Hahn, S., and Klevit, R.E. (2018). Gcn4-mediator specificity is mediated by a large and dynamic fuzzy protein-protein complex. Cell Rep. *22*, 3251–3264.

Uversky, V.N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. Curr. Opin. Struct. Biol. *44*, 18–30.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of intrinsically disordered regions and proteins. Chem. Rev. *114*, 6589–6631.

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.

Warfield, L., Tuttle, L.M., Pacheco, D., Klevit, R.E., and Hahn, S. (2014). A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. Proc. Natl. Acad. Sci. USA *111*, E3506–E3513.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a structural regulator of enhancer-promoter loops. Cell *171*, 1573–1588.

Wheeler, R.J., and Hyman, A.A. (2018). Controlling compartmentalization by non-membrane-bound organelles. Philos. Trans. R. Soc. Lond. B Biol. Sci. *26*, 373.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Winters, A.C., and Bernt, K.M. (2017). MLL-rearranged leukemias—An update on science and clinical approaches. Front Pediatr. *5*, 4.

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. *16*, 18–29.

Yin, J.W., and Wang, G. (2014). The Mediator complex: A master coordinator of transcription and cell lineage development. Development *141*, 977–987.

Yuan, W., Condorelli, G., Caruso, M., Felsani, A., and Giordano, A. (1996). Human p300 protein is a coactivator for the transcription factor MyoD. J. Biol. Chem. *271*, 9009–9013.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
MED1	Abcam	ab64965
OCT4	Santa Cruz	sc-5279X
Goat anti-Rabbit IgG Alexa Fluor 488	Life Technologies	A11008
Goat anti-Rabbit IgG Alexa Fluor 568	Life Technologies	A11011
Goat anti-Mouse IgG Alexa Fluor 674	Thermo Fisher	A21235
Med1	Bethyl	A300-793A-4
Oct4	Santa Cruz	sc-8628x
Beta-Actin	Santa Cruz	sc-7210
HA	abcam	ab9110
Bacterial and Virus Strains		
LOBSTR cells	Cheeseman Lab (WI/MIT)	N/A
Chemicals, Peptides, and Recombinant Proteins		
Beta-Estradiol	Sigma	E8875
TMR-Poly-P Peptide	MIT core facility	N/A
TMR-Poly-E Peptide	MIT core facility	N/A
Critical Commercial Assays		
Dual-glo Luciferase Assay System	Promega	E2920
AllPrep DNA/RNA Mini Kit	QIAGEN	80204
NEBuilder HiFi DNA Assembly Master Mix	NEB	E2621S
Power SYBR Green mix	Life Technologies	4367659
Deposited Data		
Oct4-degron + DMSO ChIP-seq	This Paper	GEO: GSM3401065
Oct4-degron + dTag ChIP-seq	This Paper	GEO: GSM3401066
Oct4-degron + DMSO ChIP-seq	This Paper	GEO: GSM3401067
Oct4-degron + dTag ChIP-seq	This Paper	GEO: GSM3401068
Oct4-degron + DMSO ChIP-Seq Input	This Paper	GEO: GSM3401069
Oct4-degron + dTag ChIP-Seq Input	This Paper	GEO: GSM3401070
Oct4-degron + DMSO RNA-seq	This Paper	GEO: GSM3401252 GEO: GSM3401253
Oct4-degron + dTag RNA-seq	This Paper	GEO: GSM3401254 GEO: GSM3401255
ES Cell RNA-seq	This Paper	GEO: GSM3401256 GEO: GSM3401257
Differentiating ES Cell RNA-seq	This Paper	GEO: GSM3401258 GEO: GSM3401259
Oct4 ChIP-Seq	Whyte et al., 2013	GEO: GSM1082340
Med1 ChIP-seq	Whyte et al., 2013	GEO: GSM560348
Experimental Models: Cell Lines		
V6.5 murine embryonic stem cells	Jaenisch laboratory	N/A
HEK293T cells	ATCC	CRL-3216
U2OS-268 cells	Spector laboratory	N/A
Oligonucleotides		
mir290_Neg_F GGACTCCATCCCTAGTATTTGC	Operon	N/A
mir290_Neg_R GCTAATCACAAATTTGCTCTGC	Operon	N/A
mir290_OCT4_F CCACCTAAACAAAGAACAGCAG	Operon	N/A
mir290_OCT4_R TGTACCCTGCCACTCAGTTTAC	Operon	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
mir290_MED1_F AAGCAGGGTGGTAGAGTAAGGA	Operon	N/A
mir290_MED1_R ATTCCCGATGTGGAGTAGAAGT	Operon	N/A
Recombinant DNA		
pETEC-OCT4-GFP	This Paper	N/A
pETEC-MED1-IDR-GFP	Sabari et al., 2018.	N/A
pETEC-MED1-IDR-mCherry	Sabari et al., 2018.	N/A
pETEC-MED1-IDRXL-mCherry	This Paper	N/A
pETEC-OCT4-aromaticmutant-GFP	This Paper	N/A
pETEC-OCT4-acidicmutant-GFP	This Paper	N/A
pETEC-p53-GFP	This Paper	N/A
pETEC-yeast-MED15-mCherry	This Paper	N/A
pETEC-GCN4-GFP	This Paper	N/A
pETEC-GCN4-aromaticmutant-GFP	This Paper	N/A
pETEC-cMYC-GFP	This Paper	N/A
pETEC-NANOG-GFP	This Paper	N/A
pETEC-SOX2-GFP	This Paper	N/A
pETEC-RARa-GFP	This Paper	N/A
pETEC-GATA2-GFP	This Paper	N/A
pETEC-ER-GFP	This Paper	N/A
Lac-CFP-Empty	This Paper	N/A
Lac-GFP-Gcn4-AD	This Paper	N/A
Lac-GFP-Gcn4-AD-aromaticmutant	This Paper	N/A
pGL3BEC	Modified from Promega	N/A
pRLSV40	Promega	N/A
pGal-DBD	This Paper	N/A
pGal-DBD-Oct4-C-AD	This Paper	N/A
pGal-DBD-Oct4-C-AD-acidicmutant	This Paper	N/A
pGal-DBD-GCN4-AD	This Paper	N/A
pGal-DBD-GCN4-AD-aromaticmutant	This Paper	N/A
pUC19-OCT4-FKBP-BFP	This Paper	N/A
pUC19-OCT4-FKBP-mcherry	This Paper	N/A
pX330-GFP-OCT4	This Paper	N/A
Software and Algorithms		
Fiji image processing package	Schindelin et al., 2012	https://fiji.sc/
MetaMorph acquisition software	Molecular Devices	https://www.moleculardevices.com/products/cellular-imaging-systems/acquisition-and-analysis-software/metamorph-microscopy
localCIDER package	Holehouse et al., 2017	N/A
PONDR	http://www.pondr.com/	N/A
Other		
Esrrb RNA FISH probe	Stellaris	N/A
Nanog RNA FISH probe	Stellaris	N/A
miR290 RNA FISH probe	Stellaris	N/A
Trim28 RNA FISH probe	Stellaris	N/A
Nanog DNA FISH probe	Agilent	N/A
Mir290 DNA FISH probe	Agilent	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cells

V6.5 murine embryonic stem 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. HEK293T cells were purchased from ATCC (ATCC CRL-3216). Cells were negative for mycoplasma.

Cell culture conditions

V6.5 murine embryonic stem (mES) cells were grown in 2i + LIF conditions. mES cells were always 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). For differentiation mESCs were cultured in serum media as follows: DMEM (Invitrogen, 11965-092) supplemented with 15% fetal bovine serum (Hyclone, characterized SH3007103), 100 mM nonessential amino acids (Invitrogen, 11140-050), 2 mM L-glutamine (Invitrogen, 25030-081), 100 U/mL penicillin, 100 mg/mL streptomycin (Invitrogen, 15140-122), and 0.1mM b-mercaptoethanol (Sigma Aldrich). HEK293T cells were purchased from ATCC (ATCC CRL-3216) and cultured in DMEM, high glucose, pyruvate (GIBCO 11995-073) with 10% fetal bovine serum (Hyclone, characterized SH3007103), 100 U/mL Penicillin-Streptomycin (GIBCO 15140), 2 mM L-glutamine (Invitrogen, 25030-081). Cells were negative for mycoplasma.

METHOD DETAILS

Immunofluorescence with RNA FISH

Coverslips were coated at 37°C with 5ug/mL poly-L-ornithine (Sigma-Aldrich, P4957) for 30 minutes and 5 μg/mL of Laminin (Corning, 354232) for 2 hours. Cells were plated on the pre-coated 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 μM RNA probe (Table S4, 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 37C. After washing with Wash buffer A for 30 minutes at 37°C, the nuclei were stained with 20 μm/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).

Immunofluorescence with DNA FISH

Immunofluorescence was performed as previously 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 7 μL of FISH Hybridization Buffer (Agilent G9400A), 1 μL of FISH probe (see below for region) and 2 μL of water. 5 μL of mixture was added on a slide and coverslip was placed on top (cell-side toward the hybridization mixture). Coverslips were 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 the 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 with Hoechst in PBS for 5 minutes at RT. Coverslips were washed three times in PBS, mounted on slides using Vectashield and

sealed with nail polish. Images were acquired on 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.

DNA FISH probes were custom designed and generated by Agilent to target Nanog and MiR290 super enhancers.
Nanog
Design Input Region – mm9
chr6 122605249 – 122705248
Design Region – mm9
chr6: 122605985-122705394
Mir290
Design Region – mm10
chr7: 3141151 – 3241381

Tissue culture

V6.5 murine embryonic stem cells (mESCs) were a gift from the Jaenisch lab. Cells were grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates in 2i media (DMEM-F12 (Life Technologies, 11320082), 0.5X B27 supplement (Life Technologies, 17504044), 0.5X N2 supplement (Life Technologies, 17502048), an extra 0.5mM L-glutamine (GIBCO, 25030-081), 0.1mM b-mercaptoethanol (Sigma, M7522), 1% Penicillin Streptomycin (Life Technologies, 15140163), 0.5X nonessential amino acids (GIBCO, 11140-050), 1000 U/ml LIF (Chemico, ESG1107), 1 μM PD0325901 (Stemgent, 04-0006-10), 3 μM CHIR99021 (Stemgent, 04-0004-10)). Cells were grown at 37°C with 5% CO2 in a humidified incubator. For confocal imaging, cells were grown on glass coverslips (Carolina Biological Supply, 633029), coated with 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 2hrs-16hrs at 37°C. For passaging, cells were washed in PBS (Life Technologies, AM9625), 1000 U/mL LIF. 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. For differentiation of mESCs, 6000 cells were plated per well of a 6 well tissue culture dish, or 1000 cells were plated per well of a 24 well plate with a laminin coated glass coverslip. After 24 hours, 2i media was replaced with FBS media (above) without LIF. Media was changed daily for 5 days, cells were then harvested.

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 hr, 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 (Table S2) 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.

Chromatin immunoprecipitation (ChIP) qPCR and sequencing

mES were grown to 80% confluence in 2i media. 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 r*cf.*, 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 r*cf.* for 5 minutes at 4°C. 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 r*cf*. for 10 minutes at 4°C. Input material was reserved and the remainder was incubated overnight at 4°C with magnetic beads bound with antibody (Table S2) to enrich for DNA fragments bound by the indicated factor. 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 μL TE was added and then RNA was degraded by the addition of 2.5 μL of 33 mg/mL RNase A (Sigma, R4642) and incubation at 37°C for 2 hours. Protein was degraded by the addition of 10 μL 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 μL TE and used for either qPCR or sequencing. For ChIP-qPCR experiments, qPCR was performed using Power SYBR Green mix (Life Technologies #4367659) on either a QuantStudio 5 or a QuantStudio 6 System (Life Technologies).

RNA-Seq

RNA-Seq was performed in the indicated cell line with the indicated treatment, and used to determine expressed genes. RNA was isolated by AllPrep Kit (QIAGEN 80204) and polyA selected libraries were prepared using the KAPA mRNA HyperPrep Kit (Kapa Biosystems KK8505) according to manufacturer's protocol, and single-end sequenced on a Hi-seq 2500 instrument.

Protein purification

cDNA encoding the genes of interest or their IDRs were cloned into a modified version of a T7 pET expression vector. The base vector was engineered to include a 5′ 6xHIS followed by either 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. Vectors expressing mEGFP or mCherry alone contain the linker sequence followed by a STOP codon. Mutant sequences were synthesized as geneblocks (IDT) and inserted into the same base vector as described above. 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 was inoculated into LB media containing kanamycin and chloramphenicol and grown overnight at 37°C. Cells containing the MED1-IDR constructs 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 18 hours. Cells were collected and stored frozen at −80°C. Cells containing all other constructs were treated in a similar manner except they were grown for 5 hours at 37°C after IPTG induction.

Pellets of 500ml of cMyc and Nanog cells were resuspended in 15ml of denaturing buffer (50mM Tris 7.5, 300mM NaCl, 10mM imidazole, 8M Urea) containing cOmplete protease inhibitors (Roche,11873580001) and sonicated (ten cycles of 15 s on, 60 s off). The lysates were cleared by centrifugation at 12,000 g 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. The slurry was poured into a column, washed with 15 volumes of the lysis buffer and eluted 4 X with denaturing buffer containing 250mM imidazole. Each fraction was run on a 12% gel and proteins of the correct size were dialyzed first against buffer (50mM Tris pH 7.5, 125Mm 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. 500ml cell pellets were resuspended in 15ml of Buffer A (50mM Tris pH7.5, 500 mM NaCl) containing 10mM imidazole and cOmplete protease inhibitors, sonicated, lysates cleared by centrifugation at 12,000 g for 30 minutes at 4°C, added to 1ml of pre-equilibrated Ni-NTA agarose, and rotated at 4°C for 1.5 hours. The slurry was poured into a column, washed with 15 volumes of Buffer A containing 10mM imidazole and protein was eluted 2 X with Buffer A containing 50mM imidazole, 2 X with Buffer A containing 100mM imidazole, and 3 X with Buffer A containing 250mM imidazole. Alternatively, the resin slurry was centrifuged at 3,000rpm for 10 minutes, washed with 15 volumes of Buffer and proteins were eluted by incubation for 10 or more minutes rotating with each of the buffers above (50mM, 100mM and 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.

In vitro droplet assay

Recombinant 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 proteins were added to solutions at varying concentrations with indicated final salt and 10% PEG-8000 as crowding agent in Droplet Formation Buffer (50mM Tris-HCl pH 7.5, 10% glycerol, 1mM DTT). The protein solution was immediately loaded onto a homemade chamber comprising a glass slide with a coverslip attached by two parallel strips of double-sided tape. Slides were then imaged with an Andor confocal microscope with a 150x objective. Unless indicated, images presented are of droplets settled on the glass coverslip. For experiments with fluorescently labeled polypeptides, the indicated decapeptides were synthesized by the Koch Institute/MIT Biopolymers & Proteomics Core Facility with a TMR fluorescent tag. The protein of interest was added Buffer D with 125mM NaCl and 10% Peg-8000 with the indicated polypeptide and imaged as described above. For FRAP of *in vitro* droplets 5 pulses of laser at a 50us dwell time was applied to the droplet, and recovery was imaged on an Andor microscope every 1 s for the indicated time periods. For estrogen stimulation experiments, fresh B-Estradiol (E8875 Sigma) was reconstituted to 10mM in 100% EtOH then diluted in 125mM NaCl droplet formation buffer to

100uM. One microliter of this concentrated stock was used in a 10uL droplet formation reaction to achieve a final concentration of 10uM.

Genome editing and protein degradation

The CRISPR/Cas9 system was used to genetically engineer ESC lines. Target-specific oligonucleotides were cloned into a plasmid carrying a codon-optimized version of Cas9 with GFP (gift from R. Jaenisch). The sequences of the DNA targeted (the protospacer adjacent motif is underlined) are listed in the same table. For the generation of the endogenously tagged lines, 1 million Med1-mEGFP tagged mES cells were transfected with 2.5 mg Cas9 plasmid containing the guide sequence below (pX330-GFP-Oct4) and 1.25 mg non-linearized repair plasmid 1 (pUC19-Oct4-FKBP-BFP) and 1.25 mg non-linearized repair plasmid 2 (pUC19-Oct4-FKBP-mcherry) (Table S3). Cells were sorted after 48 hours for the presence of GFP. Cells were expanded for five days and then sorted again for double positive mCherry and BFP cells. Forty thousand mCherry+/BFP+ sorted cells were plated in a six-well plate in a serial dilution. The cells were grown for approximately one week in 2i medium and then individual colonies were picked using a stereoscope into a 96-well plate. Cells were expanded and genotyped by PCR, degradation was confirmed by western blot and IF. Clones with a homozygous knock-in tag were further expanded and used for experiments. A clonal homozygous knock-in line expressing FKBP tagged Oct4 was used for the degradation experiments. Cells were grown in 2i and then treated with dTAG-47 at a concentration of 100 nM for 24 hours, then harvested.

Oct4 Guide sequence
tgcattcaaactgaggcacc*NGG(PAM)

GAL4 transcription assay

Transcription factor constructs were assembled in a mammalian expression vector containing an SV40 promoter driving expression of a GAL4 DNA-binding domain. Wild-type and mutant activation domains of *Oct4* and *Gcn4* were fused to the C terminus of the DNA-binding domain by Gibson cloning (NEB 2621S), joined by the linker GAPGSAGSAAGGSG. These transcription factor constructs were transfected using Lipofectamine 3000 (Thermofisher L3000015) into HEK293T cells (ATCC CRL-3216) or V6.5 mouse embryonic stem cells, that were grown in white flat-bottom 96-well assay plates (Costar 3917). The transcription factor constructs were co-transfected with a modified version of the PGL3-Basic (Promega) vector containing five GAL4 upstream activation sites upstream of the firefly luciferase gene. Also co-transfected was pRL-SV40 (Promega), a plasmid containing the Renilla luciferase gene driven by an SV40 promoter. 24 hours after transfection, luminescence generated by each luciferase protein was measured using the Dual-glo Luciferase Assay System (Promega E2920). The data as presented has been controlled for Renilla luciferase expression.

Lac binding assay

Constructs were assembled by NEB HIFI cloning in pSV2 mammalian expression vector containing an SV40 promoter driving expression of a CFP-LacI fusion protein. The activation domains and mutant activation domains of *Gcn4* were fused by the c-terminus to this recombinant protein, joined by the linker sequence GAPGSAGSAAGGSG. U2OS-268 cells containing a stably integrated array of ~51,000 Lac-repressor binding sites (a gift of the Spector laboratory) were transfected using lipofectamine 3000 (Thermofisher L3000015). 24 hours after transfection, cells were plated on fibronectin-coated glass coverslips. After 24 hours on glass coverslips, cells were fixed for immunofluorescence with a MED1 antibody (Table S2) as described above and imaged, by spinning disk confocal microscopy.

Purification of CDK8-Mediator

The CDK8-Mediator samples were purified as described (Meyer et al., 2008) 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 20 mM KCl, 20mM HEPES, 0.1mM EDTA, 2mM $MgCl_2$, 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 (Meyer et al., 2008), eluted material was loaded onto a 2.2mL centrifuge tube containing 2mL 0.15M KCl HEMG (20mM HEPES, 0.1mM EDTA, 2mM $MgCl_2$, 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 with or without 10 μM indicated GFP-tagged protein. Droplet reactions contained 10% PEG-8000 and 140mM salt.

QUANTIFICATION AND STATISTICAL ANALYSIS

Experimental design

All experiments were replicated. For the specific number of replicates done see either the figure legends or the specific section below. No aspect of the study was done blinded. Sample size was not predetermined and no outliers were excluded.

Average image and radial distribution analysis

For analysis of RNA FISH with immunofluorescence custom in-house MATLAB scripts were written to process and analyze 3D image data gathered in FISH (RNA/DNA) and IF channels. FISH foci were manually identified in individual z stacks through intensity thresholds, centered along a box of size $l = 2.9\ \mu m$, and stitched together in 3-D across z stacks. The called FISH foci are cross-referenced against a manually curated list of FISH foci to remove false positives, which arise due to extra-nuclear signal or blips. For every RNA FISH focus identified, signal from the corresponding location in the IF channel is gathered in the *l* x *l* square centered at the RNA FISH focus at every corresponding z-slice. The IF signal centered at FISH foci for each FISH and IF pair are then combined and an average intensity projection is calculated, providing averaged data for IF signal intensity within a *l* x *l* square centered at FISH foci. The same process was carried out for the FISH signal intensity centered on its own coordinates, providing averaged data for FISH signal intensity within a *l* x *l* square centered at FISH foci. As a control, this same process was carried out for IF signal centered at randomly selected nuclear positions. Randomly selected nuclear positions were identified for each image set by first identifying nuclear volume and then selecting positions within that volume. Nuclear volumes were determined from DAPI staining through the z stack image, which was then processed through a custom CellProfiler pipeline (included as auxiliary file). Briefly, this pipeline rescales the image intensity, condenses the image to 20% of original size for speed of processing, enhances detected speckles, filters median signal, thresholds bodies, removes holes, filters the median signal, dilates the image back to original size, watersheds nuclei, and converts the resulting objects into a black and white image. This black and white image is used as input for a custom R script that uses readTIFF and im (from spatstat) to select 40 random nuclear voxels per image set. These average intensity projections were then used to generate 2D contour maps of the signal intensity or radial distribution plots. Contour plots are generated using in-built functions in MATLAB. The intensity radial function ($\langle r \rangle$) is computed from the average data. 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 IF channel, we used chroma.js (an online color generator) to generate colors across 15 bins, with the key transition colors chosen as black, blueviolet, mediumblue, lime. This was done to ensure that the reader's eye could more readily detect the contrast in signal. The generated colormap was employed to 15 evenly spaced intensity bins for all IF plots. The averaged IF centered at FISH or at randomly selected nuclear locations are plotted using the same color scale, set to include the minimum and maximum signal from each plot. For DNA FISH analysis FISH foci were manually identified in individual z stacks through intensity thresholds in FIJI and marked as a reference area. The reference areas were then transferred to the MED1 IF channel of the image and the average IF signal within the FISH focus was determined. The average signal across 5 images comprising greater than 10 cells per image was averaged to calculate the mean MED1 IF intensity associated with the DNA FISH focus.

Chromatin immunoprecipitation PCR and sequencing (ChIP) analysis

Values displayed in the figures were normalized to the input. The average WT normalized values and standard deviation are displayed. The primers used are listed below. ChIP values at the region of interest (ROI) were normalized to input values (fold input) and for the *mir290* enhancer an additional negative region (negative norm). Values are displayed as normalized to the ES state in differentiation experiments and to DMSO control in OCT4 degradation experiments (control normalization). qPCR reactions were performed in technical triplicate.

$$\text{Fold input} = 2^{(Ct\_input - Ct\_ChIP)}$$

$$\text{Negative norm} = \frac{Fold\ input_{ROI}}{Fold\ input_{neg}}$$

$$\text{Control norm}(\text{Differentiation}) = \frac{Neg\ norm_{Differentiated}}{Neg\ norm_{ES}}$$

ChIP qPCR Primers
Mir290

mir290_Neg_F GGACTCCATCCCTAGTATTTGC
mir290_Neg_R GCTAATCACAAATTTGCTCTGC
mir290_OCT4_F CCACCTAAACAAAGAACAGCAG
mir290_OCT4_R TGTACCCTGCCACTCAGTTTAC
mir290_MED1_F AAGCAGGGTGGTAGAGTAAGGA
mir290_MED1_R ATTCCCGATGTGGAGTAGAAGT

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. ChIP-Seq tracks shown in Figure 1 are derived from GEO: GSM1082340 (OCT4) and GEO: GSM560348 (MED1) from Whyte et al. (2013). Super-enhancers and typical enhancers and their associated genes in cells grown in 2i conditions were downloaded from Sabari et al., 2018. Distributions of occupancy fold-changes were calculated using bamToGFF (https://github.com/BradnerLab/pipeline) to quantify coverage in super-enhancers and typical enhancers from cells grown in 2i conditions. Reads overlapping each typical and super-enhancer were determined using bamToGFF with parameters -e 200 -f 1 -t TRUE and were subsequently normalized to the millions of mapped reads (RPM). RPM-normalized input read counts from each condition were then subtracted from RPM-normalized ChIP-Seq read counts from the corresponding condition. Values from regions wherein this subtraction resulted in a negative number were set to 0. Log2 fold-changes were calculated between DMSO-treated (normal OCT4 amount) and dTAG-treated (depleted OCT4); one pseudocount was added to each condition.

Super-enhancer identification

Super-enhancers were identified as described in Whyte et al. (2013). Peaks of enrichment in MED1 were identified using MACS with –p 1e-9 –keep-dup = 1 and input control. MED1 aligned reads from the untreated condition and corresponding peaks of MED1 were used as input for ROSE (https://bitbucket.org/young_computation/) with parameters -s 12500 -t 2000 -g mm9 and input control. A custom gene list was created by adding D7Ertd143e, and removing Mir290, Mir291a, Mir291b, Mir292, Mir293, Mir294, and Mir295 to prevent these nearby microRNAs that are part of the same transcript from being multiply counted. Stitched enhancers (super-enhancers and typical enhancers) were assigned to the single expressed RefSeq transcript whose promoter was nearest the center of the stitched enhancer. Expressed transcripts were defined as above.

RNA-Seq analysis

For analysis, raw reads were aligned to the mm9 revision of the mouse reference genome using hisat2 with default parameters. Gene name-level read count quantification was performed with htseq-count with parameters -I gene_id –stranded = reverse -f bam -m intersection-strict and a GTF containing transcript positions from Refseq, downloaded 6/6/18. Normalized counts, normalized fold-changes, and differential expression p values were determined using DEseq2 using the standard workflow and both replicates of each condition.

Enrichment and charge analysis of OCT4

Amino acid composition plots were generated using R by plotting the amino acid identity of each residue along the amino acid sequence of the protein. Net charge per residue for OCT4 was determined by computing the average amino acid charge along the OCT4 amino acid sequence in a 5 amino acid sliding window using the localCIDER package (Holehouse et al., 2017).

Disorder enrichment analysis

A list of human transcription factors protein sequences is used for all analysis on TFs, as defined in (Saint-andré et al., 2016). The reference human proteome (Uniprot UP000005640) is used to distill the list (down to ∼1200 proteins), mostly removing non-canonical isoforms. Transcriptional coactivators and Pol II associated proteins were identified in humans using the GO enrichments IDS GO:0003713 and GO:0045944. The reference human proteome defined above was used to generate list of all human proteins, and peroxisome and Golgi proteins were identified from Uniprot reviewed lists. For each protein, D2P2 was used to assay disorder propensity for each amino acid. An amino acid in a protein is considered disordered if at least 75% of the algorithms employed by D2P2 (Oates et al., 2013) predict the residue to be disordered. Additionally, for transcription factors, all annotated PFAM domains were identified (5741 in total, 180 unique domains). Cross-referencing PFAM annotation for known DNA-binding activity, a subset of 45 unique high-confidence DNA-binding domains were identified, accounting for ∼85% of all identified domains. The vast majority of TFs (> 95%) had at least one identified DNA-binding domain. Disorder scores were computed for all DNA-binding regions in every TF, as well as the remaining part of the sequence, which includes most identified trans-activation domains.

Imaging analysis of *in vitro* droplets

To analyze *in-vitro* phase separation imaging experiments, custom MATLAB scripts were written to identify droplets and characterize their size and shape. For any particular experimental condition, intensity thresholds based on the peak of the histogram and size thresholds (2 pixel radius) were employed to segment the image. Droplet identification was performed on the "scaffold" channel (MED1 in case of MED1 + TFs, GCN4 for GCN4+MED15), and areas and aspect ratios were determined. To calculate enrichment for the *in vitro* droplet assay, droplets were defined as a region of interest in FIJI by the scaffold channel, and the maximum signal of the client within that droplet was determined. Scaffolds chosen were MED1, Mediator complex, or GCN4. This was divided by the background client signal in the image to generate a Cin/out. Enrichment scores were calculated by dividing the Cin/out of the experimental condition by the Cin/out of a control fluorescent protein (either GFP or mCherry).

DATA AND SOFTWARE AVAILABILITY

Datasets

Figure	Dataset type	IP target	Sample	GEO
Figure 2B	ChIP-Seq	OCT4	Oct4-degron + DMSO	GEO: GSM3401065
Figure 2B	ChIP-Seq	OCT4	Oct4-degron + dTag	GEO: GSM3401066
Figure 2B	ChIP-Seq	MED1	Oct4-degron + DMSO	GEO: GSM3401067
Figure 2B	ChIP-Seq	MED1	Oct4-degron + dTag	GEO: GSM3401068
Figure 2B	ChIP-Seq Input	N/A	Oct4-degron + DMSO	GEO: GSM3401069
Figure 2B	ChIP-Seq Input	N/A	Oct4-degron + dTag	GEO: GSM3401070
Figure 2B	RNA-Seq	N/A	Oct4-degron + DMSO	GEO: GSM3401252 GEO: GSM3401253
Figure 2B	RNA-Seq	N/A	Oct4-degron + dTag	GEO: GSM3401254 GEO: GSM3401255
Figure 2H	RNA-Seq	N/A	ES Cell	GEO: GSM3401256 GEO: GSM3401257
Figure 2H	RNA-Seq	N/A	Differentiating ES Cell	GEO: GSM3401258 GEO: GSM3401259

The overall accession number for the sequencing data reported in this paper is: GEO: GSE120476.

Supplemental Figures

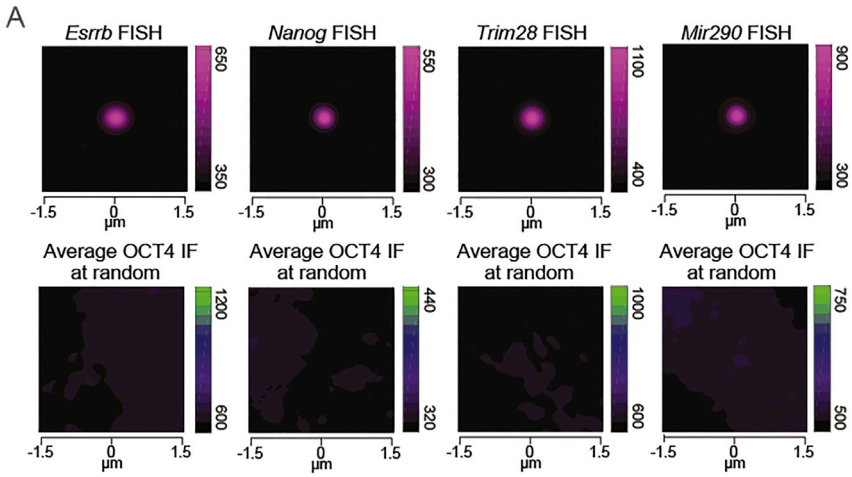


Figure S1. Random Focus Analysis, Related to Figure 1

Average fluorescence centered at the indicated RNA FISH focus (top panels) versus a randomly distributed IF foci ± 1.5 microns in X and Y (bottom panels). Color scale bars present arbitrary units of fluorescence intensity.

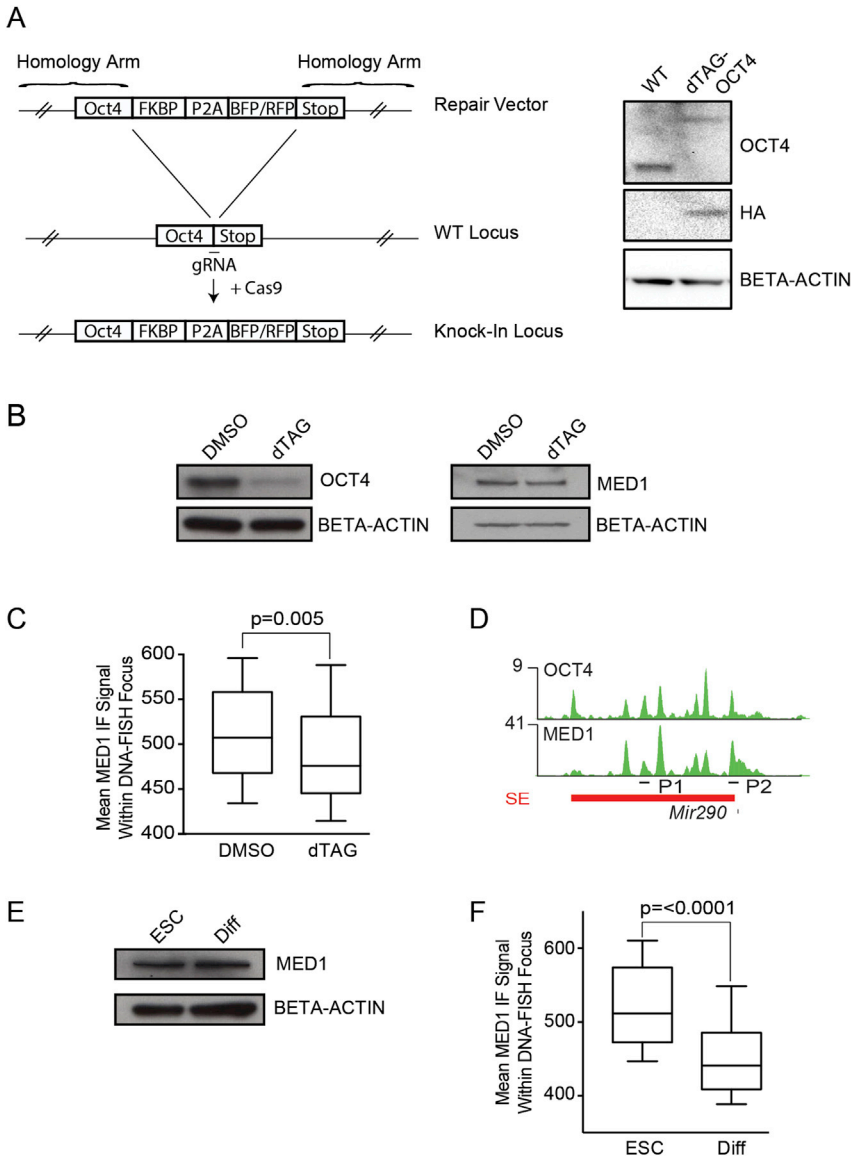


Figure S2. OCT4 Degradation and ES Cell Differentiation, Related to Figure 2

A. Schematic of the Oct4-FKBP cell-engineering strategy. V6.5 mouse ES cells were transfected with a repair vector and Cas9 expressing plasmid to generate knock-in loci with either BFP or RFP for selection (Left). WT or untreated OCT4-dTAG ES cells blotted for OCT4 showing expected shift in size, HA (on FKBP), and ACTIN (Right).

B. Western blot against OCT4 (left panels), MED1 (right panels), and BETA-ACTIN in the OCT4 degron line (dTAG), either treated with dTag47 or vehicle (DMSO).

C. Mean intensity of the MED1 IF signal within the Nanog DNA FISH focus in DMSO treated, versus dTAG treated OCT4-degron cells. n = 5 images, error bars are distribution between the 10th and 90th percentile.

D. Schematic showing the position of primers used for OCT4 (P1) and MED1 (P2) ChIP-qPCR in differentiated and ES cells at the MiR290 locus.

E. Western blot against MED1 and BETA-ACTIN in ES cells or cells differentiated by LIF withdrawal.

F. Mean intensity of MED1 IF signal within MiR290 DNA FISH focus in ES cells versus cells differentiated by LIF withdrawal. n = 5 images, error bars are distribution between the 10th and 90th percentile.

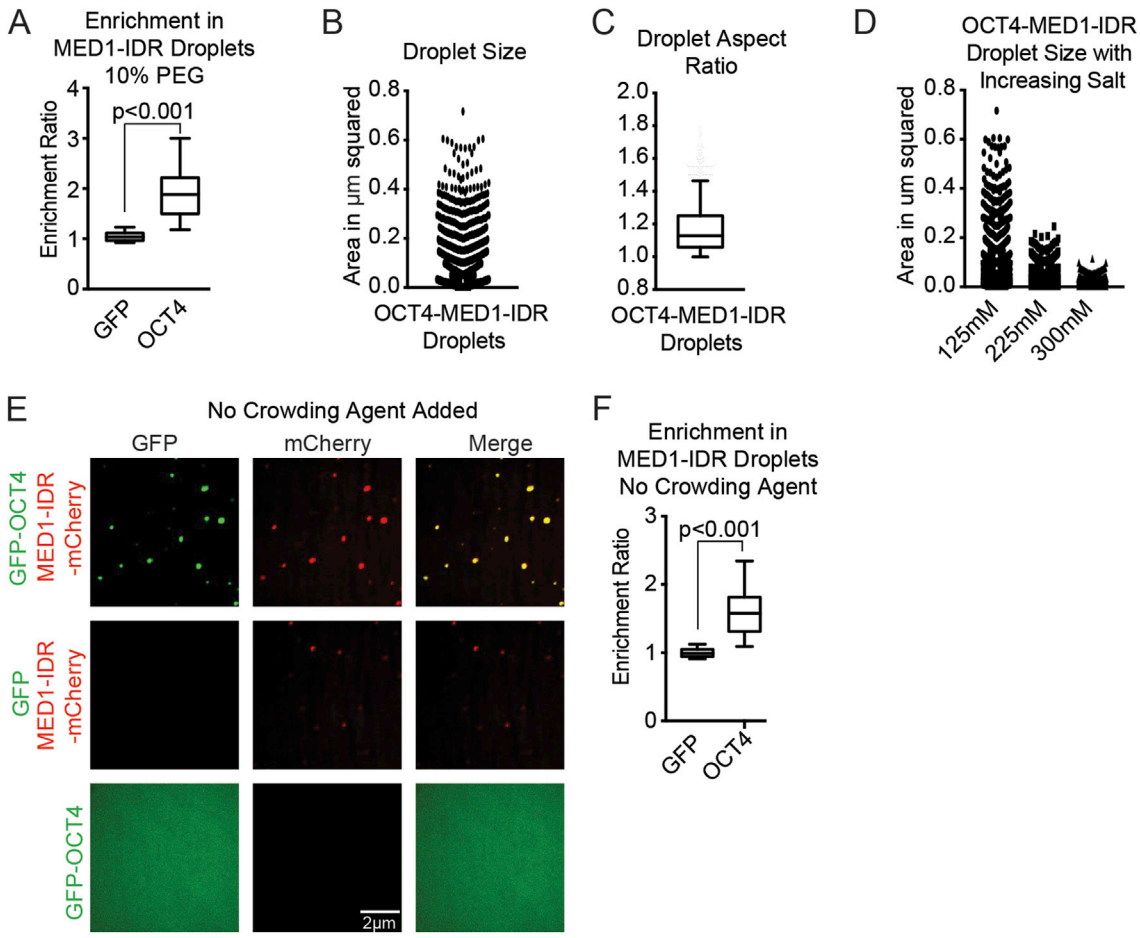


Figure S3. MED1 and OCT4 Droplet Formation, Related to Figure 3

A. Enrichment ratio of OCT4-GFP versus GFP in MED1-IDR-mCherry droplets formed in droplet formation buffer with 10% PEG-8000 at 125mM NaCl. n > 20, error bars represent the distribution between the 10th and 90th percentile.

B. Area in micrometers-squared of MED1-IDR-OCT4 droplets formed in 10% PEG-8000 at 125mM salt with 10uM of each protein.

C. Aspect ratio of MED1-IDR-OCT4 droplets formed in 10% PEG-8000 at 125mM with 10uM of each protein. n > 20, error bars represent the distribution between the 10th and 90th percentile.

D. Area in micrometers-squared of MED1-IDR-OCT4 droplets formed in 10% PEG-8000 at 125mM, 225uM, or 300uM salt, with 10uM of each protein.

E. Fluorescence microscopy of droplet formation without crowding agents at 50mM NaCl for the indicated protein or combination of proteins (at 10uM each), imaged in the channel indicated at the top of the panel.

F. Enrichment ratio of OCT4-GFP versus GFP in MED1-IDR-mCherry droplets formed in droplet formation buffer without crowding agent at 50mM NaCl. n > 20, error bars represent the distribution between the 10th and 90th percentile.

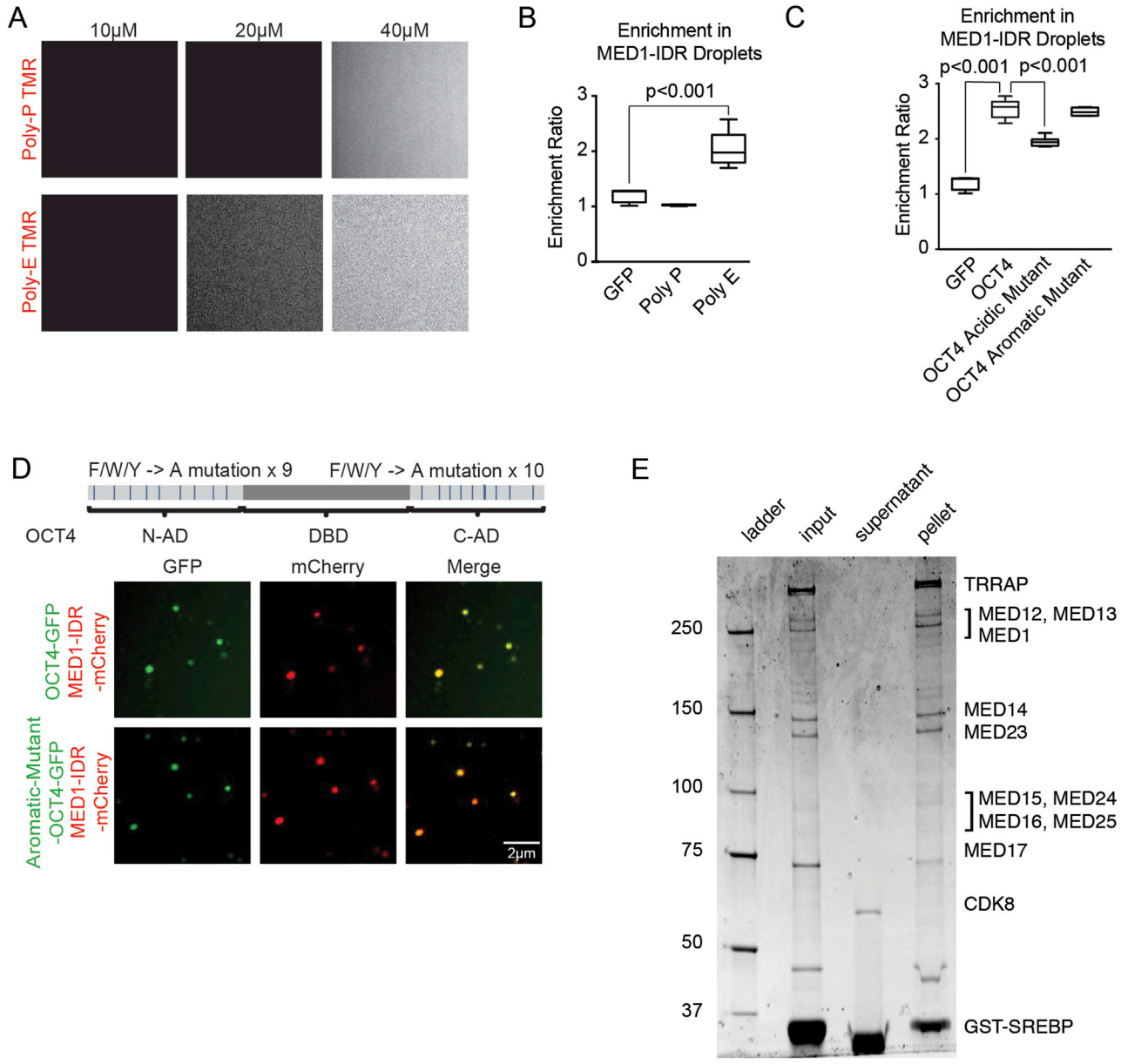


Figure S4. Phase Separation of Mutant OCT4, Related to Figure 4

A. Fluorescent microscopy of the indicated TMR-labeled polypeptide, at the indicated concentration in droplet formation buffers with 10% PEG-8000 and 125mM NaCl.

B. Enrichment ratios of the indicated polypeptide within MED1-IDR-mCherry droplets. n > 20, error bars represent the distribution between the 10th and 90th percentile.

C. Enrichment ratios of the indicated protein within MED1-IDR-mCherry droplets. n > 20, error bars represent the distribution between the 10th and 90th percentile.

D. (Upper panel) Schematic of OCT4 protein, aromatic residues in the activation domains (ADs) are marked by blue horizontal lines. All 9 aromatic residues in the N-terminal Activation Domain (N-AD) and 10 aromatic residues in the C-terminal Activation Domain (C-AD) were mutated to alanine to generate an OCT4-aromatic mutant. (Lower panel) Representative confocal images of droplet formation showing that the OCT4 aromatic mutant is still incorporated into MED1-IDR droplets. MED1-IDR-mCherry and OCT4-GFP or MED1-IDR-mCherry and OCT4-aromatic mutant-GFP were added to droplet formation buffers with 125mM NaCl at 10uM each with 10% PEG-8000 and visualized on a fluorescent microscope with the indicated filters.

E. Droplets of intact Mediator complex were collected by pelleting and equal volumes of input, supernatant, and pellet were run on an SDS-PAGE gel and stained with sypro ruby. Mediator subunits present in the pellet are annotated on the rightmost column.

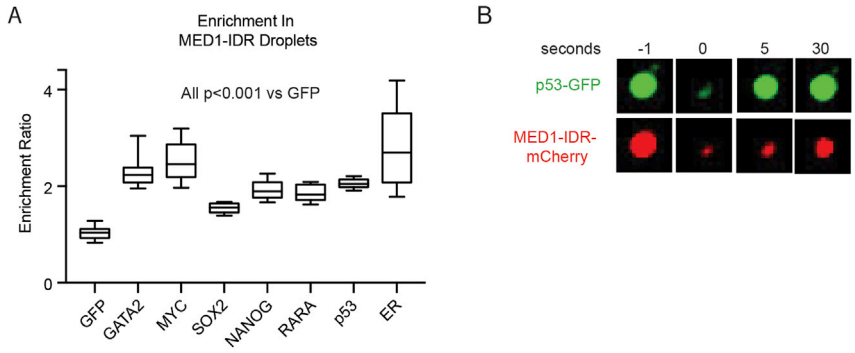


Figure S5. Diverse TFs Phase Separate with Mediator, Related to Figure 5

A. Enrichment ratios of the indicated GFP-fused TF in MED1-IDR-mCherry droplets. n > 20, error bars represent the distribution between the 10th and 90th percentile.

B. FRAP of heterotypic p53-GFP/MED1-IDR-mCherry droplets formed in droplet formation buffers with 10% PEG-8000 and 125mM NaCL, imaged every second over 30 s.

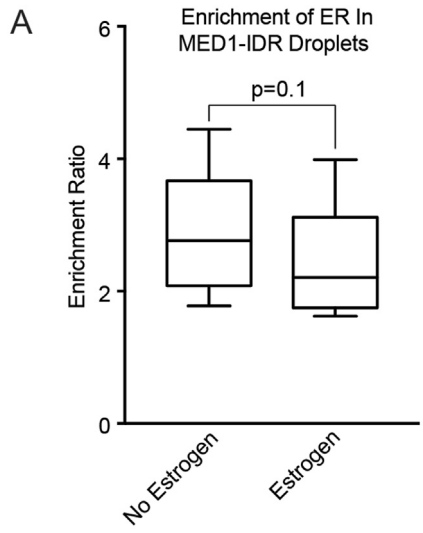


Figure S6. ER Phase Separates with MED1, Related to Figure 6

A. Enrichment ratio of ER-GFP in MED1-IDR-mCherry droplets in the presence or absence of 10uM estrogen. Droplets were formed in 10% PEG-8000 with 125mM NaCl. n > 20, error bars represent the distribution between the 10th and 90th percentile.

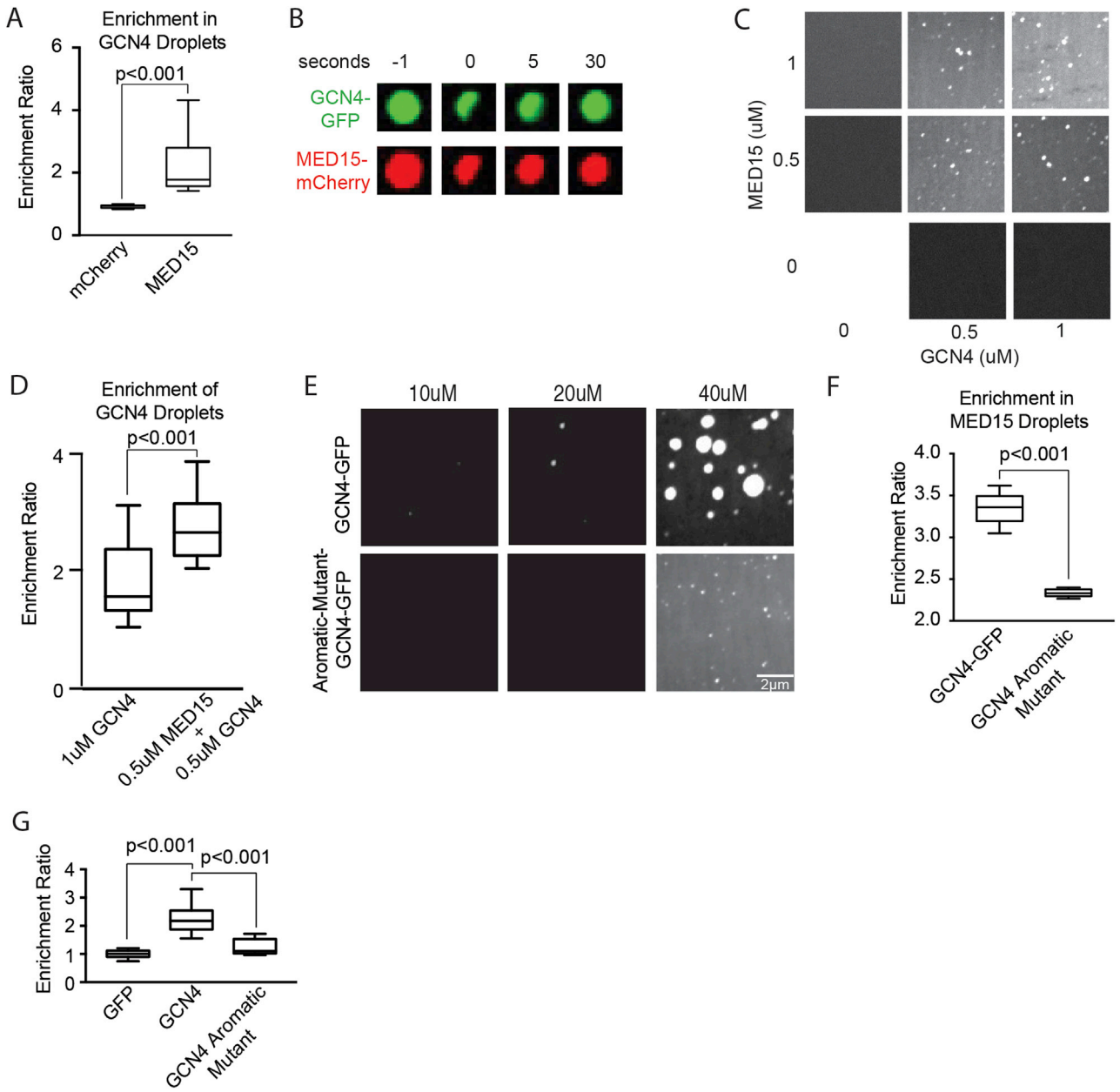


Figure S7. GCN4 and MED15 Form Phase-Separated Droplets, Related to Figure 7

A. Enrichment ratio of mCherry or MED15-mCherry in GCN4-GFP droplets, in droplet formation buffer with 10% PEG-8000 and 125mM NaCl. n > 20, error bars represent the distribution between the 10th and 90th percentile.

B. FRAP of heterotypic GCN4-GFP/MED15-IDR-mCherry droplets formed in droplet formation buffers with 10% PEG-8000 and 125mM NaCl, imaged every second over 30 s.

C. Phase diagram of GCN4-GFP and MED15-mCherry added at the indicated concentrations to droplet formation buffers with 10% PEG-8000 and 125mM salt.

D. Enrichment ratio of GCN4 droplets from S7C. n > 20, error bars represent the distribution between the 10th and 90th percentile.

E. Fluorescent imaging of GCN4-GFP or the aromatic mutant of GCN4-GFP at the indicated concentration in 10% PEG-8000 and 125mM NaCl. Shown are images from GFP channel.

F. Enrichment ratio of GCN4-GFP or the aromatic mutant of GCN4-GFP in MED15-mCherry droplets, formed in droplet formation buffer with 10% PEG-8000 and 125mM salt. n > 20, error bars represent the distribution between the 10th and 90th percentile.

G. Enrichment ratio of GFP, GCN4-GFP or GCN4-aromatic mutant-GFP in Mediator complex droplets. n > 20, error bars represent the distribution between the 10th and 90th percentile.