Genome Organization in Transcriptional Regulation

By

Charles H. Li

B.A., New York University (2012)

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 2021

Signature of Author:_____

Department of Biology
September 28th, 2020

Certified by:_____

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

Accepted by:_____

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

Genome Organization in Transcriptional Regulation

By

Charles H. Li

Submitted to the Department of Biology on September 28th, 2020 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the Massachusetts Institute of Technology

ABSTRACT

Transcriptional regulation of gene expression plays critical roles in the control of cell identity, development, and disease. Genome organization contributes to transcriptional regulation in multiple ways. At a fundamental level, the genome is organized into distinct active and repressive chromatin states that facilitate transcriptional regulation. These chromatin states are established and maintained at specific genomic regions via the interconnected activities of transcription factors and epigenetic pathways. An additional layer of genome organization is the three-dimensional structure of the genome within the nucleus. Transcriptional regulation occurs within a hierarchy of genome structures that are formed by the activities of structuring factors. Studies described in this thesis identify the transcription factor YY1 as a general structural regulator of enhancer-promoter loops (Weintraub et al., 2017). In recent years, the study of biomolecular condensates has led to a dramatic shift in our understanding of the mechanisms contributing to transcriptional regulation and to genome structure. Distinct chromatin condensates organize the genome by compartmentalizing components associated with transcriptionally active euchromatin and repressive heterochromatin. Whether disruption of chromatin condensates can lead to transcriptional dysregulation in human disease is not well understood. Our finding that MeCP2 is a key component of heterochromatin condensates and that Rett syndrome patient mutations affecting MeCP2 cause condensate disruption (Li et al., 2020), demonstrates a link between chromatin condensate disruption and human disease. These studies reveal important mechanisms of genome organization contributing to transcriptional regulation, and provide new insights into human disease that might be leveraged to provide therapeutic benefit for patients in the future.

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

ACKNOWLEDGEMENTS

I am grateful for having worked alongside the many talented members of the Young lab. Their collective guidance, support, and help have greatly shaped my development in the lab. A few lab members that I've worked closely with deserve special acknowledgement. I benefited from working closely with Abe Weintraub and Alicia Zamudio as a team on the YY1 project. Abe's positive attitude and willingness to embrace challenges were a driving force throughout the project. Alicia's critical eye in designing experiments and dedication to getting the job done were instrumental in successfully completing the project. I'd like to thank Eliot Coffey for working with me on the MeCP2 project. Eliot's enthusiasm for science and rigorous approach to experiments have made it a pleasure to work together. Denes Hnisz and Alla Sigova deserve thanks for being excellent mentors when I first started in the lab. Tony Lee has always been a great source of guidance. Many current and former lab members (Ann, Eric, John M., Ben, Susana, Jessie, Ozgur, Victoria, Ale, Lena, Ike, Jon H., Ido, Brian, Jurian, and others) have made the lab a fun and productive environment. Nancy Hannett and Katie Lazuk deserve immense thanks for keeping the lab running.

I am extremely grateful to Rick Young for taking me as a student in his lab. Rick's insistence on not only doing the best science, but also training effective leaders, has been a positive source of personal and professional growth in my life. I'm sure that the many lessons I've learned from Rick will continue to be a guiding force into the future.

I would like to thank my thesis committee, Phil Sharp, Rudolf Jaenisch, and Bob Kingston, for the time and effort they have dedicated towards guiding my growth as a scientist. Phil has always been a great source of insightful questions and comments that have helped guide my research. Rudolf has been an excellent mentor and collaborator on many projects. I have benefited greatly from Rudolf's excitement for discussing new ideas and his ability to bring together teams to do important science. Also deserving special thanks are the many members of the Jaenisch lab whom I've had the benefit and pleasure to learn from and work with throughout the years.

I'd also like to thank my previous scientific mentors, Dave Allis and Steven Josefowicz. My time as a research assistant in the Allis lab allowed me to explore and develop my scientific interests. Working closely with Steve taught me the fun of doing science as a team and helped inspire me to pursue a PhD.

None of this would have been possible without my family. My parents have always encouraged me to pursue science and provided ample opportunities for me to do so. From them, I have learned to persevere in the face of adversity, a skill that has been critical in grad school and elsewhere. I'm thankful for my brother Thomas who has always been supportive and a good source of advice outside of science.

Finally, I am thankful for my girlfriend, Limor. From meeting in college at NYU to moving to Boston for our PhDs, Limor has been a constant partner by my side. Her friendship and support have helped make challenges a little easier, and life a lot more fun.

STATEMENT ON WORK PRESENTED

Chapter 1
I wrote Chapter 1 with input and edits from Richard Young.

Chapter 2
I wrote the manuscript in collaboration with Abraham Weintraub and Richard Young, and with input from the other authors. I performed or contributed to the design of all experiments and analyses presented in the manuscript.

Chapter 3
I wrote the manuscript in collaboration with Eliot Coffey and Richard Young, and with input from the other authors. I performed or contributed to the design of all experiments and analyses presented in the manuscript.

Chapter 4
I wrote Chapter 4 with input and edits from Richard Young.

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

1. OVERVIEW

Transcriptional regulation of gene expression plays critical roles in the control of cell identity, development, and disease[1]. Genome organization contributes to transcriptional regulation in multiple ways. This thesis will examine the roles of genome organization in the transcriptional regulation of gene expression. In this introduction, I will discuss how organization of the genome into distinct chromatin states contributes to transcriptional regulation. I will then discuss the interplay between genome structure and transcription. Finally, I will discuss how recent advances in the study of biomolecular condensates have altered our understanding of genome organization and transcriptional regulation.

The genome is organized into distinct chromatin states that facilitate transcriptional regulation[2–4]. At the simplest level, the genome can be divided into two chromatin states: euchromatin and heterochromatin. Heterochromatin is highly compacted, gene-poor, and associated with transcriptional repression, while euchromatin is less compacted, gene-rich, and associated with actively transcribed genes[5–7]. Chromatin states are established and maintained at specific genomic loci via the interconnected activities of transcription factors and epigenetic pathways. Together these mechanisms contribute to transcriptional regulation by modulating accessibility of the transcriptional apparatus to the genome and recruitment of cofactors that promote either transcriptional activation or repression.

Advances in the study of genome organization have led to an understanding that genome structure is linked to transcriptional regulation. Genome structure refers to the three-dimensional organization of the genome within the nucleus. The genome is organized into a hierarchy of chromosome structures[8] – ranging from entire chromosomes that occupy distinct territories in the nucleus to individual enhancer-promoter loops. These genome structures contribute to the regulation of gene expression. Insulated neighborhoods – chromosome loops structured by the transcription factor CTCF and the cohesin complex – have emerged as a key structural and functional unit of gene control, with the vast majority of enhancer-promoter loops occurring within the boundaries of an insulated neighborhood[9–12]. Genome structure contributes to transcriptional regulation by constraining enhancers and promoters to interact within the hierarchy of genome structures. The observation that CTCF structures insulated neighborhood loops led us to investigate whether transcription factors can also structure enhancer-promoter loops, resulting in our discovery that YY1 is a structural regulator of enhancer-promoter loops[13] (Chapter 2). Notably, transcription itself can contribute to the formation of genome structures[14], indicating a more complex interplay between these two processes.

In recent years, the study of biomolecular condensates has led to a dramatic shift in our understanding of the mechanisms contributing to transcriptional regulation and to genome structure. Condensates are non-membrane bound compartments that organize cellular processes[15–18]. It is now recognized that transcriptional condensates located at transcribed genes compartmentalize components associated with transcriptionally active euchromatin[19–24]. Similarly, heterochromatin forms condensates that compartmentalize components involved in transcriptional repression[25–27]. In Chapter 3, I present our study examining the role of MeCP2 – a major factor involved in transcriptional repression – in organizing heterochromatin condensates[28]. MeCP2 forms heterochromatin condensates capable of selectively partitioning and excluding factors involved in transcriptional condensates[28], suggesting that condensate-mediated selective partitioning may contribute to the separation of distinct chromatin states in

cells. Importantly, we find that Rett syndrome patient mutations affecting MeCP2 cause condensate disruption, thus linking chromatin condensate dysregulation with human disease[28] (Chapter 3).

2. CHROMATIN STATES

Transcriptional regulation of gene expression plays important roles in cell identity, development, and disease[1]. The genome is organized into distinct chromatin states that facilitate transcriptional regulation. At a basic level, the genome is organized into two chromatin states – euchromatin and heterochromatin – which were initially described in 1928 by Emil Heitz[29]. Heterochromatin, which is highly compacted and gene-poor, is associated transcriptional repression; while euchromatin, which is less compacted and gene-rich, is associated with actively transcribed genes[5–7]. It is now understood that euchromatin and heterochromatin can be further sub-divided into more specific chromatin states that play distinct roles in genome organization and transcriptional regulation[2–4].

Chromatin states are established by the combined activities of transcription factors and epigenetic pathways. Transcription factors are bifunctional proteins that bind DNA regulatory elements at specific genomic loci and recruit additional factors involved in transcriptional regulation (Fig. 1a). As such, transcription factors are generally composed of two major domains: a DNA-binding domain and an effector domain. DNA binding domains (DBDs) recognize and bind to specific DNA sequence motifs found within DNA regulatory elements[30,31], thus facilitating the association of chromatin states with specific genomic loci. DBDs tend to adopt structured protein conformations that can be categorized into one of several evolutionarily conserved families[30,31]. This tendency to adopt structured conformations has facilitated the study of DBDs using crystallography and other approaches, which have led to an understanding of how different classes of DBDs bind and interact with DNA. Effector domains can facilitate either transcriptional activation or repression, and are thus categorized as either activation or repression domains. These domains generally do not adopt structured protein conformations, but rather display characteristics of intrinsically disordered regions, which has thus far limited understanding of their specific mechanisms of action[32–35]. However, effector domains generally function in the recruitment of cofactors that mediate their ability to regulate transcription[30,32,33]. Notably, effector domains can recruit cofactors involved in regulating epigenetic pathways[30,36], thus linking TFs and epigenetic pathways in the establishment of chromatin states.

Epigenetic pathways also contribute to chromatin states. These pathways are characterized by distinct chromatin modifications to histone proteins and DNA. Chromatin modifications contribute to chromatin states in two ways. First, chromatin modifications can affect the accessibility of chromatin to be bound by other factors by altering the physical structure of chromatin[37]. Second, chromatin modifications can facilitate recruitment of “reader” proteins that recognize and bind specific chromatin modifications[38] (Fig. 1b). Reader proteins function similarly to transcription factors in their ability to recruit cofactors that promote either transcriptional activation or repression. Among key cofactors recruited by readers are enzymes, termed “writers” and “erasers”, that are responsible for generating or removing specific chromatin modifications[38] (Fig. 1b). These chromatin-modifying enzymes facilitate maintenance of chromatin states by generating chromatin modifications associated with the chromatin state, while removing modifications associated with other chromatin states. Since epigenetic pathways can modulate the binding of transcription factors to DNA[39], the activities of TFs and epigenetic pathways function together to establish and maintain chromatin states.

Figure 1

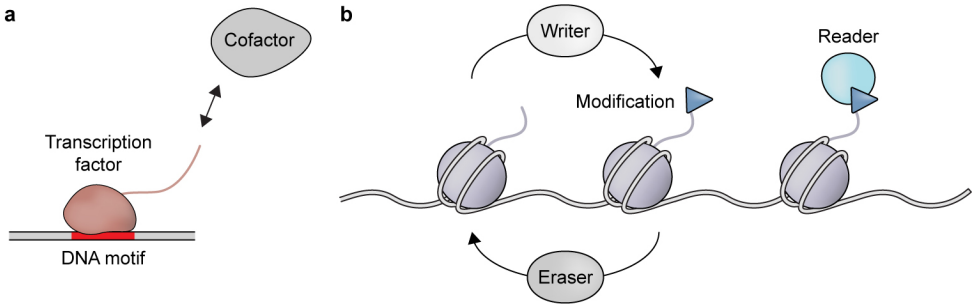


Figure 1. Transcription factors and epigenetic pathways
(a) Transcription factors typically have structured DNA-binding domains that bind specific DNA motifs and intrinsically disordered effector domains that recruit cofactors.
(b) Epigenetic pathways are characterized by chromatin modifications. Writer and eraser enzymes are responsible for generating and removing these modifications. Reader proteins bind specific modifications.

Transcriptionally active genes and their regulatory elements tend to be associated with active euchromatin states. Transcription factors mediate the establishment of active euchromatin states at gene promoters and enhancers – two classes of DNA regulatory elements involved in transcriptional control – via their ability to recognize and bind to specific DNA sequences found at these genomic loci. Activating transcription factors function to promote transcription by recruiting coactivators, such as the Mediator complex[40,41], that act to recruit RNA polymerase II to gene promoters and to stimulate transcriptional activity (Fig. 2). Notably, nascent RNA produced by transcription is capable of stabilizing TF binding to DNA[42], suggesting a positive feedback mechanism to maintain the active euchromatin state by reinforcing TF binding. Transcription factors also participate in the recruitment of cofactors capable of modifying chromatin, such as the histone acetyltransferase p300[43]. Interactions between TFs and chromatin-modifying enzymes link specific epigenetic pathways to the active euchromatin state.

Figure 2

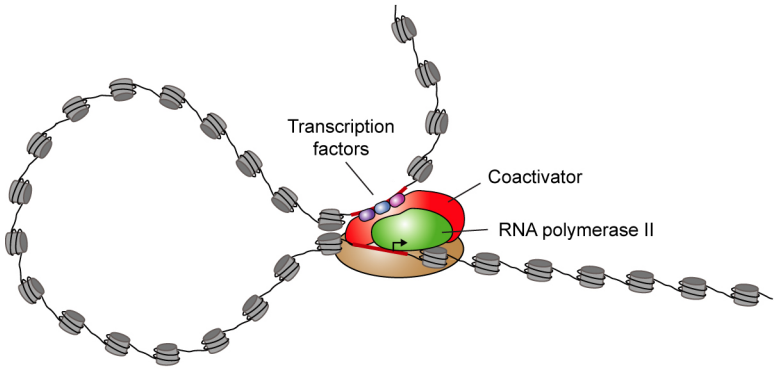


Figure 2. Transcription factors, coactivators, and RNA polymerase II in transcription
Transcription factors contribute to gene regulation by recruiting coactivators that recruit RNA polymerase II to genes and stimulate transcription. Figure adapted from Lee et al., Cell 2013.

Epigenetic pathways regulating the active euchromatin state are associated with two major types of histone modifications – acetylation and methylation – which occur on specific histone tail lysine residues[44–46]. Acetylated histones function to directly promote accessible "open" chromatin structure that is amenable to interaction of TFs and the transcriptional machinery with the underlying DNA[47,48]. Specific histone acetylation and methylation modifications associated with active euchromatin also function in recruitment of reader proteins involved in transcriptional regulation[49]. Some of these readers are coactivators that function directly in promoting transcription by RNA polymerase. For example, the coactivator BRD4 binds acetylated histones via its two bromodomains and can recruit pause-release factors that promote productive transcriptional elongation[50,51]. Coactivators recruited by reader proteins can also contribute to maintaining chromatin modifications associated with the active euchromatin state[52].

Heterochromatin can be subdivided into two distinct chromatin states – constitutive heterochromatin and facultative heterochromatin – both of which contribute to transcriptional repression. Constitutive heterochromatin occurs at genomic loci that are stably repressed across many cell types. Constitutive heterochromatin is typically associated with repression of repetitive DNA elements, such as satellite DNA and transposons[6,53]; however, it also functions in transcriptional repression of cell-type specific genes during differentiation[54,55]. In addition to its role in silencing of genes and repetitive elements, constitutive heterochromatin also plays important roles in ensuring genome integrity by structuring centromeric DNA and suppressing recombination between repetitive DNA elements[56,57]. Facultative heterochromatin is less stably repressed and is more typically associated with transcriptional repression of developmentally regulated genes. A defining feature of facultative heterochromatin is its ability to interconvert between transcriptionally repressed heterochromatin and transcriptionally active euchromatin states[58]. Context-dependent interconversion between these two states allows regulation of gene expression in response to changes in cell state[58]. Although both constitutive and facultative heterochromatin function in transcriptional repression, the two types of heterochromatin differ in the key pathways that contribute to their formation and function. I will focus the following discussion on constitutive heterochromatin, due to its relevance to the studies presented in Chapter 3.

Constitutive heterochromatin in mammals is controlled by two major epigenetic pathways that are characterized by distinct chromatin modifications – histone H3 lysine 9 trimethylation (H3K9me3) and DNA methylation[6,56,57,59] (Fig. 3). These modifications are specifically recognized and bound by chromatin reader proteins that mediate transcriptional repression. Most notably, HP1α is a reader of the H3K9me3 modification, while MeCP2 is a reader of DNA methylation[60–62]. Both HP1α and MeCP2 are general chromatin regulators that are broadly expressed in many tissues, implicated in global gene control, and essential for normal development[63–67]. HP1α and MeCP2 function in transcriptional repression by directly facilitating chromatin compaction at heterochromatin, thus causing physical exclusion of the transcriptional apparatus[68,69]. HP1α and MeCP2 also function via recruitment of corepressors associated with heterochromatin[70–73]. Among heterochromatin cofactors recruited by HP1α and MeCP2 are key chromatin-modifying enzymes that are responsible for the H3K9me3 and DNA methylation modifications associated with heterochromatin[74–76]. These interactions suggest a model whereby chromatin modifications, readers, and writer enzymes form a positive feedforward cycle promoting heterochromatin maintenance and spreading[57]. Crosstalk between these two major epigenetic pathways allows their joint regulation of heterochromatin[77–79].

Figure 3

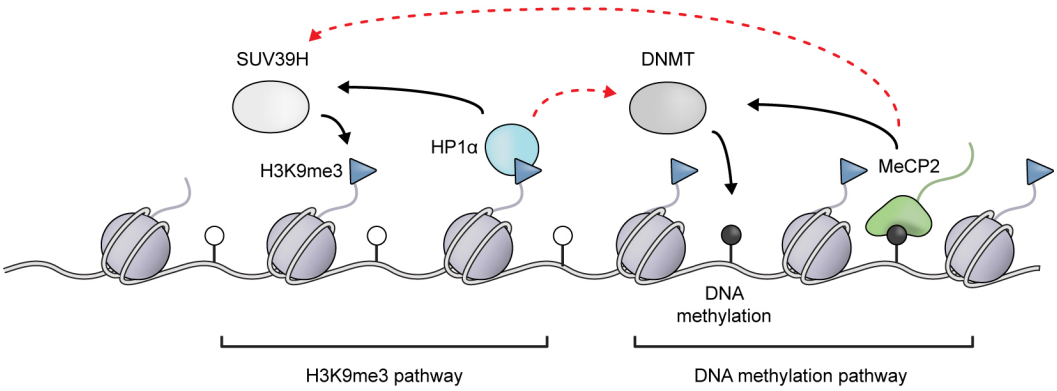


Figure 3. Epigenetic pathways controlling constitutive heterochromatin
Constitutive heterochromatin is controlled by two major epigenetic pathways characterized by the chromatin modifications H3K9me3 and DNA methylation. Reader-writer interactions help maintain each epigenetic pathway (black arrows). Crosstalk between the pathways facilitates joint regulation (red arrows).

While the epigenetic pathways that control constitutive heterochromatin facilitate heterochromatin maintenance and function, the processes that specify establishment of heterochromatin at specific genomic loci in mammalian cells are less understood. There is some evidence that sequence-specific TFs contribute to heterochromatin function[80], suggesting that TFs may play a role in establishment of heterochromatin at specific genomic loci. Additionally, transcription occurs to some extent at heterochromatin loci producing non-coding RNAs that contribute to heterochromatin formation[81]. In multiple organisms, these RNAs contribute to heterochromatin formation by participating in RNA-interference (RNAi) pathways[82], although the relevance of RNAi machinery in establishment of mammalian heterochromatin is less clear. Nevertheless, RNA is a critical component of mammalian constitutive heterochromatin[83] and heterochromatin transcription is essential for heterochromatin establishment during development[84,85], suggesting that production of these repetitive RNAs contributes to heterochromatin formation at specific genomic loci.

3. GENOME STRUCTURE

Advances in the study of genome organization have led to an appreciation of the interplay between genome structure and transcriptional regulation. Genome structure refers to the three-dimensional organization of the genome within the nucleus. The human genome is about two meters in length and must be packaged within a cell nucleus of roughly ten microns in diameter. This packaging occurs via organization of the genome into a hierarchy of chromosome structures. Studies over the last decade have revealed that these genome structures play important roles transcriptional regulation. However, evidence also indicates that transcription itself may be contributing to the formation and maintenance of genome structures[14], suggesting a more complex interplay between these two processes.

The genome is organized within the nucleus in a hierarchy of chromosome structures[8] (Fig. 4). At the highest level, individual chromosomes occupy distinct regions of the nucleus, termed chromosome territories[86]. Within chromosomes, genomic regions can be assigned to one of two compartments, wherein regions associated with the same compartment interact with each other, but not with genomic regions associated with the other compartment[87]. The two compartments are associated with features of active (compartment A) and repressive (compartment B) chromatin states. These compartments are composed of megabase-sized domains, termed toplogically associated domains (TADs)[88–90]. TADs display increased interaction frequency between regions inside the TAD and reduced interactions that cross TAD boundaries. TAD boundaries are commonly occupied by CTCF and cohesin, two major regulators of genome structure. Within TADs are insulated neighborhoods, which are chromosome loops formed by the interaction of two CTCF-bound insulator DNA elements and reinforced by the cohesin complex[9,11,12]. Insulated neighborhoods are a key structural and functional unit of gene control that contain genes and their associated enhancers, which can physically and functionally interact via the formation of enhancer-promoter DNA loops[11].

Figure 4

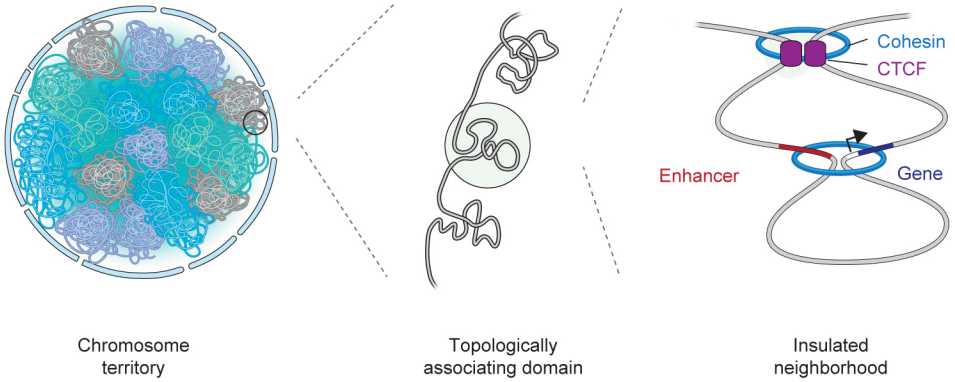


Figure 4. The genome is organized in a hierarchy of chromosome structures
The hierarchy of genome structure involves chromosome territories, toplogically associating domains (TADs), insulated neighborhoods, and enhancer-promoter loops. Figure adapted from Hnisz et al., Annu. Rev. Cancer Biol. 2018.

Structuring factors function in the establishment and maintenance of genome structures. CTCF and cohesin have emerged as two major regulators of genome structure in mammalian cells[91–95]. CTCF is a zinc-finger transcription factor that participates in transcriptional regulation by structuring chromosome loops[91–93]. CTCF is the primary DNA-binding factor that occupies insulators – a key class of DNA regulatory element involved in transcriptional regulation[96–100]. Insulators function to block enhancers from acting on gene promoters when they are positioned between enhancer and promoter elements[101,102]. CTCF contributes to insulator function by structuring chromosome loops, termed insulated neighborhoods, via the dimerization of two CTCF molecules each bound at an insulator element[9–12]. Insulated neighborhoods contain genes and their associated DNA regulatory elements, and contribute to transcriptional regulation by constraining the activities of these regulatory elements. The cohesin complex is another factor contributing to genome structure and transcriptional regulation[93–95]. Cohesin is a ring-shaped protein complex[103,104] capable of topological entrapment of DNA[105]. Cohesin co-

occupies insulator elements bound by CTCF[106–108], where it contributes to genome structure by reinforcing CTCF-mediated insulated neighborhood chromosome loops[9,11,109]. Cohesin also contributes to transcriptional regulation by structuring enhancer-promoter loops[110]. Although CTCF and cohesin are the two best characterized structuring factors, additional factors have also been proposed to participate in genome structure[111–113]. Notably, in Chapter 2, I present our finding that the zinc-finger transcription factor YY1 is a key structuring factor for enhancer-promoter loops, where it functions in a manner analogous to the role of CTCF in structuring insulated neighborhoods[13].

Genome structure contributes to transcriptional regulation by constraining the activity of enhancers to act on their intended target genes. The ability of enhancers to regulate genes over great genomic distances via formation of enhancer-promoter loops, raises the question of how specific regulation is achieved. There are many active gene promoters and enhancers in any given cell type[4,114,115], suggesting an incredible potential for unintended enhancer-promoter looping resulting in transcriptional dysregulation. Several lines of evidence suggest that genome structure contributes to transcriptional regulation by constraining enhancers and promoters to specifically interact within the hierarchy of chromosome structures. First, the vast majority of enhancer-promoter interactions occur within insulated neighborhood structures[11]. Second, experimental perturbation of both insulated neighborhood boundaries and structuring factors results in loss of chromosome loop structures and transcriptional dysregulation[9,12,116–120]. Finally, disruption of genome structure is observed in several human diseases, where it causes altered enhancer-promoter looping and aberrant gene expression[10,121–124]. Thus, genome structure contributes to transcriptional regulation by helping to resolve the problem of enhancer-promoter looping specificity. In Chapter 4-1, I discuss reasons why this is only a partial solution and potential additional mechanisms that could be contributing.

Genome structure may also contribute to transcriptional regulation via the separation of distinct chromatin states. Genome structures tend to be associated with either active or repressive chromatin states. Insulated neighborhoods can contain super-enhancers associated with transcriptionally active euchromatin or Polycomb proteins associated with facultative heterochromatin[9]. Similarly, TADs tend to be associated with either transcriptionally active or repressive chromatin states[88]. The sorting of TADs into A/B compartments, wherein TADs associated with different chromatin states disfavor interaction with each other[87], suggests that genome structures could act in separating distinct chromatin states. However, it remains unclear whether genome structure plays a causal role in this separation, as depletion of key structuring factors CTCF and cohesin does not appear to disrupt separation of genomic loci associated with active and repressive chromatin states[118,119]. It is possible that other regulators of genome structure are responsible for this activity. However, separation could also be a product of the underlying chromatin states themselves, which I discuss in Chapter 4-4.

Notably, evidence suggests that transcription also plays a critical role in the establishment and maintenance of genome structure[14]. TAD boundaries are not only occupied by CTCF, but are also enriched for transcribed housekeeping genes, transfer RNAs, and retrotransposons[88,125], suggesting that transcription from these loci could contribute to TAD structure. Interestingly, in prokaryotes, transcription of highly expressed genes located at the boundaries of chromosome structures is required for the establishment and maintenance of chromosome structures[126–128], suggesting that transcription could contribute to a similar process in mammalian cells. Furthermore, recent studies have suggested that chromosome loops are dynamic and are formed via an active loop extrusion process[129–131]. In this process, cohesin is loaded on to the genome where it initially forms a small chromosome loop, and sliding of cohesin along the chromosome results in extrusion of a progressively larger loop, until cohesin encounters CTCF-

bound insulator sites, which block further extrusion. An extrusion model is consistent with the tendency for the cohesin loader NIBPL to be enriched at active gene promoters and enhancers[110], but not at insulated neighborhood boundaries. Notably, RNA polymerase can act to push cohesin along the genome during transcription[132,133], suggesting that transcription could contribute to genome structure by extrusion of cohesin from sites of loading to insulated neighborhood boundaries. Together these observations indicate a complex interplay between genome structure and transcriptional regulation that remains to be fully understood.

4. CHROMATIN CONDENSATES

In the last few years, advances in the study of biomolecular condensates have led to a paradigm shift in our understanding of genome organization and transcriptional regulation. Biomolecular condensates are non-membrane bound compartments that organize cellular processes[15–17] (Fig. 5). While a few of these compartments, such as the nucleolus, have long been studied, it has only recently become appreciated that many key functional processes in the cell are organized by condensates. In particular, condensates are now thought to contribute substantially to genome organization and transcriptional regulation, with components associated with transcriptionally active and repressive chromatin states each being compartmentalized within chromatin condensates.

Figure 5

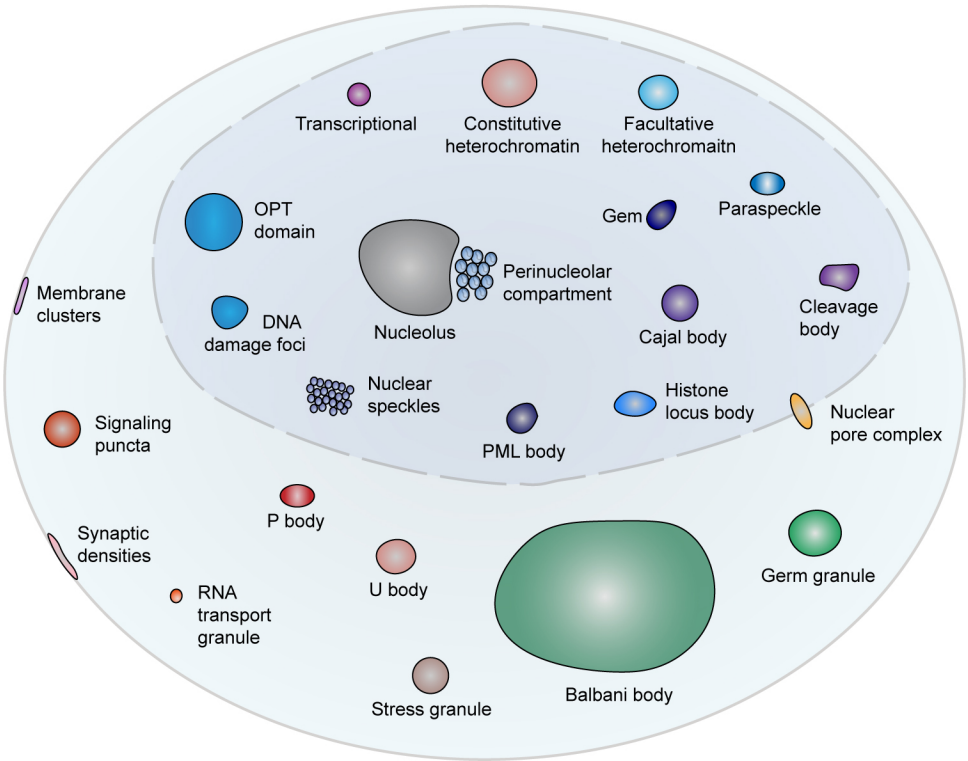


Figure 5. Biomolecular condensates in the cell
Many key cellular processes are compartmentalized in biomolecular condensates. Figure adapted from Banani et al., Nat. Rev. Mol. Cell Biol. 2017.

Condensates share several features that contribute to their roles in genome organization and transcriptional regulation. First, condensates compartmentalize cellular processes[16–18,134]. Compartmentalization allows biomolecules involved in a specific process to be separated from the many other processes that occur in the nucleus. This enhances the ability of molecules to find and interact with their partner molecules engaged in the same functional process, while reducing unintended interference between distinct functional processes. Second, condensates selectively partition biomolecules[16,18,134]. The physicochemical properties of each condensate allow different biomolecules to prefer being inside the condensate, outside the condensate, or neutral. This property is critical for functional compartmentalization, and modulating selective partitioning behavior can facilitate condensate regulation. Third, condensates concentrate biomolecules[16–18,134]. Concentration of biomolecules involved in the same process within a condensate is mediated by selective partitioning. Increased concentration can facilitate biochemical interactions and enzymatic reactions involved in functional processes.

Principles from soft matter physics provide a conceptual framework for understanding biomolecular condensates. Condensate formation is thought to occur via liquid-liquid phase separation (LLPS), a physical process by which molecules in a solution undergo a sharp transition from a single phase to a two-phase system, consisting of a condensed phase, in which the molecules are concentrated, and a dilute phase[15–18]. Transition to a two-phase system is favorable when interaction energies between molecules in the system exceed the entropy cost of de-mixing[15,16,18]. This can occur when interaction energies favor interactions between like molecules, while disfavoring interactions between unlike molecules[15,16,18]. Interaction energies are determined by the interaction valency – the number of adhesive interactions that occur between like molecules – and the concentration of components in the system, which together govern the critical point at which phase separation occurs in a switch-like fashion[15,16,18] (Fig. 6). As expected from LLPS theory, biomolecular condensate formation is determined by interaction valency and concentration of components. All major biomolecules in the nucleus (DNA, RNA, and proteins) can engage in multivalent interactions that drive condensate formation. Condensate-forming biomolecules, when present at high enough concentration, engage in multivalent interactions that form dynamic interaction networks leading to condensate formation.

Biomolecular condensate formation often involves two types of molecular interactions. Molecular interactions that are relatively low affinity and more dynamic are a notable feature of condensate-forming proteins[19,21,135–139]. These interactions are frequently associated with intrinsically disordered regions (IDRs) of proteins and depend on greater multivalency to promote condensate formation. However, proteins that contain stably structured interaction domains, which engage in conventional interactions with relatively higher affinity, are also found within condensates. These structured domains often engage in well-studied interactions that are thought to function via canonical mechanisms. However, structured interaction domains can also contribute to condensate formation[140–142]. Importantly, condensates models incorporate these conventional mechanisms of interaction – rather than being mutually exclusive with conventional mechanisms – while adding the view that large numbers of molecules, using multiple weak and dynamic interactions, concentrate and compartmentalize diverse components engaged in the same functional processes. Notably, structured and disordered domains are often present together in proteins that contribute to chromatin condensate formation[19,21,22,25,28].

Figure 6

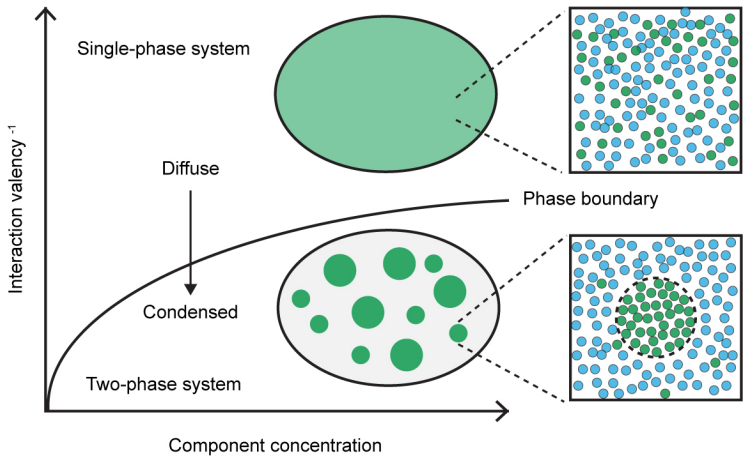


Figure 6. Phase separation is driven by valency and concentration
Phase separation is the physical process by which molecules in a solution transition from a single phase to a two-phase system. The point at which phase separation occurs is determined by interaction valency and component concentration. Figure adapted from Brangwynne et al., J. Cell Biol. 2013.

Several additional concepts have emerged that are useful for understanding the formation and regulation of biomolecular condensates. First, protein IDRs can contain multiple amino acid patches, termed “stickers”, that engage in intermolecular interactions contributing to multivalency, and these stickers are separated by flexible “spacer” regions that facilitate their interaction activities[143–145]. Second, condensate formation is driven by “scaffold” components, which are necessary for condensate formation and maintenance[146]. Once established, condensates can selectively partition and concentrate “client” components that contribute to condensate function but are not necessary for condensate formation or maintenance[146]. Third, regulation of condensates can occur by modulating valency. Post-translational modifications can modulate interaction valency by adding or removing interaction sites, which can result in altered condensate formation and composition[23,24,137,140,147,148].

Recent studies have demonstrated that transcriptional condensates compartmentalize components associated with transcriptionally active euchromatin (Fig. 7). RNA polymerase II (RNA Pol II), transcription factors (TFs), and coactivators are compartmentalized in transcriptional condensates associated with transcribed genes[19–24]. Transcriptional condensates tend to occur at genes controlled by super-enhancers[19] – clusters of enhancers occupied by disproportionate amounts of the transcriptional apparatus[149,150]. Localization of transcriptional condensates at super-enhancers likely results from the combined activities of sequence-specific TFs and epigenetic pathways associated with transcriptionally active euchromatin. TF binding to specific DNA sequences at super-enhancers can function to crowd TFs to the point of condensate formation[151], thus linking transcriptional condensates to these genomic loci. TF activation domains – many of which are intrinsically disordered regions (IDRs) – engage in multivalent interactions with coactivators to form transcriptional condensates[21,22,151]. Transcriptional condensates can be reinforced by interactions between coactivators and the histone modifications that are commonly associated with the active euchromatin state[152]. Transcriptional condensates selectively partition and concentrate RNA Pol II, in part via multivalent interactions with the intrinsically disordered C-terminal domain (CTD) of RNA Pol

II[23]. Regulation of RNA Pol II partitioning into transcriptional condensates occurs via CTD phosphorylation, which results in transition of RNA Pol II from transcriptional condensates to splicing factor condensates, which function during transcriptional elongation[23]. Thus, transcriptional condensates compartmentalize key components involved in transcriptional activation and provide an additional layer of regulation for the control of gene expression.

Figure 7

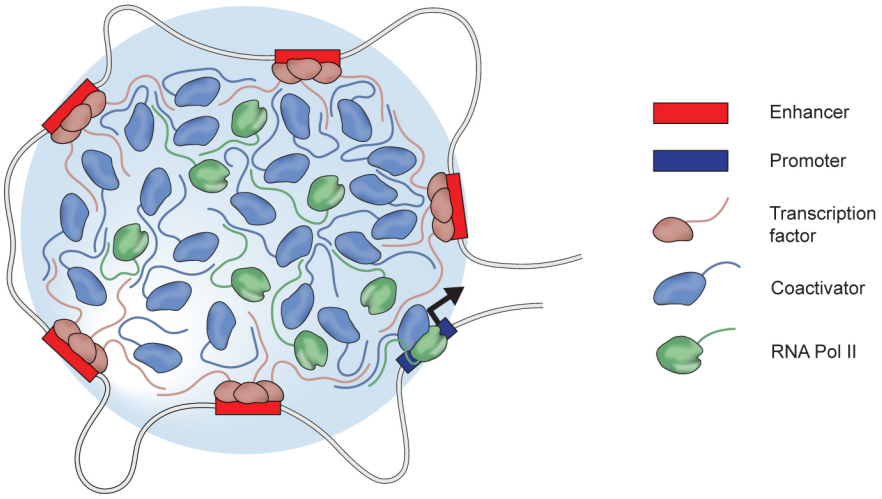


Figure 7. Model of a transcriptional condensate
Transcriptional condensates compartmentalize key components associated with transcriptionally active euchromatin. Figure adapted from Boija et al., Cell 2018.

Compartmentalization of the transcription apparatus in condensates has been proposed to resolve several mysteries related to transcriptional activation[153]. First, transcriptional condensates can explain the disproportionate enrichment of the transcription apparatus at super-enhancers[149,150], as condensates can highly concentrate component molecules. Second, condensate formation provides an explanation for the ability of a small mutation to nucleate formation of a super-enhancer at an oncogene in cancer[154], as a small increase in valency – resulting from mutations that create a TF binding site – can trigger condensate formation if the system is poised on the threshold of condensation. Third, transcriptional condensates can explain preferential sensitivity of super-enhancers to transcriptional inhibitors[155], as condensate dissolution can occur with sharp switch-like transitions. Fourth, condensates provide an explanation for how TF activation domains function. Activation domains are intrinsically disordered and engage in multivalent interactions with diverse coactivators and other components of the transcriptional apparatus to promote condensate formation[21]. Finally, transcriptional condensates can explain observations of coordinate transcriptional activation of two genes by a single enhancer[156], as compartmentalization of both genes and the enhancer in a single transcriptional condensate could allow simultaneous regulation of both genes.

Recent studies have also demonstrated that transcriptionally repressed chromatin states are similarly compartmentalized in condensates. Key components of both constitutive heterochromatin and facultative heterochromatin are capable of condensate formation[25–27], although I will focus here on constitutive heterochromatin condensates (Fig. 8). Heterochromatin condensates appear to be scaffolded by components of the two major epigenetic pathways controlling heterochromatin. HP1α, a key component of one of these pathways, is thought to be a critical scaffolding factor for heterochromatin condensates. HP1α binds to H3K9me3-modified chromatin[60,61], and chromatin bearing this modification can stimulate HP1α condensate formation[157], suggesting that heterochromatin condensates form at specific chromatin loci marked by H3K9me3. Heterochromatin condensates selectively partition and concentrate other heterochromatin components[25,157], including the histone methyltransferase SUV39H1 responsible for generating the heterochromatin-associated H3K9me3 modification[158]. These observations suggest that heterochromatin condensate formation and maintenance is facilitated by a self-reinforcing loop between H3K9me3, HP1α, and SUV39H1, which is consistent with conventional models of heterochromatin formation. In Chapter 3, I present our study demonstrating that MeCP2, a key component of the second major epigenetic pathway controlling heterochromatin, also participates in scaffolding heterochromatin condensates via its interaction with methylated DNA[28].

Figure 8

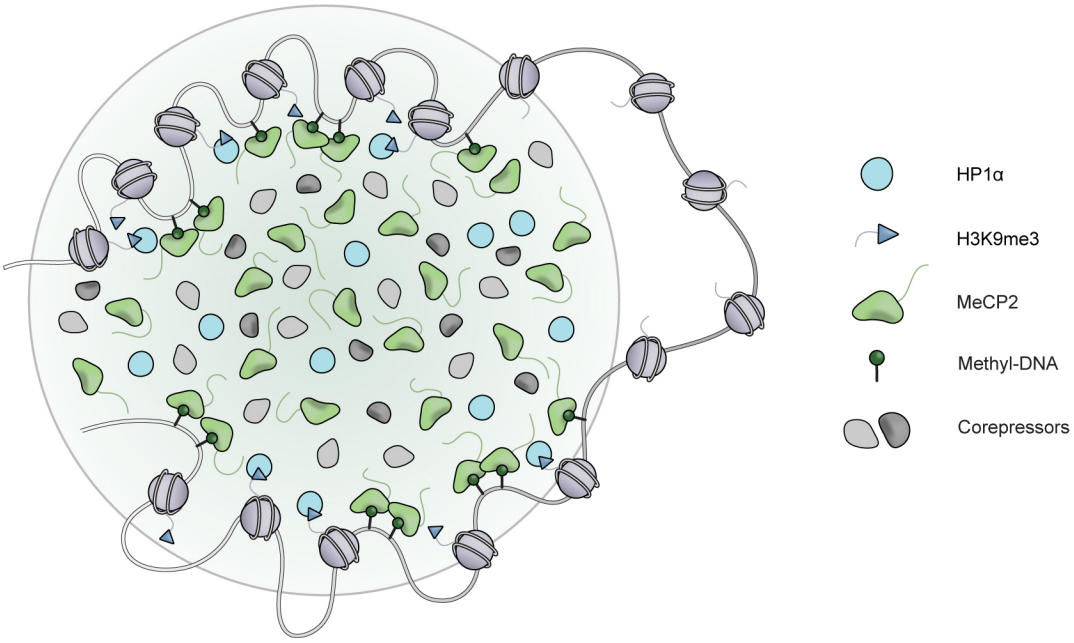


Figure 8. Model of a heterochromatin condensate
Heterochromatin condensates are scaffolded by components of the two major epigenetic pathways controlling mammalian constitutive heterochromatin.

Despite knowledge that both epigenetic pathways involving HP1α and MeCP2 contribute to heterochromatin condensates, it remains to be determined how heterochromatin condensates are initially localized to specific genomic regions where they occur. Mechanisms involving sequence-specific TF binding[80] or local RNA production[6] at heterochromatin-associated genomic loci could potentially contribute to the establishment of heterochromatin condensates at specific genomic regions. Notably, a recent study implicates heterochromatin-associated major satellite repeat RNAs in promoting heterochromatin phase separation[159], providing some support for an RNA-mediated mechanism for establishing heterochromatin condensates.

The selective partitioning properties of constitutive heterochromatin condensates likely contribute to transcriptional repression. Heterochromatin condensates can selectively concentrate key corepressors involved in transcriptional repression. MeCP2 condensates are capable concentrating the NCoR corepressor complex[28], a key MeCP2 cofactor that mediates transcriptional repression via establishment of repressive chromatin states by histone deacetylation[71]. Heterochromatin condensates can also selectively partition and exclude components involved in transcriptional activation. HP1α facilitates exclusion of the general transcription factor TFIIB from heterochromatin condensates[157], while we observe that MeCP2 is capable of excluding key components of transcriptional condensates such as BRD4[28] (Chapter 3). Exclusion of the transcriptional apparatus by selective condensate partitioning provides a new model for heterochromatin repression that differs from the conventional model of physical inaccessibility resulting from densely compacted and static heterochromatin. Importantly, a condensate model for heterochromatin does not necessarily exclude the conventional model, but rather adds an additional layer of regulation that could also contribute to heterochromatin repression. In Chapter 4-5, I further discuss how immiscibility between distinct chromatin condensates resulting from selective condensate partitioning properties could contribute to genome organization and separation of transcriptionally active and repressive chromatin states.

An important question that emerges from new knowledge of genome organization by chromatin condensates is whether dysregulation of chromatin condensates can contribute to human disease. Studies examining condensate dysregulation in disease have thus far focused on a few cytoplasmic condensate-forming proteins implicated in neurodegeneration[160]. In Chapter 3, I present our study demonstrating that Rett syndrome-causing patient mutations affecting MeCP2 cause heterochromatin condensate disruption, thus linking dysregulation of chromatin condensates with human disease[28]. Importantly, we find that mutations affecting both the structured and intrinsically disordered domains of MeCP2 can contribute to heterochromatin condensate disruption[28], indicating that condensate dysregulation can be a common consequence of mutations in multiple protein domains. In Chapter 4-6, I further discuss the role of condensates in human disease and therapeutics.

5. REFERENCES

1. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
2. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
3. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
4. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
5. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nat. Rev. Genet.* **8**, 35–46 (2007).
6. Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics and Chromatin* **8**, 1–17 (2015).
7. Wang, J., Jia, S. T. & Jia, S. New insights into the regulation of heterochromatin. *Trends Genet.* **32**, 284–294 (2016).
8. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
9. Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
10. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 3812–3814 (2016).
11. Hnisz, D., Day, D. S. & Young, R. A. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* **167**, 1188–1200 (2016).
12. Ji, X. *et al.* 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
13. Weintraub, A. S. *et al.* YY1 is a structural regulator of enhancer-promoter loops. *Cell* **172**, 1573–1588 (2017).
14. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
15. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
16. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
17. Boeynaems, S. *et al.* Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
18. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
19. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
20. Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
21. Boija, A. *et al.* Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 (2018).
22. Zamudio, A. V. *et al.* Mediator condensates localize signaling factors to key cell identity genes. *Mol. Cell* **76**, 753–766 (2019).
23. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
24. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
25. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
26. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
27. Plys, A. J. *et al.* Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813 (2019).
28. Li, C. H. *et al.* MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* (2020).

29. Heitz, E. Das heterochromatin der moose. *I. Jahrb. Wiss. Bot.* **69**, 762–818 (1928).
30. Weirauch, M. T. & Hughes, T. R. *A handbook of transcription factors*. *Subcellular Biochemistry* **52**, (2011).
31. Lambert, S. A. *et al.* The human transcription factors. *Cell* **172**, 650–665 (2018).
32. Triezenberg, S. J. Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.* **5**, 190–196 (1995).
33. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
34. Minezaki, Y., Homma, K., Kinjo, A. R. & Nishikawa, K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* **359**, 1137–1149 (2006).
35. Dyson, H. J. & Wright, P. E. Role of intrinsic protein disorder in the function and interactions of the transcriptional coactivators CREB-binding protein (CBP) and p300. *J. Biol. Chem.* **291**, 6714–6722 (2016).
36. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* **43**, 73–81 (2017).
37. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* **20**, 259–266 (2013).
38. Nicholson, T. B., Veland, N. & Chen, T. *Writers, readers, and erasers of epigenetic marks*. *Epigenetic Cancer Therapy* (Elsevier Inc., 2015).
39. Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, 551–565 (2016).
40. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
41. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772 (2010).
42. Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–982 (2015).
43. Shikama, N., Ivon, J. & La Thangue, N. The p300/CBP family: integrating signals with transcription factors and chromatin. *Trends Cell Biol.* **7**, 230–236 (1997).
44. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
45. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–83 (2011).
46. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
47. Annunziato, A. T., Frado, L. L. Y., Seale, R. L. & Woodcock, C. L. F. Treatment with sodium butyrate inhibits the complete condensation of interphase chromatin. *Chromosoma* **96**, 132–138 (1988).
48. Tse, C., Sera, T., Wolffe, A. P. & Hansen, J. C. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA Polymerase III. *Mol. Cell. Biol.* **18**, 4629–4638 (1998).
49. Musselman, C. A., Lalonde, M. E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **19**, 1218–1227 (2012).
50. Itzen, F., Greifenberg, A. K., Bösken, C. A. & Geyer, M. Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. *Nucleic Acids Res.* **42**, 7577–7590 (2014).
51. Jang, M. K. *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).
52. Wu, T., Kamikawa, Y. F. & Donohoe, M. E. Brd4's bromodomains mediate histone H3 acetylation and chromatin remodeling in pluripotent cells through P300 and Brg1. *Cell Rep.* **25**, 1756–1771 (2018).
53. Bulut-Karslioglu, A. *et al.* Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 (2014).
54. Becker, J. S. *et al.* Genomic and proteomic resolution of heterochromatin and its restriction of alternate fate genes. *Mol. Cell* **68**, 1023–1037 (2017).
55. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).

56. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: guardian of the genome. *Annu. Rev. Cell Dev. Biol.* **34**, 265–288 (2018).
57. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
58. Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* **28**, 1–13 (2007).
59. Nishibuchi, G. & Déjardin, J. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. *Chromosom. Res.* **25**, 77–87 (2017).
60. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
61. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
62. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).
63. Eissenberg, J. C. & Elgin, S. C. R. HP1a: a structural chromosomal protein regulating transcription. *Trends Genet.* **30**, 103–110 (2014).
64. Ip, J. P. K., Mellios, N. & Sur, M. Rett syndrome: insights into genetic, molecular and circuit mechanisms. *Nat. Rev. Neurosci.* **19**, 368–382 (2018).
65. Ausió, J., de Paz, A. M. artíne. & Esteller, M. MeCP2: the long trip from a chromatin protein to neurological disorders. *Trends Mol. Med.* **20**, 487–498 (2014).
66. Lyst, M. J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* **16**, 261–274 (2015).
67. Guy, J., Cheval, H., Selfridge, J. & Bird, A. The role of MeCP2 in the brain. *Annu. Rev. Cell Dev. Biol.* **27**, 631–652 (2011).
68. Georgel, P. T. *et al.* Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.* **278**, 32181–32188 (2003).
69. Azzaz, A. M. *et al.* Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation. *J. Biol. Chem.* **289**, 6850–6861 (2014).
70. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
71. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013).
72. Kruusvee, V. *et al.* Structure of the MeCP2–TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. *Proc. Natl. Acad. Sci.* **114**, E3243–E3250 (2017).
73. Smothers, J. F. & Henikoff, S. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr. Biol.* **10**, 27–30 (2000).
74. Aagaard, L. *et al.* Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J.* **18**, 1923–1938 (1999).
75. Kimura, H. & Shiota, K. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.* **278**, 4806–4812 (2003).
76. Rajavelu, A. *et al.* Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. *Nucleic Acids Res.* **46**, 9044–9056 (2018).
77. Lehnertz, B. *et al.* Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
78. Hashimshony, T., Zhang, J., Keshet, I., Bustin, M. & Cedar, H. The role of DNA methylation in setting up chromatin structure during development. *Nat. Genet.* **34**, 187–192 (2003).
79. Fuks, F. *et al.* The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* **278**, 4035–4040 (2003).
80. Bulut-Karslioglu, A. *et al.* A transcription factor-based mechanism for mouse heterochromatin formation. *Nat. Struct. Mol. Biol.* **19**, 1023–1032 (2012).
81. Grewal, S. I. S. & Elgin, S. C. R. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406 (2007).
82. Martienssen, R. A., Kloc, A., Slotkin, R. K. & Tanurdžíc, M. Epigenetic inheritance and reprogramming in plants and fission yeast. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 265–271 (2008).

83. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* **30**, 329–334 (2002).
84. Santenard, A. *et al.* Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat. Cell Biol.* **12**, 853–862 (2010).
85. Probst, A. V. *et al.* A Strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev. Cell* **19**, 625–638 (2010).
86. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
87. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
88. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
89. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
90. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
91. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev.* **30**, 881–891 (2016).
92. Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
93. Merkenschlager, M. & Nora, E. P. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17–43 (2016).
94. Peters, J. M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev.* **22**, 3089–3114 (2008).
95. Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
96. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387–396 (1999).
97. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2gene. *Nature* **405**, 482–485 (2000).
98. Saitoh, N. *et al.* Structural and functional conservation at the boundaries of the chicken β-globin domain. *EMBO J.* **19**, 2315–2322 (2000).
99. Kanduri, C. *et al.* Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.* **10**, 853–856 (2000).
100. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486–489 (2000).
101. Phillips-Cremins, J. E. & Corces, V. G. Chromatin insulators: linking genome organization to cellular function. *Mol. Cell* **50**, 461–474 (2013).
102. West, A. G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes Dev.* **16**, 271–288 (2002).
103. Haering, C. H., Löwe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, 773–788 (2002).
104. Gruber, S., Haering, C. H. & Nasmyth, K. Chromosomal cohesin forms a ring. *Cell* **112**, 765–777 (2003).
105. Haering, C. H., Farcas, A.-M. M., Arumugam, P., Metson, J. & Nasmyth, K. The cohesin ring concatenates sister DNA molecules. *Nature* **454**, 297–301 (2008).
106. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
107. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
108. Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8309–8314 (2008).
109. DeMare, L. E. *et al.* The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* **23**, 1224–34 (2013).
110. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–5 (2010).

111. Hirano, T. Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev.* **26**, 1659–1678 (2012).
112. Hergeth, S. P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep.* **16**, 1439–1453 (2015).
113. Van Bortle, K. & Corces, V. G. tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription* **3**, 277–284 (2012).
114. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
115. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
116. Hnisz, D. *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015).
117. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
118. Nora, E. P. *et al.* Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
119. Rao, S. S. P. *et al.* Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
120. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
121. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2015).
122. Liu, J. & Krantz, I. D. Cohesin and human disease. *Annu. Rev. Genomics Hum. Genet.* **9**, 303–320 (2008).
123. Remeseiro, S., Cuadrado, A. & Losada, A. Cohesin in development and disease. *Development* **140**, 3715–8 (2013).
124. Ball, A. R., Chen, Y. Y. & Yokomori, K. Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 191–202 (2014).
125. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).
126. Le, T. B. K. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731–734 (2013).
127. Le, T. B. & Laub, M. T. Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. *EMBO J.* **35**, 1582–1595 (2016).
128. Badrinarayanan, A., Le, T. B. K. & Laub, M. T. Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.* **31**, 171–199 (2015).
129. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
130. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, 201518552 (2015).
131. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6**, 1–33 (2017).
132. Glynn, E. F. *et al.* Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. *PLoS Biol.* **2**, (2004).
133. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573–578 (2004).
134. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular condensates in the nucleus. *Trends Biochem. Sci.* (2020).
135. Patel, A. *et al.* A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).
136. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
137. Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
138. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).

139. Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci.* **112**, 7189–7194 (2015).
140. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
141. Banjade, S. *et al.* Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6426–E6435 (2015).
142. Su, X. *et al.* Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595–599 (2016).
143. Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife* **6**, e30294 (2017).
144. Holehouse, A. S. & Pappu, R. V. Functional implications of intracellular phase transitions. *Biochemistry* **57**, 2415–2423 (2018).
145. Wang, J. *et al.* A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699 (2018).
146. Banani, S. F. *et al.* Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
147. Rai, A. K., Chen, J. X., Selbach, M. & Pelkmans, L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211–216 (2018).
148. Owen, I. & Shewmaker, F. The role of post-translational modifications in the phase transitions of intrinsically disordered proteins. *Int. J. Mol. Sci.* **20**, 5501 (2019).
149. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
150. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, (2013).
151. Shrinivas, K. *et al.* Enhancer features that drive formation of transcriptional condensates. *Mol. Cell* **75**, 549–561 (2019).
152. Gibson, B. A. *et al.* Organization of chromatin by intrinsic and regulated phase separation. *Cell* **179**, 470–484 (2019).
153. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
154. Mansour, M. R. *et al.* An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–7 (2014).
155. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
156. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
157. Wang, L. *et al.* Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. *Mol. Cell* **76**, 646–659 (2019).
158. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
159. Huo, X. *et al.* The nuclear matrix protein SAFB cooperates with major satellite RNAs to stabilize heterochromatin architecture partially through phase separation. *Mol. Cell* **77**, 368–383 (2020).
160. Alberti, S. & Dormann, D. Liquid–liquid phase separation in disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).

CHAPTER 2 – YY1 IS A STRUCTURAL REGULATOR OF ENHANCER-PROMOTER LOOPS

Originally published in Cell, Volume 171, Issue 7, 1573-1588. (2017).
Reprinted with permission from Elsevier

Abraham S. Weintraub[1,2,9], Charles H. Li[1,2,9], Alicia V. Zamudio[1,2], Alla A. Sigova[1,6], Nancy M. Hannett[1], Daniel S. Day[1], Brian J. Abraham[1], Malkiel A. Cohen[1], Behnam Nabet[3-5], Dennis L. Buckley[3-4,7], Yang Eric Guo[1], Denes Hnisz[1], Rudolf Jaenisch[1,2,9], James E. Bradner[3-5,7,9], Nathanael S. Gray[4,5,9], and Richard A. Young[1,2,9,10*]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA.
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02142, USA.
[3]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA.
[4]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.
[5]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.
[6]Present address: Marauder Therapeutics, Cambridge, MA 02139, USA
[7]Present address: Novartis Institutes for Biomedical Research, Cambridge, MA 02139, USA.
[8]These authors contributed equally
[9]Senior Author
[10]Lead Contact
*Correspondence to: young@wi.mit.edu.

STATEMENT ON CONTRIBUTION

I contributed to conceptualization of the project and design of the research methodology. I performed experiments and analyses presented in Figs. 1a-c, 1h, 2, 4, 5a-d, 6, and 7, with help from the other authors. I contributed to reviewing and editing the manuscript.

1. SUMMARY

There is considerable evidence that chromosome structure plays important roles in gene control, but we have limited understanding of the proteins that contribute to structural interactions between gene promoters and their enhancer elements. Large DNA loops that encompass genes and their regulatory elements depend on CTCF-CTCF interactions, but most enhancer-promoter interactions do not employ this structural protein. Here we show that the ubiquitously expressed transcription factor Yin Yang 1 (YY1) contributes to enhancer-promoter structural interactions in a manner analogous to DNA interactions mediated by CTCF. YY1 binds to active enhancers and promoter-proximal elements, and forms dimers that facilitate the interaction of these DNA elements. Deletion of YY1 binding sites or depletion of YY1 protein disrupts enhancer-promoter looping and gene expression. We propose that YY1-mediated enhancer-promoter interactions are a general feature of mammalian gene control.

2. INTRODUCTION

Cell-type specific gene expression programs in humans are generally controlled by gene regulatory elements called enhancers[1–5]. Transcription factors (TFs) bind these 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[6–11]. Despite the fundamental importance of proper gene control to cell identity and development, the proteins that contribute to structural interactions between enhancers and promoters are poorly understood.

There is considerable evidence that enhancer-promoter interactions can be facilitated by transcriptional cofactors such as Mediator, structural maintenance of chromosomes (SMC) protein complexes such as cohesin, and DNA binding proteins such as CTCF. Mediator can physically bridge enhancer-bound transcription factors (TFs) and the promoter-bound transcription apparatus[12–16]. Cohesin is loaded at active enhancers and promoters by the Mediator-associated protein NIPBL, and may transiently stabilize enhancer-promoter interactions[14,17]. CTCF proteins bound at enhancers and promoters can interact with one another, and may thus facilitate enhancer-promoter interactions[18,19], but CTCF does not generally occupy these interacting elements[20–23].

Enhancer-promoter interactions generally occur within larger chromosomal loop structures formed by the interaction of CTCF proteins bound to each of the loop anchors[24–27]. These loop structures, variously called 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[22,28–36]. Constraining DNA interactions within CTCF-CTCF loop structures in this manner may facilitate proper enhancer-promoter contacts.

Evidence that CTCF-CTCF interactions play important global roles in chromosome loop structures but are only occasionally directly involved in enhancer-promoter contacts[37], led us to consider the possibility that a bridging protein analogous to CTCF might generally participate in enhancer-promoter interactions. We report here that Yin Yang 1 (YY1) contributes to enhancer-promoter interactions in a manner analogous to DNA looping mediated by CTCF. YY1 and CTCF share many features: both are essential, ubiquitously expressed, zinc-coordinating proteins that bind hypo-methylated DNA sequences, form homodimers and thus facilitate loop formation. The two proteins differ in that YY1 preferentially occupies interacting enhancers and promoters, whereas CTCF preferentially occupies sites distal from these regulatory elements

that tend to form larger loops and participate in insulation. Deletion of YY1 binding sites or depletion of YY1 can disrupt enhancer-promoter contacts and normal gene expression. Thus, YY1-mediated structuring of enhancer-promoter loops is analogous to CTCF-mediated structuring of TADs, CTCF contact domains, and insulated neighborhoods. This model of YY1-mediated structuring of enhancer-promoter loops accounts for diverse functions reported previously for YY1, including contributions to both gene activation and repression and to gene dysregulation in cancer.

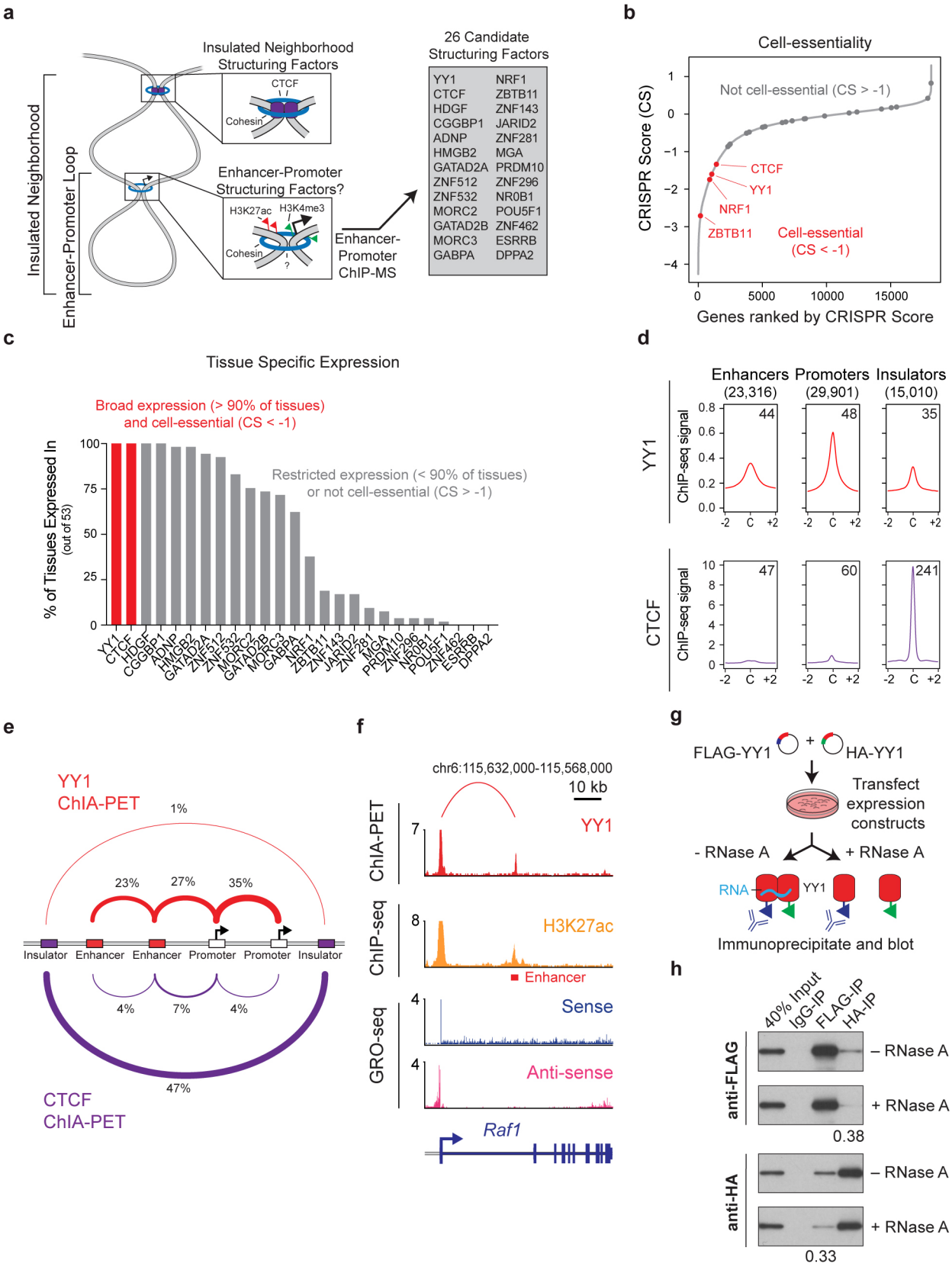
3. RESULTS

A candidate enhancer-promoter structuring factor in ES cells

We sought to identify a protein factor that might contribute to enhancer-promoter interactions in a manner analogous to that of CTCF at insulators. Such a protein would be expected to bind active enhancers and promoters, be essential for cell viability, show ubiquitous expression, and be capable of dimerization. To identify proteins that bind active enhancers and promoters, we sought candidates from chromatin immunoprecipitation with mass spectrometry (ChIP-MS), using antibodies directed towards histones with modifications characteristic of enhancer and promoter chromatin (H3K27ac and H3K4me3, respectively)[38], conducted previously in murine embryonic stem cells (mES cells)[31]. Of 26 transcription factors that occupy both enhancers and promoters (Fig. 1a), four (CTCF, YY1, NRF1 and ZBTB11) are essential based on a CRISPR cell-essentiality screen (Fig. 1b)[39] and two (CTCF, YY1) are expressed in >90% of tissues examined (Fig. 1c). YY1 and CTCF share additional features: like CTCF, YY1 is a zinc-finger transcription factor[40,41], essential for embryonic and adult cell viability[42,43] and capable of forming homodimers[44,45] (Table S1). YY1, however, tends to occupy active enhancers and promoters, as well as some insulators, whereas CTCF preferentially occupies insulator elements (Fig. 1d, Fig. S1a-c).

If YY1 contributes to enhancer-promoter interactions, then chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)[46] for YY1 should show that YY1 is preferentially associated with these interactions. CTCF ChIA-PET, in contrast, should show that CTCF is preferentially associated with insulator DNA interactions. We generated ChIA-PET data for YY1 and CTCF in mES cells and compared these two datasets. The results showed that the majority of YY1-associated interactions connect active regulatory elements (enhancer-enhancer, enhancer-promoter, and promoter-promoter, which we will henceforth call enhancer-promoter interactions), whereas the majority of CTCF-associated interactions connect insulator elements (Fig. 1e, Fig. S1d). Some YY1-YY1 interactions involved simple enhancer-promoter contacts, as seen in the *Raf1* locus (Fig. 1f) and others involved more complex contacts among super-enhancer constituents and their target promoters, as seen in the *Klf9* locus (Fig. S1e). Super-enhancers were generally occupied by YY1 at relatively high densities and exhibited relatively high YY1-YY1 interaction frequencies (Fig. S1e-h). For both YY1 and CTCF, there was also evidence of enhancer-insulator and promoter-insulator interactions, but these were more pronounced for CTCF (Fig. S1d).

Figure 1


a
Insulated Neighborhood
Enhancer-Promoter Loop
Insulated Neighborhood Structuring Factors
CTCF
Cohesin
Enhancer-Promoter Structuring Factors?
H3K27ac
H3K4me3
Cohesin
?
Enhancer-Promoter ChIP-MS
26 Candidate Structuring Factors
YY1 NRF1
CTCF ZBTB11
HDGF ZNF143
CGGBP1 JARID2
ADNP ZNF281
HMGB2 MGA
GATAD2A PRDM10
ZNF512 ZNF296
ZNF532 NR0B1
MORC2 POU5F1
GATAD2B ZNF462
MORC3 ESRRB
GABPA DPPA2
b
Cell-essentiality
Not cell-essential (CS > -1)
CRISPR Score (CS)
CTCF
YY1
NRF1
ZBTB11
Cell-essential (CS < -1)
Genes ranked by CRISPR Score
c
Tissue Specific Expression
Broad expression (> 90% of tissues) and cell-essential (CS < -1)
Restricted expression (< 90% of tissues) or not cell-essential (CS > -1)
% of Tissues Expressed In (out of 53)
d
Enhancers (23,316)
Promoters (29,901)
Insulators (15,010)
YY1
CTCF
ChIP-seq signal
e
YY1 ChIA-PET
1%
23%
27%
35%
Insulator Enhancer Enhancer Promoter Promoter Insulator
4%
7%
4%
47%
CTCF ChIA-PET
f
chr6:115,632,000-115,568,000
10 kb
ChIA-PET
YY1
ChIP-seq
H3K27ac
Enhancer
GRO-seq
Sense
Anti-sense
Raf1
g
FLAG-YY1
HA-YY1
Transfect expression constructs
- RNase A
+ RNase A
RNA
YY1
Immunoprecipitate and blot
h
40% Input
IgG-IP
FLAG-IP
HA-IP
anti-FLAG
– RNase A
+ RNase A
0.38
anti-HA
– RNase A
+ RNase A
0.33

Previous studies have reported that YY1 can form dimers[44]. To confirm that YY1 dimerization occurs, FLAG-tagged and HA-tagged versions of YY1 protein were expressed in cells, nuclei were isolated and the tagged YY1 proteins in nuclear extracts were immunoprecipitated with either anti-FLAG or anti-HA antibodies. The results show that the FLAG-tagged and HA-tagged YY1 proteins interact (Fig. 1g, h, Fig. S1i, j), consistent with prior reports that YY1 proteins oligomerize[44]. Other highly expressed nuclear proteins such as OCT4 did not co-precipitate, indicating that the assay was specific (Fig. S1j). We previously reported that YY1 can bind both DNA and RNA independently, and that YY1 binding of active regulatory DNA elements is enhanced by the binding of RNA species that are transcribed at these loci[47]. It is therefore possible that YY1-YY1 interactions may be enhanced by the ability of each of the YY1 proteins to bind RNA species. Indeed, when we repeated the experiment described above with nuclear extracts containing the tagged YY1 proteins, and a portion of the sample was treated with RNase A prior to immunoprecipitation with anti-tag antibodies, there was a ~60% reduction in the amount of co-immunoprecipitated YY1 partner protein (Fig. 1g, h). These results suggest that stable YY1-YY1 interactions may be facilitated by RNA.

Figure 1. YY1 is a candidate enhancer-promoter structuring factor
(a) Model depicting an enhancer-promoter loop contained within a larger insulated neighborhood loop. Candidate enhancer-promoter structuring transcription factors were identified by ChIP-MS of histones with modifications characteristic of enhancer and promoter chromatin.
(b) CRISPR scores (CS) of all genes in KBM7 cells from Wang et al. (2015). Candidate enhancer-promoter structuring factors identified by ChIP-MS are indicated as dots and those identified as cell-essential (CS < -1) are shown in red.
(c) Histogram showing the number of tissues in which each candidate enhancer-promoter structuring factor is expressed across 53 tissues surveyed by GTEx. Candidates that are both broadly expressed (expressed in greater than 90% of tissues surveyed) and cell-essential are shown in red.
(d) Metagene analysis showing the occupancy of YY1 and CTCF at enhancers, promoters, and insulator elements in mouse ESCs.
(e) Summary of the classes of high-confidence interactions identified by YY1 and CTCF ChIA-PET in mES cells.
(f) Example of a YY1-YY1 enhancer-promoter interaction at the *Raf1* locus in mES cells.
(g) Model depicting co-immunoprecipitation assay to detect YY1 dimerization and evaluate dependence on RNA for YY1 dimerization.
(h) Western blot results showing co-immunoprecipitation of FLAG-tagged YY1 and HA-tagged YY1 protein from nuclear lysates prepared from transfected cells. Quantification of the remaining signal normalized to input after RNase A treatment for the co-immunoprecipitated tagged YY1 is displayed under the relevant bands.

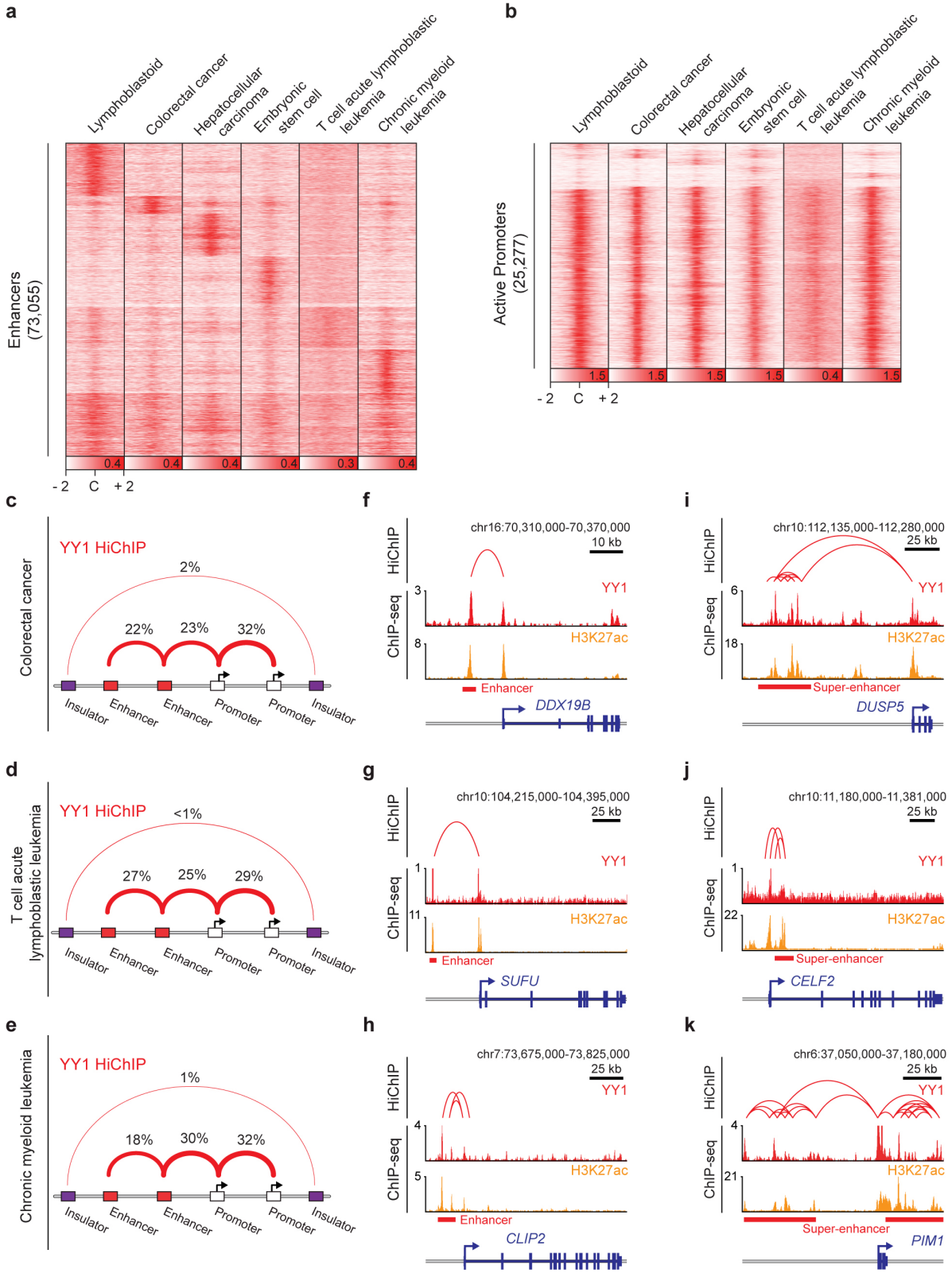
YY1 generally occupies enhancers and promoters in mammalian cells

YY1 is ubiquitously expressed in mammalian cells, so we investigated whether YY1 generally occupies enhancers and promoters in a broad spectrum of mammalian cell types. Examination of sites bound by YY1 across human cell types showed that YY1 does generally occupy enhancers and promoters genome-wide and, as expected, enhancer occupancy tends to be cell-type-specific (Fig. 2a, b, Fig. S2a-f). As with mES cells, YY1 was also found at a subset of insulators in the human cells (Fig. S2a-f). Examination of YY1 ChIP-seq data in multiple murine cell types confirmed that YY1 generally occupies enhancers and promoters, and is present at some insulators (Fig. S2g-j). These results indicate that YY1 generally occupies enhancer and promoter elements in mammalian cells.

To determine whether YY1 is associated with sites of enhancer-promoter interactions in human cells, we conducted YY1 HiChIP experiments[48] in three different cell types. These experiments revealed that YY1 is predominantly associated with enhancer-promoter interactions (Fig. 2c-k). YY1 was also associated with some insulator-enhancer and insulator-promoter interactions, suggesting that the factor may also occasionally participate in such interactions (Fig. S2k-m). In summary, the HiChIP results indicate that YY1 generally occupies sites involved in enhancer-promoter interactions, and occasionally occupies sites of insulator interactions, in mammalian cells.

Figure 2. YY1 generally occupies enhancers and promoters in mammalian cells
(a-b) Heatmaps displaying the YY1 occupancy at enhancers (a) and active promoters (b) in six human cell types.
(c-e) Summaries of the major classes of high-confidence interactions identified with YY1 HiChIP in three human cell types.
(f-k) Examples of YY1-YY1 enhancer-promoter interactions in three human cell types: colorectal cancer (f and i), T cell acute lymphoblastic leukemia (g and j), and chronic myeloid leukemia (h and k). Displayed examples show YY1-YY1 enhancer-promoter interactions involving typical enhancers (f-h) and involving super-enhancers (i-k).

Figure 2

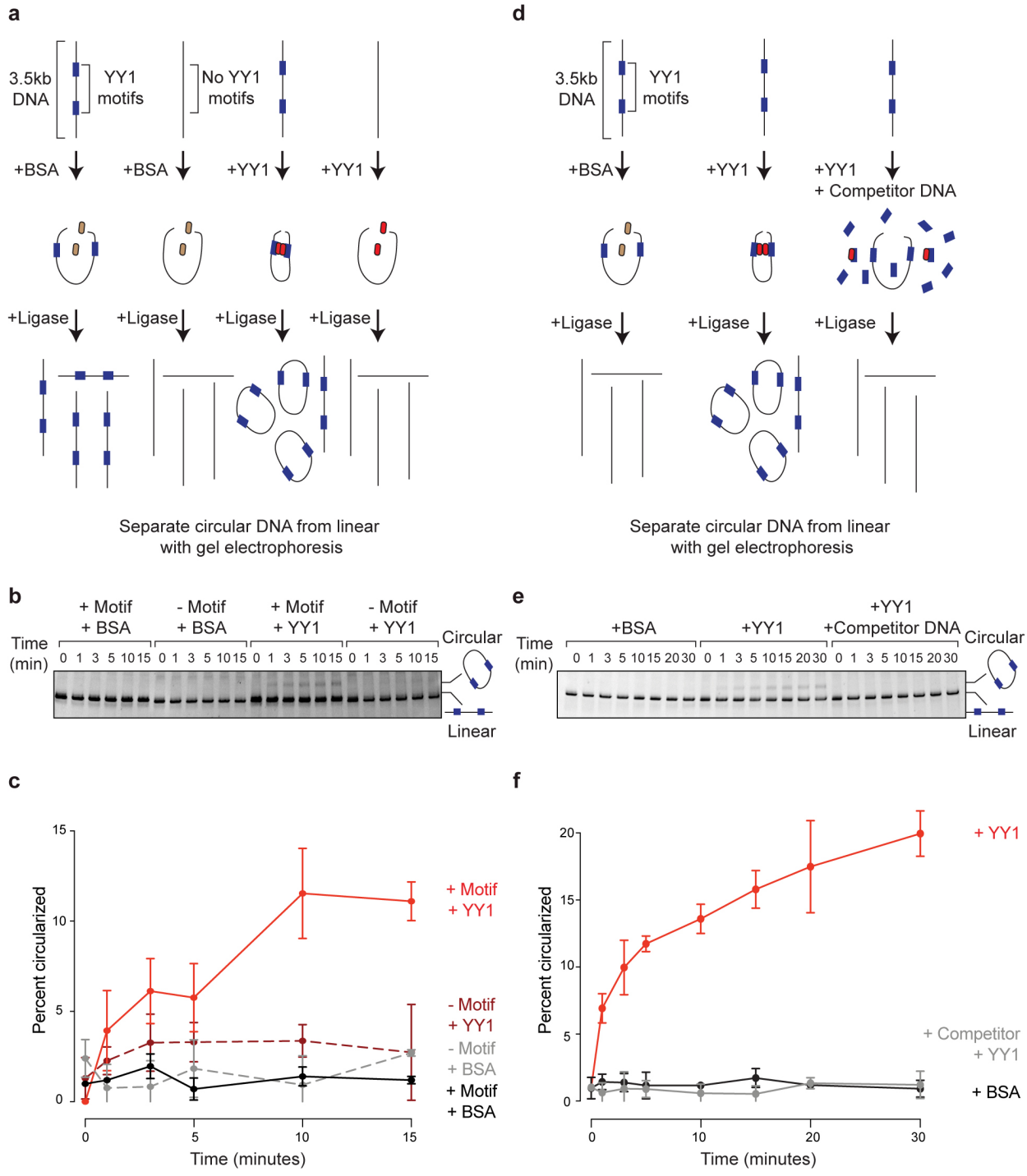

a
Lymphoblastoid
Colorectal cancer
Hepatocellular carcinoma
Embryonic stem cell
T cell acute lymphoblastic leukemia
Chronic myeloid leukemia
Enhancers (73,055)
0.4
0.4
0.4
0.4
0.3
0.4
- 2
C
+ 2
b
Active Promoters (25,277)
1.5
1.5
1.5
1.5
0.4
1.5
c
Colorectal cancer
YY1 HiChIP
2%
22%
23%
32%
Insulator
Enhancer
Enhancer
Promoter
Promoter
Insulator
d
T cell acute lymphoblastic leukemia
YY1 HiChIP
<1%
27%
25%
29%
e
Chronic myeloid leukemia
YY1 HiChIP
1%
18%
30%
32%
f
chr16:70,310,000-70,370,000
10 kb
HiChIP
YY1
ChIP-seq
H3K27ac
Enhancer
DDX19B
g
chr10:104,215,000-104,395,000
25 kb
SUFU
h
chr7:73,675,000-73,825,000
25 kb
CLIP2
i
chr10:112,135,000-112,280,000
25 kb
Super-enhancer
DUSP5
j
chr10:11,180,000-11,381,000
25 kb
CELF2
k
chr6:37,050,000-37,180,000
25 kb
PIM1

YY1 can enhance DNA interactions in vitro

CTCF proteins can form homodimers and larger oligomers, and thus when bound to two different DNA sites can form a loop with the intervening DNA[45]. The observation that YY1 is bound to interacting enhancers and promoters, coupled with the evidence that YY1-YY1 interactions can occur in vitro and in cell extracts, is consistent with the idea that YY1-YY1 interactions can contribute to loop formation between enhancers and promoters. To obtain evidence that YY1 can have a direct effect on DNA interactions, we used an in vitro DNA circularization assay to determine if purified YY1 can enhance the rate of DNA interaction in vitro. The rate of DNA circularization catalyzed by T4 DNA ligase has been used previously to measure persistence length and other physical properties of DNA[49]. We reasoned that if YY1 bound to DNA is capable of dimerizing and thereby forming DNA loops, then incubating a linear DNA template containing YY1 binding sites with purified YY1 protein should bring the ends into proximity and increase the rate of circularization (Fig. 3a, d). Recombinant YY1 protein was purified and shown to have DNA binding activity using a mobility shift assay (Fig. S3a, b). This recombinant YY1 was then tested in the DNA circularization assay; the results showed that YY1 increased the rate of circularization and that this depended on the presence of YY1 motifs in the DNA (Fig. 3b, c). The addition of an excess of a competing 200 base pair DNA fragment containing the YY1 consensus binding sequence abrogated circularization of the larger DNA molecule (Fig. 3d-f). The addition of bovine serum albumin (BSA) did not increase the rate of DNA ligation (Fig. 3c, f). These results support the idea that YY1 can directly facilitate DNA interactions.

Figure 3. YY1 can enhance DNA interactions in vitro
(a and d) Models depicting the in vitro DNA circularization assays used to detect the ability of YY1 to enhance DNA looping interactions.
(b and e) Results of the in vitro DNA circularization assay visualized by gel electrophoresis. The dominant lower band reflects the starting linear DNA template, while the upper band corresponds to the circularized DNA ligation product.
(c and f) Quantifications of DNA template circularization as a function of incubation time with T4 DNA ligase. Values correspond to the percent of DNA template that is circularized and represents the mean and standard deviation of four experiments.

Figure 3


a
3.5kb DNA
YY1 motifs
No YY1 motifs
+BSA
+BSA
+YY1
+YY1
+Ligase
+Ligase
+Ligase
+Ligase
Separate circular DNA from linear with gel electrophoresis
d
3.5kb DNA
YY1 motifs
+BSA
+YY1
+YY1 + Competitor DNA
+Ligase
+Ligase
+Ligase
Separate circular DNA from linear with gel electrophoresis
b
+ Motif + BSA
- Motif + BSA
+ Motif + YY1
- Motif + YY1
Time (min)
0 1 3 5 10 15 0 1 3 5 10 15 0 1 3 5 10 15 0 1 3 5 10 15
Circular
Linear
e
+BSA
+YY1
+YY1 +Competitor DNA
Time (min)
0 1 3 5 10 15 20 30 0 1 3 5 10 15 20 30 0 1 3 5 10 15 20 30
Circular
Linear
c
Percent circularized
Time (minutes)
+ Motif + YY1
- Motif + YY1
- Motif + BSA
+ Motif + BSA
f
Percent circularized
Time (minutes)
+ YY1
+ Competitor + YY1
+ BSA

Enhancer-promoter interactions depend on YY1 in living cells

To test whether enhancer-promoter interactions in living cells depend on YY1 binding sites in these elements, a CRISPR/Cas9 system was used to generate a small deletion of a YY1 binding motif in the regulatory regions of two genes (Fig. 4a). Deletion of the optimal DNA-binding motif for YY1 in the promoter of the *Raf1* gene resulted in decreased YY1 binding at the promoter, reduced contact frequency between the enhancer and promoter, and a decrease in *Raf1* mRNA levels (Fig. 4b, Fig. S4a). Deletion of the optimal DNA-binding motif for YY1 in the promoter of the *Etv4* gene also resulted in decreased YY1 binding and decreased enhancer-promoter contact frequency, although it did not significantly affect the levels of *Etv4* mRNA (Fig. 4c, Fig. S4b). These results suggest that the YY1 binding sites contribute to YY1 binding and enhancer-promoter contact frequencies at both *Raf1* and *Etv4*, although the reduction in looping frequencies at *Etv4* was not sufficient to have a significant impact on *Etv4* mRNA levels. The lack of an effect on *Etv4* mRNA levels may be a consequence of the residual YY1 that is bound to the *Etv4* promoter region, where additional CCAT motifs are observed (Fig. 4c). Indeed, when YY1 protein is depleted (see below; Fig. S6e), the levels of both *Raf1* and *Etv4* mRNA decrease.

Previous studies have reported that YY1 is an activator of some genes and a repressor of others but a global analysis of YY1 dependencies has not been described with a complete depletion of YY1 in mES. We used an inducible degradation system[50–52] to fully deplete YY1 protein levels and measured the impact on gene expression in mES cells genome-wide through RNA-seq analysis (Fig. 5a, b). Depletion of YY1 led to significant (adjusted p-value < 0.05) changes in expression of 8,234 genes, divided almost equally between genes with increased expression and genes with decreased expression (Fig. 5c, Table S2, S3). The genes that experienced the greatest changes in expression with YY1 depletion were generally occupied by YY1 (Fig. 5d).

Previous studies have shown that YY1 is required for normal embryonic development[42]. We therefore investigated whether the loss of YY1 leads to defects in embryonic stem (ES) cell differentiation into the three germ layers (Fig. 5e). Murine ES cells, and isogenic cells that were subjected to inducible degradation of YY1, were stimulated to form embryoid bodies (Fig. 5f) and the cells in these bodies were subjected to immunohistochemistry staining and single-cell RNA-seq to monitor expression of differentiation-specific factors. The results showed that cells lacking YY1 showed pronounced defects in expression of the master transcription factors that drive normal differentiation (Fig. 5g, h, Fig. S5).

Figure 4

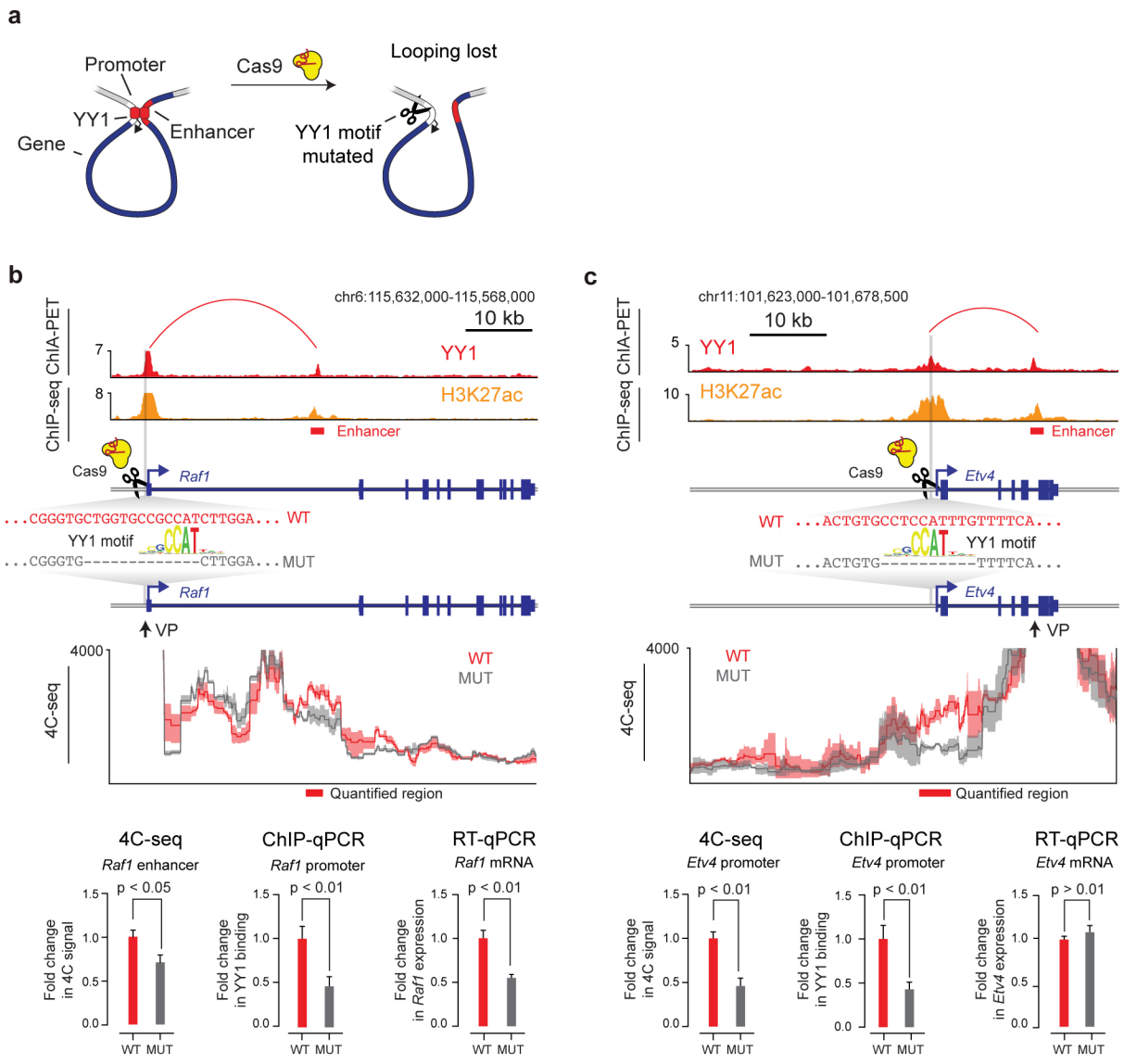
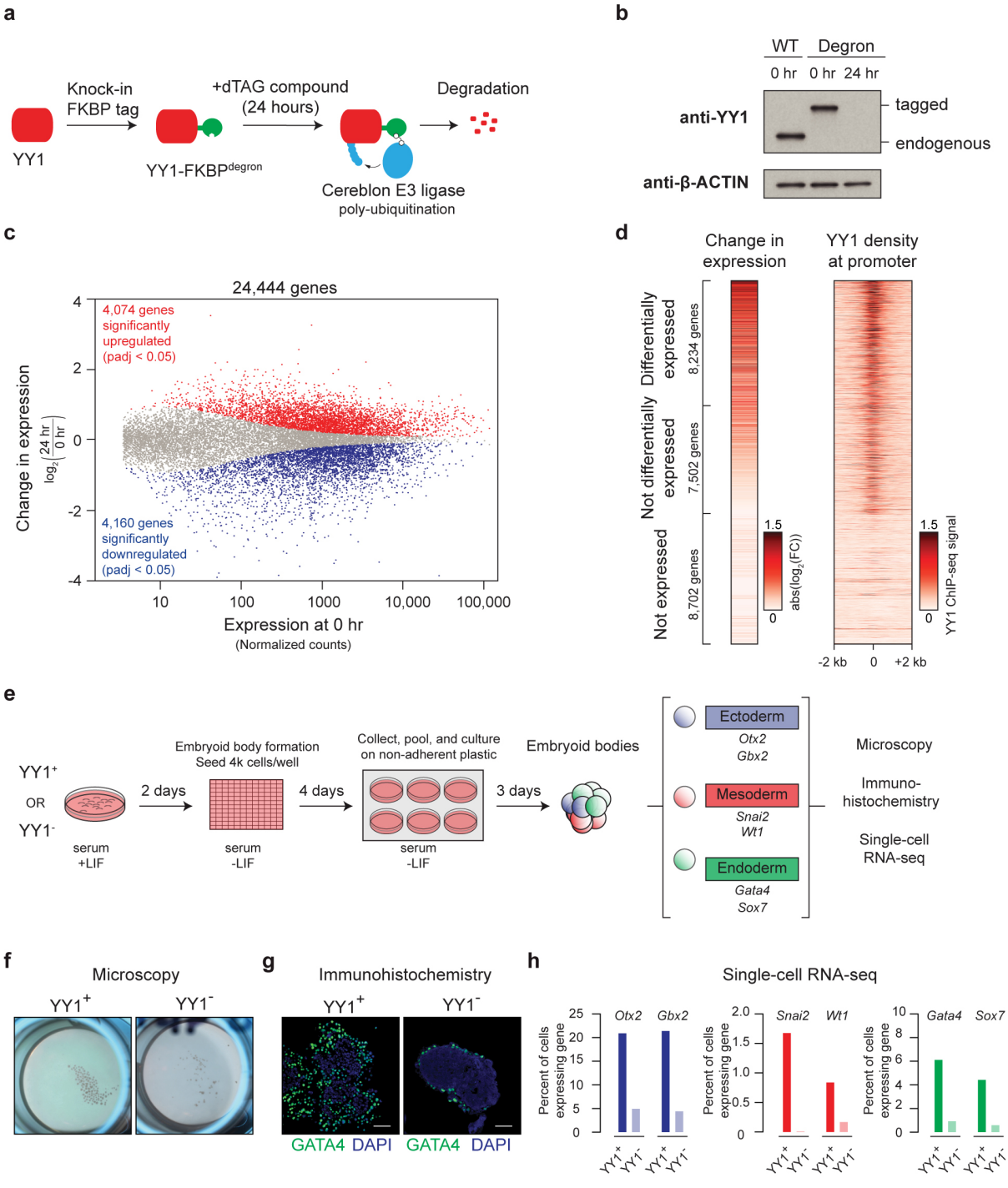


Figure 4. Deletion of YY1 binding sites causes loss of enhancer-promoter interactions
(a) Model depicting CRISPR/Cas9-mediated deletion of a YY1 binding motif in the regulatory region of a gene.
(b and c) CRISPR/Cas9-mediated deletion of YY1 binding motifs in the regulatory regions of two genes, *Raf1* (b) and *Etv4* (c), was performed and the effects on YY1 occupancy, enhancer-promoter looping, and mRNA levels were measured. The positions of the targeted YY1 binding motifs, the genotype of the wildtype and mutant lines, and the 4C-seq viewpoint are indicated. The mean 4C-seq signal is represented as a line (individual replicates are shown in Fig. S4) and the shaded area represents the 95% confidence interval. Three biological replicates were assayed for 4C-seq and ChIP-qPCR experiments, and six biological replicates were assayed for RT-qPCR experiments. Error bars represent the standard deviation. All p-values were determined using the Student's t test.

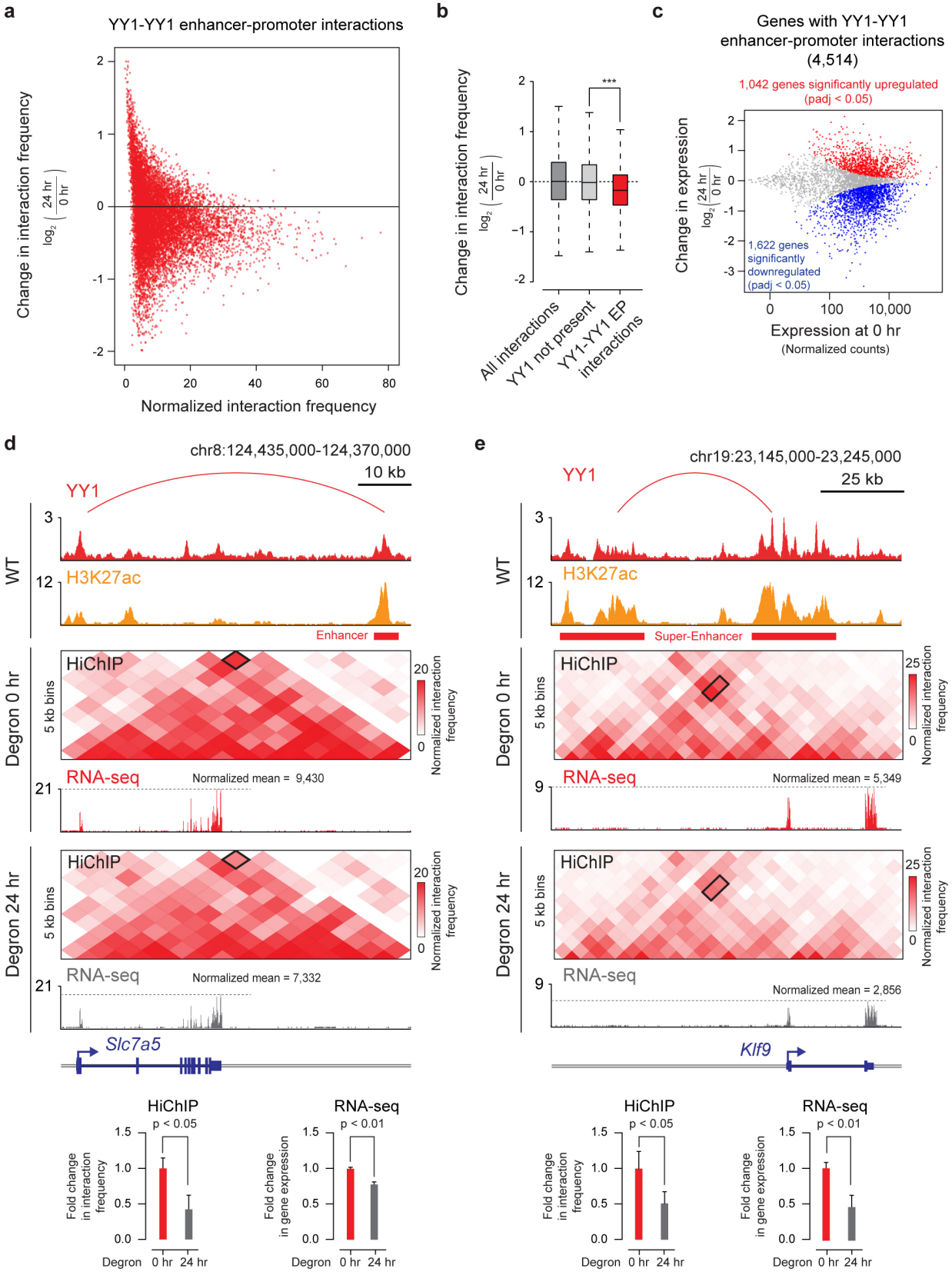
Figure 5


a
Knock-in FKBP tag
+dTAG compound (24 hours)
Degradation
YY1
YY1-FKBP^degron
Cereblon E3 ligase
poly-ubiquitination
b
WT
Degron
0 hr
0 hr
24 hr
anti-YY1
tagged
endogenous
anti-β-ACTIN
c
24,444 genes
4,074 genes significantly upregulated (padj < 0.05)
4,160 genes significantly downregulated (padj < 0.05)
Change in expression
log2(24 hr / 0 hr)
Expression at 0 hr
(Normalized counts)
d
Change in expression
YY1 density at promoter
Differentially expressed
8,234 genes
Not differentially expressed
7,502 genes
Not expressed
8,702 genes
abs(log2(FC))
YY1 ChIP-seq signal
-2 kb
0
+2 kb
e
YY1+
OR
YY1-
serum +LIF
2 days
Embryoid body formation
Seed 4k cells/well
serum -LIF
4 days
Collect, pool, and culture on non-adherent plastic
serum -LIF
3 days
Embryoid bodies
Ectoderm
Otx2
Gbx2
Mesoderm
Snai2
Wt1
Endoderm
Gata4
Sox7
Microscopy
Immuno-histochemistry
Single-cell RNA-seq
f
Microscopy
YY1+
YY1-
g
Immunohistochemistry
YY1+
YY1-
GATA4 DAPI
GATA4 DAPI
h
Single-cell RNA-seq
Percent of cells expressing gene
Otx2
Gbx2
Snai2
Wt1
Gata4
Sox7
YY1+ YY1-

We next investigated whether changes in DNA looping occur upon global depletion of YY1 in mES cells. HiChIP for H3K27ac, a histone modification present at both enhancers and promoters, was performed before and after YY1 depletion to detect differences in enhancer-promoter interaction frequencies. Prior to YY1 depletion, the results of the HiChIP experiment showed interactions between the various elements that were similar to the earlier YY1 ChIA-PET results (Fig. S6a, b). After YY1 depletion, the interactions between YY1-occupied enhancers and promoters decreased significantly (Fig. 6a, b). The majority (60%) of genes connected by YY1 enhancer-promoter loops showed significant changes in gene expression (Fig. 6c; Fig. S6d). Examination of the HiChIP DNA interaction profiles at specific genes confirmed these effects. For example, with YY1 depletion the *Slc7a5* promoter and its enhancer showed a ~50% reduction in interaction frequency, and *Slc7a5* expression levels were reduced by ~27% (Fig. 6d). Similarly, after YY1 depletion the *Klf9* promoter and its super-enhancer showed a ~40% reduction in interaction frequency and *Klf9* expression levels were reduced by ~50% (Fig. 6e).

Figure 5. Depletion of YY1 disrupts gene expression
(a) Model depicting dTAG system used to rapidly deplete YY1 protein.
(b) Western blot validation of knock-in of FKBP degron tag and ability to inducibly degrade YY1 protein.
(c) Change in gene expression (log_2 fold-change) upon degradation of YY1 for all genes plotted against the expression in untreated cells. Genes that displayed significant changes in expression (FDR adjusted p-value < 0.05) are colored with upregulated genes plotted in red and downregulated genes plotted in blue.
(d) Heatmaps displaying the change in expression of each gene upon degradation of YY1 and wild type YY1 ChIP-seq signal in a ±2kb region centered on the TSS of each gene. Each row represents a single gene and genes are ranked by their adjusted p-value for change in expression upon YY1 degradation.
(e) Model depicting experimental outline to test the effect of YY1 degradation on embryonic stem cell differentiation into the three germ layers via embryoid body formation from untreated cells ($YY1^+$) and cells treated with dTAG compound to degrade YY1 ($YY1^-$).
(f) Microscopy images of embryoid bodies formed from $YY1^+$ and $YY1^-$ cells
(g) Immunohistochemistry images of embryoid bodies formed from $YY1^+$ and $YY1^-$ cells. GATA4 is displayed in green and DNA stained using DAPI is displayed in blue. The scale bar represents 50 μm.
(h) Quantification of single-cell RNA-seq results for embryoid bodies formed from $YY1^+$ and $YY1^-$ cells. The percentage of cells expressing various differentiation-specific genes is displayed for $YY1^+$ and $YY1^-$ embryoid bodies.

Figure 6


a
YY1-YY1 enhancer-promoter interactions
Change in interaction frequency
log2 (24 hr / 0 hr)
Normalized interaction frequency
b
Change in interaction frequency
***
All interactions
YY1 not present
YY1-YY1 EP interactions
c
Genes with YY1-YY1 enhancer-promoter interactions (4,514)
1,042 genes significantly upregulated (padj < 0.05)
1,622 genes significantly downregulated (padj < 0.05)
Change in expression
Expression at 0 hr
(Normalized counts)
d
chr8:124,435,000-124,370,000
10 kb
YY1
WT
H3K27ac
Enhancer
Degron 0 hr
HiChIP
5 kb bins
Normalized interaction frequency
RNA-seq
Normalized mean = 9,430
Degron 24 hr
Normalized mean = 7,332
Slc7a5
HiChIP
p < 0.05
Fold change in interaction frequency
Degron 0 hr 24 hr
RNA-seq
p < 0.01
Fold change in gene expression
e
chr19:23,145,000-23,245,000
25 kb
Super-Enhancer
Normalized mean = 5,349
Normalized mean = 2,856
Klf9

Rescue of enhancer-promoter interactions in cells

The ability of an artificially tethered YY1 protein to rescue defects associated with a YY1 binding site mutation would be a strong test of the model that YY1 mediates enhancer-promoter interactions (Fig. 7a). We carried out such a test with a dCas9-YY1 fusion protein targeted to a site adjacent to a YY1 binding site mutation in the promoter-proximal region of *Etv4* (Fig. 7b, c). We found that artificially tethering YY1 protein to the promoter led to increased contact frequency between the *Etv4* promoter and its enhancer and caused increased transcription from the gene (Fig. 7d). These results support the model that YY1 is directly involved in structuring enhancer-promoter loops.

To more globally test if YY1 can rescue the loss of enhancer-promoter interactions after YY1 degradation, we subjected mES cells to YY1 degradation with the dTAG method and then washed out the dTAG compound and allowed YY1 to be restored to normal levels (Fig. 7e, Fig. S7a, b). Enhancer-promoter frequencies were monitored with H3K27ac HiChIP. Consistent with our previous experiment (Fig. 6), the loss of YY1 caused a loss in enhancer-promoter interactions, but the recovery of YY1 levels was accompanied by a substantial increase in enhancer-promoter interactions (Fig. 7f). These results were comparable to the effects observed with the rescue of CTCF-CTCF interactions in a similar experiment described recently (Fig. 7f, Fig. S7c)[53], and support the model that YY1 contributes to structuring of a large fraction of enhancer-promoter loops genome-wide.

Figure 6. Depletion of YY1 disrupts enhancer-promoter looping
(a) Scatter plot displaying for all YY1-YY1 enhancer-promoter interactions the change in normalized interaction frequency (log_2 fold change) upon degradation of YY1, as measured by H3K27ac HiChIP, and plotted against the normalized interaction frequency in untreated cells.
(b) Change in normalized interaction frequency (log_2 fold change) upon degradation of YY1 for three different classes of interactions: all interactions, interactions not associated with YY1 ChIP-seq peaks, and YY1-YY1 enhancer-promoter interactions.
(c) Scatter plot displaying for each gene associated with a YY1-YY1 enhancer-promoter interaction the change in gene expression (log_2 fold-change) upon degradation of YY1 plotted against the expression in untreated cells. Genes that showed significant changes in expression (FDR adjusted p-value < 0.05) are colored with upregulated genes plotted in red and downregulated genes plotted in blue.
(d and e) Effect of YY1 degradation at the *Slc7a5* locus (d) and *Klf9* locus (e) on enhancer-promoter interactions and gene expression. The top of each panel shows an arc representing an enhancer-promoter interaction detected in the HiChIP data. Signal in the outlined pixels was used to quantify the change in normalized interaction frequency upon YY1 degradation. Three biological replicates were assayed per condition for H3K27ac HiChIP and two biological replicates were assayed for RNA-seq. Error bars represent the standard deviation. P-values for HiChIP were determined using the Student's t test. P-values for RNA-seq were determined using a Wald test.

Figure 7

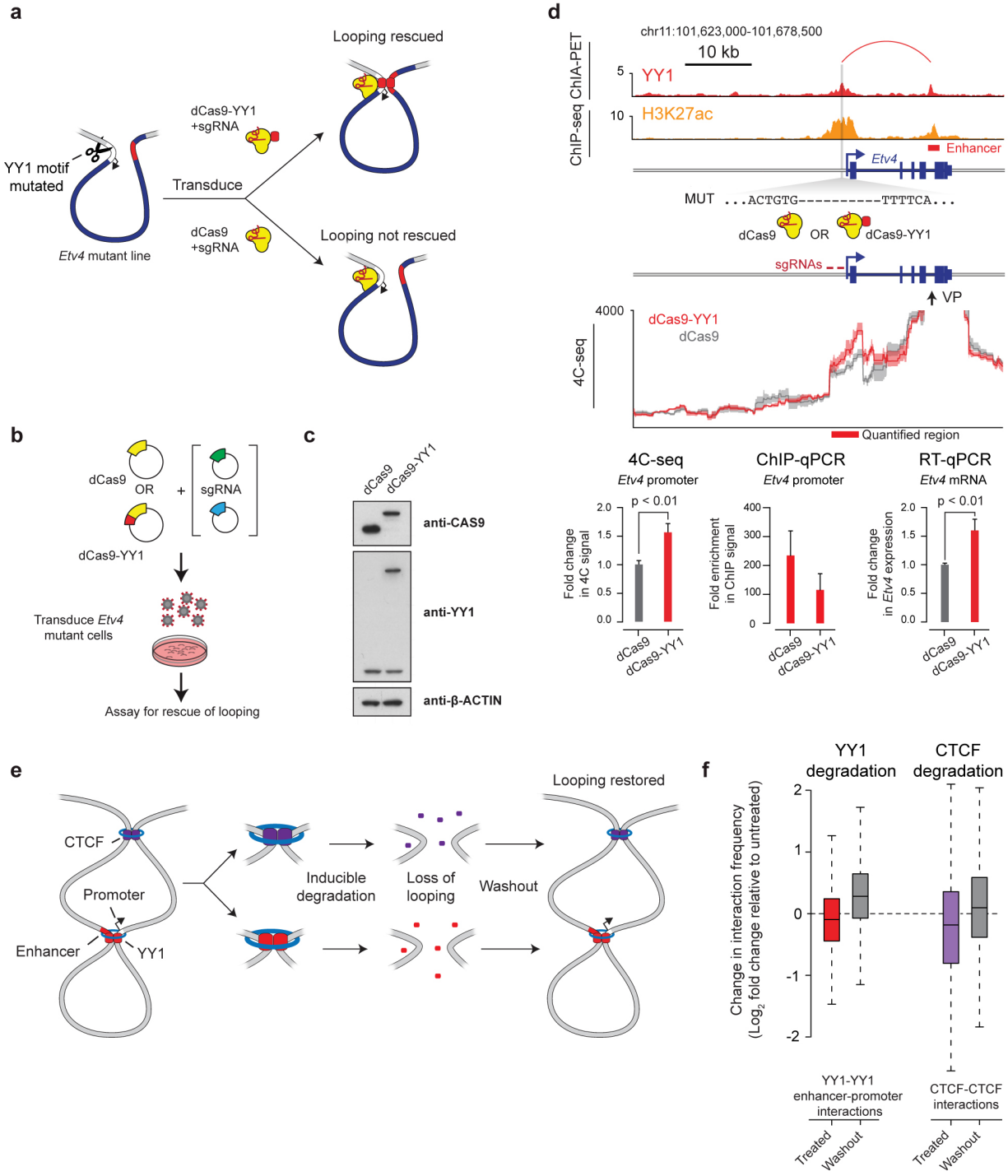

a
Looping rescued
dCas9-YY1
+sgRNA
YY1 motif
mutated
Transduce
dCas9
+sgRNA
Looping not rescued
Etv4 mutant line
b
dCas9
OR
+
sgRNA
dCas9-YY1
Transduce Etv4
mutant cells
Assay for rescue of looping
c
dCas9
dCas9-YY1
anti-CAS9
anti-YY1
anti-β-ACTIN
d
chr11:101,623,000-101,678,500
10 kb
ChIP-seq ChIA-PET
5
YY1
10
H3K27ac
Enhancer
Etv4
MUT ...ACTGTG----------TTTTCA...
dCas9
OR
dCas9-YY1
sgRNAs
VP
4000
dCas9-YY1
dCas9
4C-seq
Quantified region
4C-seq
Etv4 promoter
p < 0.01
Fold change in 4C signal
ChIP-qPCR
Etv4 promoter
Fold enrichment in ChIP signal
RT-qPCR
Etv4 mRNA
p < 0.01
Fold change in Etv4 expression
dCas9
dCas9-YY1
e
CTCF
Promoter
Enhancer
YY1
Inducible degradation
Loss of looping
Washout
Looping restored
f
YY1 degradation
CTCF degradation
Change in interaction frequency (Log2 fold change relative to untreated)
YY1-YY1 enhancer-promoter interactions
CTCF-CTCF interactions
Treated
Washout

Figure 7. Rescue of enhancer-promoter interactions in cells
(a) Model depicting use of dCas9-YY1 to artificially tether YY1 to a site adjacent to the YY1 binding site mutation in the promoter-proximal region of *Etv4* in order to determine if artificially tethered YY1 can rescue enhancer-promoter interactions.
(b) Model depicting dCas9-YY1 rescue experiments. *Etv4* promoter-proximal YY1 binding motif mutant cells were transduced with lentivirus to stably express either dCas9 or dCas9-YY1, and two sgRNAs to direct their localization to the sequences adjacent to the deleted YY1 binding motif in the *Etv4* promoter-proximal region. The ability to rescue enhancer-promoter looping was assayed by 4C-seq.
(c) Western blot results showing that *Etv4* promoter-proximal YY1 binding motif mutant cells transduced with lentivirus to stably express either dCas9 or dCas9-YY1 successfully express dCas9 or dCas9-YY1.
(d) Artificial tethering of YY1 using dCas9-YY1 was performed at sites adjacent to the YY1 binding site mutation in the promoter-proximal region of *Etv4*. The effects of tethering YY1 using dCas9-YY1 on enhancer-promoter looping and expression of the *Etv4* gene were measured and compared to dCas9 alone. The genotype of the *Etv4* promoter-proximal YY1 binding motif mutant cells and the 4C-seq viewpoint (VP) is shown. The 4C-seq signal is displayed as the smoothed average reads per million per base pair. The mean 4C-seq signal is represented as a line and the shaded area represents the 95% confidence interval. Three biological replicates were assayed for 4C-seq and CAS9 ChIP-qPCR experiments, and six biological replicates were assayed for RT-qPCR experiments. Error bars represent the standard deviation. All p-values were determined using the Student's t test.
(e) Model depicting the loss of looping interactions after the inducible degradation of the structuring factors CTCF and YY1 followed by restoration of looping upon washout of degradation compounds.
(f) Change in normalized interaction frequency (log_2 fold change) after YY1 and CTCF degradation (treated) and recovery (washout) relative to untreated cells. For YY1 degradation, change in normalized interaction frequency is plotted for YY1-YY1 enhancer-promoter interactions. For CTCF degradation, change in normalized interaction frequency is plotted for CTCF-CTCF interactions.

4. DISCUSSION

We describe here evidence that the transcription factor YY1 contributes to enhancer-promoter structural interactions. For a broad spectrum of genes, YY1 binds to active enhancers and promoters and is required for normal levels of enhancer-promoter interaction and gene transcription. YY1 is ubiquitously expressed, occupies enhancers and promoters in all cell types examined, is associated with sites of DNA looping in cells where such studies have been conducted, and is essential for embryonic and adult cell viability, so it is likely that YY1-mediated enhancer-promoter interactions are a general feature of mammalian gene control.

Evidence that CTCF-CTCF interactions play important roles in chromosome loop structures, but are only occasionally involved in enhancer-promoter interactions, led us to consider the possibility that a bridging protein analogous to CTCF might generally participate in enhancer-promoter interactions. CTCF and YY1 share many features: they are DNA-binding zinc-finger factors[40,41] that selectively bind hypo-methylated DNA sequences[54,55], are ubiquitously expressed (Fig. 1c)[56], essential for embryonic viability[42,43], and capable of dimerization (Fig. 1g, h, Fig. S1i, j)[44,45]. The two proteins differ in several important ways. CTCF-CTCF interactions occur predominantly between sites that can act as insulators and to a lesser degree between enhancers and promoters (Fig. 1e, Fig. S1a-d). YY1-YY1 interactions occur predominantly between enhancers and promoters and to a lesser extent between insulators (Fig. 1e, Fig. S1a-d). At insulators, CTCF binds to a relatively large and conserved sequence motif (when compared to those bound by other TFs); these same sites tend to be bound in many different cell types, which may contribute to the observation that TAD boundaries tend to be preserved across cell types. At enhancers and promoters, YY1 binds to a relatively small and poorly conserved sequence motif within these regions, where RNA species are produced that can facilitate stable YY1 DNA binding[47]. The cell-type-specific activity of enhancers and promoters thus contributes to the observation that YY1-YY1 interactions tend to be cell-type-specific.

The model that YY1 contributes to structuring of enhancer-promoter loops can account for the many diverse functions previously reported for YY1, including activation and repression, differentiation, and cellular proliferation. For example, following its discovery in the early 1990's[41,57,58], YY1 was intensely studied and reported to act as a repressor for some genes and an activator for others; these context-specific effects have been attributed to many different mechanisms[59–61]. There are many similar reports of context-specific activation and repression by CTCF[37,62]. Although it is reasonable to assume that YY1 and CTCF can act directly as activators or repressors at some genes, the evidence that these proteins contribute to structuring of DNA loops makes it likely that the diverse active and repressive roles that have been attributed to them are often a consequence of their roles in DNA structuring. In this model, the loss of CTCF or YY1 could have positive or negative effects due to other regulators that were no longer properly positioned to produce their regulatory activities.

Previous studies have hinted at a role for YY1 in long distance DNA interactions. CTCF, YY1 and cohesin have been implicated in the formation of DNA loops needed for V(D)J rearrangement at the immunoglobulin locus during B cell development[63–65]. B cell-specific deletion of YY1 causes a decrease in the contraction of the IgH locus, thought to be mediated by DNA loops, and a block in the development of B cells[65]. Knockdown of YY1 has also been shown to reduce intrachromosomal interactions between the Th2 LCR and the *IL4* promoter[66]. As this manuscript was completed, a paper appeared reporting that YY1 is present at the base of interactions between neuronal precursor cell specific enhancers and genes and that YY1 knockdown causes a loss of these interactions[67]. The results described here argue that YY1 is more of a general structural regulator of enhancer-promoter interactions for a large population of genes, both cell-type specific and otherwise, in all cells. Thus, the tendency of YY1 to be involved in cell-type specific loops is a reflection of the cell-type specificity of enhancers and, consequently, their interactions with genes that can be expressed in a cell-specific or a more general manner.

YY1 plays an important role in human disease; YY1 haploinsufficiency has been implicated in an intellectual disability syndrome and YY1 overexpression occurs in many cancers. A cohort of patients with various mutations in one allele and exhibiting intellectual disability have been described as having a “YY1 Syndrome”, and lymphoblastoid cell lines from these patients show reduced occupancy of regulatory regions and small changes in gene expression at a subset of genes associated with YY1 binding[68]. These results are consistent with the model we describe for YY1 in global enhancer-promoter structuring, and with the idea that higher neurological functions are especially sensitive to such gene dysregulation. YY1 is over-expressed in a broad spectrum of tumor cells, and this over-expression has been proposed to cause unchecked cellular proliferation, tumorigenesis, metastatic potential, resistance to immune-mediated apoptotic stimuli and resistance to chemotherapeutics[59,69]. The mechanisms that have been reported to mediate these effects include YY1-mediated downregulation of p53 activity, interference with poly-ADP-ribose polymerase, alteration in c-Myc and NF-κB expression, regulation of death genes and gene products, differential YY1 binding in the presence of inflammatory mediators and YY1 binding to the oncogenic c-Myc transcription factor[59,69]. Although it is possible that YY1 carries out all these functions, it’s role as a general enhancer-promoter structuring factor is a more parsimonious explanation of these pleotropic phenotypes.

Many zinc-coordinating transcription factors are capable of homo- and hetero-dimerization[70,71] and because these comprise the largest class of transcription factors in mammals[72], we suggest that a combination of cell-type-specific and cell-ubiquitous transcription factors make a substantial and underappreciated contribution to enhancer-promoter loop structures. There are compelling studies of bacterial and bacteriophage transcription factors that contribute to looping of regulatory DNA elements through oligomerization[73,74], and reports of several eukaryotic factors with similar capabilities[75]. Nonetheless, most recent study of eukaryotic enhancer-promoter interactions has focused on cofactors that lack DNA binding capabilities and bridge enhancer-bound transcription factors and promoter-bound transcription apparatus[12–16,76], with the notable exception of the proposals that some enhancer-promoter interactions are determined by the nature of transcription factors bound at the two sites[77]. We predict that future studies will reveal additional transcription factors that belong in the class of DNA binding proteins whose predominant role is to contribute to chromosome structure.

5. SUPPLEMENTAL FIGURES

Figure S1

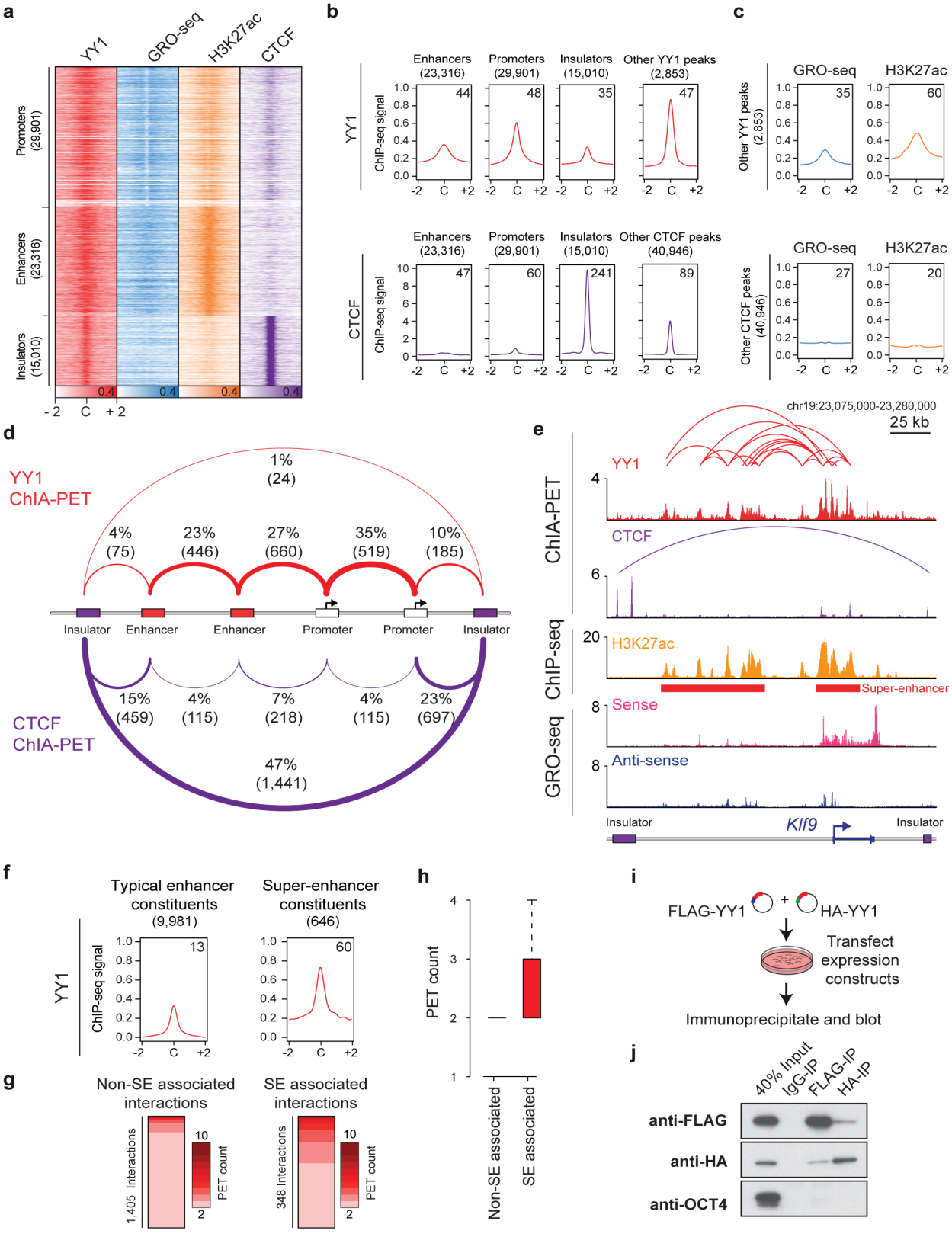


Figure S1. YY1-associated interactions connect enhancers and promoters, related to Figure 1
(a) Heatmap displaying YY1, H3K27ac, and CTCF ChIP-seq signal and GRO-seq signal at promoters, enhancers, and insulators in mouse embryonic stem cells (mES cells). ChIP-seq and GRO-seq signal is plotted as reads per million per base pair in a ±2kb region centered on each promoter, enhancer, and insulator.
(b) Expanded metagene analysis showing the occupancy of YY1 and CTCF at enhancers, promoters, and insulator elements in mES cells. In addition, occupancy of YY1 was plotted at YY1 peaks that were not classified as an enhancer, promoter, or insulator, and occupancy of CTCF was plotted at CTCF peaks that were not classified as an enhancer, promoter, or insulator. ChIP-seq profiles are shown as mean reads per million per base pair for elements of each class in a ±2kb region centered on each region. The number of enhancers, promoters, and insulators surveyed are noted in parentheses. To facilitate comparisons of the same factor between different regions the total ChIP-seq signal in the region was quantified and is displayed in the top right corner of the plot for each metagene analysis.
(c) Metagene analysis showing GRO-seq signal and H3K27ac ChIP-seq signal at YY1 and CTCF peaks in mES cells that were not classified as part of an enhancer, promoter, or insulator. ChIP-seq profiles are shown as mean reads per million per base pair for elements of each class in a ±2kb region centered on each region. The number of YY1 and CTCF peaks surveyed are noted in parentheses. To facilitate comparisons of the same factor between different regions the total ChIP-seq signal in the region was quantified and is displayed in the top right corner of the plot for each metagene analysis.
(d) Expanded summary of the major classes of high-confidence interactions identified in YY1 and CTCF ChIA-PET datasets presented in Fig. 1e. Interactions are classified based on the presence of enhancer, promoter, and insulator elements at the anchors of each interaction. Interactions are displayed as arcs between these elements and the thickness of the arcs approximately reflects the percentage of interactions of that class relative to the total number of interactions that were classified.
(E) An example of extensive YY1-associated enhancer-promoter interactions. The high-confidence YY1 interactions are depicted as red arcs, while high-confidence CTCF interactions are depicted as blue arcs. ChIP-seq binding profiles for YY1, CTCF, and H3K27ac, and stranded GRO-seq signal are displayed as reads per million per base pair at the *Klf9* locus in mES cells. The *Klf9* gene is indicated in the gene model and the interacting super-enhancers are labeled under the H3K27ac ChIP-seq track.
(F) Metagene analysis showing the occupancy of YY1 at typical enhancer constituents and super-enhancer constituents. ChIP-seq profiles are shown in mean reads per million per base pair for elements of each class in a ±2kb region centered on each region. To facilitate comparisons of the same factor between different regions the total ChIP-seq signal in the region was quantified and is displayed in the top right corner of the plot for each metagene analysis. The number of elements surveyed is listed at the top of the plot. Both plots are floored at the minimum amount of typical enhancer constituent signal.
(G) Heatmaps displaying for each high-confidence YY1 interaction the number of PETs that support the interaction, for interactions that have at least one anchor overlapping a super-enhancer (right) and for interactions that have no ends overlapping a super-enhancer (left). Each row represents an interaction and the color intensity of each row represents the PET count for that interaction. (H) Box plot displaying the PET counts of high confidence YY1 ChIA-PET interactions that are either not associated with super-enhancers or associated with super-enhancers.
(I) Model depicting co-immunoprecipitation assay to detect YY1 dimerization.
(J) Western blot results showing co-immunoprecipitation of FLAG-tagged YY1 and HA-tagged YY1 protein from nuclear lysates prepared from transfected cells. Interaction between FLAG-tagged YY1 and HA-tagged YY1 protein is observed, while interaction with OCT4 protein is not observed.
The sources of the datasets used in this figure are listed in Table S4.

Figure S2

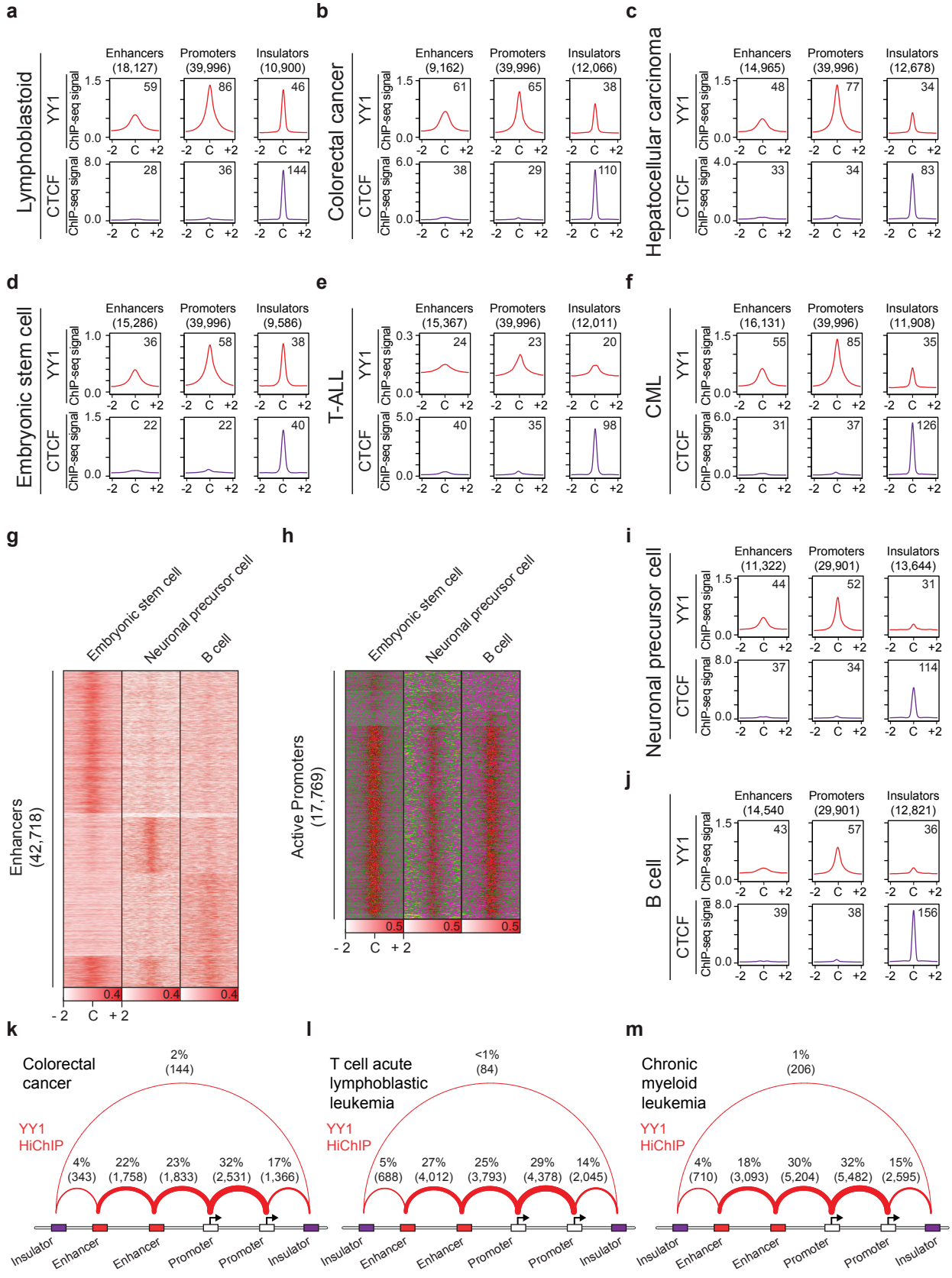

a
Lymphoblastoid
Enhancers (18,127)
Promoters (39,996)
Insulators (10,900)
YY1 ChIP-seq signal
CTCF ChIP-seq signal
b
Colorectal cancer
Enhancers (9,162)
Promoters (39,996)
Insulators (12,066)
c
Hepatocellular carcinoma
Enhancers (14,965)
Promoters (39,996)
Insulators (12,678)
d
Embryonic stem cell
Enhancers (15,286)
Promoters (39,996)
Insulators (9,586)
e
T-ALL
Enhancers (15,367)
Promoters (39,996)
Insulators (12,011)
f
CML
Enhancers (16,131)
Promoters (39,996)
Insulators (11,908)
g
Embryonic stem cell
Neuronal precursor cell
B cell
Enhancers (42,718)
h
Active Promoters (17,769)
i
Neuronal precursor cell
Enhancers (11,322)
Promoters (29,901)
Insulators (13,644)
j
B cell
Enhancers (14,540
Promoters (29,901)
Insulators (12,821)
k
Colorectal cancer
2% (144)
YY1 HiChIP
4% (343)
22% (1,758)
23% (1,833)
32% (2,531)
17% (1,366)
Insulator
Enhancer
Enhancer
Promoter
Promoter
Insulator
l
T cell acute lymphoblastic leukemia
<1% (84)
5% (688)
27% (4,012)
25% (3,793)
29% (4,378)
14% (2,045)
m
Chronic myeloid leukemia
1% (206)
4% (710)
18% (3,093)
30% (5,204)
32% (5,482)
15% (2,595)

Figure S2. YY1 is associated with enhancers and promoters in cancer cells, related to Figure 2
Metagene analysis showing the occupancy of YY1 and CTCF at enhancers, gene promoters, and insulator elements in six human cell types. ChIP-seq profiles are shown in mean reads per million per base pair for elements of each class in a ±2kb region centered on each region. For each cell type the numbers enhancers, promoters, and insulators surveyed are listed below the listed elements. To facilitate comparisons of the same factor between different regions the total ChIP-seq signal in the region was quantified and is displayed in the top right corner of the plot for each metagene analysis.
(a) Metagene analysis in human lymphoblastoid cells.
(b) Metagene analysis in human colorectal cancer cells.
(c) Metagene analysis in human hepatocellular carcinoma cells.
(d) Metagene analysis in human embryonic stem cells.
(e) Metagene analysis in human T cell acute lymphoblastic leukemia (T-ALL) cells.
(f) Metagene analysis in human chronic myeloid leukemia (CML) cells.
(g) Heatmaps displaying the YY1 occupancy at enhancers and promoters in three murine cell types. YY1 ChIP-seq signal is plotted as reads per million per base pair in a ±2kb region centered on each an enhancer and promoter. Each column represents a different cell type. Each row represents an enhancer that was identified in at least one of the displayed cell types
(h) Heatmaps displaying the YY1 occupancy at active promoters in three murine cell types. YY1 ChIP-seq signal is plotted as reads per million per base pair in a ±2kb region centered on each promoter. Each column represents a different cell type. Each row represents a promoter that was identified as an active promoter in at least one of the displayed cell types
Metagene analysis showing the occupancy of YY1 and CTCF at enhancers, gene promoters, and insulator elements in two murine cell types. ChIP-seq profiles are shown in mean reads per million per base pair for elements of each class in a ±2kb region centered on each region. For each cell type the numbers of enhancers, promoters, and insulators surveyed are listed below the listed elements. To facilitate comparisons of the same factor between different regions the total ChIP-seq signal in the region was quantified and is displayed in the top right corner of the plot for each metagene analysis.
(i) Metagene analysis in murine neuronal precursor cells (NPC).
(j) Metagene analysis in murine B cells.
(k-m) Expanded summary of the major classes of high-confidence interactions identified in YY1 HiChIP datasets presented in Fig. 2c-e. Interactions are classified based on the presence of enhancer, promoter, and insulator elements at the anchors of each interaction. Interactions are displayed as arcs between these elements and the thickness of the arcs approximately reflects the percentage of interactions of that class relative to the total number of interactions that were classified.
The sources of the datasets used in this figure are listed in Table S4.

Figure S3

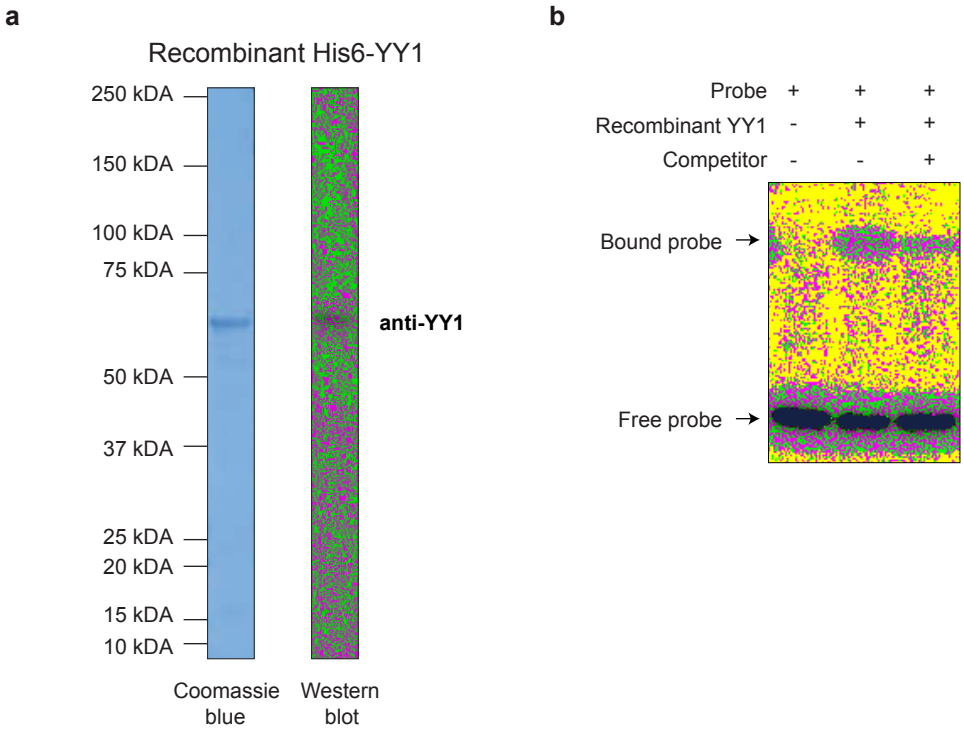


Figure S3. YY1 can enhance DNA interactions in vitro, related to Figure 3
(a) Purity of recombinant His6-YY1 protein was validated by gel electrophoresis of the purified material followed by Coomassie blue staining and western blot analysis with anti-YY1 antibody.
(b) Activity of purified recombinant YY1 protein was validated by EMSA. Purified YY1 was incubated with biotinylated DNA probe in the presence or absence of a non-biotinylated competitor DNA. Activity of the recombinant protein was assessed by the ability to bind DNA and was determined by resolution on a native gel. Unbound “free” biotinylated probe is found at the bottom of the gel, while probe bound by YY1 migrates slower and appears as a higher band. Addition of competitor DNA abrogates this effect indicating that the activity is specific.

Figure S4

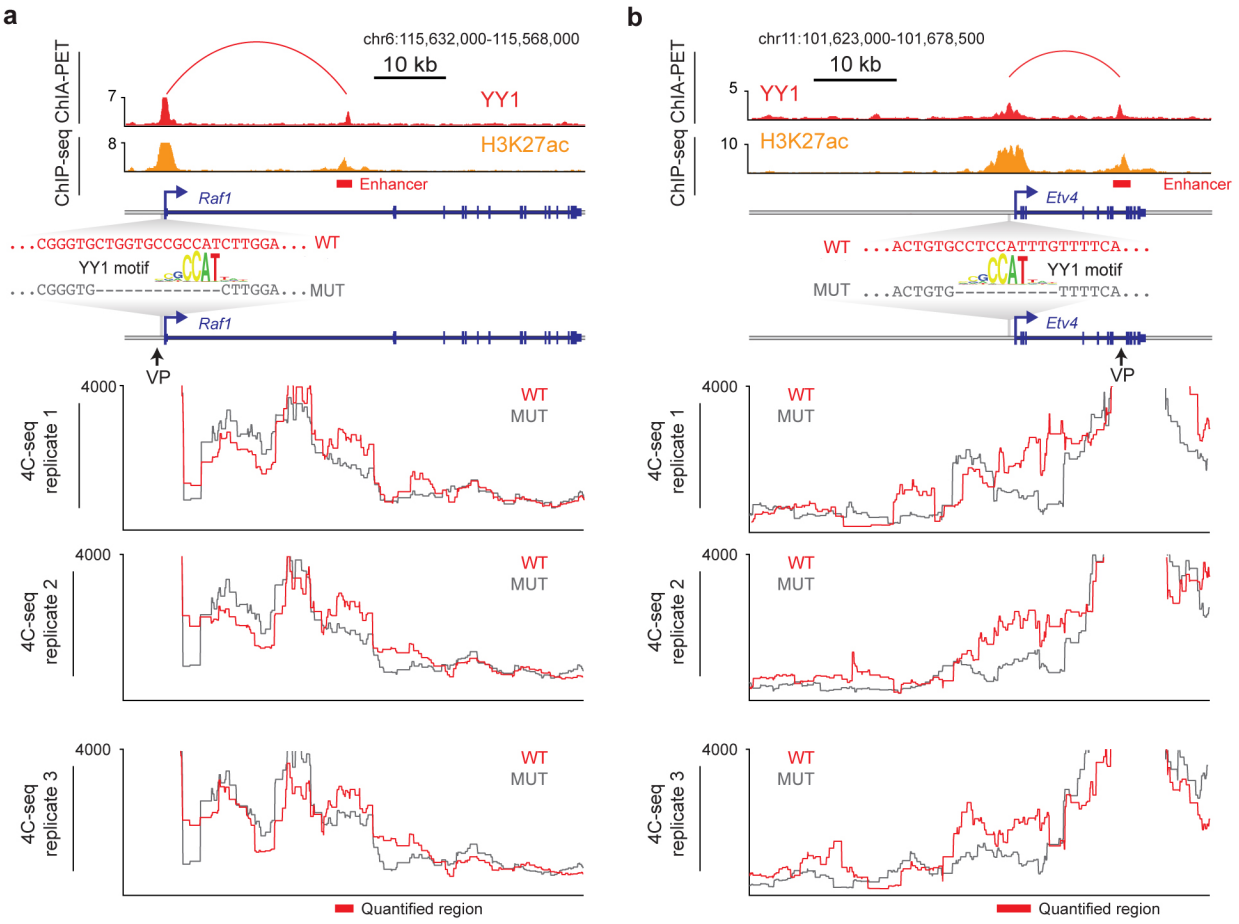


Figure S4. Loss of YY1 binding causes loss of enhancer-promoter interactions, related to Figure 4
(a and b) CRISPR/Cas9-mediated deletion of YY1 binding motifs in the regulatory regions of two genes, *Raf1* (a) and *Etv4* (b). The top of each panel shows a high-confidence YY1-YY1 enhancer-promoter interaction and ChIP-seq binding profiles for YY1 and H3K27ac displayed as reads per million per base pair. Position of the targeted YY1 DNA binding motif and the genotype of the wildtype and mutant lines are shown. The bottom of each panel shows chromatin interaction profiles in wildtype and mutant cells anchored on the indicated viewpoint (VP) for three biological replicates. 4C-seq signal is displayed as smoothed reads per million per base pair.
The sources of the datasets used in this figure are listed in Table S4.

Figure S5

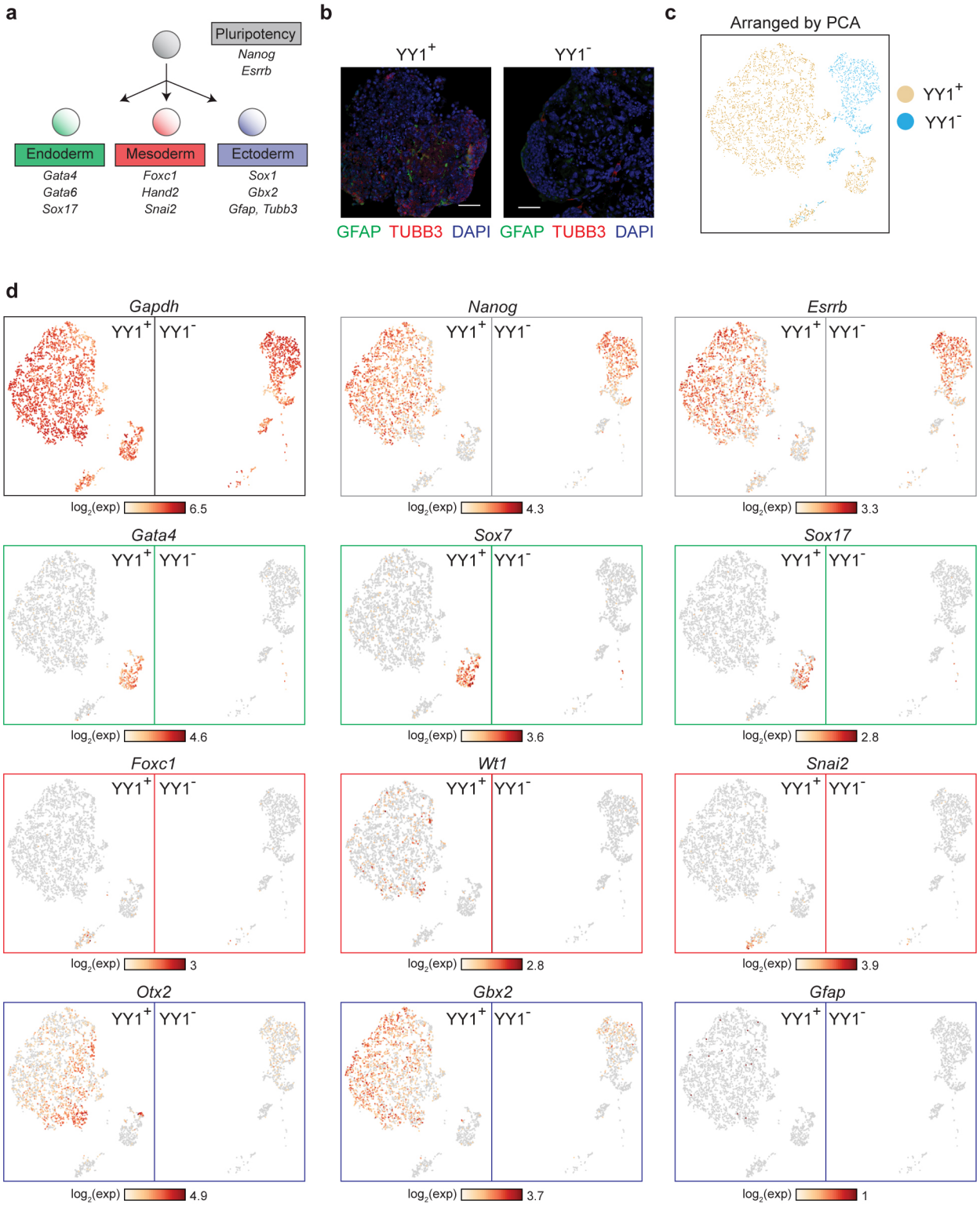

a
Pluripotency
Nanog
Esrrb
Endoderm
Gata4
Gata6
Sox17
Mesoderm
Foxc1
Hand2
Snai2
Ectoderm
Sox1
Gbx2
Gfap, Tubb3
b
YY1+
YY1-
GFAP TUBB3 DAPI GFAP TUBB3 DAPI
c
Arranged by PCA
YY1+
YY1-
d
Gapdh
Nanog
Esrrb
Gata4
Sox7
Sox17
Foxc1
Wt1
Snai2
Otx2
Gbx2
Gfap
log2(exp)
6.5
4.3
3.3
4.6
3.6
2.8
3
2.8
3.9
4.9
3.7
1

Figure S5. Depletion of YY1 impairs ES cell differentiation, related to Figure 5
(a) Model depicting differentiation of pluripotent ES cells into cells of the three germ layers. Pluripotency and differentiation specific markers that were examined are indicated.
(b) Immunohistochemistry images of embryoid bodies formed from untreated cells (YY1+) and cells treated with dTAG compound to degrade YY1 (YY1-). GFAP and TUBB3, which are expressed in cells belonging to the ectoderm lineage, are displayed in green and red, respectively. DNA stained using DAPI is displayed as blue.
(c) Principle component analysis (PCA) based representation of single-cell RNA-seq data for embryoid bodies formed from untreated cells (YY1+) and cells treated with dTAG compound to degrade YY1 (YY1-). Each dot represents a single-cell and dots are arranged based on PCA. Cells from YY1+ embryoid bodies are shown in beige and cells from YY1- embryoid bodies are shown in blue.
(d) Expression of pluripotency and differentiation specific genes (Fig. S5a) as measured by single-cell RNA-seq of embryoid bodies formed from untreated cells (YY1+) and cells treated with dTAG compound to degrade YY1 (YY1-). Each dot represents a single-cell and dots are shaded based on their normalized expression value.
The sources of the datasets used in this figure are listed in Table S4.

Figure S6

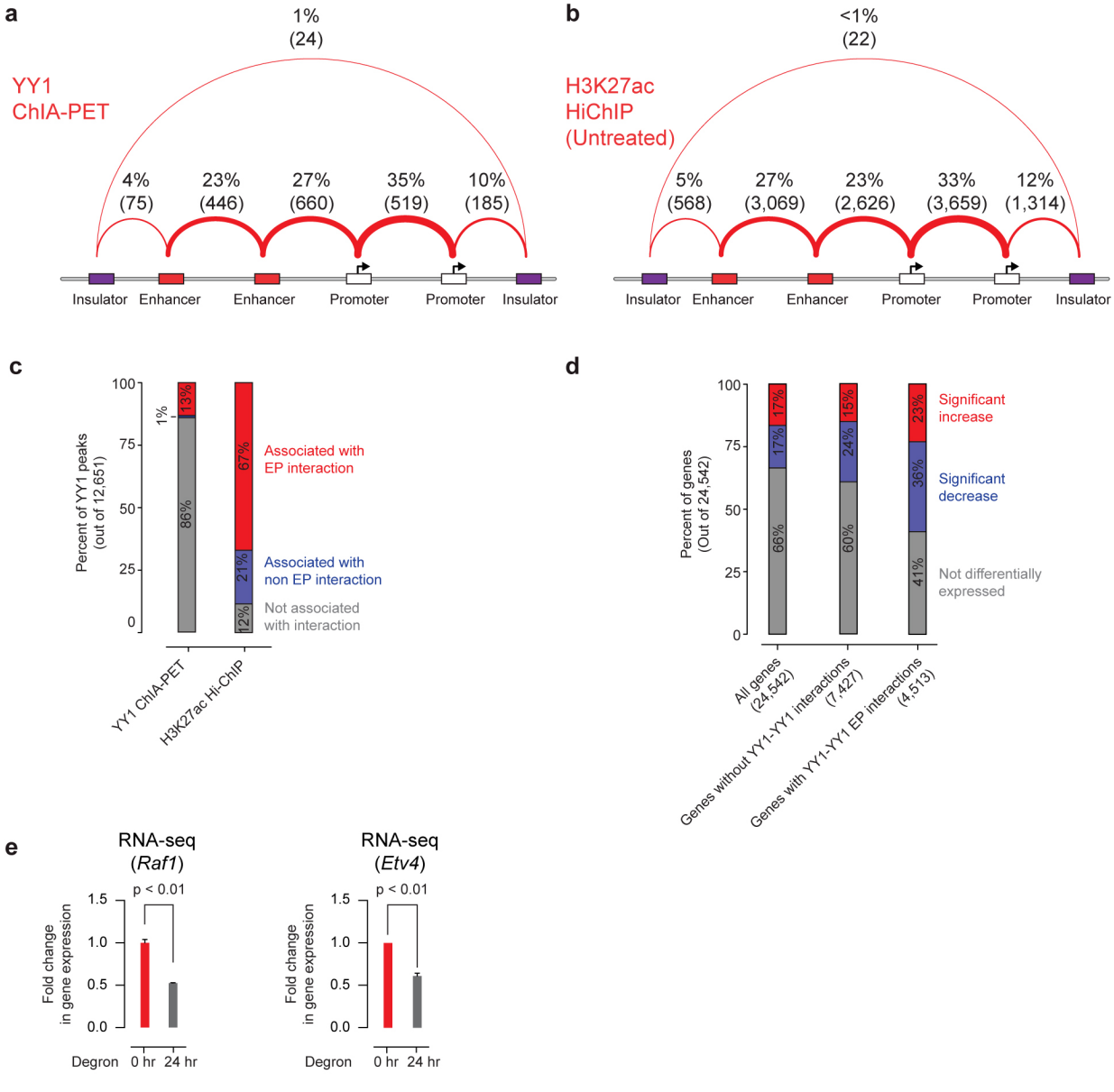

a
1%
(24)
YY1
ChIA-PET
4%
(75)
23%
(446)
27%
(660)
35%
(519)
10%
(185)
Insulator
Enhancer
Enhancer
Promoter
Promoter
Insulator
b
<1%
(22)
H3K27ac
HiChIP
(Untreated)
5%
(568)
27%
(3,069)
23%
(2,626)
33%
(3,659)
12%
(1,314)
Insulator
Enhancer
Enhancer
Promoter
Promoter
Insulator
c
Percent of YY1 peaks
(out of 12,651)
100
75
50
25
0
1%
13%
86%
67%
21%
12%
Associated with
EP interaction
Associated with
non EP interaction
Not associated
with interaction
YY1 ChIA-PET
H3K27ac Hi-ChIP
d
Percent of genes
(Out of 24,542)
100
75
50
25
0
17%
17%
66%
15%
24%
60%
23%
36%
41%
Significant
increase
Significant
decrease
Not differentially
expressed
All genes
(24,542)
Genes without YY1-YY1 interactions
(7,427)
Genes with YY1-YY1 EP interactions
(4,513)
e
RNA-seq
(Raf1)
p < 0.01
Fold change
in gene expression
1.5
1.0
0.5
0.0
Degron
0 hr
24 hr
RNA-seq
(Etv4)
p < 0.01
Fold change
in gene expression
1.5
1.0
0.5
0.0
Degron
0 hr
24 hr

Figure S6. Depletion of YY1 disrupts enhancer-promoter looping, related to Figure 6
(a and b) Summaries of the major classes of high-confidence interactions identified by YY1 ChIA-PET (a) and H3K27ac HiChIP (b). Interactions are classified based on the presence of enhancer, promoter, and insulator elements at the anchors of each interaction. Interactions are displayed as arcs between these elements and the thickness of the arcs approximately reflects the percentage of interactions of that class relative to the total number of interactions that were classified.
(c) Percent of YY1 ChIP-seq peaks in mES cells that are associated with enhancer-promoter interactions, associated with non-enhancer-promoter interactions, and not associated with a detected interaction for high confidence interactions identified by YY1 ChIA-PET and H3K27ac HiChIP.
(d) Percent of genes that significantly increase in expression, significantly decrease in expression, or are not differentially expressed in response to YY1 degradation for three classes of genes: all genes, genes involved in enhancer-promoter interactions that do not have YY1 peaks at both ends, and genes involved in YY1-YY1 enhancer-promoter interactions.
(e) Expression of Raf1 and Etv4 genes before (0 hr) and after YY1 degradation (24 hr) as measure by RNA-seq. Error bars represent the standard deviation of two biological replicates. P-values were determined using a Wald test.
The sources of the datasets used in this figure are listed in Table S4.

Figure S7

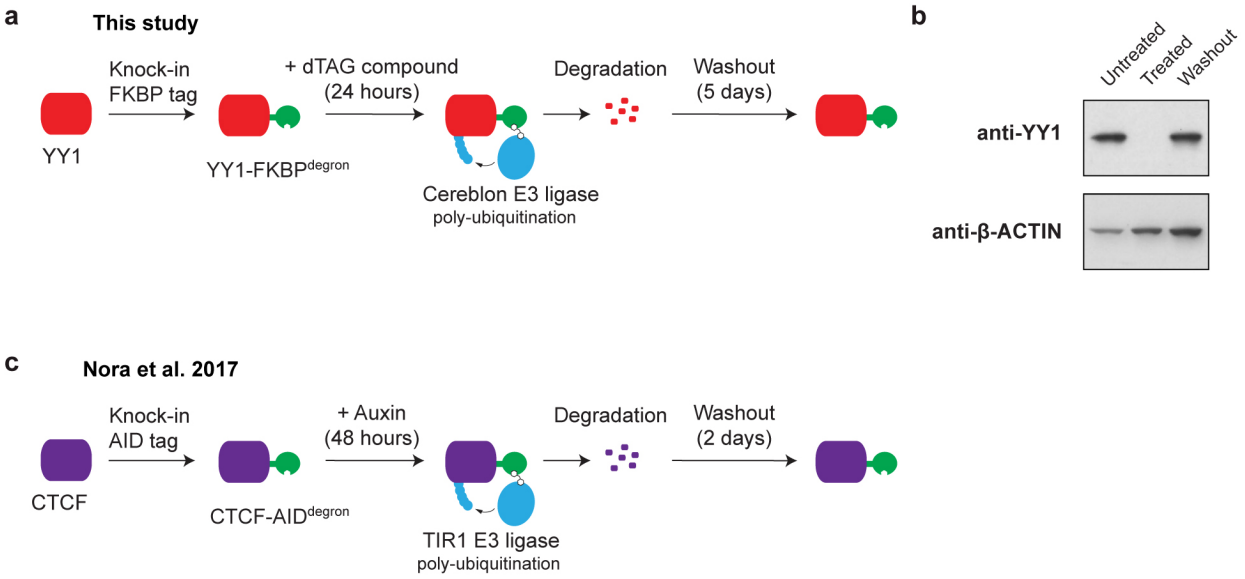


Figure S7. Rescue of enhancer-promoter interactions in cells, related to Figure 7
(a) Model depicting dTAG system used to rapidly degrade YY1 protein. The FKBP degron tag was knocked-in to both alleles of the endogenous *Yy1* gene locus. Addition of dTAG compound results in recruitment of the cereblon E3 ligase to FKBP degron-tagged YY1 protein, resulting in rapid proteasome-mediated degradation. The effects of YY1 degradation were examined 24 hours after treatment with dTAG compound. Washout of the dTAG compound for 5 days allowed recovery of YY1 protein.
(b) Western blot validation of YY1 degradation after 24 hour treatment with dTAG compound and YY1 recovery after 5 day washout of the dTAG compound.
(c) Model depicting AID degradation system used to rapidly degrade CTCF protein in Nora et al. (2017). The AID tag was knocked-in at the endogenous *Ctcf* gene locus. Addition of auxin results in the recruitment of the TIR1 E3 ligase to AID-tagged CTCF protein, resulting in proteasome-mediated degradation. The effects of CTCF degradation were examined 48 hours after treatment with dTAG compound. Washout of auxin for 2 days allowed recovery of CTCF protein.

6. METHODS

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

Cell Lines

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. Cells were negative for mycoplasma (tested every three months).

Cell Culture Conditions

V6.5 murine embryonic stem (mES) cells were grown in serum + LIF on irradiated murine embryonic fibroblasts (MEFs) or in 2i + LIF conditions. For all experiments except for the washout experiment (Fig. 7) cells were grown in serum + LIF on irradiated MEFs and then passaged twice off of MEFs before harvesting. Genome editing was done in 2i + LIF conditions. Cells were always grown on 0.2% gelatinized (Sigma, G1890) tissue culture plates. For the washout experiment (Fig. 7) cells were grown on 2i + LIF.

The media used for general culturing in serum + LIF conditions is as follows: DMEM-KO (Invitrogen, 10829-018) supplemented with 15% fetal bovine serum (Hyclone, characterized SH3007103), 1,000 U/ml LIF (ESGRO, ESG1106), 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 8 ul/mL of 2-mercaptoethanol (Sigma, M7522).

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), 0.5 mM L-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).

Prior to differentiation mESCs were cultured in serum + LIF 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), 0.1mM beta-mercaptoethanol (Sigma Aldrich) and $2x10^6$ units of leukemia inhibitory factor (LIF)

The media used for embryoid body formation (serum - LIF) is 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).

HCT-116 (male) cells were purchased from ATCC (CCL-247) 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 (tested every 3 months).

Jurkat (male) cells were purchased from ATCC (TIB-152) and cultured in RPMI-1640 (Gibco 61870-127) with 10% fetal bovine serum (Hyclone, characterized SH3007103), 100 U/mL Penicillin-Streptomycin (Gibco 15140). Cells were negative for mycoplasma (tested every 3 months).

K562 cells (female) were purchased from ATCC (CCL-243) and cultured in in RPMI-1640 (Gibco 61870-127) with 10% fetal bovine serum (Hyclone, characterized SH3007103), 100 U/mL Penicillin-Streptomycin (Gibco 15140). Cells were negative for mycoplasma (tested every 3 months).

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 (tested every 3 months).

METHOD DETAILS

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.

Recombinant YY1 purification and characterization

YY1 purification

YY1 protein was purified using methods established by the Lee Lab[78] and previously described[47]. A plasmid containing N-terminal His6-tagged human YY1 coding sequence (a gift from Dr. Yang Shi) was transformed into BL21-CodonPlus (DE3)-RIL cells (Stratagene, 230245). A fresh bacterial colony was inoculated into LB media containing ampicillin and chloramphenicol and grown overnight at 37°C. These bacteria were diluted 1:10 in 500 mL pre-warmed LB with ampicillin and chloramphenicol and grown for 1.5 hours at 37°C. After induction of YY1 expression with 1mM IPTG, cells were grown for another 5 hours, collected, and stored frozen at -80°C until ready to use.

Pellets from 500mL cells were resuspended in 15mL of Buffer A (6M GuHCl, 25mM Tris, 100mM NaCl, pH8.0) containing 10mM imidazole, 5mM 2-mercaptoethanol, cOmplete protease inhibitors (Roche, 11873580001) and sonicated (ten cycles of 15 seconds on, 60 seconds off). The lysate was cleared by centrifugation at 12,000g for 30 minutes at 4°C and added to 1mL of Ni-NTA agarose (Invitrogen, R901-15) pre-equilibrated with 10X volumes of Buffer A. Tubes containing this agarose lysate slurry were rotated at room temperature for 1 hour. The slurry was poured into a column, and the packed agarose washed with 15 volumes of Buffer A containing 10mM imidazole. Protein was eluted with 4 X 2 mL Buffer A containing 500mM imidazole.

Fractions were run out by SDS-PAGE gel electrophoresis and stained with Coomassie Brilliant Blue (data not shown). Fractions containing protein of the correct size and high purity were

combined and diluted 1:1 with elution buffer. DTT was added to a final concentration of 100mM and incubated at 60°C for 30 minutes. The protein was refolded by dialysis against 2 changes of 1 Liter of 25mM Tris-HCl pH 8.5, 100mM NaCl, 0.1mM $ZnCl_2$, and 10mM DTT at 4°C followed by 1 change of the same dialysis buffer with 10% glycerol. Protein was stored in aliquots at -80°C.

YY1 characterization

The purity of the recombinant YY1 was assessed by SDS-PAGE gel electrophoresis followed by Coomassie Brilliant Blue staining and Western blotting (Fig. S3a). The activity of the recombinant protein was assessed by EMSA (Fig. S3b).

EMSA was performed using the LightShift Chemiluminescent EMSA Kit (Thermo Scientific #20148) following the manufacturer's recommendations. Briefly, recombinant protein was incubated with a biotinylated probe in the presence or absence of a cold competitor. Reactions were separated using a native gel and transferred to a membrane. Labeled DNA was detected using chemiluminescence.

To generate the biotin labeled probe, 30-nucleotide-long 5' biotinylated single stranded oligonucleotides (IDT) were annealed in 10 mM Tris pH 7.5, 50 mM NaCl, and 1mM EDTA at a 50 uM concentration. The same protocol was used to generate the cold competitor. The probe was serially diluted to a concentration of 10 fmol/µL and cold competitor to a concentration of 2 pmol/µL. 2 µL of diluted probe and cold competitor were used for each binding reaction for a final amount of 20 fmol labeled probe and 4 pmol cold competitor (200 fold excess) in each reaction.

Binding reactions were set-up in a 20 µL volume containing 1x Binding Buffer (10 mM Tris, 50 mM KCl, 1 mM DTT; pH 7.5), 2.5% Glycerol, 5 mM $MgCl_2$, 50 ng/µL Poly dI dC, 0.05% Np-40, 0.1 mM $ZnCl_2$, 10 mM HEPES, and 2 ug of recombinant YY1 protein. Binding reactions were pre-incubated for 20 mins at room temperature with or without the cold competitor. Labeled probe was then added to binding reactions and incubated for 80 minutes at room temperature. After the 80 min incubation 5x Loading Buffer (Thermo Scientific #20148) was added to the reaction and run on a 4-12% TBE gel using 0.5x TBE at 40 mA for 2.5 hrs at 4°C. The TBE gel was pre-run for 1 hr at 4°C. DNA was then electrophoretically transferred to a Biodyne B Nylon Membrane (pre-soaked in cold 0.5x TBE for 10 mins) at 380 mA for 30 mins at 4°C. The DNA was then crosslinked to the membrane by placing the membrane on a Dark Reader Transilluminator for 15 mins. The membrane was allowed to air dry at room temperature overnight and chemiluminescence detected the following day.

Detection of biotin-labeled DNA was done as follows. The membrane was blocked for 20 mins using Blocking Buffer (Thermo Scientific 20148). The membrane was then incubated in conjugate/blocking buffer (Thermo Scientific 20148) for 15 mins. The membrane was then washed four times with 1x Wash Buffer (Thermo Scientific 20148) for 5 mins. The membrane was then incubated in Substrate Equilibration Buffer (Thermo Scientific 20148) for 5 mins and then incubated in Substrate Working Solution (Thermo Scientific 20148) for 5 mins. The membrane was then imaged using a CCD camera using a 120 second exposure. All of these steps were performed at room temperature.

Genome Editing

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 oligos used for the cloning are included in Table S3. The sequences of the DNA targeted (the protospacer adjacent motif is underlined) are listed below:

Locus	Targeted DNA
Raf1_promoter	5'-ACTCCCGCCATCCAAGATGGCGG-3'
Etv4_promoter	5'-GAGCTACTTGAAAACAAATGGAGG-3'
YY1_stop_codon	5'-GTCTTCTCTCTTCTTTTCACTGG-3'

For the motif deletions, five hundred thousand mES cells were transfected with 2.5 µg plasmid and sorted 48 hours later for the presence of GFP. Thirty thousand GFP-positive sorted cells were plated in a six-well plate in a 1:2 serial dilution (first well 15,000 cells, second well 7,500 cells, etc.). The cells were grown for approximately one week in 2i + LIF. Individual colonies were picked using a stereoscope into a 96-well plate. Cells were expanded and genotyped by PCR and Sanger sequencing. Clones with deletions spanning the motif were further expanded and used for experiments.

For the generation of the endogenously tagged lines, five hundred thousand mES cells were transfected with 2.5 ug Cas9 plasmid and 1.25 ug non-linearized repair plasmid 1 (pAW62.YY1.FKBP.knock-in.mCherry) and 1.25 ug non-linearized repair plasmid 2 (pAW63.YY1.FKBP.knock-in.BFP). 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. Thirty thousand mCherry+/BFP+ sorted cells were plated in a six-well plate in a 1:2 serial dilution (first well 15,000 cells, second well 7,500 cells, etc). 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 (YY1_gPCR_3F/3R, Table S3). Clones with a homozygous knock-in tag were further expanded and used for experiments.

Chromatin Immunoprecipitation (ChIP)

ChIP was performed as previously described[79] with a few adaptations. mES cells were depleted of MEFs by splitting twice onto newly gelatinized plates without MEFs. Approximately 50 million mES cells were crosslinked for 15 minutes at room temperature by the addition of one-tenth volume of fresh 11% formaldehyde solution (11% formaldehyde, 50 mM HEPES pH 7.3, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0) to the growth media followed by 5 min quenching with 125 mM glycine. Cells were rinsed twice with 1X PBS and harvested using a silicon scraper and flash frozen in liquid nitrogen. Jurkat cells were crosslinked for 10 minutes in media at a concentration of 1 million cells/mL. Frozen crosslinked cells were stored at −80°C. 100µl of Protein G Dynabeads (Life Technologies 10009D) were washed 3X for 5 minutes with 0.5% BSA (w/v) in PBS. Magnetic beads were bound with 10 µg of anti-YY1 antibody (Santa Cruz, sc-281X) overnight at 4°C, and then washed 3X with 0.5% BSA (w/v) in PBS.

Cells were prepared for ChIP as follows. All buffers contained freshly prepared 1× 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, 1× protease inhibitors) and rotated for 10

minutes at 4°C, then spun at 1350 rcf for 5 minutes at 4°C. The pellet was resuspended in lysis buffer II (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× protease inhibitors) and rotated for 10 minutes at 4°C and spun at 1350 rcf for 5 minutes at 4°C. The pellet was resuspended in sonication buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 0.1% SDS, and 1% Triton X-100, 1× protease inhibitors) and then sonicated on a Misonix 3000 sonicator for 10 cycles at 30 seconds each on ice (18-21 W) with 60 seconds on ice between cycles. Sonicated lysates were cleared once by centrifugation at 16,000 rcf for 10 minutes at 4°C. 50 µL was reserved for input, and then the remainder was incubated overnight at 4°C with magnetic beads bound with antibody 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 200 µL 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). Values displayed in the figures were normalized to the input, a negative control region, and wild-type values according to the following formulas:

Input norm = $2^{(Ct\_input - Ct\_ChIP)}$

Neg norm = $\frac{Fold_{ROI}}{Fold_{neg}}$

WT norm = $\frac{Neg\ norm_{mut}}{Neg\ norm_{WT}}$

qPCRs were performed in technical triplicate, and ChIPs were performed in biological triplicate. Values were comparable across replicates. The average WT norm values and standard deviation are displayed (Fig. 4a, b). The primers used are listed in Table S3.

For ChIP-seq experiments, purified ChIP DNA was used to prepare Illumina multiplexed sequencing libraries. Libraries for Illumina sequencing were prepared following the Illumina TruSeq DNA Sample Preparation v2 kit. Amplified libraries were size-selected using a 2% gel cassette in the Pippin Prep system from Sage Science set to capture fragments between 200 and 400 bp. Libraries were quantified by qPCR using the KAPA Biosystems Illumina Library Quantification kit according to kit protocols. Libraries were sequenced on the Illumina HiSeq 2500 for 40 bases in single read mode.

ChIA-PET

ChIA-PET was performed using a modified version[36] of a previously described protocol[46]. mES cells (~ 500 million cells, grown to ~80% confluency) were crosslinked with 1% formaldehyde at room temperature for 15 min and then neutralized with 125mM glycine. Crosslinked cells were washed three times with ice-cold PBS, snap-frozen in liquid nitrogen, and stored at -80°C before further processing. Nuclei were isolated as previously described above, and chromatin was fragmented using a Misonix 3000 sonicator. Either CTCF or YY1 antibodies were used to enrich protein-bound chromatin fragments exactly as described in the ChIP-seq section. A portion of ChIP DNA was eluted from antibody-coated beads for concentration quantification and for enrichment analysis using qPCR. For ChIA-PET library construction ChIP DNA fragments were end-repaired using T4 DNA polymerase (NEB # M0203) followed by A-tailing with Klenow (NEB M0212). Bridge linker oligos (Table S5) were annealed to generate a double stranded bridge linker with T-overhangs. 800 ng of bridge linker was added and the proximity ligation was performed overnight at 16°C in 1.5 mL volume. Unligated DNA was then digested with exonuclease and lambda nuclease (NEB M0262S, M0293S). DNA was eluted off the beads by incubation at 65°C for 1 hour with intermittent vortexing in 200 µL 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. Precipitated DNA was resuspended in Nextera DNA resuspension buffer (Illumina FC-121-1030). The DNA was then tagmented with the Nextera Tagmentation kit (Illumina FC-121-1030). 5 µL of transposon was used per 50 ng of DNA. The tagmented library was purified with a Zymo DNA Clean & Concentrator (Zymo D4003) and bound to streptavidin beads (Life Technologies #11205D) to enrich for ligation junctions (containing the biotinylated bridge linker). 12 cycles of the polymerase chain reaction were performed to amplify the library using standard Nextera primers (Illumina FC-121-1030). The amplified library was size-selected (350-500 bp) and sequenced using paired-end sequencing on an Illumina Hi-Seq 2500 platform

HiChIP

HiChIP was performed as described[48] with a few modifications. Ten million cells cross-linked for 10 min at room temperature with 1% formaldehyde in growth media and quenched in 0.125 M glycine. After washing twice with ice-cold PBS, the supernatant was aspirated and the cell pellet was flash frozen in liquid nitrogen and stored at −80°C.

Cross-linked cell pellets were thawed on ice, resuspended in 800 µL of ice-cold Hi-C lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 0.2% IGEPAL CA-630 with 1x cOmplete protease inhibitor (Roche, 11697498001)), and incubated at 4°C for 30 minutes with rotation. Nuclei were pelleted by centrifugation at 2500 rcf for 5 min at 4°C and washed once with 500 µL of ice-cold Hi-C lysis buffer. After removing supernatant, nuclei were resuspended in 100 µL of 0.5% SDS and incubated at 62°C for 10 minutes. SDS was quenched by adding 335 µL of 1.5% Triton X-100 and incubating for 15 minutes at 37°C. After the addition of 50 µL of 10X NEB Buffer 2 (NEB, B7002) and 375 U of MboI restriction enzyme (NEB, R0147), chromatin was digested at 37°C for 2 hours with rotation. Following digestion, MboI enzyme was heat inactivated by incubating the nuclei at 62°C for 20 min.

To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 52 µL of fill-in master mix, containing 37.5 µL of 0.4 mM biotin-dATP (Invitrogen, 19524016), 1.5 µL of 10 mM dCTP (Invitrogen, 18253013), 1.5 µL of 10 mM dGTP (Invitrogen, 18254011), 1.5 µL of 10 mM dTTP (Invitrogen, 18255018), and 10 µL of 5 U/µL DNA Polymerase I, Large (Klenow) Fragment (NEB, M0210), was added and the tubes were incubated at 37ºC for 1 hour with rotation. Proximity ligation was performed by addition of 947 µL of ligation master mix, containing 150 µL of 10X NEB T4 DNA ligase buffer (NEB, B0202), 125 µL of 10% Triton X-100, 7.5 µL of 20 mg/mL BSA (NEB, B9000), 10 µL of 400 U/µL T4 DNA ligase (NEB, M0202), and 655.5 µL of water, and incubation at room temperature for 4 hours with rotation.

After proximity ligation, nuclei were pelleted by centrifugation at 2500 rcf for 5 minutes and resuspended in 1 mL of ChIP sonication buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate, and 0.1% SDS with protease inhibitor). Nuclei were sonicated using a Covaris S220 for 6 minutes with the following settings: fill level 8, duty cycle 5, peak incidence power 140, cycles per burst 200. Sonicated chromatin was clarified by centrifugation at 16,100 rcf for 15 min at 4°C and supernatant was transferred to a tube. 60 µL of protein G magnetic beads were washed three times with sonication buffer, resuspended in 50 µL of sonication buffer. Washed beads were then added to the sonicated chromatin and incubated for 1 hour at 4°C with rotation. Beads were then separated on a magnetic stand and the supernatant was transferred to a new tube. 7.5 µg of H3K27ac antibody (Abcam, ab4729) or 7.5 ug of YY1 antibody (Abcam, ab109237) was added to the tube and the tube was incubated overnight at 4°C with rotation. For YY1 six reactions were carried out and pooled prior to tagmentation. The next day, 60 µL of protein G magnetic beads were washed three time in 0.5% BSA in PBS and washed once with sonication buffer before being resuspended in 100 µL of sonication buffer and added to each sample tube. Samples were incubated for 2 hours at 4ºC with rotation. Beads were then separated on a magnetic stand and washed three times with 1 mL of high salt sonication buffer (50 mM HEPES-KOH pH 7.5, 500 mM NaCl, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS) followed by three times with 1 mL of LiCl wash buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS) and once with 1 mL of TE with salt (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 50 mM NaCl). Beads were then resuspended in 200 µL of elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS) and incubated at 65°C for 15 minutes to elute. To purify eluted DNA, 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 45 minutes. Samples were then incubated at 65°C for 5 hours to reverse cross-links. DNA was then purified using Zymo DNA Clean and Concentrate 5 columns (Zymo, D4013) according to manufacturer's protocol and eluted in 14 µL water. The amount of eluted DNA was quantified by Qubit dsDNA HS kit (Invitrogen, Q32854).

Tagmentation of ChIP DNA was performed using the Illumina Nextera DNA Library Prep Kit (Illumina, FC-121-1030). First, 5 µL of streptavidin C1 magnetic beads (Invitrogen, 65001) was washed with 1 mL of tween wash buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA pH 8.0, 1 M NaCl, 0.05% Tween-20) and resuspended in 10 µL of 2X biotin binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, 2 M NaCl). 54.19 ng purified DNA was added in a total volume of 10 µL of water to the beads and incubated at room temperature for 15 minutes with agitation every 5 minutes. After capture, beads were separated with a magnet and the supernatant was discarded. Beads were then washed twice with 500 µL of tween wash buffer, incubating at 55°C for 2 minutes with shaking for each wash. Beads were resuspended in 25 µL of Nextera Tagment DNA buffer. To tagment the captured DNA, 3.5 µL of Nextera Tagment DNA Enzyme

1 was added with 21.5 µL of Nextera Resuspension Buffer and samples were incubated at 55°C for 10 minutes with shaking. Beads were separated on a magnet and supernatant was discarded. Beads were washed with 500 µL of 50 mM EDTA at 50°C for 30 minutes, then washed three times with 500 µL of tween wash buffer at 55°C for 2 minutes each, and finally washed once with 500 µL of 10 mM Tris-HCl pH 7.5 for 1 minute at room temperature. Beads were separated on a magnet and supernatant was discarded.

To generate the sequencing library, PCR amplification of the tagmented DNA was performed while the DNA is still bound to the beads. Beads were resuspended in 15 µL of Nextera PCR Master Mix, 5 µL of Nextera PCR Primer Cocktail, 5 µL of Nextera Index Primer 1, 5 µL of Nextera Index Primer 2, and 20 µL of water. DNA was amplified with 8 cycles of PCR. After PCR, beads were separated on a magnet and the supernatant containing the PCR amplified library was transferred to a new tube, purified using the Zymo DNA Clean and Concentrate-5 (Zymo D4003T) kit according to manufacturer's protocol, and eluted in 14 µL water. Purified HiChIP libraries were size selected to 300-700 bp using a Sage Science Pippin Prep instrument according to manufacturer's protocol and subject to paired-end sequencing on an Illumina HiSeq 2500. Libraries were initially sequenced with 100x100 bp paired-end sequencing. A second round of sequencing was done on the same libraries with 50x50 bp paired-end sequencing.

4C-seq

A modified version of 4C-seq[80,81] was developed. The major change was the proximity 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[35,82].

Approximately 5 million mES cells were trypsinized and then resuspended in 5 mL 10% FBS/PBS. 5 mL of 4% formaldehyde in 10% FBS/PBS was added and cells were crosslinked for 10 minutes. Glycine was added to a final concentration of 0.125 M and cells were centrifuged at 300 rcf for 5 minutes. Cells were washed twice with PBS, transferred to a 1.5 mL Eppendorf tube, snap frozen and stored at -80.

Pellets were gently resuspended in Hi-C lysis buffer (10 mM Tris-HCl pH 8, 10 mM NaCl, 0.2% Igepal) with 1x cOmplete protease inhibitors (Roche 11697498001). Cells were incubated on ice for 30 minutes then washed once with 500 µL of ice-cold Hi-C lysis buffer with no protease inhibitors. Pellets were resuspended in 50 µL of 0.5% SDS and incubated at 62°C for 7 minutes. 145 µL of H2O and 25 µL of 10% Triton X-100 were added and tubes incubated at 37°C for 15 minutes. 25 µL of the appropriate 10X New England Biolabs restriction enzyme buffer and 200 units of enzyme were added and the chromatin was incubated at 37°C degrees in a thermomixer at 500 RPM for four hours, 200 more units of enzyme was added and the reaction was incubated overnight at 37°C degrees in a thermomixer at 500 RPM, then 200 more units were added and the reaction was incubated another four hours at 37°C degrees in a thermomixer at 500 RPM. DpnII (NEB) was used as the primary cutter for both *Raf1* and *Etv4*. Restriction enzyme was inactivated by heating to 62°C for 20 minutes while shaking at 500 rpm. Proximity ligation was performed in a total of 1200 µL with 2000 units of T4 DNA ligase (NEB M020) for six hours at room temperature. After ligation samples were spun down for 5 minutes at 2500 rcf and resuspended in 300 µL 10 mM Tris-HCl, 1% SDS and 0.5 mM NaCl with 1000 units of Proteinase K. Crosslinks were reversed by incubation overnight at 65°C.

Samples were then phenol-chloroform extracted and ethanol precipitated and the second digestion was performed overnight in 450 µL with 50 units of restriction enzyme. BfaI (NEB R0568S) was used for *Etv4* and CviQI (NEB R0639S) was used for *Raf1*. Samples were phenol-chloroform extracted and ethanol precipitated and the second ligation was performed in 14 mL total with 6700 units of T4 DNA ligase (NEB M020) at 16°C overnight. Samples were ethanol precipitated, resuspended in 500 µL Qiagen EB buffer, and purified with a Qiagen PCR purification kit.

PCR amplification was performed with 16 50 µL PCR reactions using Roche Expand Long Template polymerase (Roche 11759060001). Reaction conditions are as follows: 11.2 µL Roche Expand Long Template Polymerase, 80 µL of 10 X Roche Buffer 1, 16 µL of 10 mM dNTPs (Promega PAU1515), 112 µL of 10 µM forward primer, 112 µL of 10 µM reverse primer (Table S5), 200 ng template, and milli-q water until 800 µL total. Reactions were mixed and then distributed into 16 50 µL reactions for amplification. Cycling conditions were a "Touchdown PCR" based on reports that this decreases non-specific amplification of 4C libraries[83]. The conditions are: 2' 94°C, 10'' 94°C, 1' 63°C, 3' 68 °C, repeat steps 2-4 but decrease annealing temperature by one degree, until 53°C is reached at which point the reaction is cycled an additional 15 times at 53°C, after 25 total cycles are performed the reaction is held for 5' at 68°C and then 4°C. Libraries were cleaned-up using a Roche PCR purification kit (Roche 11732676001) using 4 columns per library. Reactions were then further purified with Ampure XP beads (Agencourt A63882) with a 1:1 ratio of bead solution to library following the manufactures instructions. Samples were then quantified with Qubit and the KAPA Biosystems Illumina Library Quantification kit according to kit protocols. Libraries were sequenced on the Illumina HiSeq 2500 for 40 bases in single read mode.

RNA isolation, RT-qPCR, and sequencing

RNA was isolated using the RNeasy Plus Mini Kit (QIAGEN, 74136) according to manufacturer's instructions.

For RT-qPCR assays, reverse transcription was performed using SuperScript III Reverse Transcriptase (Invitrogen, 18080093) with oligo-dT primers (Promega, C1101) according to manufacturers' instructions. Quantitative real-time PCR was performed on Applied Biosystems 7000, QuantStudio 5, and QuantStudio 6 instruments using TaqMan probes for *Raf1* (Applied Biosystems, Mm00466513_m1) and *Etv4* (Applied Biosystems, Mm00476696_m1) in conjunction with TaqMan Universal PCR Master Mix (Applied Biosystems, 4304437) according to manufacturer's instructions.

For RNA-seq experiments, stranded polyA selected libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina, RS-122-2101) according to manufacturer's standard protocol. Libraries were subject to 40 bp single end sequencing on an Illumina HiSeq 2500 instrument.

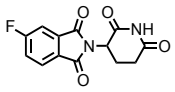
YY1 degradation

A clonal homozygous knock-in line expressing FKBP tagged YY1 was used for the degradation experiments. Cells were grown two passages off MEFs and then treated with dTAG-47 at a concentration of 500 nM for 24 hours.

dTAG-47 Washout Experiments

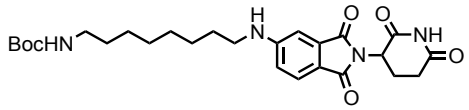
The homozygous knock-in line expressing FKBP tagged YY1 was cultured on 2i + LIF media. Cells were treated with dTAG-47 at a concentration of 500 nM for 24 hours. After 24 hours of drug treatment, cells were washed three times with PBS and passaged onto a new plate. Cells were then fed daily and passaged onto a new plate every 48 hours until YY1 protein levels were restored (5 days after drug withdrawal). Cells were then harvested for protein or RNA extraction or cross-linked for ChIP or HiChIP.

dTAG-47 synthesis



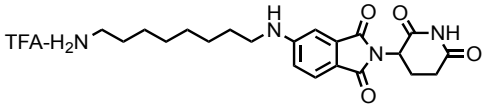
2-(2,6-dioxopiperidin-3-yl)-5-fluoroisoindoline-1,3-dione
4-fluorophthalic anhydride (3.32 g, 20 mmol, 1 eq) and 3-aminopiperidine-2,6-dione hydrochloride salt (3.620 g, 22 mmol, 1.1 eq) were dissolved in AcOH (50 mL) followed by potassium acetate (6.08 g, 62 mmol, 3.1 eq). The mixture was fitted with an air condenser and heated to 90 °C. After 16 hours, the mixture was diluted with 200 mL water and cooled over ice. The slurry was then centrifuged (4000 rpm, 20 minutes, 4 °C) and decanted. The remaining solid was then resuspended in water, centrifuged and decanted again. The solid was then dissolved in MeOH and filtered through a silica plug (that had been pre-wetted with MeOH), washed with 50% MeOH/DCM and concentrated under reduced pressure to yield the desired product as a grey solid (2.1883 g, 7.92 mmol, 40%).
^{1}H NMR (500 MHz, DMSO-d_6) δ 11.13 (s, 1H), 8.01 (dd, *J* = 8.3, 4.5 Hz, 1H), 7.85 (dd, *J* = 7.4, 2.2 Hz, 1H), 7.72 (ddd, *J* = 9.4, 8.4, 2.3 Hz, 1H), 5.16 (dd, *J* = 12.9, 5.4 Hz, 1H), 2.89 (ddd, *J* = 17.2, 13.9, 5.5 Hz, 1H), 2.65 – 2.51 (m, 2H), 2.07 (dtd, *J* = 12.9, 5.3, 2.2 Hz, 1H).

LCMS 277.22 (M+H).

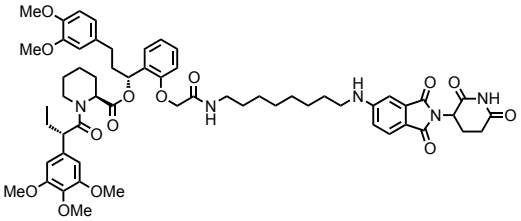


tert-butyl (8-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-5-yl)amino)octyl)carbamate
2-(2,6-dioxopiperidin-3-yl)-5-fluoroisoindoline-1,3-dione (294 mg, 1.06 mmol, 1 eq) and *tert*-butyl (8-aminooctyl)carbamate (286 mg, 1.17 mmol, 1.1 eq) were dissolved in NMP (5.3 mL, 0.2M). DIPEA (369 µL, 2.12 mmol, 2 eq) was added and the mixture was heated to 90 °C. After 19 hours, the mixture was diluted with ethyl acetate and washed with water and three times with brine. The organic layer was dried over sodium sulfate, filtered and concentrated under reduced pressure. Purification by column chromatography (ISCO, 12 g column, 0-10% MeOH/DCM, 30 minute gradient) gave the desired product as a brown solid (0.28 g, 0.668 mmol, 63%).
^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.12 (s, 1H), 7.62 (d, *J* = 8.3 Hz, 1H), 7.02 (s, 1H), 6.81 (d, *J* = 7.2 Hz, 1H), 4.93 (dd, *J* = 12.3, 5.3 Hz, 1H), 4.51 (s, 1H), 3.21 (t, *J* = 7.2 Hz, 2H), 3.09 (d, *J* = 6.4 Hz, 2H), 2.90 (dd, *J* = 18.3, 15.3 Hz, 1H), 2.82 – 2.68 (m, 2H), 2.16 – 2.08 (m, 1H), 1.66 (p, *J* = 7.2 Hz, 2H), 1.37 (d, *J* = 62.3 Hz, 20H).

LCMS 501.41 (M+H).



5-((8-aminooctyl)amino)-2-(2,6-dioxopiperidin-3-yl)isoindoline-1,3-dione trifluoroacetate
tert-butyl (8-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-5-yl)amino)octyl)carbamate (334.5 g, 0.668 mmol, 1 eq) was dissolved in TFA (6.7 mL) and heated to 50 °C. After 1 hour, the mixture was cooled to room temperature, diluted with DCM and concentrated under reduced pressure. The crude material was triturated with diethyl ether and dried under vacuum to give a dark yellow foam (253.1 mg, 0.492 mmol, 74%).
^{1}H NMR (500 MHz, Methanol-d_4) δ 7.56 (d, *J* = 8.4 Hz, 1H), 6.97 (d, *J* = 2.1 Hz, 1H), 6.83 (dd, *J* = 8.4, 2.2 Hz, 1H), 5.04 (dd, *J* = 12.6, 5.5 Hz, 1H), 3.22 (t, *J* = 7.1 Hz, 2H), 2.94 – 2.88 (m, 2H), 2.85 – 2.68 (m, 3H), 2.09 (ddd, *J* = 10.4, 5.4, 3.0 Hz, 1H), 1.70 – 1.61 (m, 4H), 1.43 (d, *J* = 19.0 Hz, 8H).
LCMS 401.36 (M+H).



(2*S*)-(1*R*)-3-(3,4-dimethoxyphenyl)-1-(2-(2-((8-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-5-yl)amino)octyl)amino)-2-oxoethoxy)phenyl)propyl 1-((*S*)-2-(3,4,5-trimethoxyphenyl)butanoyl)piperidine-2-carboxylate (dTAG47)
5-((8-aminooctyl)amino)-2-(2,6-dioxopiperidin-3-yl)isoindoline-1,3-dione trifluoroacetate salt (10.3 mg, 0.020 mmol, 1 eq) was added to 2-(2-((*R*)-3-(3,4-dimethoxyphenyl)-1-(((*S*)-1-((*S*)-2-(3,4,5-trimethoxyphenyl)butanoyl)piperidine-2-carbonyl)oxy)propyl)phenoxy)acetic acid (13.9 mg, 0.020 mmol, 1 eq) as a 0.1 M solution in DMF (200 microliters) at room temperature. DIPEA (10.5 microliters, 0.060 mmol, 3 eq) and HATU (7.6 mg, 0.020 mmol, 1 eq) were then added. After 29.5 hours, the mixture was diluted with EtOAc, and washed with 10% citric acid (aq), brine, saturated sodium bicarbonate, water and brine. The organic layer was dried over sodium sulfate, filtered and condensed. Purification by column chromatography (ISCO, 4 g silica column, 0-10% MeOH/DCM, 25 minute gradient) gave the desired product as a yellow solid (14.1 mg, 0.0131 mmol, 65%).
^{1}H NMR (500 MHz, Methanol-d_4) δ 7.55 (d, *J* = 8.4 Hz, 1H), 7.26 – 7.20 (m, 1H), 6.99 – 6.93 (m, 1H), 6.89 (t, *J* = 7.7 Hz, 2H), 6.82 (dd, *J* = 8.4, 2.3 Hz, 2H), 6.77 (d, *J* = 7.5 Hz, 1H), 6.74 (d, *J* = 1.9 Hz, 1H), 6.63 (d, *J* = 9.6 Hz, 2H), 6.12 (dd, *J* = 8.1, 6.0 Hz, 1H), 5.40 (d, *J* = 4.3 Hz, 1H), 5.03 (dd, *J* = 13.1, 5.5 Hz, 1H), 4.57 (d, *J* = 14.9 Hz, 1H), 4.46 – 4.39 (m, 1H), 4.11 (d, *J* = 13.6 Hz, 1H), 3.86 (t, *J* = 7.3 Hz, 1H), 3.80 – 3.76 (m, 7H), 3.71 – 3.65 (m, 8H), 3.14 (ddt, *J* = 17.2, 13.3, 7.1 Hz, 4H), 2.90 – 2.80 (m, 1H), 2.77 – 2.40 (m, 6H), 2.24 (d, *J* = 13.8 Hz, 1H), 2.12 – 1.97 (m, 3H), 1.92 (dq, *J* = 14.0, 7.8 Hz, 1H), 1.67 (ddt, *J* = 54.1, 14.7, 7.1 Hz, 5H), 1.50 (dd, *J* = 46.1, 14.1 Hz, 3H), 1.38 (dt, *J* = 14.5, 7.1 Hz, 4H), 1.28 – 1.17 (m, 6H), 0.87 (t, *J* = 7.3 Hz, 3H).
^{13}C NMR (126 MHz, MeOD) δ 174.78, 174.69, 172.53, 171.71, 170.50, 169.66, 169.31, 156.22, 155.41, 154.62, 150.36, 148.83, 138.05, 136.90, 136.00, 134.93, 130.54, 128.40, 126.21, 123.14, 121.82, 117.94, 116.62, 113.58, 113.05, 112.73, 106.59, 70.69, 68.05, 61.06, 56.59, 56.51, 56.45, 53.42, 50.99, 50.31, 45.01, 44.09, 40.07, 37.44, 32.22, 32.17, 30.38, 30.32, 30.18, 29.84, 29.32, 28.05, 27.80, 27.58, 26.38, 23.87, 21.95, 12.57.
LCMS: 1077.35 (M+H)

In vitro DNA circularization assay

First, two plasmids (pAW49, pAW79) were generated. pAW49 contains YY1 binding sites separated by ~3.5 kb of intervening DNA. pAW79 is identical except it contains filler DNA instead of the YY1 motifs. The intervening DNA was chosen based on looking at YY1 ChIP-seq and motif distribution in mES cells to identify regions that lacked YY1 occupancy and YY1 binding motifs. The YY1 binding motifs were chosen based on successful EMSAs[47]. Approximately 200 bp of sequence was added between the binding motifs and the termini in order to provide flexibility for the termini to ligate. The plasmid was built using Gibson assembly.

Next, a PCR was run using plasmid as a template to generate a linear piece of DNA (Table S5). This PCR product was PCR purified (Qiagen 28104) and then digested with BamHI (NEB R3136) and PCR purified. The BamHI digested template was used in the ligation assay.

The ligation assay was carried out as follows. Reactions were prepared on ice in 66 µL with the following components:
BSA control: 0.25 nM DNA, 1x T4 DNA ligase buffer (NEB B0202S), H_2O, 0.12 µg/µL of BSA
YY1: 0.25 nM DNA, 1x T4 DNA ligase buffer (NEB B0202S), H_2O, 0.12 µg/µL of YY1
YY1 + competitor: 0.25 nM DNA, 1x T4 DNA ligase buffer (NEB B0202S), H2O, 0.12 µg/µL of YY1, 100 nM competitor DNA (Table S5)

Assuming an extinction coefficient for YY1 of 19940 M^{-1} cm^{-1} and 75% purity, that gives an approximate YY1 molar concentration of ~ 3 uM.

Reactions were incubated at 20°C for 20 minutes to allow binding of YY1 to the DNA. For each timepoint 6 µL of the reaction was withdrawn and quenched in a total volume of 9 µL with a final concentration of 30 mM EDTA, 1x NEB loading dye (NEB, B7024S), 1 ug/µL of proteinase K, and heated at 65°C for 5 minutes. Timepoint 0 was taken and then 600 units of T4 DNA ligase (NEB M0202) was added and the reaction was carried out at 20°C. Indicated timepoints were taken and then samples were run on a 4-20% TBE gradient gel for three hours at 120 V. The gel was stained with SYBR Gold (Life Technologies S11494) and imaged with a CCD camera.

Quantification was done using Image Lab version 5.2.1 (Bio-Rad Laboratories). First, band density of the starting product and ligation product were measured. Then the percent circularized was calculated: (ligation product)/(ligation product + starting band)*100. In Fig. 3, to facilitate visualization overexposed gels are shown. For the quantification exposures were used that did not have any overexposed pixels.

Co-immunoprecipitation

V6.5 mESCs were transfected with pcDNA3_FLAG_YY1 and pcDNA3_FLAG_HA using Lipofectamine 3000 (Life Technologies #L3000001) according to the manufacturer's instructions. Briefly, cells were split and 8 million cells were plated onto a gelatinized 15 cm plate. 7.5 µg of each plasmid was mixed with 30 µL P3000 reagent and 75 µL Lipofectamine 3000 reagent (Life Technologies #L3000001) in 1250 µL of DMEM (Life technologies #11995-073). After ~12-16 hours media was changed.

Cells were harvested 48 hours after transfection by washing twice with ice-cold PBS and collected by scraping in ice-cold PBS. Harvested cells were centrifuged at 1,000 rcf for 3 minutes to pellet cells. Supernatant was discarded and cell pellets were flash frozen and stored

at -80°C until ready to prepare nuclear extract. For each 15 cm plate of cells, frozen cell pellets were resuspended in 5 mL of ice-cold hypotonic lysis buffer (20 mM HEPES-KOH pH 7.5, 20% glycerol, 10 mM NaCl, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.5 mM DTT and protease inhibitor (Roche, 11697498001)) and incubated on ice for 10 minutes to extract nuclei. Nuclei were pelleted by centrifugation at 14,000 rcf for 10 minutes at 4°C. Supernatant was discarded and nuclei were resuspended in 0.5 mL of ice-cold nuclear extraction buffer (20 mM HEPES-KOH pH 7.5, 20% glycerol, 250 mM NaCl, 0.1% Triton X-100, 1.5 mM $MgCl_2$ and protease inhibitor) and incubated for 1 hour at 4°C with rotation. Lysates were clarified by centrifugation at 14,000 rcf for 10 minutes at 4°C. Nuclear extract, supernatant, was transferred to a new tube and diluted with 1 mL of ice-cold dilution buffer (20 mM HEPES-KOH pH 7.5, 10% glycerol, 100 mM NaCl, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT and protease inhibitor). Protein concentration of extracts was quantified by BCA assay (Thermo Scientific, 23225) and protein concentration was adjusted to 400 µg/mL by addition of appropriate volume of 1:2 nuclear extraction buffer:dilution buffer. For RNase A-treated nuclear extract experiments, 250 µL of nuclear extract (100 µg) was treated by addition of 7.5 µL of 33 mg/mL RNase A (Sigma, R4642) or 18.75 µL of 20 U/µL SUPERase In RNase Inhibitor (Invitrogen, AM2696) followed by incubation at 37°C for 10 minutes. For all experiments, an aliquot of extract was saved and stored at -80°C for use as an input sample after immunoprecipitation.

To prepare beads for immunoprecipitation of FLAG-tagged and HA-tagged YY1 from nuclear extract, 50 µL of protein G magnetic beads per immunoprecipitation was washed three times with 1 mL of blocking buffer (0.5% BSA in PBS), rotating for 5 minutes at 4°C for each wash. After separation on a magnet, beads were resuspended in 250 µL of blocking buffer. After addition of 5 µg of anti-FLAG (Sigma, F7425)), anti-HA (Abcam, ab9110), or normal IgG (Millipore, 12-370) antibody, beads were allowed to incubate for at least 1 hour at 4°C with rotation to bind antibody. After incubation, beads were washed three times with 1 mL of blocking buffer, rotating for 5 minutes at 4°C for each wash.

Washed beads were separated on a magnet and the supernatant was discarded before resuspending in 250 µL of nuclear extract (100 µg). Beads were allowed to incubate with extract overnight at 4°C with rotation. The following morning, beads were washed five times with 1 mL of ice-cold wash buffer, rotating for 5 minutes at 4°C for each wash. Washed beads were resuspended in 100 µL of 1X XT sample buffer (Biorad, 1610791) with 100 mM DTT and incubated at 95°C for 10 min. Beads were separated on a magnet and supernatant containing immunoprecipitated material was transferred to a new tube.

To assay immunoprecipitation results by western blot, 10 µL of each samples was run on a 4-20% Bis-Tris gel (Bio-rad, 3450124) using XT MOPS running buffer (Bio-rad, 1610788) at 80 V for 20 minutes, followed by 150 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, 20% 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, shaking. Membrane was then incubated with 1:50,000 anti-FLAG-HRP (Sigma, A8592), 1:25:000 anti-HA-HRP (Cell Signaling, 2999), or anti-OCT3/4 (C-10, Santa Cruz sc-5279) 1:2000 antibody 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 min at room temperature shaking for each wash. Membranes were developed with ECL substrate (Thermo Scientific, 34080) and imaged using a CCD camera or exposed using film.

Embryoid Body Formation

Prior to differentiation, YY1-FKBP tagged knock-in mESCs were cultured in serum + LIF on irradiated MEFs. Starting 48 hours prior to the differentiation and continuing throughout the entire experiment the $YY1^-$ condition were exposed to 500 nM dTAG-47. 4,000 cells (either $YY1^-$ or $YY1^+$) were then plated into each well of a 96-well plate (Nunclon Sphera, ThermoFisher) in Embryoid Body formation media (serum - LIF). Three plates were generated for each condition. The EBs were cultured in 96-well plates for 4 days and then pooled and cultured in ultra-low attachment culture plates (Costar, Corning). After three days, cells were harvested for single-cell RNA-seq (day 7 of differentiation). Cells were harvested for single-cell RNA-seq by dissociation with Accutase for 30 minutes at 37°C. The cells were then resuspended in PBS with 0.04% BSA and then prepared for sequencing (see section on single-cell RNA-seq). Immunohistochemistry was performed after four days (day 8 of differentiation).

Immunohistochemistry

Cells were fixed in 4% paraformaldehyde in PBS and embedded in paraffin. Cells were sectioned and stained according to standard protocols using TUJI (Biolegend 801201, 1:1000), GFAP (Dako Z0344, 1:200), and Gata-4 (Abcam ab84593 1:100) primary antibodies and appropriate Alexa Fluor dye conjugated secondary antibodies (1:1000, ThermoFisher) and DAPI. Slides were mounted with Fluoro-mount G (Electron Microscopy Science) and imaged using a Zeiss LSM 710 laser scanning confocal microscope. In all images scale bars are 50 μm.

Single-cell RNA-seq library preparation

Single-cell RNA-seq libraries were prepared using the Chromium Controller (10X Genomics). Briefly, single cells in 0.04% BSA in PBS were separated into droplets and then reverse transcription and library construction was performed according to the 10X Chromium Single Cell 3' Reagent Kit User Guide and sequenced on an Illumina Hi-seq 2500.

dCas9-YY1 tethering

First two lentiviral constructs were generated by modifying lenti dCAS-VP64_Blast (lenti dCAS-VP64_Blast was a gift from Feng Zhang (Addgene 61425)[84]. The VP64 was removed to generate dCas9 alone (pAW91) or the human YY1 cDNA was inserted to the C-terminus to generate dCas9-YY1 (pAW90).

For virus production, HEK293T cells grown to 50-75% confluency on a 15 cm dish and then transfected with 15 ug of pAW90 or pAW91, 11.25 μg psPAX (Addgene 12260), and 3.75 μg pMD2.G (Addgene 12259). psPAX and pMD2.G were kind gifts of Didier Trono. After 12 hours, media was replaced. Viral supernatant was collected 24 hours after media replacement (36 hrs post transfection) and fresh media was added. Viral supernatant was collected again 48 hours after the media replacement (60 hours post transfection). Viral supernatant was cleared of cells by either centrifugation at 500 x g for 10 minutes. The virus was concentrated with Lenti-X concentrator (Clontech 631231) per manufacturers' instructions. Concentrated virus was resuspended in mES media (serum + LIF) and added to 5 million cells in the presence of polybrene (Millipore TR-1003) at 8 ug/mL. After 24 hours, viral media was removed and fresh media containing Blasticidin (Invitrogen ant-bl-1) at 10 ug/mL. Cells were selected until all cells on non-transduced plates died.

Two additional lentiviral constructs were generated (pAW12.lentiguide-GFP, pAW13.lentiguide-mCherry) by modifying lentiGuide-puro (lentiGuide-Puro was a gift from Feng Zhang (Addgene plasmid # 52963)[85] to remove the puromycin and replace it either GFP or mCherry. The tethering guide RNAs (Table S5, etv4_p_sgT1_F&R, etv4_p_sgT2_F&R) were then cloned into pAW12 and pAW13. Virus was generated as described above and mES cells were transduced. Double positive cells were identified and collected by flow cytometry and expanded. These expanded cell lines were analyzed by 4C-seq, ChIP-qPCR (anti-Cas9, CST 14697), and RT-qPCR exactly as described elsewhere in the methods.

QUANTIFICATION AND STATISTICAL ANALYSIS

ChIP-MS data analysis

Previously published ChIP-MS data[86] was downloaded. For each mark the $\log_2$ ratio of the immunoprecipitation over the input and over IgG was calculated. Then a high confidence set of proteins was identified by filtering out all proteins that had a $\log_2$ fold change less than or equal to one in either the input or IgG control. Then we filtered for transcription factors using the annotation provided in the original table to end up with the 26 candidates displayed in Fig. 1.

Tissue specific expression analysis

In order to identify candidate structuring factors that are broadly expressed across many tissues, tissue specific expression data from RNA-seq was downloaded from the Genotype-Tissue Expression (GTEx) Project (release V6p). Genes were considered to be expressed in a particular tissues if the median reads per million per kilobase for that tissue was greater than 5 (RPKM > 5). Broadly expressed genes were identified as genes that were expressed in greater than 90% of the 53 tissues surveyed by GTEx.

Definition of regulatory regions

Throughout the manuscript multiple analyses rely on overlaps with different regulatory regions, namely enhancers, promoters, and insulators. Here we explain how these regulatory regions were defined.

Promoters
Promoters were defined as +/- 2 kilobases from the transcription start site.

Active Promoters
Active promoters were defined as +/- 2 kilobases from the transcription start site that overlapped with a H3K27ac peak.

Enhancers
Enhancers were defined as H3K27ac peaks that did not overlap with a promoter.

Insulators
Insulators were defined by downloading called insulated neighborhoods[26] (available at: http://younglab.wi.mit.edu/insulatedneighborhoods.htm). Each row represents an insulated neighborhood (defined as a SMC1 cohesin ChIA-PET interaction with both anchors overlapping a CTCF peak). The file contains six columns, columns 1-3 contain the coordinates for the left interaction anchors of the insulated neighborhoods, and columns 4-6 contain the coordinates for

the right interaction anchors of the insulated neighborhoods. Columns 1-3 and 4-6 were concatenated and then filtered to identify the unique anchors. The unique loop anchors regions correspond to SMC1 ChIA-PET peaks. Insulators elements were identified as the subset of CTCF ChIP-seq peaks that overlapped the unique anchors.

Super-enhancers
Oct4/Sox2/Nanog/Med1 super-enhancers and constituents were downloaded[87].
Typical-enhancer constituents
Oct4/Sox2/Nanog/Med1 typical-enhancer constituents were downloaded[87].

ChIP-seq data analysis

Alignment
Reads from ChIP-seq experiments were aligned to the mm9 revision of the mouse reference genome using only annotated chromosomes 1-19, chrX, chrY, and chrM or to the hg19 revision of the human genome using only annotated chromosomes 1-22, chrX, chrY, and chrM. Alignment was performed using Bowtie[88] with parameters –best –k 1 –m 1 –sam and –l set to read length.

Read pileup for display
Wiggle files representing counts of ChIP-Seq reads across the reference genome were created using MACS[89] with parameters –w –S –space=50 –nomodel –shiftsize=200. Resulting wiggle files were normalized for sequencing depth by dividing the read counts in each bin by the millions of mapped reads in each sample and were visualized in the UCSC genome browser[90].

Gene list and promoter list
For mouse data analysis 36,796 RefSeq transcripts were downloaded in the GTF format from the UCSC genome browser on February 1, 2017. For human data analysis, 39,967 RefSeq transcripts were downloaded on December 7th, 2016 in the GTF format from the UCSC genome browser on February 1, 2017. For each transcript, a promoter was created that is a 4,000 bp window centered on the transcription start site.

Peak calling
Regions with an exceptionally high coverage of ChIP-Seq reads (i.e. peaks) were identified using MACS with parameters –keep-dup=auto –p1e-9 and with corresponding input control.

Heatmaps and Metagenes
Profiles of ChIP-seq and GRO-seq signal at individual regions of interest were created by quantifying the signal in reads per million per base pair (rpm/bp) in bins that equally divide each region of interest using bamToGFF (https://github.com/BradnerLab/pipeline) with parameters –m 200 –r –d. Reads used for quantification were removed of presumed PCR duplicate reads using samtools v0.1.19-44428cd rmdup[91]. Promoters with the same gene id, chromosome, start, and end coordinates were collapsed into one instance.

Heatmaps of ChIP-seq profiles were used to display ChIP-seq signal at enhancer and active promoters. Each row of a heatmap represents an individual region of interest with the ChIP-seq signal profile at that region displayed in rpm/bp in a ±2kb region centered on the region of interest. For each heatmap the number of regions of interest are displayed in parentheses in the figure panel. For murine ES cell heatmaps, ChIP-seq signal was quantified in 200 bins per region of interest. For human tissues and non-ES cell murine tissues, heatmaps were generated by quantifying ChIP-seq signal in 50 bins per region of interest.

Metagene plots were used to display the average ChIP-seq signal across related regions of interest. Metagene plots were generated for enhancer, promoter, and insulator elements, separately. The average profile (metagene) was calculated by calculating the mean ChIP-seq or GRO-seq signal profiles across the related regions of interest. For each metagene plot, the average profile is displayed in rpm/bp in a ±2kb region centered on the regions of interest. The number of enhancers, promoters, and insulators surveyed are noted in parentheses. To facilitate comparisons of the ChIP-seq signal from a single factor between different sets of regions, the total ChIP-seq signal for each metagene analysis was quantified and is displayed in the top right corner of each metagene plot. We note that different antibodies have different immunoprecipitation efficiencies resulting in different signal intensities. Therefore, we believe that quantitative comparisons should be made across different sites in the same ChIP rather than across different ChIPs at the same site.

RNA-seq data analysis

RNA-seq Analysis

RNA-seq data was aligned and quantified using kallisto (version 0.43.0)[92] with the following parameters: -b 100 --single -l 180 -s 20 using the mm9 RefSeq transcriptome (downloaded on February 1, 2017). The output files represent the estimated transcript counts.

Differential gene expression analysis was performed using DESeq2 (version 1.14.1)[93]. Analysis was performed on the gene level. To calculate the gene-level read counts, the estimated transcript counts were summed across all the isoforms of the gene. This was then input into deseq2 and adjusted p values were calculated using the default settings. Log_2 fold changes and adjusted p values are included in Table S2. An FDR value of 0.05 was used as a cut off for significant differential expression. For Fig. 5c, the values on the y axis are the deseq2-calculated log_2 fold change values. The values on the x axis are the DESeq2 calculated baseMean values.

For Fig. 5d, the absolute value of fthe DESeq2 calculated log_2 fold change is plotted on the left side. On the right side the YY1 density at the promoter is plotted. Because the analysis is done on the gene level, the YY1 promoter signal for genes with multiple isoforms was averaged.

For the GO analysis the list of differentially expressed genes (Table S3) was input into the PANTHER GO analysis web tool (http://pantherdb.org/, Version 11.1)[94,95] and a statistical overrepresentation test was performed using the default settings.

RNA-seq Display

For displaying RNA-seq tracks, the RNA-seq data was mapped with Tophat to the mm9 RefSeq transcriptome (downloaded on February 1, 2017) using the following parameters: -n 10 tophat -p 10 --no-novel-juncs –o. Wiggle files representing counts of RNA-Seq reads across the reference genome were created using MACS[89] with parameters –w –S –space=50 –nomodel –shiftsize=200. Resulting wiggle files were normalized for sequencing depth by dividing the read counts in each bin by the millions of mapped reads in each sample and were visualized in the UCSC genome browser[90].

Single-cell RNA-seq Analysis

Sequencing data was demultiplexed using the 10X Genomics Cell Ranger software (version 2.0.0) and aligned to the mm10 transcriptome. Unique molecular identifiers were collapsed into a gene-barcode matrix representing the counts of molecules per cell as determined and filtered by Cell Ranger using default parameters. Normalized expression values were generated using Cell Ranger using the default parameters. For Fig. 5h, the number of cells with a >1 normalized expression value for the specified transcript were counted. For Fig S5c, the cells were arranged by principal component analysis using the default Cell Ranger parameters. In Fig. S5d, cells were split into the two panels based on what condition they came from. The arrangement is the same as in Fig. S5c. Individual cells are then colored by normalized expression level.

4C-seq data analysis

4C-seq Analysis

The 4C-seq samples were first processed by removing their associated read primer sequences (Table S5) from the 5' end of each FASTQ read. To improve mapping efficiency of the trimmed reads by making the read longer, the restriction enzyme digest site was kept on the trimmed read. After trimming the reads, the reads were mapped using bowtie with options –k 1 –m 1 against the mm9 genome assembly. All unmapped or repetitively mapping reads were discarded from further analysis. The mm9 genome was then "digested" *in silico* according to the restriction enzyme pair used for that sample to identify all the fragments that could be generated by a 4C experiment given a restriction enzyme pair. All mapped reads were assigned to their corresponding fragment based on where they mapped to the genome. The digestion of a sample in a 4C experiment creates a series of "blind" and "non-blind" fragments as described by the Tanay and De Laat labs[80]. In brief, "blind" fragments lack a secondary restriction enzyme site whereas "non-blind" fragments contain a secondary restriction enzyme site. Because of this we expect to only observe reads derived from non-blind fragments. We therefore only used reads derived from non-blind fragments.

Experiments were conducted in biological triplicate and the mutant and WT samples were quantile normalized with each other.

If no reads were detected at a non-blind fragment for a given sample when reads were detected in at least one other sample, we assigned a "0" to that non-blind fragment for the sample(s) missing reads.

4C-seq Display

To display 4C-seq genomic coverage tracks, we first smoothed the normalized 4C-seq signal using a 5kb running mean at 50bp steps across the genome for each sample. Individual replicates are displayed in Fig S4. Next, biological replicates of the same condition were combined and the mean and 95% confidence interval of the 4C-seq signal for each bin across the genome was calculated. In Fig 4 and Fig 7, the 4C-seq signal tracks display the mean 4C-seq signal along the genome as a line and the 95% confidence interval as the shaded area around the line. For each 4C-seq signal track, the viewpoint used in the 4C-seq experiment is indicated as an arrow labeled VP.

To quantify the change in 4C-seq signal in a specific region of interest, the normalized 4C-seq signal (non-smoothed) was counted for each sample and the mean and standard deviation of

the quantified signal was calculated for biological replicates of the same condition. The mean and standard deviation of the quantified signal was normalized to the appropriate control condition (either WT or dCas9) before plotting. Below each 4C-seq signal track, the quantified region is indicated as a red bar labeled “Quantified region”. The coordinates of the quantified region for *Raf1* are chr6:115598005-115604631, and for *Etv4* are chr11:101644625-101648624.

ChIA-PET data analysis

ChIA-PET Read Processing

For each ChIA-PET dataset, raw reads were processed in order to identify a set of putative interactions that connect interaction anchors for further statistical modeling and analysis. First, paired-end tags (PETs), each containing two paired reads, were analyzed for the presence of the bridge-linker sequence and trimmed to facilitate read mapping. PETs containing at least one instance of the bridge-linker sequence in either of the two reads were kept for further processing and reads containing the bridge-linker sequence were trimmed immediately before the linker sequence using cutadapt with options “-n 3 -O 3 -m 15 -a forward=ACGCGATATCTTATCTGACT -a reverse=AGTCAGATAAGATATCGCGT” (http://cutadapt.readthedocs.io/en/stable/). PETs that did not contain an instance of the bridge-linker sequence were not processed further. Trimmed read were mapped individually to the mm9 mouse reference genome using Bowtie[88] with options “-n 1 -m 1 -p 6. After alignment, paired reads were re-linked with an in-house script using read identifiers. To avoid potential artifacts arising from PCR bias, redundant PETs with identical genomic mapping coordinates and strand information were collapsed into a single PET. Potential interaction anchors were determined by identifying regions of local enrichment in the individually mapped reads using MACS[89] with options “-g mm -p 1e-9 --nolambda --nomodel --shiftsize=100”. PETs with two mapped reads that each overlapped a different potential interaction anchor by at least 1 bp were used to identify putative interactions between the overlapped interaction anchors. Each putative interaction represents a connection between two interaction anchors and is supported by the number of PETs (PET count) that connect the two interaction anchors.

ChIA-PET Statistical Analysis Overview

In processing our chromatin interaction data, we sought to identify the putative interactions that represent structured chromatin contacts, defined as chromatin contacts that are structured by forces other than the fiber dynamics resulting from the linear genomic distance between the two contacting regions. In contrast, we sought to filter out putative interactions that likely result from PETs arising from non-structured chromatin contacts, defined as contacts resulting from the close linear genomic proximity of the two contacting regions, or from technical artifacts of the ChIA-PET protocol. We expect that putative interactions that represent structured chromatin contacts should be detected with greater frequency, or PET count, than expected given the linear genomic distance between the two contacting regions, allowing us to distinguish between these two classes of interactions.

To this end, we developed Origami, a statistical method to identify high confidence interactions that are likely to represent structured chromatin contacts. Conceptually, Origami uses a semi-Bayesian two-component mixture model to estimate the probability that a putative interaction corresponds to one of two groups: structured chromatin contacts, or non-structured chromatin contacts and technical artifacts. Origami estimates this as a probability score for each putative interaction by modeling the relationship between PET count, linear genomic distance between interaction anchors, and read depth at the interaction anchors. High confidence interactions are

then identified as the subset of putative interactions that are likely to represent structured chromatin contacts, by requiring high confidence interactions to have a probability score > 0.9.

All the methods below were developed within the *origami* software that is available at https://github.com/younglab/origami. The version used was version 1.1 (tagged on GitHub repository as v1.1). The software below was run with the following parameters: --iterations=10000 --burn-in=100 --prune=0 --min-dist=4000 --peak-count-filter=5.

Origami Statistical Model

We developed Origami, a method to analyze ChIA-PET data, in order to identify putative interactions that likely represent structured chromatin contacts, and to filter out putative interactions that likely represent non-structured chromatin contacts that occur as a result of the close linear genomic proximity of contacting regions and interactions that represent technical artifacts of the ChIA-PET protocol. This includes modeling of the relationship between the number of PETs observed to support each interaction (I_i), linear genomic distance between interaction anchors (d_i), and the sequencing depth at the interaction anchors, to estimate the probability that each putative interaction (i) represents a structured chromatin contact given the observed PET count (I_i).

We initially assume that putative interactions classify into one of two groups, $j \in \{0,1\}$, such that each putative interaction, $i \in \{1..N\}$, has a latent group identity Z_i that corresponds to a value of j. Group 1 is designated as the set of putative interactions resulting from structured chromatin contacts that we expect to detect with greater frequencies than expected given the linear genomic distance between the contacting regions. Group 0 is designated as the set of putative interactions resulting from non-structured chromatin contacts due to close linear genomic proximity of the contacting regions, or from technical artifacts of the ChIA-PET protocol.

We developed a semi-Bayesian two-component mixture model to estimate the probability that each putative interaction represents a structured chromatin contact. For each group, we modeled the likelihood to observe the PET count (I_i) under that group as a Poisson process with two underlying factors. These factors are the number of PETs observed as a result of being part of the group (G_{ij}), and the number of PETs observed as a result of the linear genomic distance between the anchors given the group (D_{ij}). We modeled the number of PETs observed as a result of being part of the group (G_{ij}) as a Poisson process with mean, λ_j. We modeled the number of PETs observed as a result of the linear genomic distance between the anchors given the group (D_{ij}) as a Poisson process with mean, μ_{ij}. Since these two factors are thought to be independent[96], the total Poisson process is the summation of these two underlying factors.

We modeled the data variables under the following distributions:

$$I_i \sim \sum_{j \in \{0,1\}} w_{ij} * (G_{ij} + D_{ij})$$
$$G_{ij} \sim Poisson(\lambda_j)$$
$$D_{ij} \sim Poisson[\mu_{ij}(d_i)]$$
$$I_i \sim Poisson[\lambda_j + \mu_{ij}(d_i)]$$

We modeled our parameters with the following prior distributions:

$$\lambda_j \sim Gamma(1,1)$$
$$w_{i1} \sim Beta(1 + a_i, 1 + b_i)$$

Since w_{i1} is a binomial probability, $w_{i0} = 1 - w_{i1}$.

From these priors and likelihood distributions, the posterior distributions of these parameters are as follows:

$$\lambda_j \sim Gamma[1 + \sum_{Z_i=j} G_{ij}, 1 + \#(Z_i = j)]$$
$$w_{i1} \sim Beta[1 + \alpha_i + Zi, 1 + \beta_i + (1 - Zi)]$$

Aside from D_{ij} and μ_{ij}, we estimated the parameters using the iterative process Markov Chain Monte Carlo (MCMC) with Gibbs Sampling with the appropriate posterior to sample from[97].

To estimate μ_{ij}, we modeled the function between D_{ij} and the linear genomic distance (d_i) on the log10 scale using a smoothed cubic spline (via smooth.spline in R), taking μ_{ij} to be the expected number of PETs to be observed due to distance (D_{ij}) given the linear genomic distance (d_i), for each putative interaction (i).

The constants α_i and β_i were set to be as minimally informative as possible. The constant α_i was set equal to the number of putative interactions sharing one anchor with i that have PET counts less than I_i. The constant β_i was set equal to the number of putative interactions sharing one anchor with i that have PET counts greater than I_i plus the ratio of the depth score (s_i) to the median depth score with all values <1 floored to 0. The depth score (s_i) for each putative interaction is defined as the product of the number of reads that map to its interaction anchors.

Origami Implementation

We implemented the model described above by Markov Chain Monte Carlo simulation. By iteratively estimating the group identity (Z_i) of each putative interaction, we sought to explore the probability space for Z_i and determined a probability score (p_i) for each putative interaction that reflects the probability that the interaction results from a structured chromatin contact (belongs to group 1). The steps in our implementation are as follows.

For each putative interaction, we recorded the number of PETs observed that support the interaction (I_i), the linear genomic distance of the interaction between the outermost basepairs of the putative interaction's two anchors (d_i), and a depth score (s_i), which is defined as the product of the number of the reads in the dataset that map to each anchor of the putative interaction.

To seed the parameters of the model for the first iteration, the following was performed. The mixing weights (w_{ij}) were set to be equal at 0.5 for each interaction. The group process means (λ_j) were assigned values of 5 and 1 for group 1 and 0, respectively. The distance process mean (μ_{ij}) was initially set to 0 for all interactions.

Additionally values of α_i and β_i were computed for each interaction, but not used in the first iteration. In all subsequent iterations, α_i and β_i, are used in updating the values of the mixing weights (w_{ij}). The parameter α_i was set equal to the number of putative interactions sharing one anchor with i that have PET counts less than I_i. The parameter β_i was set equal to the

number of putative interactions sharing one anchor with i that have PET counts greater than I_i plus the ratio of the depth score (s_i) over the median depth score for all putative interactions, where when this ratio is less than 1 it is floored to 0.

For each putative interaction, we estimated the likelihood (l_{ij}) that the putative interaction is observed with PET count (I_i), given that the putative interaction belongs to group 1 and group 0, as follows.

$$l_{ij} = dPoisson(I_i; \lambda_j + \mu_{ij})$$

Where dPoisson is the density function of the Poisson distribution for the mean $\lambda_j + \mu_{ij}$ and evaluated on I_i.

We calculated the relative weighted likelihood (r_i) of each putative interaction belonging to group 1. To do this we multiplied each of the two likelihoods calculated for each putative interaction by their respective mixing weights (w_{ij}) and evaluated as follows.

$$r_i = \frac{w_{i1} * L_{i1}}{(w_{i1} * L_{i1}) + (w_{i0} * L_{i0})}$$

We update the group identity (Z_i) of each interaction by drawing from the binomial distribution with a probability of r_i as follows.

$$Z_i = rBinomial(1, r_i)$$

Where rBinomial means we randomly draw 1 or 0 with the probability of r_i for drawing 1.

We update the mixing weights (w_{ij}) using our newly updated group identies (Z_i), by drawing from the Beta distribution in the following way.

$$w_{i1} = rBeta[1 + \alpha_i + Zi, 1 + \beta_i + (1 - Zi)]$$

Where rBeta means we randomly draw from the beta distribution with the above parameters. Since w_{i1} is a binomial probability, $w_{i0} = 1 - w_{i1}$.

In order to estimate the PET counts for G_{ij} and D_{ij}, we randomly sampled the number of PETs for G_{ij} and D_{ij} by taking advantage of the fact that when two Poisson variables are known to sum to a given count, then the distribution of either variable follows a binomial distribution with probability $\lambda_j/(\lambda_j + \mu_{ij})$. Accordingly, we estimated the PET counts for G_{ij} and D_{ij} in the following way:

$$G_{ij} = rBinomial(I_i, \frac{\lambda_j}{\lambda_j + \mu_{ij}})$$

$$D_{ij} = I_i - G_{ij}$$

Where rBinomial means we randomly draw up to I_i PETs with the probabilty $\lambda_j/(\lambda_j + \mu_{ij})$ of drawing each PET.

We update the group process mean (λ_j) using the following identity, requiring that $\lambda_1 > \lambda_0$ in order to maintain identifiability of the two groups (although during our runs this constraint was not necessary).

$\lambda_j = rGamma(1 + \sum_{Z_i=j} G_{ij}\,, 1 + \#(Z_i = j))$

Where rGamma means we randomly draw from the Gamma distribution with the above parameters.

To update the distance process means (μ_{ij}), we calculated the function between D_{ij} and the $log_{10}(d_i + 1)$, using a smoothed cubic spline (via smooth.spline in R). To simplify estimation of μ_{ij}, we chose to take the maximum likelihood estimate of this process.

We iterated steps 4-10 in the following way. We performed an initial 1,000 iterations as a burn-in, which were discarded. Then we performed 10,000 iterations.

We estimated the probability that each putative interaction belongs to group 1 by calculating a probability score (p_i) for each putative interaction that equals the mean value of Z_i across the 10,000 iterations. High confidence interactions were identified as putative interactions with $p_i > 0.9$.

$$p_i = \frac{1}{\#(iterations)} \sum Z_i \approx P(Zi = 1)$$

HiChIP data analysis

HiChIP Processing

The HiChIP samples were processed by first identifying reads with a restriction fragment junction (i.e. a site where ligation occurred). Reads containing the restriction fragment junction were trimmed such that the information 5' to the junction was kept. Reads without restriction fragment junctions were left untrimmed. Reads were then mapped using bowtie with options –k 1 –m 1 against the mm9 genome assembly. All unmapped or repetitively mapping reads were discarded from further analysis. Reads were joined back together in pairs by their read identifier. The genome was binned and for every pair of bins the number of PETs joining them was calculated. These data were then used as input into the Origami pipeline described above to identify significant bin to bin interaction pairs.

HiChIP Analysis

Quantitative analysis of HiChIP and Hi-C data (Fig 6, 7) was done as follows. High confidence interactions were identified by Origami. A union of high confidence interactions was then created for each experiment.

Experiment	Figure	Condition	Replicate
Degron	6, S6	noDrug	1
Degron	6, S6	noDrug	2
Degron	6, S6	noDrug	3
Degron	6, S6	yesDrug	1
Degron	6, S6	yesDrug	2
Degron	6, S6	yesDrug	3
Washout	7	Untreated (UT)	1
Washout	7	Untreated (UT)	2
Washout	7	Untreated (UT)	3
Washout	7	Treated (TR)	1
Washout	7	Treated (TR)	2
Washout	7	Washout (WO)	1
Washout	7	Washout (WO)	2
Washout	7	Washout (WO)	3
CTCF Washout	7	Untreated (UT)	1
CTCF Washout	7	Untreated (UT)	2
CTCF Washout	7	Treated (TR)	1
CTCF Washout	7	Treated (TR)	2
CTCF Washout	7	Washout (WO)	1
CTCF Washout	7	Washout (WO)	2

For example, the degron high confidence set would consist of the union of the 6 degron samples listed above. The PET counts were then normalized to each other using DESeq2[93]. The mean of each group was then calculated and then the fold change was then calculated by taking the ratio of the perturbed condition to the non-perturbed condition (i.e. yesDrug to noDrug or TR/UT;WO/UT) with a pseudo-count of 0.5 added to both. This complete set of significant interactions is what is displayed in Fig. 6b as “All Interactions.”

For subset analysis the anchor of each interaction was classified by overlapping with known genomic features as defined earlier. This resulted in a binary score for whether an anchor overlapped with an enhancer, promoter, insulator, YY1, or CTCF. The interactions were then subset to identify the following groups:

YY1 not present (Fig 6): no YY1 at either end of the interaction.
YY1 enhancer-promoter interactions (Fig 6, Fig 7): YY1 at both ends AND an enhancer or promoter at both ends.
CTCF-CTCF interaction: CTCF at both ends.

The $\log_2$ fold change for these groups is plotted in Fig. 6b, 7f.

The analysis in Fig. 6c was done by identifying the gene at the end of YY1 enhancer-promoter loops. This was done by intersecting promoters (as defined above) with the significant loop anchors. Genes with multiple promoters were collapsed after the intersection to generate a list of genes at the end of YY1 enhancer-promoter loops. The DESeq2 calculated $\log_2$ fold change for these genes is then plotted in Fig. 6c. Genes are colored based on the DESeq2 calculated adjusted p value (as in Fig. 5).

HiChIP Display

HiChIP interaction matrices displayed in Fig. 6d, e. For these interaction matrices, all putative interactions are displayed and the intensity of each pixel represents the mean of the DESeq2 normalized interaction frequency of all biological replicates of that condition. In Fig. 6d, e, the outlined pixel, which reflects the frequency of interaction between sites at the base of the diagonals, was used to quantify the change in normalized interaction frequency upon YY1 degradation.

In Fig. 2, high-confidence HiChIP interactions are displayed as arcs. For display, the interactions displayed were filtered to remove bin to adjacent bin contacts and non-enhancer-promoter interactions. Arcs were centered on the relevant genomic feature within the bin (for example a ChIP-seq peak summit or transcription start site).

Interaction classification

High-confidence ChIA-PET and HiChIP interactions were classified based on the presence of enhancer, promoter, and insulator elements at the anchors of each interaction as defined above. In the case where an interaction anchor overlapped both an enhancer and an insulator or a promoter and an insulator a hierarchy where anchors were considered first as promoters, then enhancers, then insulators. For example, if there is an interaction where the left anchor is insulator/promoter and the right anchor is enhancer/insulator it would be counted as an enhancer-promoter interaction and not an insulator-insulator interaction.

To display summaries of the classes of high-confidence interactions, each class of interactions is displayed as an arc between the relevant enhancer, promoter, and insulator elements. The thickness of the arcs approximately reflects the percentage of interactions of that class relative to the total number of interactions that were classified. In the main figures, enhancer-enhancer, enhancer-promoter, promoter-promoter, and insulator-insulator interaction classes are displayed. Extended summaries that additionally include enhancer-insulator and promoter-insulator interactions are displayed in the supplemental figures.

Figure Display

In certain figure panels displaying genome tracks, enhancer elements are indicated as red boxes labelled “Enhancer”. These regions represent the authors’ interpretation of the ChIP-seq data and are distinct from the algorithmically defined enhancers used in the quantitative genome-wide analysis.

Statistical Analysis

In order to use the unpaired t-test we made two assumptions.
1) Populations are distributed according to a Gaussian distribution. For most experiments three replicates were used, and so sample sizes were too small to reliably calculate departure from normality (i.e. with a D'Agostino test).
2) The two populations have the same variance. A test for variance was not carried out.

Full p values are listed here*:

Figure	Sub panel	Test	Biological Replicates	P value
4b	4C-seq	Student's T-Test	3	0.011
4b	ChIP-qPCR	Student's T-Test	3	0.0066
4b	RT-qPCR	Student's T-Test	6	<0.0001
4c	4C-seq	Student's T-Test	3	0.0013
4c	ChIP-qPCR	Student's T-Test	3	0.0048
4c	RT-qPCR	Student's T-Test	6	0.0394
6b		Welch Two Sample T-Test	3	< 2.2e-16
6d	HiChIP	Student's T-Test	3	0.0162
6d	RNA-seq	Wald	2	7.22E-13
6e	HiChIP	Student's T-Test	3	0.0446
6e	RNA-seq	Wald	2	1.25E-58
7d	4c-seq	Student's T-Test	3	0.004717003
7d	RT-qPCR	Student's T-Test	6	<0.0001
S6d	*Raf1*	Wald	2	1.63E-53
S6d	*Etv4*	Wald	2	2.88E-34

*note that the Student's T-test was conducted using GraphPad Prism which sets a lower limit at 0.0001, the Welch Two Sample T-test was conducted using R which sets a lower limit at 2.2e-16, Wald test was conducted using DESeq2 in R which does not have a lower limit on the p value.

Data and Software Availability

All datasets used are summarized in Table S4.
Origami: https://github.com/younglab/origami using version v1.1-alpha-2.

The data associated with this study have been deposited in the Gene Expression Omnibus (GEO) under ID code GSE99521.

Supplemental Tables

Supplemental tables are available online at:
http://www.cell.com/cell/fulltext/S0092-8674(17)31317-X
Table S1: Comparison of YY1 and CTCF, related to Figure 1
Table S2: RNA-seq data, related to Figure 5
Table S3: GO Analysis, related to Figure 5
Table S4: Datasets used in the study, related to Figure 1, Figure 2, and Figures 4-7
Table S5: Oligos used in the study, related to STAR methods

Supplemental information

Supplemental Information includes STAR Methods, 7 Figures, and 5 Tables

Author contributions

Conceptualization, A.S.W., C.H.L., A.S., R.A.Y.; Methodology, A.S.W, C.H.L.; Software, A.S.W., C.H.L., D.S.D, B.J.A.; Formal Analysis, A.S.W., B.J.A., D.S.D., Investigation, A.S.W., C.H.L., A.Z., A.S., N.H., M.A.C, D.H.; Resources, N.M.H., B.N., D.L.B, R.J., J.E.B., N.S.G., Y.E.G.; Writing – Original Draft, A.S.W., C.H.L., R.A.Y.; Writing – Review & Editing, all authors; Visualization, A.S.W.; Supervision, R.A.Y., A.S.W.; Funding Acquisition, R.A.Y.

Acknowledgements

We thank the Whitehead Institute Genome Technology Core, FACS facility, and Johanna Goldmann for their assistance. This work is supported by NIH grants HG002668/GM123511 (R.A.Y.), Ludwig Graduate Fellowship funds (A.S.W.), NSF GRFP (A.V.Z.), ACS New England Division Postdoctoral Fellowship PF-16-146-01-DMC (D.S.D.), Margaret and Herman Sokol Postdoctoral Award (D.H.), ACS Postdoctoral Fellowship PF-17-010-01-CDD (B.N.), Merck Fellow of the Damon Runyon Cancer Research Foundation DRG-2196-14 (D.L.B), Hope Funds for Cancer Research Grillo-Marxuach Family Fellowship (B.J.A.), Cancer Research Institute Irvington Fellowship (Y.E.G.). The Whitehead Institute filed a patent application based on this study. R.A.Y., J.E.B., and N.S.G. are the founders of Syros Pharmaceuticals. B.J.A. is a shareholder in Syros. R.A.Y. is a founder of Marauder Therapeutics. J.E.B. is a Scientific Founder of SHAPE Pharmaceuticals, Acetylon Pharmaceuticals, Tensha Therapeutics (now Roche) and C4 Therapeutics. J.E.B. is now an executive and shareholder in Novartis AG. N.S.G. is a founder and member of the Scientific Advisory Board of C4 and Petra Therapeutics. Datasets generated can be found in GEO (GSE99521). Accession numbers of datasets used are listed in Supplemental Table S4. Oligos used can be found in Table S5.

7. REFERENCES

1. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).
2. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
3. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).
4. Ong, C. T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* **12**, 283–293 (2011).
5. Ren, B. & Yue, F. Transcriptional enhancers: bridging the genome and phenome. *Cold Spring Harb. Symp. Quant. Biol.* **80**, 17–26 (2016).
6. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
7. Fraser, J., Williamson, I., Bickmore, W. A. & Dostie, J. An overview of genome organization and how we got there: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* **79**, 347–372 (2015).
8. Heard, E. & Bickmore, W. The ins and outs of gene regulation and chromosome territory organisation. *Curr. Opin. Cell Biol.* **19**, 311–316 (2007).
9. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
10. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
11. Spitz, F. Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles. *Semin. Cell Dev. Biol.* **57**, 57–67 (2016).
12. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
13. Jeronimo, C. *et al.* Tail and kinase modules differently regulate core Mediator recruitment and function in vivo. *Mol. Cell* **64**, 455–466 (2016).
14. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–5 (2010).
15. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772 (2010).
16. Petrenko, N., Jin, Y., Wong, K. H. & Struhl, K. Mediator undergoes a compositional change during transcriptional activation. *Mol. Cell* **64**, 443–454 (2016).
17. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20**, 578–588 (2010).
18. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
19. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
20. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).
21. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
22. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
23. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
24. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
25. Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 771–775 (2014).

26. Hnisz, D., Day, D. S. & Young, R. A. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* **167**, 1188–1200 (2016).
27. Merkenschlager, M. & Nora, E. P. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17–43 (2016).
28. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
29. Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
30. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 3812–3814 (2016).
31. Ji, X. *et al.* 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
32. Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
33. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–21 (2015).
34. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
35. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
36. Tang, Z. *et al.* CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
37. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
38. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
39. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1092–1096 (2015).
40. Klenova, E. M. *et al.* 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 (1993).
41. Shi, Y., Seto, E., Chang, L. S. & Shenk, T. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67**, 377–388 (1991).
42. Donohoe, M. E. *et al.* Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell. Biol.* **19**, 7237–44 (1999).
43. Heath, H. *et al.* CTCF regulates cell cycle progression of αβ T cells in the thymus. *EMBO J.* **27**, 2839–2850 (2008).
44. López-Perrote, A. *et al.* Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. *J. Biol. Chem.* **289**, 22614–22629 (2014).
45. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–734 (2014).
46. Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
47. Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–982 (2015).
48. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
49. Shore, D., Langowski, J. & Baldwin, R. L. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 4833–4837 (1981).
50. Erb, M. A. *et al.* Transcription control by the ENL YEATS domain in acute leukaemia. *Nature* **543**, 270–274 (2017).
51. Huang, H. T. *et al.* MELK is not necessary for the proliferation of basal-like breast cancer cells.

Elife **6**, 1–29 (2017).
52. Winter, G. E. *et al.* Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **348**, 1376–81 (2015).
53. Nora, E. P. *et al.* Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
54. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2gene. *Nature* **405**, 482–485 (2000).
55. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).
56. Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
57. Hariharan, N., Kelley, D. E. & Perry, R. P. δ, A transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9799–9803 (1991).
58. Park, K. & Atchison, M. L. Isolation of a candidate repressor/activator, NF-E1 (YY-1, Delta), that binds to the immunoglobulin K 3' enhancer and the immunoglobulin heavy-chain uE1 site. **88**, 9804–9808 (1991).
59. Gordon, S., Akopyan, G., Garban, H. & Bonavida, B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **25**, 1125–1142 (2006).
60. Shi, Y., Lee, J. S. & Galvin, K. M. Everything you have ever wanted to know about Yin Yang 1... *Biochim. Biophys. Acta - Rev. Cancer* **1332**, (1997).
61. Thomas, M. J. & Seto, E. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* **236**, 197–208 (1999).
62. Ohlsson, R., Renkawitz, R. & Lobanenkov, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–527 (2001).
63. Degner, S. C. *et al.* CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9566–9571 (2011).
64. Guo, C. *et al.* CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424–430 (2011).
65. Liu, H. *et al.* Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev.* **21**, 1179–1189 (2007).
66. Hwang, S. S. *et al.* Transcription factor YY1 is essential for regulation of the Th2 cytokine locus and for Th2 cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 276–281 (2013).
67. Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27**, 1139–1152 (2017).
68. Gabriele, M. *et al.* YY1 haploinsufficiency causes an intellectual disability syndrome featuring transcriptional and chromatin dysfunction. *Am. J. Hum. Genet.* **100**, 907–925 (2017).
69. Zhang, Q., Stovall, D. B., Inoue, K. & Sui, G. The oncogenic role of Yin Yang. *Crit. Rev. Oncog.* **16**, 163–197 (2011).
70. Amoutzias, G. D., Robertson, D. L., Van de Peer, Y. & Oliver, S. G. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* **33**, 220–229 (2008).
71. Lamb, P. & McKnight, S. L. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* **16**, 417–422 (1991).
72. Weirauch, M. T. & Hughes, T. R. *A Handbook of Transcription Factors*. *Subcellular Biochemistry* **52**, (2011).
73. Adhya, S. Multipartite genetic control elements: communication by DNA loop. *Annu. Rev. Genet.* **23**, 227–250 (1989).
74. Schleif, R. DNA looping. *Annu. Rev. Biochem.* **61**, 199–223 (1992).
75. Matthews, K. S. DNA looping. *Microbiological Reviews* **56**, 123–136 (1992).
76. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).

77. Muerdter, F. & Stark, A. Gene regulation: activation through space. *Curr. Biol.* **26**, R882–R902 (2016).
78. Jeon, Y. & Lee, J. T. YY1 Tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).
79. Lee, T., Johnston, S. & Young, R. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* **1**, 729–748 (2006).
80. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).
81. van de Werken, H. J. G. *et al. 4C technology: protocols and data analysis*. *Methods in Enzymology* **513**, (2012).
82. Nagano, T. *et al.* Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* **16**, 1–13 (2015).
83. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
84. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–8 (2014).
85. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
86. Ji, X. *et al.* Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3841–6 (2015).
87. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
88. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
89. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
90. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
91. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
92. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
93. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
94. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the panther classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).
95. Mi, H. *et al.* PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).
96. Phanstiel, D. H., Boyle, A. P., Heidari, N. & Snyder, M. P. Mango: A bias-correcting ChIA-PET analysis pipeline. *Bioinformatics* **31**, 3092–3098 (2015).
97. Gelman, A. *et al. Bayesian Data Analysis*. (2004).

CHAPTER 3 – MECP2 LINKS HETEROCHROMATIN CONDENSATES AND NEURODEVELOPMENTAL DISEASE

Originally published in Nature (2020)
Reprinted with permission from Springer Nature

Charles H. Li[1,2,7], Eliot L. Coffey[1,2,7], Alessandra Dall'Agnese[1], Nancy M. Hannett[1], Xin Tang[1], Jonathan E. Henninger[1], Jesse M. Platt[1,3], Ozgur Oksuz[1], Alicia V. Zamudio[1,2], Lena K. Afeyan[1,2], Jurian Schuijers[1,5], X. Shawn Liu[1,6], Styliani Markoulaki[1], Tenzin Lungjangwa[1], Gary LeRoy[4], Devon S. Svoboda[1], Emile Wogram[1], Tong Ihn Lee[1], Rudolf Jaenisch[1,2]*, Richard A. Young[1,2]*

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[3]Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[4]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA
[5]Present address: Center for Molecular Medicine, University Medical Center Utrecht, 3584 CG Utrecht, The Netherlands
[6]Present address: Department of Physiology and Cellular Biophysics, Columbia University Medical Center, New York, NY 10032, USA
[7]These authors contributed equally
*Correspondence: Richard A. Young (young@wi.mit.edu), Rudolf Jaenisch (jaenisch@wi.mit.edu)

STATEMENT ON CONTRIBUTION

I contributed to conceptualization of the project and design of the research methodology. I contributed to experiments and analyses presented in all figures with help from the other authors. I supervised all experiments and analyses. I contributed to writing, reviewing, and editing the manuscript

1. SUMMARY

MeCP2 (methyl CpG binding protein 2) is a key component of constitutive heterochromatin, which plays important roles in chromosome maintenance and transcriptional silencing[1–3]. Mutations in MeCP2 cause Rett syndrome (RTT)[3–5], a postnatal progressive neurodevelopmental disorder associated with severe mental disability and autism-like symptoms that manifests in girls during early childhood. Heterochromatin, long considered a dense and relatively static structure[1,2], is now understood to exhibit properties consistent with a liquid-like condensate[6,7]. Here we report that MeCP2 is a dynamic component of heterochromatin condensates in cells, is stimulated by DNA to form liquid-like condensates, contains multiple domains that contribute to condensate formation, manifests physicochemical properties that selectively concentrate heterochromatin cofactors compared to components of transcriptionally active condensates, and when altered by RTT-causing mutations is disrupted in its ability to form condensates. We propose that MeCP2 enhances heterochromatin/euchromatin separation through its condensate partitioning properties and that condensate disruption may be a common consequence of RTT patient mutations.

2. RESULTS

MeCP2 and HP1 proteins are key regulators of heterochromatin[1–4]. Recent studies have shown that HP1 proteins are dynamic components of heterochromatin in vivo and can form phase-separated condensates in vitro, arguing that heterochromatin is a dynamic liquid-like condensate[6,7]. To confirm that MeCP2 is also a dynamic component of heterochromatin, we used live-cell fluorescence microscopy to image both MeCP2 and HP1α, endogenously tagged with fluorescent proteins, in murine embryonic stem cells (mESCs) (Fig. 1a-c, Extended Data Fig. 1). The results showed that MeCP2-GFP and HP1α-mCherry occur in nuclear bodies that overlap Hoechst-dense heterochromatin foci (Fig. 1a, Extended Data Fig. 1a) and that the two proteins occur in the same heterochromatin condensates (Extended Data Fig. 1b). Fluorescence recovery after photobleaching (FRAP) of MeCP2-GFP and HP1α-mCherry puncta revealed recovery on the timescale of seconds (Fig. 1b, c, Extended Data Fig. 1c-f), consistent with characteristics of liquid-like condensates. These results show that MeCP2 is a dynamic component of heterochromatin condensates in live mESCs.

To determine whether MeCP2 is a dynamic component of heterochromatin in mammalian tissues, we generated mice expressing GFP-tagged MeCP2 protein from the endogenous locus (Extended Data Fig. 2a-c). MeCP2 is reported to be expressed in all cell types (Extended Data Fig. 2d); we studied neurons because MeCP2 is highly abundant in these cells[8] and mutations in *MECP2* cause neurodevelopmental disorders[3–5]. Imaging of MAP2-positive neurons revealed that MeCP2-GFP occurs in Hoechst-dense heterochromatin foci (Fig. 1d). FRAP of MeCP2-GFP puncta revealed rapid and complete recovery on the timescale of seconds (Fig. 1e, f). These results indicate that MeCP2 is a dynamic component of liquid-like heterochromatin condensates in murine brain cells.

Figure 1

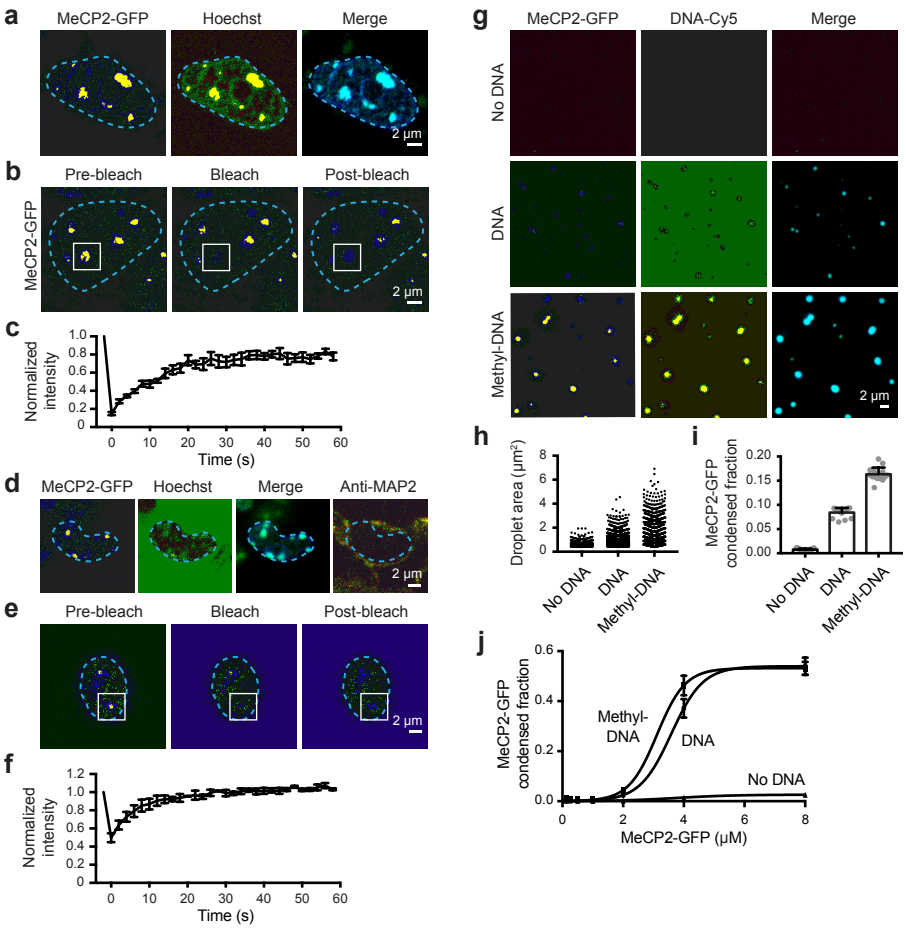


Figure 1. MeCP2 forms condensates in vivo and in vitro

a. Live-cell images of endogenous-tagged MeCP2-GFP and Hoechst staining in mESCs.
b. Live-cell images of FRAP experiments with endogenous-tagged MeCP2-GFP mESCs.
c. FRAP curves for experiments in Fig. 1b. Photobleaching occurs at t = 0 s. Mean±SD, n=7 cells.
d. Fixed-cell images of endogenous-tagged MeCP2-GFP brain sections from chimeric mice.
e. Images of FRAP experiments performed on acute brain slices from endogenous-tagged MeCP2-GFP chimeric mice.
f. FRAP curves for experiments in Fig. 1e. Photobleaching occurs at t = 0 s. Mean±SEM, n=3 cells.
g. Droplet experiments examining MeCP2 droplet formation with DNA. MeCP2-GFP at 2 μM was mixed with 160 nM unmethylated DNA (DNA), methylated DNA (methyl-DNA), or no DNA in droplet formation buffers with 100 mM NaCl.
h. Droplet areas for experiments in Fig. 1g. Fields per condition n=15.
i. MeCP2-GFP condensed fraction for experiments in Fig. 1g. Mean±SD. Fields per condition n=15.
j. MeCP2-GFP condensed fraction curves for experiments examining MeCP2 droplet formation with DNA. MeCP2-GFP was mixed with 160 nM DNA, methyl-DNA, or no DNA in droplet formation buffers with 100 mM NaCl. Mean±SD. Fields per condition n=15.

To investigate whether MeCP2 has physicochemical properties that may contribute to heterochromatin condensates in cells, we examined purified MeCP2-GFP fusion protein using in vitro droplet assays. MeCP2-GFP formed spherical droplets that displayed properties consistent with phase-separated liquid condensates, including sensitivity to protein and salt concentration, droplet fusion behavior, and dynamic rearrangement of molecules measured using FRAP (Extended Data Fig. 3a-i). Purified HP1α-mCherry also formed droplets (Extended Data Fig. 3j), consistent with prior findings[6,7]. MeCP2 binds to DNA[3,4], so we studied the effects of DNA on MeCP2 droplet formation. When DNA was added to physiologically relevant concentrations of MeCP2-GFP, MeCP2-GFP formed droplets (Fig. 1g). With the addition of methylated DNA, which MeCP2 binds with higher affinity[3,4], larger droplets were formed (Fig. 1g, h), and these droplets contained a larger fraction of MeCP2 (Fig. 1i) and increased levels of DNA (Extended Data Fig. 3k). These results were observed across a range of MeCP2 concentrations (Fig. 1j). These observations suggest that DNA can cause crowding of MeCP2 and thus lower the threshold for condensate formation, analogous to the manner in which enhancer DNA elements crowd transcription factors to lower the threshold for formation of transcriptional condensates[9].

We used droplet assays to identify domains of MeCP2 that contribute to condensate formation. Intrinsically disordered regions (IDRs) can participate in condensate formation[10], and MeCP2 contains two conserved IDRs that flank its structured methyl-DNA binding domain (MBD) (Fig. 2a, Extended Data Fig. 3l). Droplet formation assays were conducted using physiologically relevant concentrations of recombinant MeCP2-GFP domain deletion mutant proteins in the presence of DNA. While mutant proteins lacking the N-terminal IDR (ΔIDR-1) formed droplets, those lacking the C-terminal IDR (ΔIDR-2) did not (Fig. 2b-d). Furthermore, IDR-1 alone did not form droplets, whereas IDR-2 alone did form droplets, albeit with diminished size and number relative to both full-length and ΔIDR-1 proteins (Fig. 2b-d). Similar results were observed when MeCP2-GFP domain deletion mutant proteins were examined in droplet assays in the absence of DNA (Extended Data Fig. 3m-o). These results indicate that MeCP2's C-terminal IDR, which has been implicated in various functions including heterochromatin association[11], chromatin compaction[12], co-repressor recruitment[13], and transcriptional repression[14], contributes to condensate formation. Furthermore, the results indicate that the MBD also contributes to condensate formation because DNA binding lowers the threshold for condensate formation.

Specific sequence features within IDRs have been found to contribute to condensate formation[10]; several of these features occur within MeCP2's IDR-2 (Fig. 2a, Extended Data Fig. 3l), leading us to investigate whether these contribute to MeCP2 condensate behaviors. We found that deletion of basic patches within IDR-2 disrupted MeCP2 droplet formation, while deletion mutants removing aromatic residues, a histidine-rich patch, and a proline-rich patch remained capable of droplet formation (Fig. 2e-g). Droplet formation correlated with ability to repress transcription, a key MeCP2 function[14], in a transcriptional repression reporter assay (Fig. 2h, i). These results suggest that the basic patches in MeCP2's C-terminal IDR, some of which are disrupted in RTT[15], play especially important roles in MeCP2 condensate formation.

Figure 2

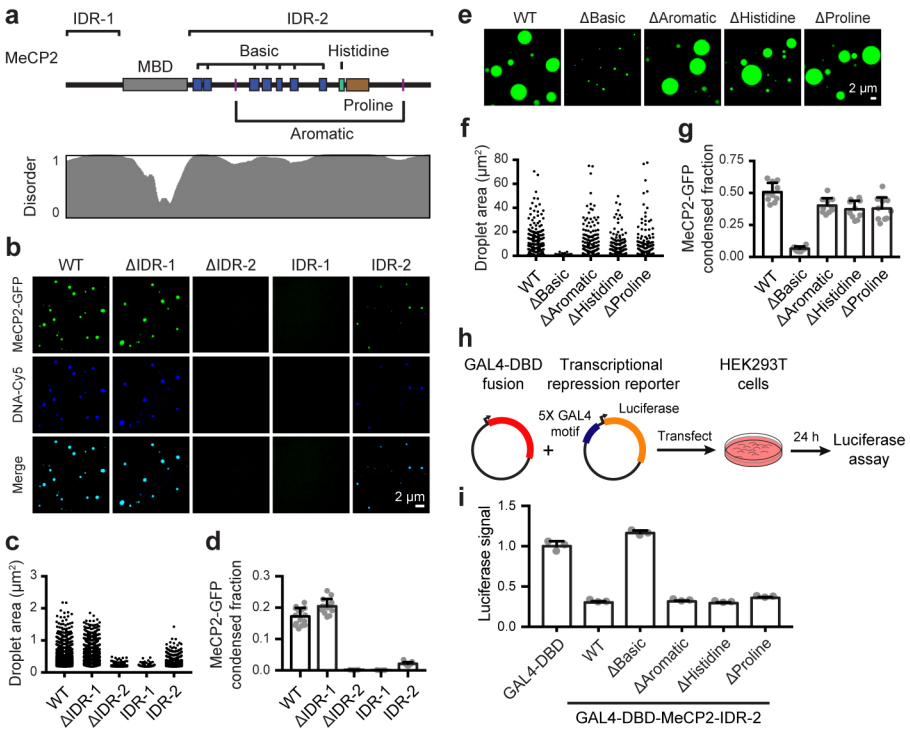


Figure 2. MeCP2 features that contribute to condensate formation

a. Schematic of MeCP2 protein indicating the MBD, IDR-1, IDR-2, and sequence features within IDR-2 previously implicated in condensate formation for other proteins. Contribution of IDR-2 sequence features to condensate formation was examined using deletion mutants that remove the basic patches (ΔBasic), aromatic residues (ΔAromatic), histidine-rich patch (ΔHistidine), and proline-rich patch (ΔProline). Predicted protein disorder is displayed below.

b. Droplet experiments examining ability of MeCP2 deletion mutants to form droplets with DNA. MeCP2-GFP deletion mutants at 2 μM were mixed with 40 nM DNA in droplet formation buffers with 100 mM NaCl.

c. Droplet areas for experiments in Fig. 2b. Fields per condition n=15.

d. MeCP2-GFP condensed fraction for experiments in Fig. 2b. Mean±SD. Fields per condition n=15.

e. Droplet experiments examining ability of MeCP2 IDR-2 sequence feature deletion mutants to form droplets. MeCP2-GFP IDR-2 sequence feature deletion mutants at 10 μM were added to droplet formation buffers with 150 mM NaCl and 10% PEG-8000.

f. Droplet areas for experiments in Fig. 2e. Fields per condition n=10.

g. MeCP2-GFP condensed fraction for experiments in Fig. 2e. Mean±SD. Fields per condition n=10.

h. Schematic of transcriptional repression reporter assay used to examine the ability of MeCP2 IDR-2 sequence features to contribute to transcriptional repression.

i. Normalized luciferase signals for reporter assay examining ability of MeCP2 IDR-2 sequence features to contribute to transcriptional repression. Luciferase signal was normalized to GAL4-DBD alone. Mean±SD, n=3 biologically independent samples per condition.

Active transcriptional condensates[16,17] do not overlap heterochromatin condensates (Extended Data Fig. 4a). While separation of euchromatin and heterochromatin can be attributed to different DNA-binding factors and differential association with nuclear lamina[18], it is possible that the condensate properties of specific proteins might also contribute to the separation of these distinct compartments[19]. To investigate this possibility, we tested whether condensates formed by MeCP2 preferentially incorporate and concentrate HP1α compared to key components of euchromatic transcriptional condensates, such as MED1 and BRD4. We found that MeCP2-GFP droplets incorporated and concentrated HP1α-mCherry to a substantially greater extent than MED1 and BRD4 IDRs (Extended Data Fig. 4b-g). Similar results were obtained in the presence or absence of DNA (Extended Data Fig. 4b-e), and in the presence of nucleosomal DNA, albeit with less efficiency (Extended Data Fig. 4f, g). Nucleosomal DNA alone did not form droplets under these conditions (Extended Data Fig. 4h), though it enhanced MeCP2 droplet formation (Extended Data Fig. 4i-k). Notably, when combined with MeCP2-GFP in the presence or absence of DNA, BRD4-IDR-mCherry was enriched in a distinct droplet phase that did not coalesce with the MeCP2-GFP droplet phase, although the two phases appeared adjacent and touching (Extended Data Fig. 4b, d, Extended Data Fig. 5). These results suggest that MeCP2 condensates may contribute to selective partitioning of components of heterochromatin and active euchromatin. There is some evidence for MeCP2 occupancy of euchromatin[8,15,20], but MeCP2 levels in active euchromatin may not be sufficient to form condensates that facilitate partitioning of heterochromatin components.

RTT patient mutations occur predominantly in MeCP2's MBD and IDR-2 (Fig. 3a), which both contribute to condensate formation. To examine whether patient mutations in these domains disrupt the ability of MeCP2 to form condensates, we examined MeCP2-GFP proteins with RTT patient mutations using droplet formation assays (Fig. 3b-g, Extended Data Fig. 6). Patient missense mutations affecting the MBD reduced the ability of MeCP2 to form droplets (Fig. 3d, e, Extended Data Fig. 6b). Similarly, patient mutations that truncate IDR-2 disrupted the ability of MeCP2 to form droplets, with mutations that truncated a greater portion of IDR-2 having a greater disruptive impact on droplet formation (Fig. 3b, c, Extended Data Fig. 6a). These results suggest that condensate disruption may be a common consequence of RTT patient mutations.

Figure 3. RTT patient mutations disrupt MeCP2 condensate formation

a. Schematic of MeCP2 protein with bar chart displaying the number of *MECP2* coding mutations in female Rett syndrome patients found in RettBASE database for each amino acid position. Positions of nonsense, frameshift, and missense mutations are shown below.
b. Droplet experiments examining effects of RTT patient truncation mutations that disrupt IDR-2 on MeCP2 droplet formation. MeCP2-GFP WT and RTT IDR-2 mutants (R168X, R255X, R270X, R294X, P389X) at indicated concentrations were mixed with 40 nM methylated DNA in droplet formation buffers with 100 mM NaCl.
c. MeCP2-GFP condensed fraction as a function of MeCP2-GFP concentration for experiments in Fig. 3b. Mean±SD. Fields per condition n=15.
d. Droplet experiments examining the effects of RTT patient missense mutations that disrupt the MBD on MeCP2 droplet formation. MeCP2-GFP WT and RTT MBD mutants (R133C and T158M) at indicated concentrations were mixed with 20 nM methylated DNA in droplet formation buffers with 100 mM NaCl.
e. MeCP2-GFP condensed fraction as a function of MeCP2-GFP concentration for experiments in Fig. 3d. Mean±SD. Fields per condition n=15.

Figure 3

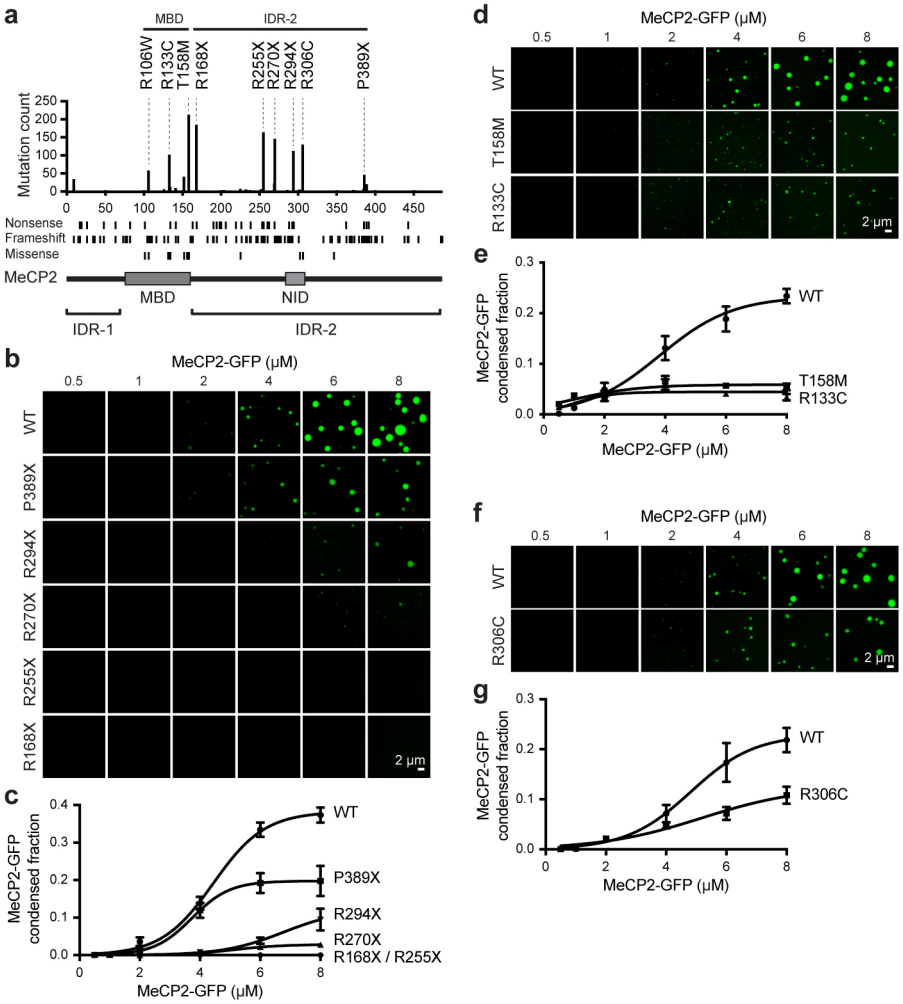


Figure 3. RTT patient mutations disrupt MeCP2 condensate formation

f. Droplet experiments examining effect of RTT patient missense mutation R306C on MeCP2 droplet formation. MeCP2-GFP WT and RTT R306C mutant at indicated concentrations were mixed with 20 nM methylated DNA in droplet formation buffers with 100 mM NaCl.

g. MeCP2-GFP condensed fraction as a function of MeCP2-GFP concentration for experiments in Fig. 3f. Mean±SD. Fields per condition n=15.

The observation that RTT patient missense mutations occur frequently in the MBD while truncation mutations occur frequently in IDR-2 (Fig. 3a) is consistent with expectations of a condensate model. Missense mutations in the structured MBD reduce condensate formation because DNA binding lowers the threshold for formation, while deletion mutations abrogate the multi-valent interactions that contribute to IDR-mediated condensate formation. Nonetheless, there are RTT patient missense mutations in IDR-2, so we investigated whether three of these (P225R, R306C, and P322L) disrupt condensate formation. All three mutations reduced the ability of MeCP2 to form droplets in vitro (Fig. 3f, g, Extended Data Fig. 6c-f). The R306C mutation was previously shown to disrupt an interaction between MeCP2's NCoR-interaction domain (NID) and TBLR1, a subunit of the NCoR co-repressor complex[13,21]. We therefore examined the ability of R306C mutant condensates to incorporate the C-terminal domain of TBLR1 (TBLR1-CTD), which directly interacts with the NID[21]. WT MeCP2 droplets readily enriched TBLR1-CTD-mCherry, whereas R306C mutant droplets showed less enrichment (Extended Data Fig. 7), suggesting that MeCP2 condensates can contribute to NID-mediated NCoR recruitment, a key MeCP2 function previously shown to be disrupted in RTT[13,21]. These results suggest that RTT patient missense mutations in IDR-2 contribute to condensate disruption.

A minimal MeCP2 fragment (Mini), which removes most of IDR-2 but retains the NID (and thus R306) (Extended Data Fig. 8a), can partially rescue RTT phenotypes in a murine model of RTT[22]. This observation led us to investigate whether MeCP2 Mini protein is capable of droplet formation. MeCP2 Mini was capable of forming droplets (Extended Data Fig. 8b-d) that could enrich DNA and HP1α-mCherry (Extended Data Fig. 8e-g), as well as TBLR1-CTD-mCherry (Extended Data Fig. 8h, i). Furthermore, live-cell imaging of mESCs expressing endogenously tagged MeCP2-GFP WT and Mini showed that both proteins partitioned similarly into heterochromatin condensates (Extended Data Fig. 8j, k). These results show that MeCP2 Mini retains condensate formation capabilities and suggest that the capability to form condensates may contribute to MeCP2 Mini's partial rescue of RTT phenotypes.

To explore the possibility that patient mutations causing IDR-2 loss lead to deficiencies in condensate incorporation in living cells, we focused on the common R168X patient mutation, which completely deletes IDR-2 and corresponds to the ΔIDR-2 deletion mutant used to examine MeCP2's condensate-forming ability in vitro (Fig. 2b-d, Extended Data Fig. 3m-o). We examined mESCs expressing endogenously tagged MeCP2-GFP WT and R168X-GFP mutant proteins (Extended Data Fig. 9). Live-cell imaging showed a striking reduction in the ability of mutant protein to partition into heterochromatin condensates (Extended Data Fig. 9a, b). Reduced partitioning was not a simple consequence of decreased mutant protein abundance (Extended Data Fig. 9c, d), as partitioning into heterochromatin condensates was not rescued by overexpression of the R168X mutant (Extended Data Fig. 10a-c). Reduced partitioning of MeCP2 into heterochromatin condensates was also observed in R168X mutant neurons (Fig. 4a, b, Extended Data Fig. 10d-f). These results indicate that mutations that occur in RTT patients reduce MeCP2's condensate interactions in cells.

Figure 4

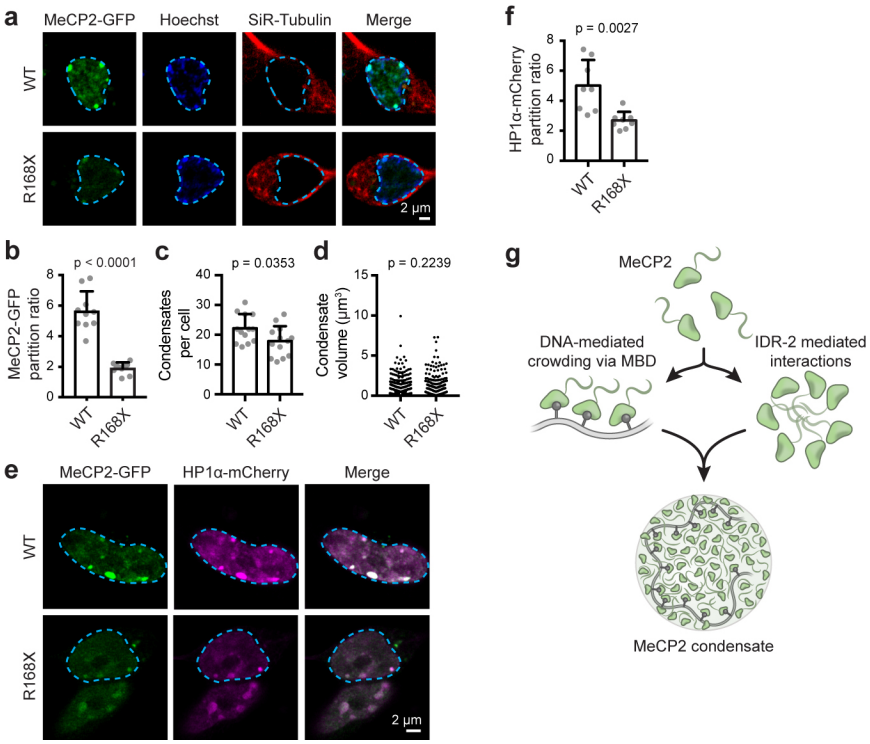


Figure 4. R168X mutant MeCP2 displays reduced partitioning into heterochromatin condensates and causes disease-relevant cellular phenotypes in neurons

a. Live-cell images of endogenous-tagged MeCP2-GFP WT and R168X mutant proteins with Hoechst and SiR-Tubulin staining in neurons.
b. Partition ratios of MeCP2-GFP proteins at heterochromatin condensates for experiments in Fig. 4a. Mean±SD, n=10 cells per condition. Two-tailed Student's t-test: p<0.0001, t=8.8921, df=18.
c. Number of heterochromatin condensates per cell in endogenous-tagged MeCP2-GFP WT and R168X mutant neurons. Mean±SD, n=13 cells per condition. Two-tailed Student's t-test: p=0.0353, t=2.2314, df=24.
d. Heterochromatin condensate volumes in endogenous-tagged MeCP2-GFP WT and R168X mutant neurons. Mean±SD, condensates per condition: WT (n=311), R168X (n=252). Two-tailed Student's t-test: p=0.2239, t=1.2176, df=561.
e. Live-cell images of endogenous-tagged MeCP2-GFP (WT or R168X mutant) and HP1α-mCherry in neurons.
f. Partition ratios of HP1α-mCherry at heterochromatin condensates for experiments in Fig. 4e. Mean±SD, n=8 cells per condition. Two-tailed Student's t-test: p=0.0027, t=3.6444, df=14.
g. Model of interactions contributing to MeCP2 condensate formation with DNA.

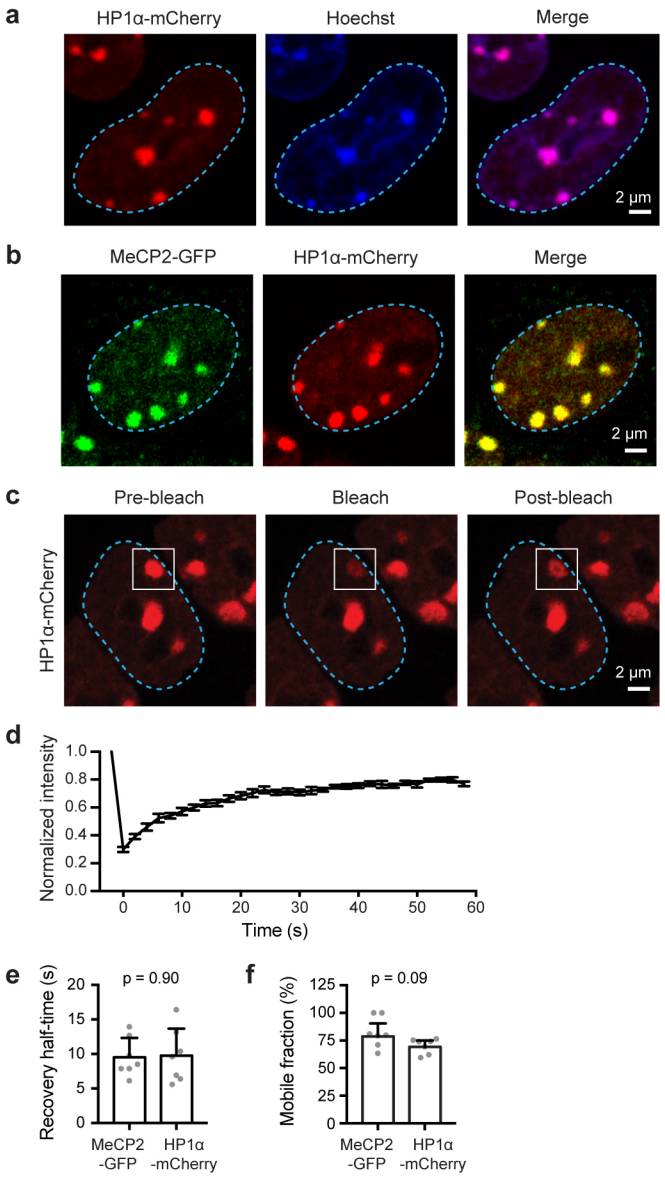
RTT is associated with various cellular phenotypes, including altered chromatin architecture[23], disrupted cofactor recruitment[13,15], and widespread transcriptional dysregulation[8,20,24]. R168X mutant mESCs showed evidence of each of these disease-associated cellular phenotypes. R168X mutant mESCs displayed changes in chromatin architecture, as heterochromatin condensates increased in number (Extended Data Fig. 9e) but decreased in volume (Extended Data Fig. 9f). Mutant cells showed reduced ability to partition HP1α cofactor into heterochromatin condensates (Extended Data Fig. 9g, h), which was not due to reduced HP1α abundance (Extended Data Fig. 9i), consistent with the ability of MeCP2 condensates to selectively partition and concentrate HP1α in vitro (Extended Data Fig. 4b-g). R168X mutant mESCs displayed evidence of widespread transcriptional dysregulation with loss of heterochromatin-associated repetitive element silencing (Extended Data Fig. 9j), reduced total RNA abundance (Extended Data Fig. 9k), and broad downregulation of euchromatic genes (Extended Data Fig. 9l). These RTT-associated cellular phenotypes were also observed in R168X mutant neurons (Fig. 4c-f, Extended Data Fig. 10g-i). Thus, the loss of the IDR-2 domain, which plays a major role in condensate formation, produced a range of cellular phenotypes associated with RTT.

3. DISCUSSION

We propose a condensate model for MeCP2 (Fig. 4g) that incorporates our conventional understanding of the mechanisms by which MeCP2 dysregulation contributes to cellular phenotypes, but adds the view that large numbers of MeCP2 molecules, using multiple weak and dynamic interactions, form membraneless bodies that can concentrate and compartmentalize additional components engaged in heterochromatin function. A recent study also reported that MeCP2 exhibits condensate properties that may be relevant to its interaction with histone H1[25]. Our results suggest a link between RTT mutations, altered MeCP2 condensate properties, and disease-associated cellular phenotypes. MeCP2's MBD and IDR-2 domains are both required for efficient condensate formation, and because multiple RTT mutations in these domains disrupt condensate formation, condensate disruption may be a common pathway for disease pathology caused by mutations in both domains. RTT mutations can also reduce MeCP2 protein levels[26], which may contribute to condensate disruption, as condensates can be highly sensitive to protein concentration[10]. RTT mutations are a leading cause of intellectual disability in females, yet evidence in animal models indicates that some symptoms may be reversible if a suitable therapy were developed[22,27,28]. We suggest that new approaches to pharmacological modification of condensate behaviors[29,30], if developed to selectively impact heterochromatin condensates, might provide therapeutic benefits for patients with RTT.

4. EXTENDED DATA FIGURES

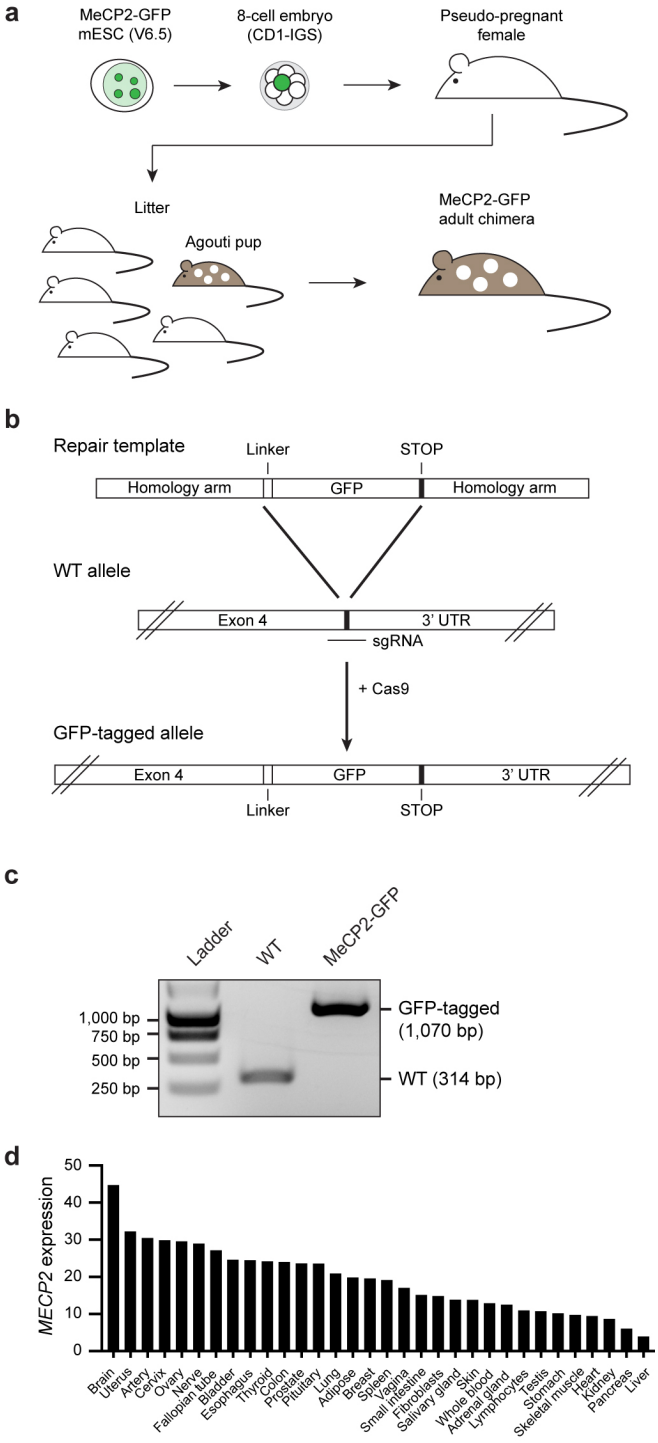
Extended Data Figure 1



Extended Data Figure 1. MeCP2 and HP1α are dynamic components of heterochromatin condensates

a. Live-cell images of endogenous-tagged HP1α-mCherry and Hoechst staining in mESCs.
b. Live-cell images of endogenous-tagged MeCP2-GFP and HP1α-mCherry in mESCs.
c. Live-cell images of FRAP experiments with endogenously tagged HP1α-mCherry mESCs.
d. FRAP curves for experiments in Extended Data Fig. 1c. Photobleaching occurs at t = 0 s. Mean±SEM, n=7 cells.
e. Half-time of photobleaching recovery for MeCP2-GFP and HP1α-mCherry at heterochromatin condensates in imaging experiments in Fig. 1b and Extended Data Fig. 1c. Mean±SEM, n=7 cells per condition. Two-tailed Student's t-test: p=0.90, t=0.13, df=12.
f. Mobile fractions of MeCP2-GFP and HP1α-mCherry within heterochromatin condensates in imaging experiments in Fig. 1b and Extended Data Fig. 1c, determined by FRAP analysis. Mean±SEM, n=7 cells per condition. Two-tailed Student's t-test: p=0.09, t=1.87, df=12.

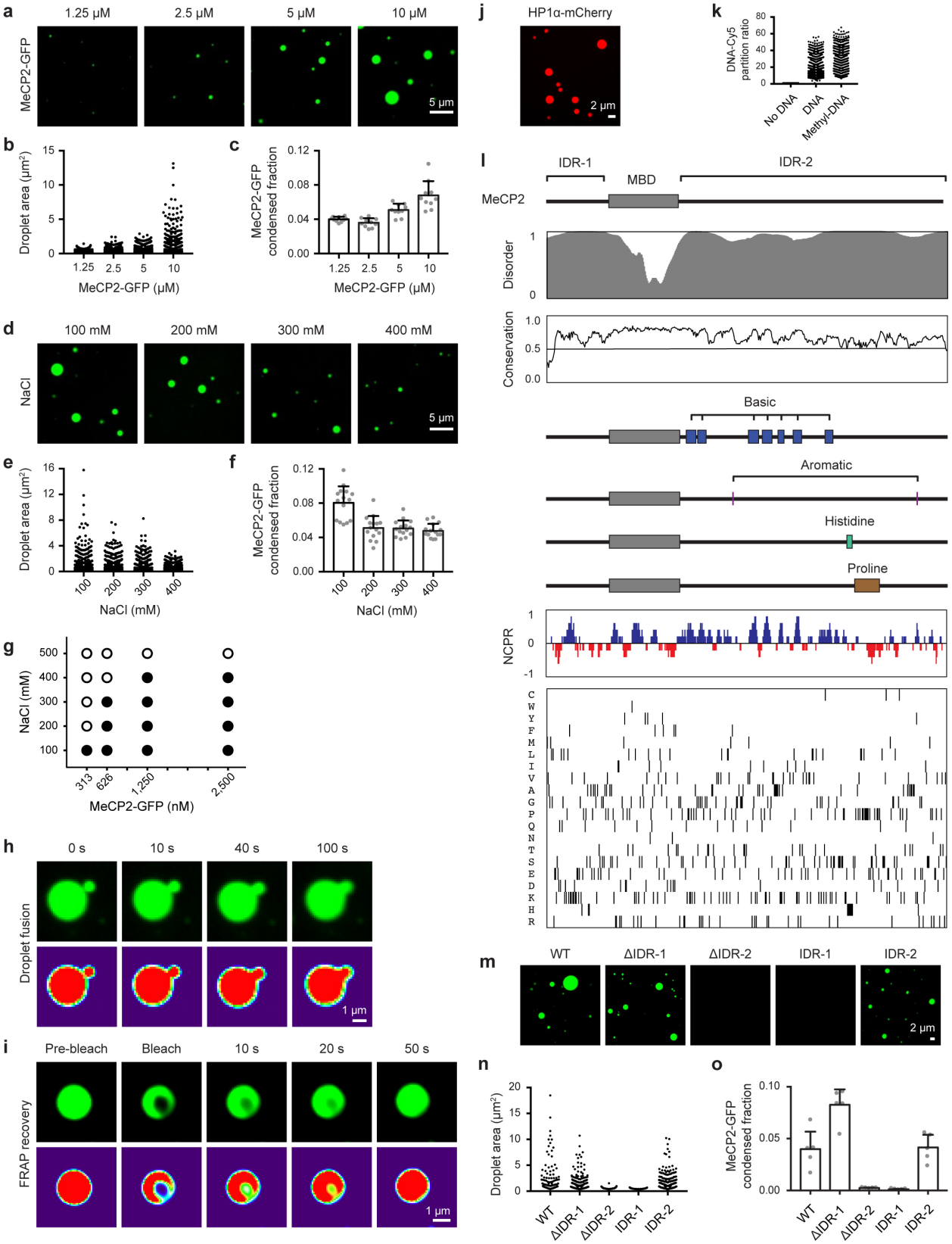
Extended Data Figure 2



Extended Data Figure 2. Generation of endogenous-tagged MeCP2-GFP chimeric mice

a. Schematic of MeCP2-GFP chimeric mouse generation using endogenous-tagged MeCP2-GFP mESCs. Endogenous-tagged MeCP2-GFP mESCs derived from V6.5 background were injected into embryos from CD1-IGS mice and multiple embryos were implanted into pseudo-pregnant female mice. Chimeric pups were distinguished from non-chimeric littermates by agouti coat color. MeCP2-GFP tagged adult chimeric mice were used for experiments.
b. Schematic of strategy used to generate endogenous-tagged MeCP2-GFP mESCs.
c. PCR genotyping of endogenous-tagged MeCP2-GFP mESCs. For gel source data, see Supplementary Figure 1.
d. *MECP2* expression values in transcripts per million (TPM) measured by RNA-seq for various human tissues surveyed by GTEx. Tissues are ordered based on expression level. TPM values greater than 1 are considered to be expressed.

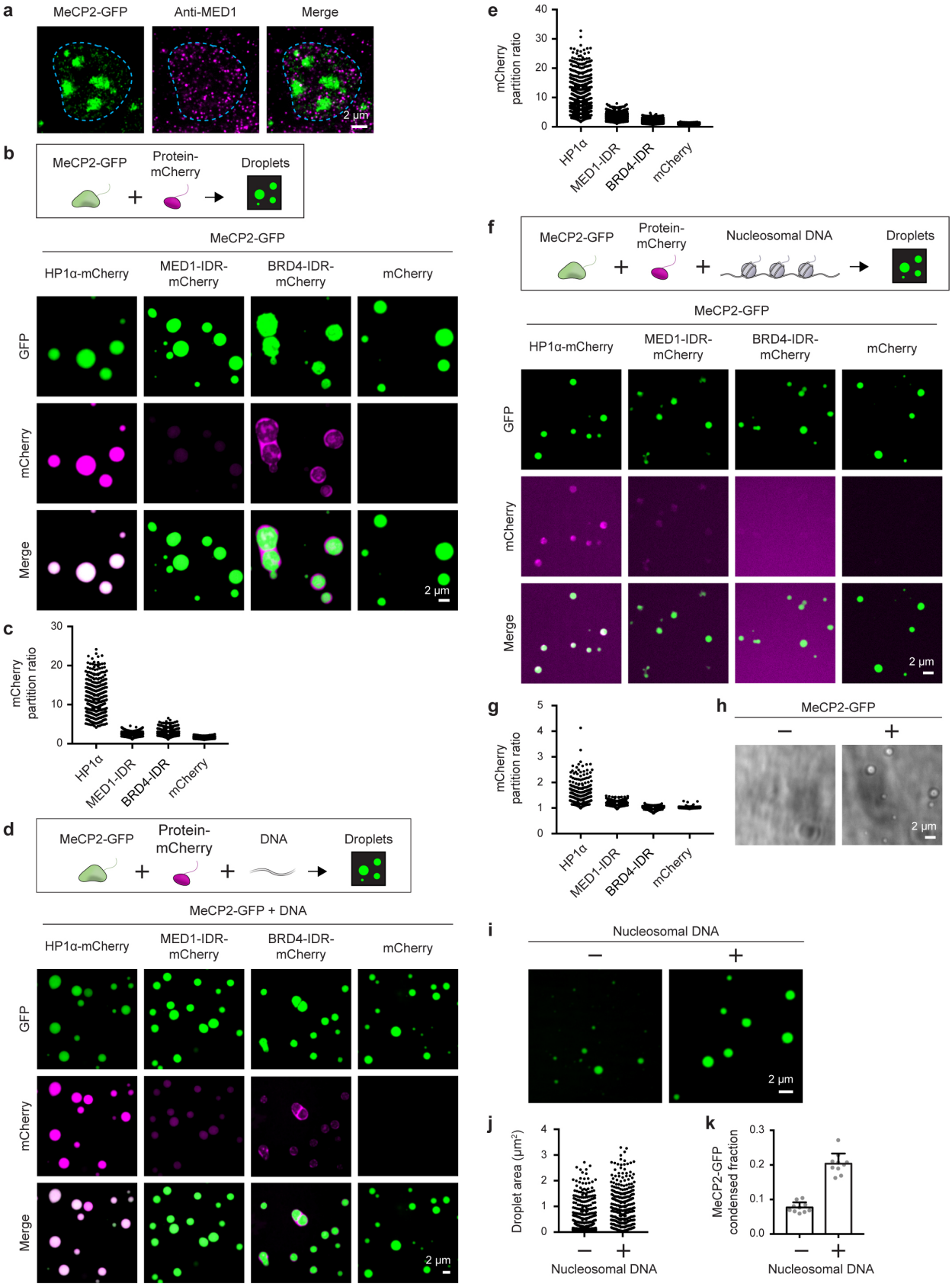
Extended Data Figure 3



Extended Data Figure 3. MeCP2 forms phase-separated droplets in vitro

a. Droplet experiments examining effect of MeCP2 concentration. MeCP2-GFP was added to droplet formation buffers with 100 mM NaCl and 10% PEG-8000.
b. Droplet areas for experiments in Extended Data Fig. 3a. Fields per condition n=10.
c. MeCP2-GFP condensed fraction for experiments in Extended Data Fig. 3a. Mean±SD. Fields per condition n=10.
d. Droplet experiments examining effect of salt concentration. MeCP2-GFP at 10 μM was added to droplet formation buffers with indicated NaCl concentrations and 10% PEG-8000.
e. Droplet areas for experiments in Extended Data Fig. 3d. Fields per condition n=15.
f. MeCP2-GFP condensed fraction for experiments in Extended Data Fig. 3d. Mean±SD. Fields per condition n=15.
g. Phase diagram of MeCP2 droplet formation. MeCP2-GFP was added to droplet formation buffers with indicated NaCl concentrations and 5% PEG-8000. Filled-in circles indicate conditions with droplets. Fields per condition n=10.
h. Droplet experiments showing MeCP2 droplet fusion. MeCP2-GFP at 10 μM was added to droplet formation buffers with 100 mM NaCl and 10% PEG-8000.
i. Droplet experiments showing MeCP2 droplet FRAP. Conditions same as in Extended Data Fig. 3h. Photobleaching at t=0 s.
j. Droplet experiments examining HP1α. HP1α-mCherry at 10 μM was added to droplet formation buffers with 100 mM NaCl and 10% PEG-8000.
k. DNA-Cy5 partition ratios in MeCP2-GFP droplets for experiments in Fig. 1g. Fields per condition n=15.
l. Expanded schematic of MeCP2 protein (Fig. 2a) with protein sequence conservation, net charge per residue (NCPR), and residue plots.
m. Droplet experiments examining MeCP2 deletion mutants. MeCP2-GFP deletion mutants at 10 μM were added to droplet formation buffers with 100 mM NaCl and 10% PEG-8000.
n. Droplet areas for experiments in Extended Data Fig. 2m. Fields per condition n=5.
o. MeCP2-GFP condensed fraction for experiments in Extended Data Fig. 2m. Mean±SD. Fields per condition n=5.

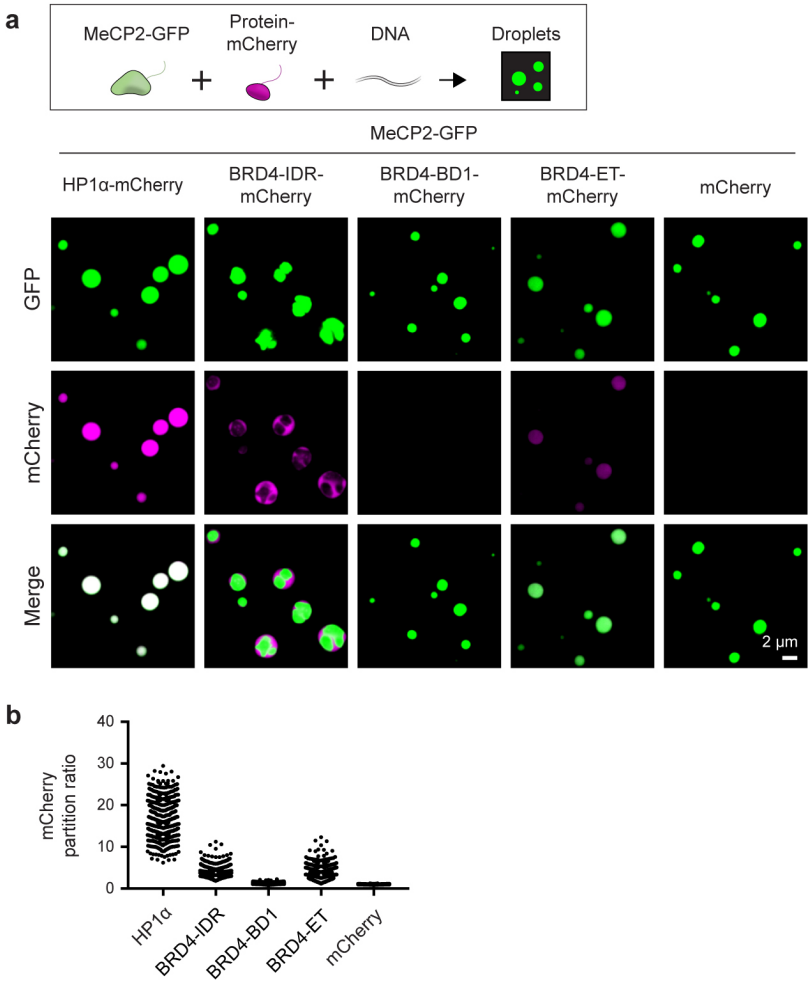
Extended Data Figure 4



Extended Data Figure 4. MeCP2 condensates preferentially concentrate HP1α compared to components of transcriptional condensates

a. Immunofluorescence images of heterochromatin condensates (MeCP2-GFP) and transcriptional condensates (Anti-MED1) in mESCs.
b. Droplet experiments examining ability of MeCP2 condensates to preferentially concentrate HP1α compared to components of transcriptional condensates. MeCP2-GFP at 7.5 μM was mixed with HP1α-mCherry, MED1-IDR-mCherry, BRD4-IDR-mCherry, or mCherry at 7.5 μM in droplet formation buffers with 150 mM NaCl and 10% PEG-8000.
c. mCherry partition ratios in MeCP2-GFP droplets for experiments in Extended Data Fig. 4b. Fields per condition n=15.
d. Droplet experiments with naked DNA examining ability of MeCP2 condensates to preferentially concentrate HP1α compared to components of transcriptional condensates. Conditions same as in Extended Data Fig. 4b, but with addition of 160 nM DNA.
e. mCherry partition ratios in MeCP2-GFP droplets for experiments in Extended Data Fig. 4d. Fields per condition n=15.
f. Droplet experiments with nucleosomal DNA examining ability of MeCP2 condensates to preferentially concentrate HP1α compared to components of transcriptional condensates. MeCP2-GFP at 5 μM was mixed with HP1α-mCherry, MED1-IDR-mCherry, BRD4-IDR-mCherry, or mCherry at 5 μM and 6 nM poly-nucleosomes in droplet formation buffers with 100 mM NaCl and 3 mM $MgCl_2$.
g. mCherry partition ratios in MeCP2-GFP droplets for experiments in Extended Data Fig. 4f. Fields per condition n=10.
h. Brightfield images examining droplet formation with nucleosomal DNA alone and with MeCP2. Poly-nucleosomes at 6 nM were mixed with 5 μM MeCP2-GFP or no MeCP2-GFP in droplet formation buffers with 100 mM NaCl and 3 mM $MgCl_2$.
i. Droplet experiments examining MeCP2 droplet formation with nucleosomal DNA. Conditions same as in Extended Data Fig. 4h.
j. Droplet areas for experiments in Extended Data Fig. 4i. Fields per condition n=10.
k. MeCP2-GFP condensed fraction for experiments in Extended Data Fig. 4i. Mean±SD. Fields per condition n=10.

Extended Data Figure 5

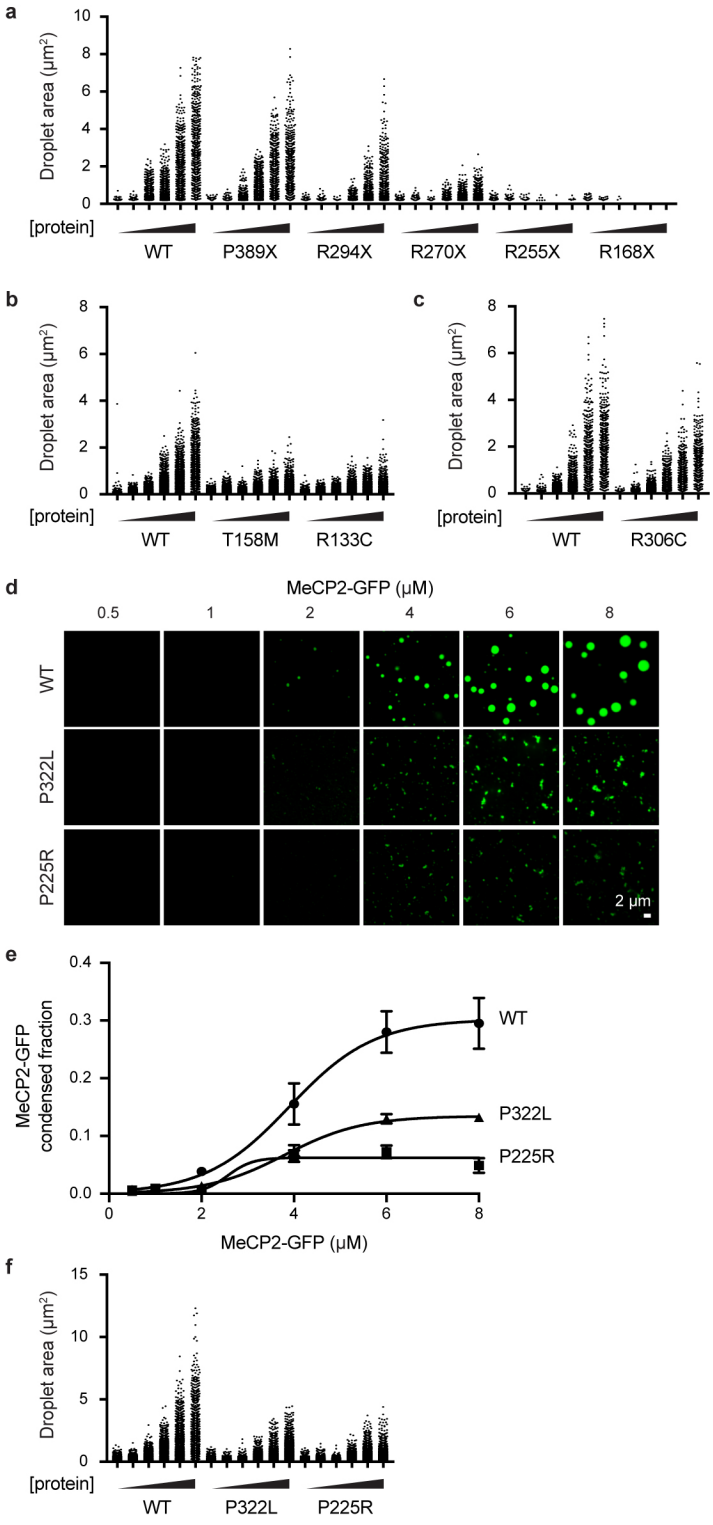


Extended Data Figure 5. MeCP2 condensate partitioning of BRD4 domains

a. Droplet experiments examining ability of MeCP2 condensates to preferentially incorporate and concentrate HP1α compared to BRD4 domains in the presence of naked DNA. BRD4-IDR, bromodomain 1 (BD1), and extra-terminal (ET) domain were examined. MeCP2-GFP at 7.5 μM was mixed with either HP1α-mCherry, BRD4-IDR-mCherry, BRD4-BD1-mCherry, BRD4-ET-mCherry, or mCherry each at 7.5 μM and 160 nM methylated DNA in droplet formation buffers with 150 mM NaCl and 10% PEG-8000.

b. mCherry partition ratios in MeCP2-GFP droplets for experiments in Extended Data Fig. 5a. Fields per condition n=15.

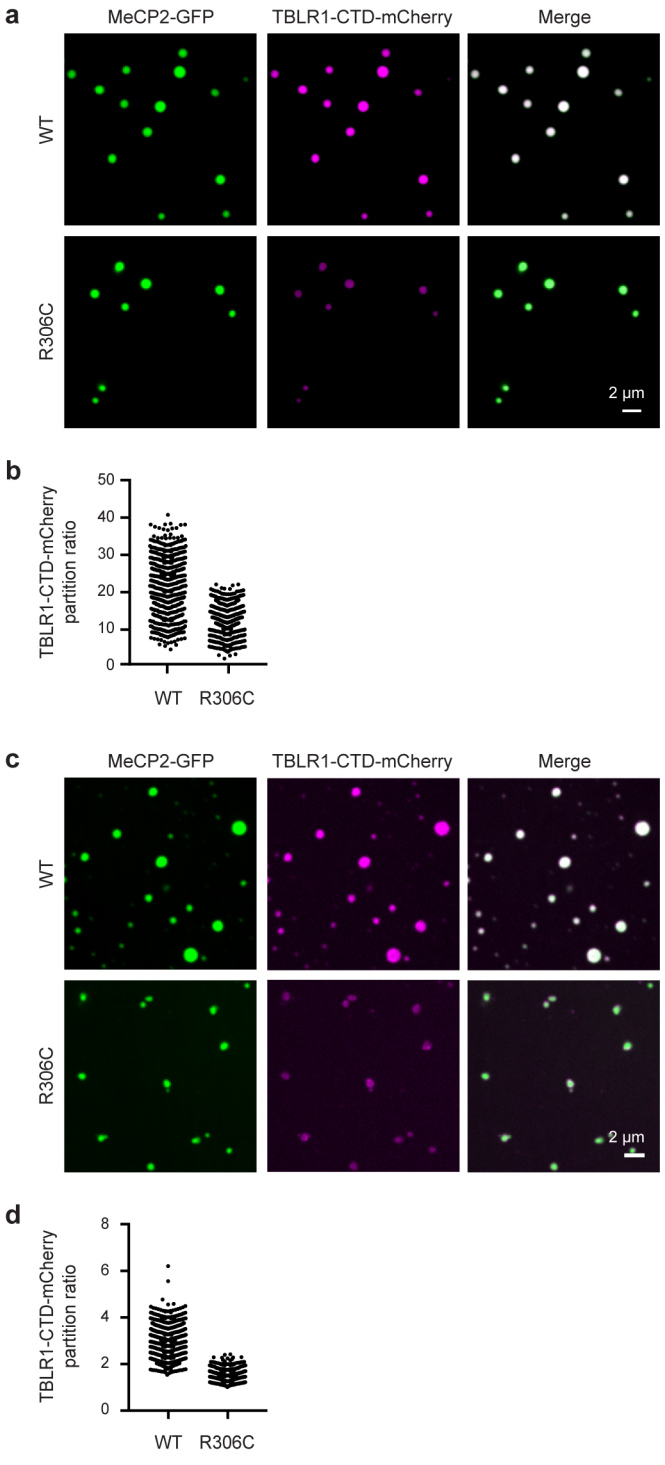
Extended Data Figure 6



Extended Data Figure 6. RTT patient mutations disrupt MeCP2 condensate formation

a. Droplet areas for experiments in Fig. 3b. Fields per condition n=15.
b. Droplet areas for experiments in Fig. 3d. Fields per condition n=15.
c. Droplet areas for experiments in Fig. 3f. Fields per condition n=15.
d. Droplet experiments examining effects of RTT patient missense mutations that disrupt IDR-2 on MeCP2 droplet formation. MeCP2-GFP WT and RTT IDR-2 mutants (P225R and P322L) at indicated concentrations were mixed with 20 nM methylated DNA in droplet formation buffers with 100 mM NaCl.
e. MeCP2-GFP condensed fraction as a function of MeCP2-GFP concentration for experiments in Extended Data Fig. 6d. Mean±SD. Fields per condition n=15.
f. Droplet areas for experiments in Extended Data Fig. 6d. Fields per condition n=15.

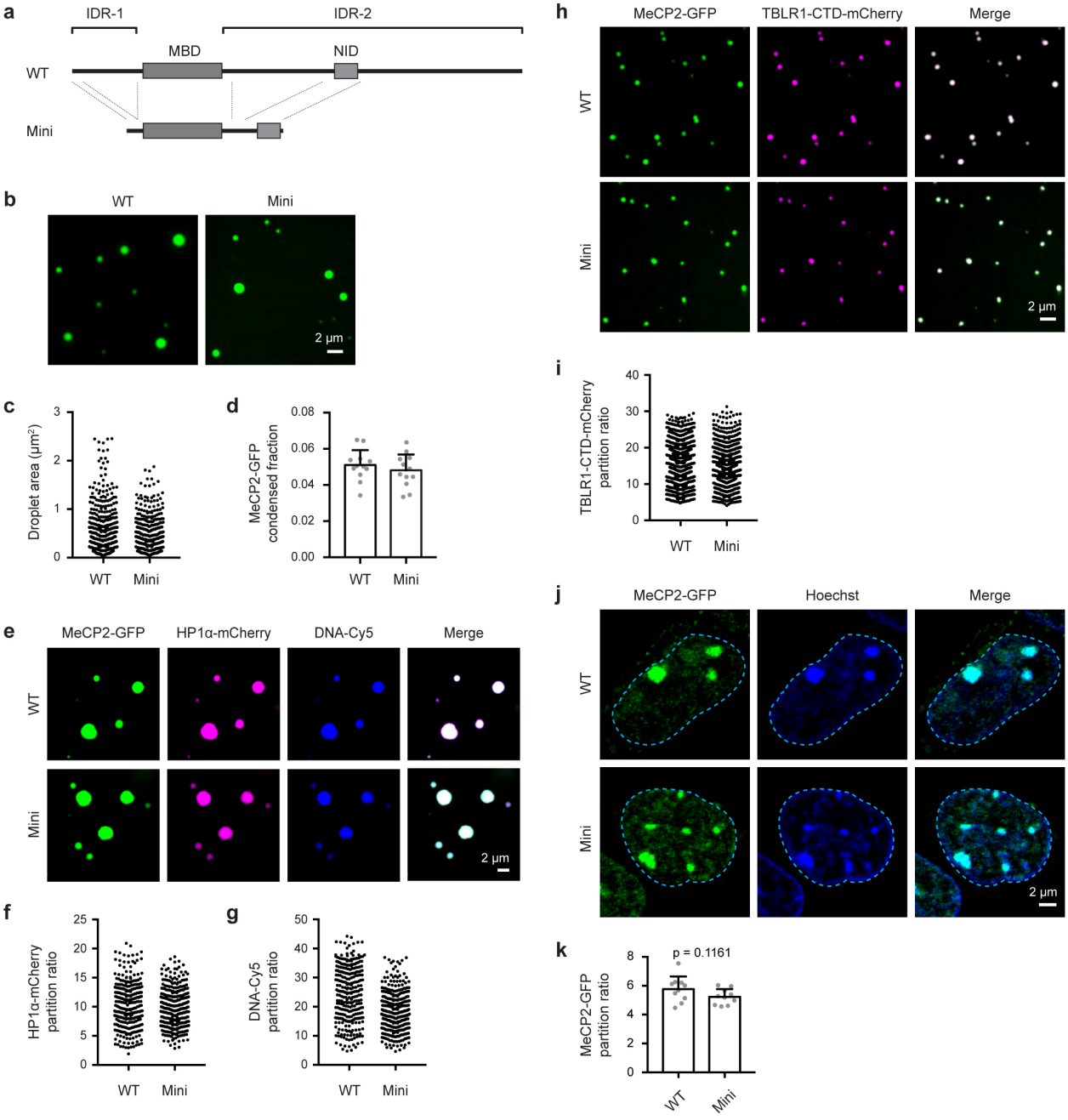
Extended Data Figure 7



Extended Data Figure 7. TBLR1 partitioning into MeCP2 droplets is disrupted by RTT mutation R306C

a. Droplet experiments examining ability of MeCP2 WT and R306C mutant condensates to enrich TBLR1-CTD. MeCP2-GFP WT or R306C mutant at 6 μM was mixed with TBLR1-CTD-mCherry at 10 μM in droplet formation buffers with 125 mM NaCl and 10% PEG-8000.
b. TBLR1-CTD-mCherry partition ratios in MeCP2-GFP WT and R306C mutant droplets for experiments in Extended Data Fig. 7a. Fields per condition n=15.
c. Droplet experiments examining ability of MeCP2 WT and R306C mutant condensates to enrich TBLR1-CTD. MeCP2-GFP WT or R306C mutant at 10 μM was mixed with TBLR1-CTD-mCherry at 4 μM in droplet formation buffers with 125 mM NaCl.
d. TBLR1-CTD-mCherry partition ratios in MeCP2-GFP WT and R306C mutant droplets for experiments in Extended Data Fig. 7c. Fields per condition n=12.

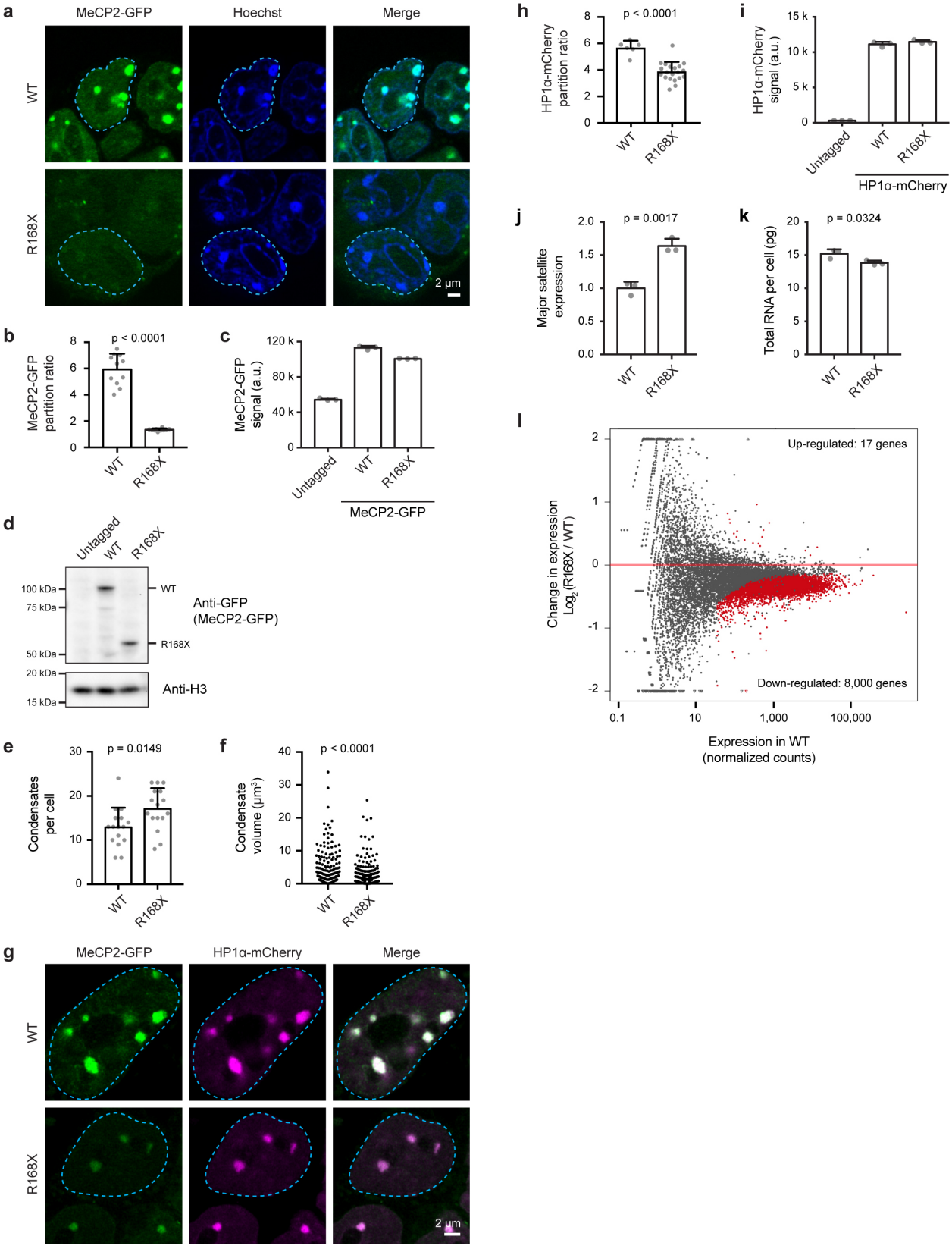
Extended Data Figure 8



Extended Data Figure 8. MeCP2 Mini forms droplets in vitro and partitions into heterochromatin condensates in mESCs

a. Schematic of MeCP2 protein with a minimal MeCP2 protein (Mini) (Tillotson et al., Nature 2017) that retains the MBD and NID displayed below.
b. Droplet experiments examining ability of Mini MeCP2 to form droplets. MeCP2-GFP WT and Mini at 4 μM were added to droplet formation buffers with 125 mM NaCl and 10% PEG-8000.
c. Droplet areas for experiments in Extended Data Fig. 8b. Fields per condition n=12.
d. MeCP2-GFP condensed fraction for experiments in Extended Data Fig. 8b. Mean±SD. Fields per condition n=12.
e. Droplet experiments examining ability of MeCP2 WT and Mini to form droplets with HP1α and DNA. MeCP2-GFP WT or Mini at 7.5 μM was mixed with 7.5 μM HP1α-mCherry and 160 nM DNA in droplet formation buffers with 150 mM NaCl and 10% PEG-8000.
f. HP1α-mCherry partition ratios in MeCP2-GFP WT and Mini droplets for experiments in Extended Data Fig. 8e. Fields per condition n=15.
g. DNA-Cy5 partition ratios in MeCP2-GFP WT and Mini droplets for experiments in Extended Data Fig. 8e. Fields per condition n=15.
h. Droplet experiments examining ability of MeCP2 WT and Mini condensates to enrich TBLR1-CTD. MeCP2-GFP WT or Mini at 4 μM was mixed with TBLR1-CTD-mCherry at 10 μM in droplet formation buffers with 125 mM NaCl and 10% PEG-8000.
i. TBLR1-CTD-mCherry partition ratios in MeCP2-GFP WT and Mini droplets for experiments in Extended Data Fig. 8h. Fields per condition n=12.
j. Live-cell microscopy of endogenous-tagged MeCP2-GFP WT and Mini proteins with Hoechst staining in mESCs.
k. Partition ratios of MeCP2-GFP proteins at heterochromatin condensates for experiments in Extended Data Fig. 8j. Mean±SD, n=10 cells per condition. Two-tailed Student's t-test: p=0.1161, t=1.6509, df=18.

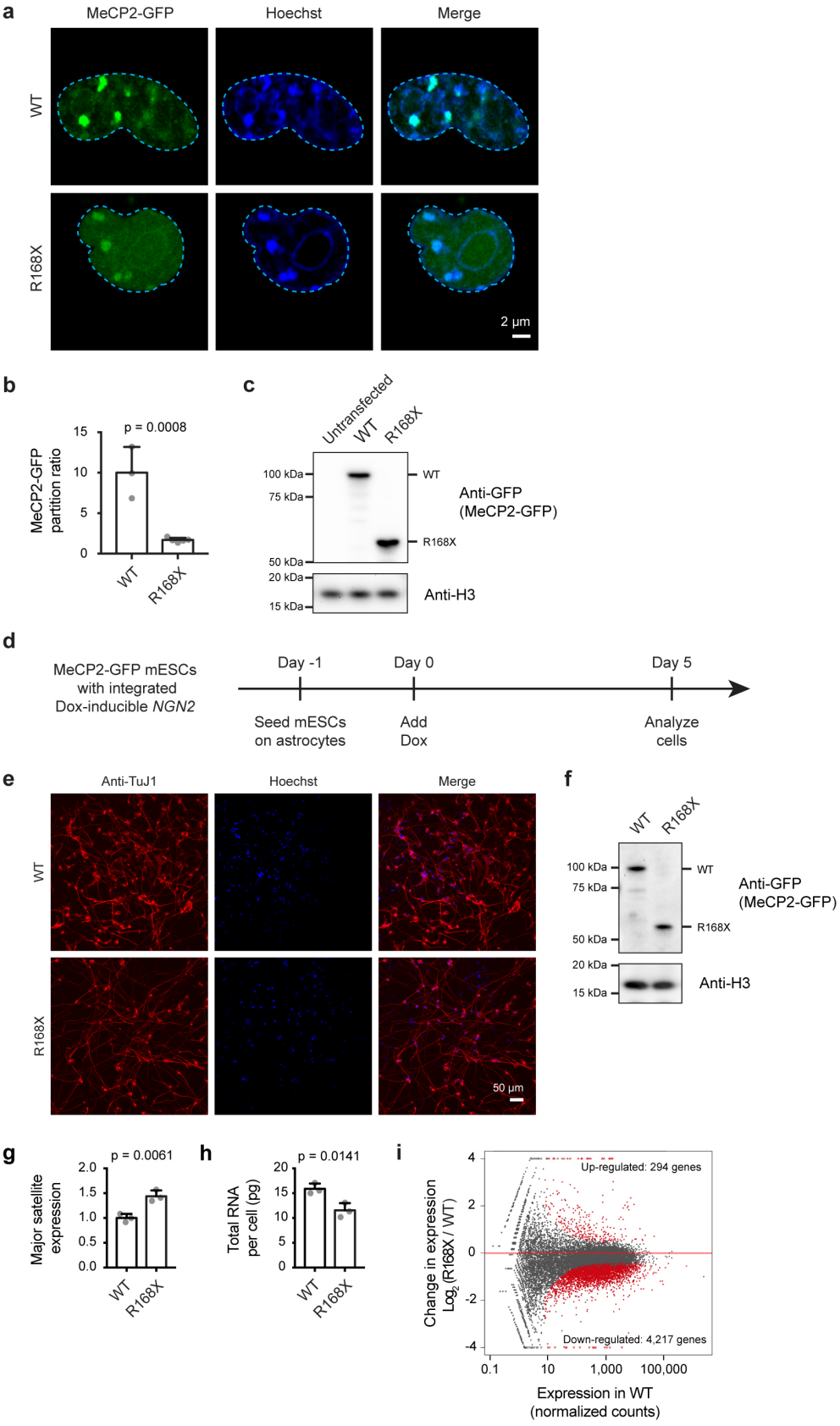
Extended Data Figure 9



Extended Data Figure 9. R168X mutant MeCP2 displays reduced partitioning into heterochromatin condensates and causes disease-relevant cellular phenotypes in mESCs

a. Live-cell images of endogenous-tagged MeCP2-GFP WT and R168X mutant proteins with Hoechst staining in mESCs.
b. Partition ratios of MeCP2-GFP proteins at heterochromatin condensates for experiments in Extended Data Fig. 9a. Mean±SD, cells per condition: WT (n=11), R168X (n=10). Two-tailed Student's t-test: $p<0.0001$, $t=12.13$, $df=19$.
c. MeCP2-GFP signal in endogenous-tagged MeCP2-GFP WT and R168X mutant mESCs measured by flow cytometry. Mean±SD, n=3 biologically independent samples per condition.
d. Western blot of endogenous-tagged MeCP2-GFP WT and R168X mutant mESCs. Anti-H3 was used as a processing control. For gel source data, see Supplementary Figure 1.
e. Number of heterochromatin condensates per cell in endogenous-tagged MeCP2-GFP WT and R168X mutant mESCs. Mean±SD, n=16 cells per condition. Two-tailed Student's t-test: $p=0.0149$, $t=2.5832$, $df=30$.
f. Heterochromatin condensate volumes in endogenous-tagged MeCP2-GFP WT and R168X mutant mESCs. Mean±SD, condensates per condition: WT (n=206), R168X (n=273). Two-tailed Student's t-test: $p<0.0001$, $t=4.2065$, $df=477$.
g. Live-cell images of endogenous-tagged MeCP2-GFP (WT or R168X mutant) and HP1α-mCherry in mESCs.
h. Partition ratios of HP1α-mCherry at heterochromatin condensates for experiments in Extended Data Fig. 9g. Mean±SD, cells per condition: WT (n=6), R168X (n=20). Two-tailed Student's t-test: $p<0.0001$, $t=5.7136$, $df=24$.
i. HP1α-mCherry signal in endogenous-tagged MeCP2-GFP (WT or R168X mutant) and HP1α-mCherry mESCs measured by flow cytometry. Mean±SD, n=3 biologically independent samples per condition.
j. Normalized major satellite repeat expression in endogenous-tagged MeCP2-GFP WT and R168X mutant mESCs. Mean±SD, n=3 biologically independent samples per condition. Two-tailed Student's t-test: $p=0.0017$, $t=7.5436$, $df=4$.
k. Total RNA per cell in endogenous-tagged MeCP2-GFP WT and R168X mutant mESCs. Mean±SD, n=3 biologically independent samples per condition. Two-tailed Student's t-test: $p=0.0324$, $t=3.2154$, $df=4$.
l. RNA-seq comparing endogenous-tagged MeCP2-GFP WT and R168X mutant mESCs. Differentially expressed genes (red dots) were determined by two-tailed Wald test with multiple test adjusted p-value<0.1. For both conditions, n=3 biologically independent samples.

Extended Data Figure 10

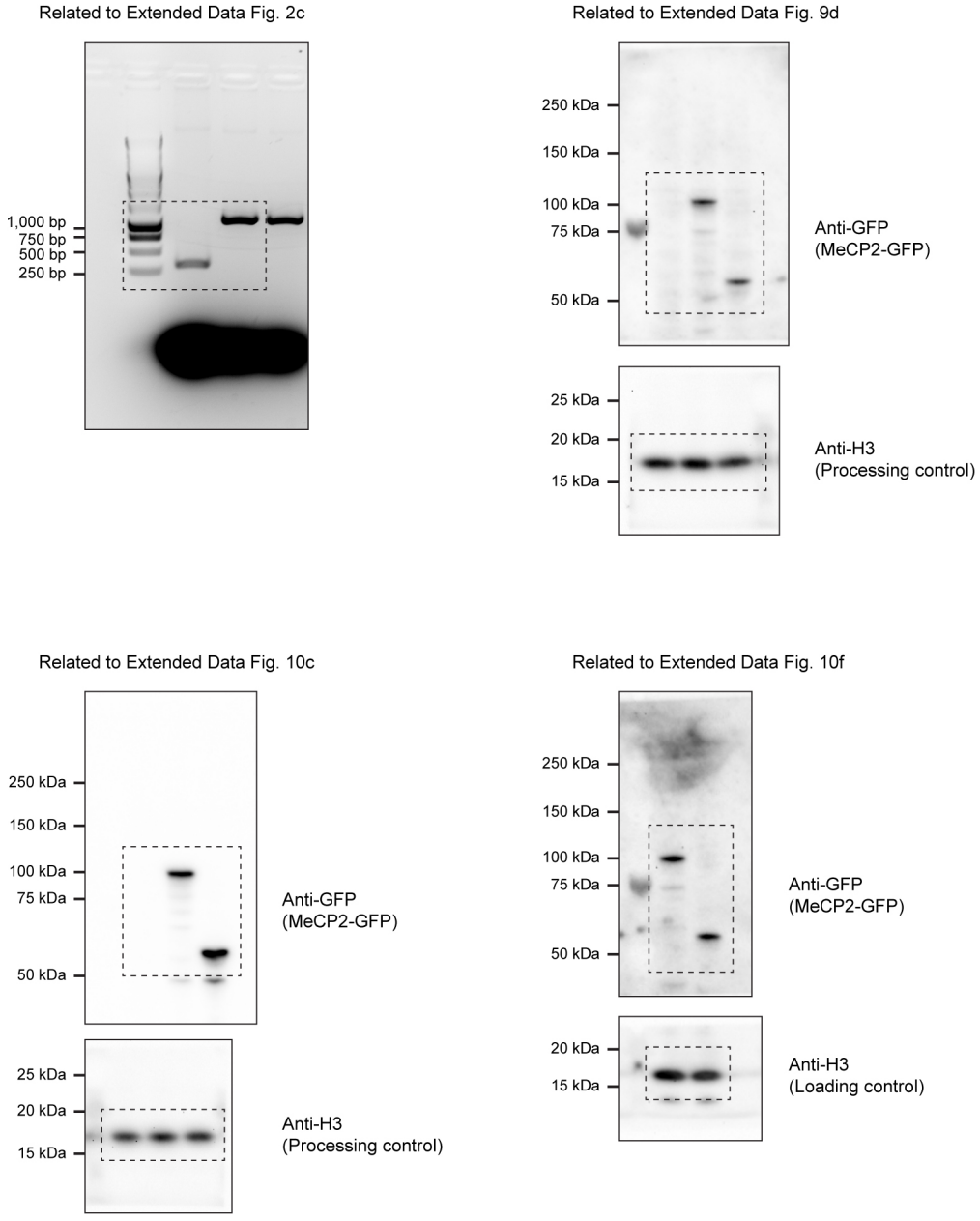


Extended Data Figure 10. R168X mutant in mESCs and neurons

a. Live-cell images of mESCs overexpressing either WT or R168X mutant MeCP2-GFP.
b. Partition ratios of MeCP2-GFP proteins at heterochromatin condensates relative to the nucleoplasm for experiments in Extended Data Fig. 10a. Mean±SD, cells per condition: WT (n=3), R168X (n=5). Two-tailed Student's t-test: p=0.0008, t=6.1529, df=6.
c. Western blot of mESCs overexpressing either WT or R168X mutant MeCP2-GFP. Anti-H3 was used as a processing control. For gel source data, see Supplementary Figure 1.
d. Schematic of generation of mESC-derived neurons. Endogenous-tagged MeCP2-GFP WT and R168X mESCs were modified for Dox-inducible *NGN2* expression using the PiggyBac system. Prior to neuronal differentiation, mESCs were seeded on astrocytes. Neuronal differentiation was induced by adding doxycycline to drive *NGN2* expression. Five days after induction of *NGN2* expression, neurons were analyzed.
e. Fixed-cell immunofluorescence images of neurons derived from MeCP2-GFP WT and R168X mutant mESCs. Anti-TuJ1 staining was used to distinguish neurons.
f. Western blot of endogenous-tagged MeCP2-GFP WT and R168X mutant neurons. Anti-H3 was used as a loading control. For gel source data, see Supplementary Figure 1.
g. Normalized major satellite repeat expression in endogenous-tagged MeCP2-GFP WT and R168X mutant neurons. Mean±SD, n=3 biologically independent samples per condition. Two-tailed Student's t-test: p=0.0061, t=5.3004, df=4.
h. Total RNA per cell in endogenous-tagged MeCP2-GFP WT and R168X neurons. Mean±SD, n=3 biologically independent samples per condition. Two-tailed Student's t-test: p=0.0141, t=4.1676, df=4.
i. RNA-seq comparing endogenous-tagged MeCP2-GFP WT and R168X mutant neurons. Differentially expressed genes (red dots) were identified using a two-tailed Wald test with multiple test adjusted p-value<0.1. For both conditions, n=3 biologically independent samples.

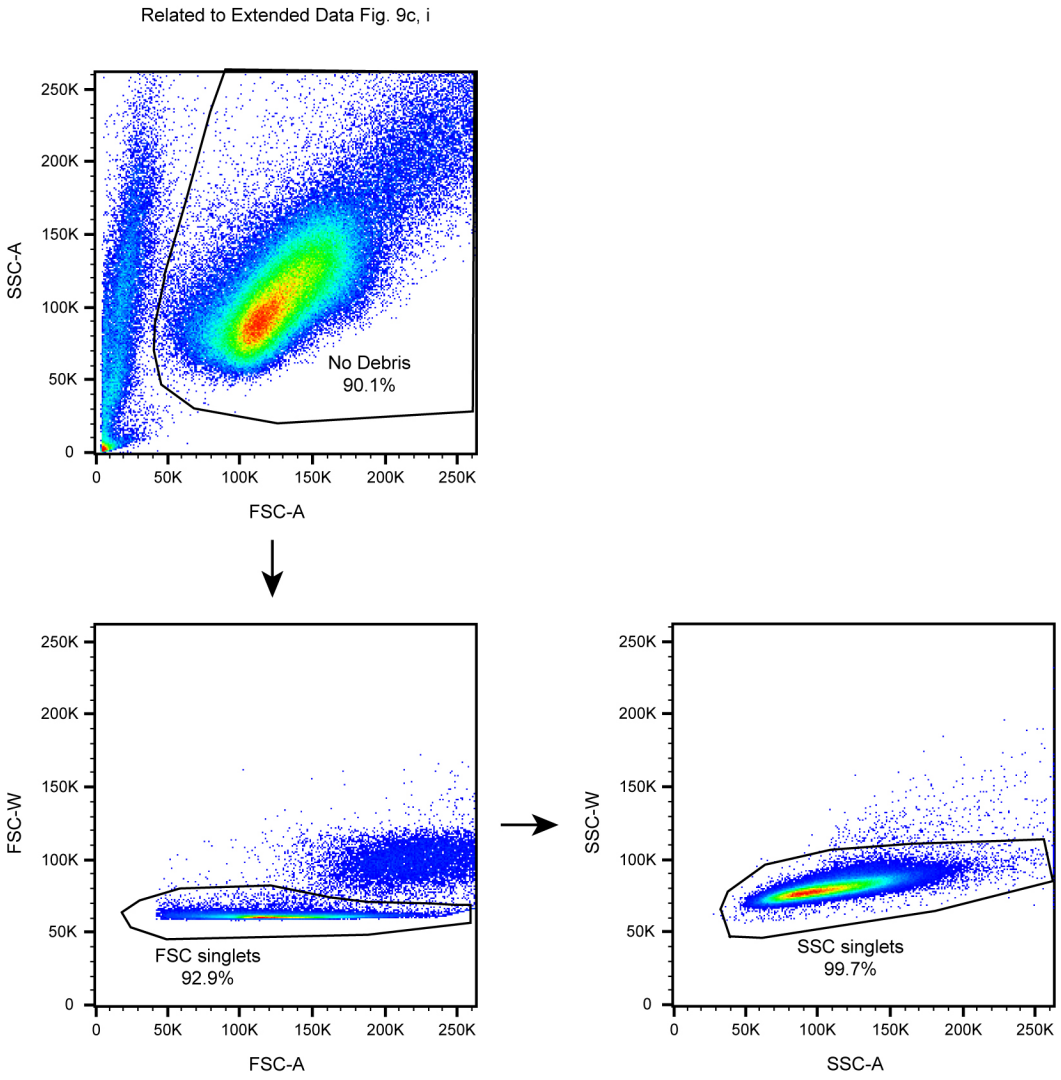
5. SUPPLEMENTARY FIGURES

Supplementary Figure 1



Supplementary Figure 1. Gel source data
Gel source data with size markers. Regions used for display in the indicated figures are denoted with dashed boxes.

Supplementary Figure 2



Supplementary Figure 2. Flow cytometry gating strategy
Example flow cytometry gating strategy. Related to Extended Data Fig. 9c, i.

6. METHODS

Cell culture

V6.5 murine embryonic stem cells (ESCs) were a gift from R. Jaenisch, and were authenticated by STR analysis compared to commercially acquired cells of the same line. MeCP2-GFP Mini (ΔNIC, Tillotson et al., Nature 2017) murine ESCs were a gift from A. Bird, and were not subject to authentication.

ESCs were cultured in 2i/LIF media on tissue culture treated plates coated with 0.2% gelatin (Sigma G1890). ESCs were grown in a humidified incubator with 5% CO_2 at 37 °C. Cells were passaged every 2-3 days by dissociation using TrypLE Express (Gibco 12604). The dissociation reaction was quenched using serum/LIF media. Cells were tested regularly for mycoplasma using the MycoAlert Mycoplasma Detection Kit (Lonza LT07-218) and found to be negative.

The composition of N2B27 media is as follows: DMEM/F12 (Gibco 11320) supplemented with 0.5X N2 supplement (Gibco 17502), 0.5X B27 supplement (Gibco 17504), 2 mM L-glutamine (Gibco 25030), 1X MEM non-essential amino acids (Gibco 11140), 100 U/mL penicillin-streptomycin (Gibco 15140), 0.1 mM 2-mercaptoethanol (Sigma M7522).

The composition of 2i/LIF media is as follows: N2B27 media, 3 μM CHIR99021 (Stemgent 04-0004), 1 μM PD0325901 (Stemgent 04-0006), and 1000 U/mL leukemia inhibitor factor (LIF) (ESGRO ESG1107).

The composition of serum/LIF media is as follows: KnockOut DMEM (Gibco 10829) supplemented with 15% fetal bovine serum (Sigma F4135), 2 mM L-glutamine (Gibco 25030), 1X MEM non-essential amino acids, 100 U/mL penicillin-streptomycin (Gibco 15140), 0.1 mM 2-mercaptoethanol (Sigma M7522), and 1000 U/mL leukemia inhibitor factor (LIF) (ESGRO ESG1107).

HEK293T cells were purchased from ATCC (ATCC CRL-3216) and cultured in DMEM (Gibco 11995-073) with 10% fetal bovine serum (Sigma F4135), 100 U/mL Penicillin-Streptomycin (Gibco 15140), 2 mM L-glutamine (Gibco 25030). Cells were not subject to authentication. Cells were tested regularly for mycoplasma using the MycoAlert Mycoplasma Detection Kit (Lonza LT07-218) and found to be negative.

Genome editing

The CRISPR/Cas9 system was used to generate genetically modified ESC lines. Target-specific sequences were cloned into a plasmid containing sgRNA backbone, a codon-optimized version of Cas9, and mCherry or BFP. For generation of the MeCP2-mEGFP and HP1a-mCherry endogenously tagged lines, homology directed repair templates were cloned into pUC19 using NEBuilder HiFi DNA Master Mix (NEB E2621S). The homology repair template consisted of mEGFP or mCherry cDNA sequence flanked on either side by 800 bp homology arms amplified from genomic DNA using PCR. The following sgRNA sequences with PAM sequence in parentheses were used for CRISPR/Cas9 targeting:
sgRNA_Mecp2_C-term: GTAAAGTCAGCTAACTCTCT (CGG)
sgRNA_Mecp2_R168: gAGGTGGTTTCTGCTCTCTCC (TGG)
sgRNA_Cbx5_C-term: gAAGAAAGCGCGAAGAGCTAA (AGG)

To generate genetically modified cell lines, 750,000 cells were transfected with 833 ng Cas9 plasmid and 1,666 ng non-linearized homology repair template using Lipofectamine 3000 (Invitrogen L3000). Cells were sorted 48 hours after transfection for the presence of either mCherry or BFP fluorescence proteins encoded on the Cas9 plasmid to enrich for transfected cells. This population was allowed to expand for 1 week before sorting a second time for the presence of GFP or mCherry. 40,000 GFP of mCherry positive cells were plated in serial dilution in a 6-well plate and allowed to expand for a week before individual colonies were manually picked into a 96-well plate. 24 colonies were screened for successful targeting using PCR genotyping to confirm insertion. PCR genotyping was performed using Phusion polymerase (Thermo Scientific F531S). Products were amplified according to kit recommendations and visualized on a 1% agarose gel. The following primers were used for PCR genotyping:
MeCP2-GFP_fwd: AGCAGCATCTGCAAAGAAGAG
MeCP2-GFP_rev: CAGAGCCCTACCCATAAGGAG
HP1α-mCherry_fwd: AACGTGAAGTGTCCACAGATTG
HP1α-mCherry_rev: TTATGGATGCGTTTAGGATGG
MeCP2-GFP_R168X_fwd: AGACACCTCCTTGGACCCTAA
MeCP2-GFP_R168X_rev: ACCCTTTTCACCTGAACACCT

Neuronal differentiation

Neurons were derived from mESCs by expression of *NGN2* to induce neuronal differentiation[31,32]. A doxycycline-inducible *NGN2* expression construct with a puromycin-resistance gene was integrated into mESCs using the PiggyBac transposon system. ESCs with successful integration of the expression construct were selected with puromycin (Gibco A1113803). Prior to induction of neuronal differentiation, mESCs were seeded in 2i/LIF media onto a layer of mouse astrocytes grown on either tissue culture treated plates or 35 mm glass plates (MatTek P35G-1.5-20-C) coated with poly-L-ornithine (Sigma P4957) and laminin (Corning 354232). 24 hours after seeding mESCs, *NGN2* expression was induced by changing to N2B27 media with 2 µg/mL doxycycline (Sigma D9891). Media was changed daily with 2 µg/mL doxycycline in N2B27 media. Neurons were harvested for experiments 5 days after induction of *NGN2* expression, and neuronal status was confirmed by immunofluorescence staining for TuJ1 (Covance MMS-435P).

Live-cell imaging

Cells were grown on 35 mm glass plates (MatTek P35G-1.5-20-C) coated with poly-L-ornithine (Sigma P4957) for 30 minutes at 37 °C followed by coating with laminin (Corning 354232) for 2 hours at 37 °C, and imaged in 2i/LIF media using an LSM880 confocal microscope with Airyscan detector (Zeiss). Cells were imaged on a 37 °C heated stage supplemented with 37 °C humidified air. Additionally, the microscope was enclosed in an incubation chamber heated to 37 °C. ZEN Black Edition v. 2.3 (Zeiss) software was used for acquisition. Images were acquired with the Airyscan detector in super-resolution (SR) mode with a Plan-Apochromat 63x/1.4 oil objective. Raw Airyscan images were processed using ZEN v. 2.3.

In order to quantify MeCP2 condensate volumes, Z-stack images were taken using the ZEN v. 2.3 software. Cells were treated with SiR-Hoechst (also known as SiR-DNA dye) (Cytoskeleton CY-SC007) to stain DNA or SiR-Tubulin (Cytoskeleton CY-SC002) to stain tubulin to facilitate cell identification and microscope focusing. Far-red (SiR-DNA) signal was used to determine the upper-and lower-Z boundaries of the nucleus. Then, images were taken in both 488 nm channel (MeCP2-GFP) and the 643 nm channel (Sir-DNA) at 0.19 micron steps up through the

nucleoplasm. Images are the result of a single Airyscan image, processed using the ZEN v. 2.3 software. Heterochromatin condensate volumes were calculated using a custom script (www.github.com/jehenninger/MECP2_neuron) in Python v. 3.4.3. To calculate heterochromatin condensate volumes, the SiR-DNA signal was used to define nuclear-boundaries for a given cells. Heterochromatin condensates were identified as signal dense objects within the nuclear boundary with an empirical cutoff of 2.35 standard deviations above the mean signal. Once identified the volume of each heterochromatin condensate was quantified.

Fluorescence recovery after photobleaching (FRAP) was used to investigate dynamic internal rearrangement and internal-external exchange of molecules within heterochromatin foci, which are properties expected for liquid-like condensates[33]. FRAP was performed on LSM880 Airyscan microscope with 488 nm and 561 nm lasers. Bleaching was performed at 100% laser power and images were collected every two seconds. Each image utilizes the LSM880 Airyscan averaging capacity and is the averaged result of two images. The combined image was then processed using ZEN v. 2.3. FIJI/ImageJ (v. 2.0.0-rc-65) was used to calculate intensity values in images. Recovery after photobleaching was calculated by first subtracting background values, and then quantifying fluorescence intensity lost within the bleached condensate normalized to signal within a condensate in a separate, neighboring cell to account for image acquisition photobleaching. Post-bleach image taken 12 seconds post-photobleaching.

Partition ratios were used to quantify the ability of a protein to partition into heterochromatin condensates relative to the nucleoplasm in live-cell imaging experiments. A partition ratio for each cell was calculated as the ratio of the average pixel intensity within heterochromatin condensates relative to the average pixel intensity within 8-12 non-heterochromatic nucleoplasmic regions. Heterochromatin condensates and the nucleoplasm were defined using Hoechst staining. A single focal plane was analyzed for each cell and cells with 2 or more heterochromatin condensates were used for analysis.

Immunofluorescence microscopy

Murine ESCs were plated onto glass coverslips in preparation for immunostaining. After 24 hours, cells were fixed with 4% paraformaldehyde in PBS, washed 3 times with PBS, and then permeabilized with 0.5% triton-X100 in PBS for 10 min at room temperature. Cells were then washed 3 times with PBS, blocked for 1 hour in 4% IgG-free BSA (Jackson ImmunoResearch 001-000-162) in PBS, and then stained over night with the indicated antibody in 4% IgG-free BSA at room temperature in a humidified chamber. Cells were then washed 3 times with PBS. Secondary antibodies were added to cells in 4% IgG-free BSA and incubated for 1 hour at room temperature. Cells were then washed twice in PBS. Cells were stained with Hoechst dye (Molecular Probes H3570) in PBS or water for 5 minutes, and then mounted in Vectashield mounting media (Vector Laboratories H-1000). Imaging was performed on an RPI spinning disk confocal at 100X magnification using the MetaMorph software v. 7.10.3.279 (Molecular Devices). Primary antibodies: anti-MAP2 (Invitrogen MA5-12823) and anti-MED1 (Abcam ab64965). Secondary antibodies: anti-mouse Alexa Fluor 568 (Invitrogen A11031) and anti-rabbit Alexa Fluor 555 (Invitrogen A21428).

Overexpression

To examine the ability of MeCP2-GFP WT and R168X mutant proteins to partition into heterochromatin condensates when present at high levels in cells, the proteins were overexpressed in cells with N-terminal GFP fusions by transfection of expression constructs. 750,000 murine ESCs were transfected with 2.5 µg MeCP2-GFP WT or R168X mutant

expression plasmid using Lipofectamine 3000 (Invitrogen L3000). The following day, cells were dissociated and seeded onto poly-L-ornithine and laminin-coated 35 mm glass-bottom dishes (MatTek P35G-1.5-20-C) for live-cell imaging the next day.

Flow cytometry

Relative expression levels of endogenous-tagged proteins were examined using flow cytometry. Murine ESCs were dissociated using TrypLE Express (Gibco 12604) and the dissociation reaction was quenched using serum/LIF media. Cells were resuspended in single cell suspension in PBS and passed through a cell strainer (Corning 352235). Cells were analyzed using a LSRII flow cytometer (BD) and data was analyzed using FlowJo v.10 (BD). Standard forward and side scatter gating was used to exclude debris and isolate singlet cells. Example flow cytometry gating strategy is shown in Supplementary Figure 2. Mean fluorescence intensity was quantified for the singlet population, to determine the relative levels of endogenous-tagged fluorescent fusion proteins (MeCP2-GFP and HP1α-mCherry).

Western blot

Western blot was used to confirm expression of MeCP2-GFP WT and R168X proteins. Cell lysates were prepared by resuspending cell pellets in 2X Laemmli buffer and incubating at room temperature for 20 minutes. Lysates were then sonicated using a probe sonicator and boiled at 95 °C for 10 minutes. Samples were run on a 4-12% Bis-Tris polyacrylamide gel (Bio-Rad 3450125) using XT MOPS running buffer (Bio-Rad 1610788) at 80 V for 20 minutes, followed by 150 V until dye front reached the end of the gel. Protein was wet transferred to a 0.45 µm PVDF membrane (Millipore IPVH00010) in ice-cold transfer buffer (25 mM Tris, 192 mM glycine, 20% 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, then incubated with 1:1,000 anti-GFP (Takara Bio 632381) or 1:1,000 anti-Histone H3 (Cell Signaling Technology 4499) antibody in 5% non-fat milk in TBST overnight at 4 °C. Primary antibodies were validated by their respective vendors. After washing with TBST, the membrane was incubated with 1:10,000 HRP-conjugated anti-Mouse IgG (GE Healthcare NXA931V) or anti-Rabbit IgG (GE Healthcare NA934V) secondary antibody diluted in 5% non-fat milk in TBST for 2 hours at room temperature. After washing with TBST, the membrane was developed with chemiluminescent HRP substrate (Millipore WBKLS0100) and imaged using a CCD camera. Images were captured and analyzed using Image Lab software v. 6.0.1 (Bio-Rad).

Chimeric mouse generation

To generate endogenous MeCP2-GFP tagged chimeric mice, we injected endogenous MeCP2-GFP tagged mESCs grown on irradiated murine embryonic fibroblasts (MEF) into 8-cell embryos or blastocysts and implanted into pseudo-pregnant female CD1-IGS mice. Chimeric mice were identified based on the color of the fur and female chimeras were used for imaging experiments at 10-weeks of age. Mouse studies were observational in nature, and not subject to randomization or blinding. Imaging experiments using mouse brain tissues were performed with a sample size of 3 cells, which is in line with other studies of condensates in cells[16]. All experiments using mice were carried out with approval from the MIT Committee on Animal Care (CAC) under protocol number 1019-059-22. Experiments were carried out under the supervision of the Division of Comparative medicine (DCM) at MIT, which provides centralized management of the animal facility at the Whitehead Institute for Biomedical Research. The mouse facility conforms to federal guidelines (Animal Welfare Assurance Number A3125-01), and MIT is accredited by the Assessment and Accreditation of Laboratory Animal Care (AAALAC). Routine

bedding, food, and water changes were performed by DCM. Mice were housed in a centrally controlled environment with a 12-hour light/12-hour dark cycle, temperature of 68-72 °F, and humidity of 30-50%.

Brain slice immunofluorescence microscopy

Adult female endogenous MeCP2-GFP chimeric mice were perfused with 10% formalin to fix brain tissues. Following fixation, mouse brains were incubated at 4 °C in 30% sucrose for 3 days. Brains were then sectioned using cryostat (Leica CM3050 S). Brain sections were then placed on a slide and stored at -20 °C. For immunofluorescence, brain sections were allowed to warm to room temperature, fixed with 4% PFA for 10 minutes, and immunofluorescence was performed as described in the above immunofluorescence microscopy methods section.

In order to determine size and number of heterochromatin condensates in mouse neurons, brain sections taken from endogenous-tagged MeCP2-GFP mice (described above) were stained with anti-MAP2 to indicate neuronal cells. Sections were then imaged with 0.2-micron Z-stacks using MetaMorph v. 7.10.3.279 (Molecular Devices). In MAP2-positive cells, endogenous MeCP2-GFP signal was used to determine size and number of heterochromatin condensates using FIJI/ImageJ v. 2.0.0-rc-65 3D object counter with automatic threshold determination.

Brain slice FRAP

Organotypic brain slices were obtained from adult endogenous MeCP2-GFP chimeric mice, based on a slight modification of a previously described method[34]. Following decapitation, the brain was extracted and placed into ice-cold dissection medium composed of hibernate A (BrainBits HA), 2% B27 supplement (Gibco 17504), 2 mM L-glutamine (Gibco 25030), and 1% penicillin/streptomycin (Gibco 15140). The cerebellum and midbrain were removed and the remaining cerebral hemispheres were separated and sliced coronally at 250 µm thickness using a McIlwain tissue chopper (Ted Pella MTC/2E). The slices were gently separated from each other in chilled dissection medium and transferred onto glass-bottom dishes in culture medium containing Neurobasal A (Gibco 10888022) with 2% B27 supplement, 2 mM L-glutamine, and 1% penicillin/streptomycin. Imaging was performed immediately after brain slice preparation. FRAP experiment was performed using the Andor Revolution spinning disk confocal with the FRAPPA module (Andor Technology). Bleaching was performed using 5-7 pulses of 20 micro-second dwell time and images were collected every second. Fluorescence intensity was measured using FIJI/ImageJ v. 2.0.0-rc-65 and analyzed as described above. Post-bleach image taken 12 seconds post-photobleaching.

Protein purification

Human cDNA was cloned into a modified version of a T7 pET expression vector. The base vector was engineered to include sequences encoding a N-terminal 6xHis followed by either mEGFP or mCherry and a 14 amino acid linker sequence “GAPGSAGSAAGGSG.” cDNA sequences, generated by PCR, were inserted in-frame after the linker sequence using NEBuilder HiFi DNA Assembly Master Mix (NEB E2621S). Mutant cDNA sequences were generated by PCR and inserted into the same base vector as described above. All expression constructs were subject to Sanger sequencing to confirm sequence identity. The following human proteins were used in experiments:
MeCP2 full length (WT): residues 1-486
MeCP2 ΔIDR-1: residues 78-486
MeCP2 ΔIDR-2 (R168X): residues 1-167

MeCP2 IDR-1: residues 1-77
MeCP2 IDR-2: residues 168-486
MeCP2 ΔBasic: residues 1-486, removing IDR-2 basic patches (residues 170-181, 184-194, 246-258, 263-274, 282-289, 301-310, and 340-348)
MeCP2 ΔAromatic: residues 1-486, removing IDR-2 aromatic residues (F226 and Y450)
MeCP2 ΔHistidine: residues 1-486, removing IDR-2 histidine-rich domain (residues 366-372)
MeCP2 ΔProline: residues 1-486, removing IDR-2 proline-rich domain (residues 376-405)
MeCP2 R133C: residues 1-486, R133C
MeCP2 T158M: residues 1-486, T158M
MeCP2 P225R: residues 1-486, P225R
MeCP2 R255X: residues 1-254
MeCP2 R270X: residues 1-269
MeCP2 R294X: residues 1-293
MeCP2 R306C: residues 1-486, R306C
MeCP2 P322L: residues 1-486, P322L
MeCP2 P389X: residues 1-288
MeCP2 Mini: same as ΔNIC mutant from Tillotson et al., Nature 2017
HP1a: residues 1-191
MED1 IDR: residues 948-1574
BRD4 IDR: residues 674-1351
BRD4 Bromo domain 1: residues 40-168
BRD4 ET domain: residues 600-683
TBLR1-CTD: residues 134-514

For protein expression, plasmids were transformed into LOBSTR cells (gift from I. Cheeseman) and grown as follows. A fresh bacterial colony was inoculated into LB media containing kanamycin and chloramphenicol and grown overnight at 37 °C. Cells were diluted 1:30 in 500 mL prewarmed LB with freshly added kanamycin and chloramphenicol and grown 1.5 hours at 37 °C. To induce expression, IPTG was added to the bacterial culture at 1 mM final concentration and growth continued for 4 hours. Induced bacteria were then pelleted by centrifugation and bacterial pellets were stored at -80 °C until ready to use.

The 500 mL cell pellets were resuspended in 15ml of Lysis Buffer (50mM Tris-HCl pH 7.5, 500 mM NaCl, and 1X cOmplete protease inhibitors) followed by sonication of ten cycles of 15 seconds on, 60 seconds off. Lysates were cleared by centrifugation at 12,000 x g for 30 minutes at 4 °C, added to 1 mL of pre-equilibrated Ni-NTA agarose, and rotated at 4 °C for 1.5 hours. The slurry was centrifuged at 3,000 rpm for 10 minutes, washed with 10 volumes of lysis buffer and proteins were eluted by incubation for 10 or more minutes rotating with lysis buffer containing 50 mM imidazole, 100 mM imidazole, or 3 X 250 mM imidazole followed by centrifugation and gel analysis. Fractions containing protein of the correct size were dialyzed against two changes of buffer containing 50 mM Tris-HCl pH 7.5, 125 mM or 500 mM NaCl, 10% glycerol and 1 mM DTT at 4 °C. Protein concentration of purified proteins was determined using the Pierce BCA Protein Assay Kit (Thermo Scientific 23225). Recombinant proteins were stored in 10% glycerol, 50 mM Tris-HCl pH 7.5, 125 mM or 500 mM NaCl, 1 mM DTT. Amicon Ultra Centrifugal filters (30K or 50K MWCO, Millipore) were used to concentrate proteins to desired working concentrations.

In vitro droplet assay

In vitro droplet assays were used to investigate the physicochemical properties of condensate-associated proteins[33,35]. In vitro droplet assays containing DNA were performed by adding recombinant protein to Buffer D (10% glycerol, 50 mM Tris-HCl pH 7.5, 1 mM DTT) containing DNA at the indicated concentration. In vitro droplet assays containing nucleosomal arrays were performed by diluting purified nucleosomes to desired concentration in buffer containing 6 mM $MgCl_2$, 2% glycerol, 50 mM Tris-HCl pH 7.5 and 1 mM DTT. Recombinant protein was mixed with buffer containing 2% glycerol, 50 mM Tris-HCl pH 7.5 and 1 mM DTT, and then combined with the diluted nucleosomes to initiate droplet formation. In vitro droplet assays containing PEG-8000 were induced by adding recombinant proteins to droplet formation buffer composed of 10% glycerol, 50 mM Tris-HCl pH 7.5, 1 mM DTT and NaCl ranging from 0 mM to 500 mM, with 10% PEG-8000 added. For phase diagram generation (Extended Data Fig. 3g) droplet formation buffer was modified to contain 5% PEG-8000. Droplet assays were performed in 8-tube PCR strip. The indicated protein amount was added to droplet formation buffers and the solution was mixed by pipetting. The reaction was incubated for 10 minutes at room temperature in the 8-well PCR strip, and then loaded onto either a custom slide chamber created from a glass coverslip mounted on two parallel strips of double-sided tape mounted on a glass microscopy slide, or a well of a glass-bottom 384-well plate (CellVis P384-1.5H-N). Reactions were incubated for 20 minutes in the imaging vessel to allow droplets in solution to settle on the glass imaging surface. The reaction was then imaged on an Andor Revolution spinning disk confocal microscope using an Andor iXion+ EM-CCD camera with a 100x or 150x objective using MetaMorph v. 7.10.3.279 (Molecular Devices). Images presented are of droplets that have settled on the glass coverslip or the glass bottom of the 384-well plate.

To analyze in-vitro phase separation imaging experiments, custom Python v. 3.4.3 scripts (www.github.com/jehenninger/in_vitro_droplet_assay) were used 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 or 9 pixels per z-slice) were employed to segment the image. Droplet identification was performed on the 488 nm channel (MeCP2-GFP) and areas and aspect ratios were determined. Hundreds of droplets, identified in between 5-15 independent fields of view from each reaction, were quantified. Exact number of visual fields and droplets used for visualization and quantification are reported in the associated figure legends of relevant panels or in the methods below. To calculate the condensed fraction, the sum total of the intensities in all droplets of a given field (I-in) and the sum total intensity in the bulk dilute phase outside the droplets were calculated for each channel. Condensed fraction was computed as (I-in)/((I-in)+(I-out)). To calculate the partition ratio, the average intensity of each droplet (C-in) and the average intensity of the bulk dilute phase outside the droplet (C-out) was calculated for each channel. The partition coefficient was computed as (C-in)/(C-out). In Fig. 1, Fig 3, and Extended Data Fig. 6, the condensed fraction curves were fitted to the data using a logistic curve[36] in Prism v. 7.0a (GraphPad).

For in vitro droplet FRAP, droplets were formed as described above. The experiment was performed using the Andor Revolution spinning disk confocal microscope with FRAPPA module (Andor Technology). Bleaching was performed using 1 pulse of 20 second dwell time and images were collected every second. Fluorescence intensity was measured using FIJI/ImageJ v. 2.0.0-rc-65 and analyzed as described above.

To generate a phase diagram for MeCP2, MeCP2-GFP droplets formation reactions were performed in a range of NaCl and protein concentration. Ten independent fields of view were

captured for each condition, and droplets were identified as described above. An average partition ratio threshold of >1.85 was used to determine if a given condition formed droplets.

Droplet numbers

For relevant figure panels, the number of droplets analyzed per condition are indicated below:
Fig. 1h, Extended Data Fig. 3k: No DNA (n=592), DNA (n=1,395), methyl-DNA (n=1,130).
Fig. 2c: WT (n=1,419), ΔIDR-1 (n=1,084), ΔIDR-2 (n=208), IDR-1 (n=112), IDR-2 (n=626).
Fig. 2f: WT (n=273), ΔBasic (n=538), ΔAromatic (n=210), ΔHistidine (n=274), ΔProline (n=193).
Extended Data Fig. 3b: MeCP2-GFP 1.25 μM (n=1,767), 2.5 μM (n=1,041), 5 μM (n=834), 10 μM (n=483).
Extended Data Fig. 3e: NaCl 100 mM (n=685), 200 mM (n=603), 300 mM (n=521), 400 mM (n=930).
Extended Data Fig. 3n: WT (n=106), ΔIDR-1 (n=228), ΔIDR-2 (n=89), IDR-1 (n=51), IDR-2 (n=247).
Extended Data Fig. 4c: HP1α-mCherry (n=476), MED1-IDR-mCherry (n=561), BRD4-IDR-mCherry (n=462), mCherry (n=413).
Extended Data Fig. 4e: HP1α-mCherry (n=1,221), MED1-IDR-mCherry (n=1,156), BRD4-IDR-mCherry (n=1,124), mCherry (n=1,143).
Extended Data Fig. 4g: HP1α-mCherry (n=456), MED1-IDR-mCherry (n=331), BRD4-IDR-mCherry (n=338), mCherry (n=402).
Extended Data Fig. 4j: No poly-nucleosomes (n=599), Poly-nucleosomes (n=351).
Extended Data Fig. 5b: HP1α-mCherry (n=496), BRD4-IDR-mCherry (n=484), BRD4-BD1-mCherry (n=596), BRD4-ET-mCherry (n=451), mCherry (n=398).
Extended Data Fig. 6a: WT 0.5 μM (n=24), WT 1 μM (n=35), WT 2 μM (n=390), WT 4 μM (n=752), WT 6 μM (n=733), WT 8 μM (n=508), P389X 0.5 μM (n=36), P389X 1 μM (n=49), P389X 2 μM (n=315), P389X 4 μM (n=680), P389X 6 μM (n=578), P389X 8 μM (n=509), R294X 0.5 μM (n=30), R294X 1 μM (n=47), R294X 2 μM (n=14), R294X 4 μM (n=200), R294X 6 μM (n=545), R294X 8 μM (n=516), R270X 0.5 μM (n=58), R270X 1 μM (n=44), R270X 2 μM (n=12), R270X 4 μM (n=158), R270X 6 μM (n=549), R270X 8 μM (n=541), R255X 0.5 μM (n=39), R255X 1 μM (n=53), R255X 2 μM (n=21), R255X 4 μM (n=7), R255X 6 μM (n=1), R255X 8 μM (n=7), R168X 0.5 μM (n=42), R168X 1 μM (n=19), R168X 2 μM (n=3), R168X 4 μM (n=1), R168X 6 μM (n=1), R168X 8 μM (n=1).
Extended Data Fig. 6b: WT 0.5 μM (n=346), WT 1 μM (n=1,304), WT 2 μM (n=1,442), WT 4 μM (n=1,117), WT 6 μM (n=1,027), WT 8 μM (n=946), T158M 0.5 μM (n=2,274), T158M 1 μM (n=1,561), T158M 2 μM (n=3,798), T158M 4 μM (n=2,085), T158M 6 μM (n=1,723), T158M 8 μM (n=1,165), R133C 0.5 μM (n=2,577), R133C 1 μM (n=1,465), R133C 2 μM (n=2,305), R133C 4 μM (n=1,937), R133C 6 μM (n=1,380), R133C 8 μM (n=764).
Extended Data Fig. 6c: WT 0.5 μM (n=31), WT 1 μM (n=90), WT 2 μM (n=1,237), WT 4 μM (n=672), WT 6 μM (n=536), WT 8 μM (n=537), R306C 0.5 μM (n=23), R306C 1 μM (n=221), R306C 2 μM (n=1,236), R306C 4 μM (n=520), R306C 6 μM (n=507), R306C 8 μM (n=465).
Extended Data Fig. 6f: WT 0.5 μM (n=1,580), WT 1 μM (n=1,700), WT 2 μM (n=1,042), WT 4 μM (n=1,202), WT 6 μM (n=1,293), WT 8 μM (n=971), P322L 0.5 μM (n=934), P322L 1 μM (n=1,688), P322L 2 μM (n=2,719), P322L 4 μM (n=4,782), P322L 6 μM (n=1,395), P322L 8 μM (n=2,731), P225R 0.5 μM (n=1,378), P225R 1 μM (n=2,061), P225R 2 μM (n=1,632), P225R 4 μM (n=4,510), P225R 6 μM (n=2,876), P225R 8 μM (n=3,015).
Extended Data Fig. 7b: MeCP2-GFP WT (n=719), MeCP2-GFP R306C (n=707).
Extended Data Fig. 7d: MeCP2-GFP WT (n=1,103), MeCP2-GFP R306C (n=535).
Extended Data Fig. 8c: MeCP2-GFP WT (n=459), MeCP2-GFP Mini (n=363).
Extended Data Fig. 8f, g: MeCP2-GFP WT (n=288), MeCP2-GFP Mini (n=341).
Extended Data Fig. 8i: MeCP2-GFP WT (n=1,109), MeCP2-GFP Mini (n=910).

Fluorescent DNA production

Fluorescent DNA for droplet assays was produced by amplifying plasmid DNA using oligonucleotide primers with 5'-Cy5 fluorophore modifications (Integrated DNA Technologies). Amplification of plasmid templates was performed using Phusion polymerase (Thermo Scientific F531S). Fluorescent PCR products were gel purified using the Monarch gel extraction kit (NEB T1020S). The 376 bp DNA sequence used in droplet assays is:
TGTAAAACGACGGCCAGTGGATCCTAGGCTTAATTTGCATTGCAGTACATTTGCATGCATG
ATATTTGCATTAAGCTTGATTTGCATGTTTCAGAATTTGCATCGGCTAGCATTTGCATGGGC
TAGAATTTGCATGCCGGATAATTTGCATGGCGATTCATTTGCATGCCAAATCATTTGCATGC
ATGAACATTTGCATGGCTTACAATTTGCATGAAACATAATTTGCATCGATCGAAATTTGCAT
GTAGCCGAATTTGCATGTAGCTAAATTTGCATGAAATCGGATTTGCATGTAGCAATATTTGC
ATCTAGCCTAATTTGCATACCCTAGCATTTGCATTAGATTCGGCGGCCGCGTCATAGCTGT
TTCCTG

To generate methylated DNA template for in vitro droplet assays, Cy5-labeled fluorescent PCR product produced as described above was treated with M.SssI methyltransferase (NEB M0226L). The reaction was performed in 50 µL and contained 160 µM s-adenosylmethionine (SAM), 1 µL of M.SssI, and 4 µg of DNA. Contents were incubated for 4 hours at 37 °C, and then M.SssI was heat-inactivated for 20 minutes at 65 °C. Resulting methylated templates were purified using the NEB Monarch DNA and PCR cleanup kit (NEB T1030S). Methylation of templates was verified by methyl-specific restriction digestion using ClaI (NEB R0197S).

Poly-nucleosome purification

Poly-nucleosome arrays were purified from mESCs using a modified protocol adapted from Loyola et al., Mol Cell 2006[37]. Briefly, nuclei were isolated from mESCs by resuspending cells in a hypotonic buffer BC50 (HEPES pH 7.5, 50 mM NaCl) + 5 mM $MgCl_2$ + 0.05% NP-40 and douncing with a Kontes glass dounce (15 strokes with each pestle A then B). The nuclei were then digested with a limited amount of micrococcal nuclease and then the samples were centrifuged at maximum speed for 10 min. To purify poly-nucleosome arrays, the supernatant was loaded on a sucrose gradient and centrifuged for 20 hours in a swinging bucket rotor (Sorvall SW28) at 18,000 rpm. The sucrose gradients (28 mL each) were 5-15% in a base buffer of HEPES pH 7.5 and 200 mM NaCl. Individual fractions corresponding to poly-nucleosome arrays were collected. To determine the length distribution of the poly-nucleosome arrays in each faction, DNA was purified from each fraction and analyzed on an agarose gel. Fractions containing nucleosomal arrays ranging between 7-20 nucleosomes in length were pooled and dialyzed against buffer BC50 + 5mM $MgCl_2$. Purified poly-nucleosomes were stored in liquid nitrogen until ready to use in droplet assays.

MeCP2 IDR-2 sequence features

Specific sequence features within protein IDRs have been found to contribute to condensate formation[36,38–43]. Sequence features within MeCP2 IDR-2 were identified and deletion mutants were used to examined for their ability to contribute to droplet formation in vitro and transcriptional repression in a reporter assay. Basic patches in IDR-2 were defined as previously described[42]. Briefly, net charge per residue (NCPR) along MeCP2 protein sequence was computed using a sliding window of 5 residues and a step size of 1 residue using localCIDER (v. 0.1.14)[44]. Stretches of 4 or more consecutive windows having a NCPR > +0.35 per window were considered to be basic patches. MeCP2 IDR-2 contained 7 basic patches corresponding to residues 170-181, 184-194, 246-258, 263-274, 282-289, 301-310, and 340-

348. Two aromatic residues (residues F226 and Y450) are present in IDR-2. A histidine-rich domain (residues 366-372) and a proline-rich domain (residues 376-405) in IDR-2 were defined based on UniProt annotations.

Transcriptional repression reporter assay

A transcriptional repression reporter assay was used to examine the ability of MeCP2 IDR-2 sequence feature deletion mutants to repress transcription. Plasmids expressing MeCP2 IDR-2 sequence feature deletion mutants as fusions with the GAL4 DNA-binding domain (GAL4-DBD) from a SV40 promoter were co-transfected with a transcriptional repression reporter plasmid, containing an array of five GAL4 DNA binding sequence motifs located upstream of a chicken beta-actin promoter driven Firefly luciferase gene. To control for transfection efficiency, a plasmid expressing Renilla luciferase under the control of the SV40 promoter was also co-transfected. HEK293T cells were transfected using Lipofectamine 3000 (Invitrogen L3000), 24 hours after plating in a 96-well white flat bottom plate (Corning 3917). 24 hours after transfection, expression of the transcriptional repression reporter Firefly luciferase and control Renilla luciferase were assays using the Dual-Glo Luciferase Assay System (Promega E2940) and measured using a plate reader. Luciferase activity was calculated for each condition by dividing the Firefly luciferase signal by the Renilla luciferase signal, and was normalized to the GAL4-DBD alone condition. Assay was performed with three biologically independent samples per condition.

Gene expression analysis

RT-qPCR was used to quantify expression of heterochromatin-associated major satellite repeats. RNA was harvested using the RNeasy Mini Plus kit (QIAGEN 74134). A reverse transcriptase reaction was then performed using SuperScript III (Invitrogen 18080). RT-qPCR reactions were performed using Power SYBR Green PCR Master Mix (Applied Biosystems 43676) and measured using a QuantStudio 5 Real-Time PCR System (Applied Biosystems). Major satellite expression level was calculated using the delta-delta Ct method using *Gapdh* as a control and normalized to expression level in the WT condition. The following primers were used.
MajorSat_for: TGGAATATGGCGAGAAAACTG
MajorSat_rev: AGGTCCTTCAGTGGGCATTT
Gapdh_for: AACTTTGGCATTGTGGAAGG
Gapdh_rev: CACATTGGGGGTAGGAACAC

RNA-seq was used to profile expression of genes. RNA was harvested from 2 million cell aliquots using the RNeasy Mini Plus kit (QIAGEN 74134). Amount of RNA extracted was quantified using a Nanodrop spectrophotometer (Thermo Scientific). A fixed amount of ERCC RNA Spike-In (Invitrogen 4456740) was added to each sample for use in cell number normalization[45]. Samples were treated with DNA-free DNA Removal Kit (Invitrogen AM1906) prior to library preparation using the KAPA RNA HyperPrep Kit with RiboErase (KAPA Biosystems K8562) and sequencing on a HiSeq2500 (Illumina).

RNA-seq reads were mapped using STAR aligner (v. 2.6.1a)[46] to the murine RefSeq mm9 reference with ERCC spike-in reference sequences added. Alignment was performed using ENCODE long RNA-seq pipeline default parameters: --outFilterType BySJout, --outSAMattributes NH HI AS NM MD, --outFilterMultimapNmax 20, --outFilterMismatchNmax 999, --outFilterMismatchNoverReadLmax 0.05, --alignIntronMin 20, --alignIntronMax 1000000, --alignMatesGapMax 1000000, --alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --sjdbScore

1. Gene expression values were quantified using RSEM (v. 1.2.31) with default parameters[47]. Differential expression analysis was performed using DESeq2 (v. 1.24.0) with default parameters[48]. Spike-in cell number normalization was performed by using ERCC spike-ins to estimate size factors used for DESeq2 library normalization. DESeq2 uses a two-tailed Wald test to identify differentially expressed genes, and the default multiple test adjusted p-value cutoff of 0.1 was used to determine differentially expressed genes.

Bioinformatic analysis

MECP2 gene expression values in transcripts per million (TPM) from RNA-seq of human tissues were acquired from the Genotype-Tissue Expression (GTEx) project release v. 7. In instances where multiple regions of the same tissue were assayed, the highest expression value was used to represent the tissue. TPM values greater than 1 were considered to be expressed. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Rett syndrome patient mutation data was acquired from RettBASE[49]. Coding mutations associated with female patients with Rett syndrome were used for analysis. A histogram of mutation count along the length of the protein was plotted and the type of mutation (nonsense, frameshift, missense) was indicated.

Predicted disorder values along the length of human MeCP2 protein was determined using PONDR VLS2 algorithm. Higher values indicate greater predicted disorder.

Protein sequence conservation along the length of human MeCP2 protein was determined by extracting protein sequences in UniProt that shared at least 50% identity with human MeCP2 sequence (UniRef50_Q9Z2D6). Extracted sequences were subject to multiple sequence alignment using Clustal Omega (v. 1.2.4). Alignments were scored for protein sequence conservation along the length of human MeCP2 using Jensen-Shannon divergence[50] . Higher values indicated greater conservation.

Statistics and Reproducibility

Relevant statistical information for each experiment are included in the associated figure legends. For t-tests, data was assumed to be normal. For RNA-seq analysis, a two-sided Wald test was used to identify differentially expressed genes and p-values were adjusted for multiple comparison.

Experiments with representative images conducted in this study were repeated multiple times independently with similar results. Live-cell imaging of endogenously tagged mESCs and neurons (Fig. 1a, 4a, 4e, Extended Data Fig. 1a, 1b, 9a, 9g) was performed at least three times, on different days with cells plated independently. Live-cell imaging of endogenously tagged MeCP2-Mini mESCs (Extended Data Fig. 8j) was performed twice. Live-cell imaging of overexpressed MeCP2-GFP in mESCs (Extended Data Fig. 10a) was performed at least three times. Immunofluorescence in mouse brain cells (Fig. 1d) was performed on two brain sections. Immunofluorescence in mESCs (Extended Data Fig. 4a) was performed twice. Immunofluorescence in differentiated neurons (Extended Data Fig. 10e) was performed twice. FRAP experiments in mESCs (Fig. 1b, Extended Data Fig. 1c) were performed independently at least twice. FRAP experiments on mouse brain sections (Fig. 1e) was performed on three brain slices. FRAP experiments on MeCP2 droplets (Extended Data Fig. 3i) were performed on 10 individual droplets. Droplet fusion (Extended Data Fig. 3h) was observed more than 10 times.

PCR genotyping of MeCP2-GFP tagged ESCs (Extended Data Fig. 2c) was performed twice. Droplet experiments in Fig. 1g, 2b, 2e, 3b, 3d, 3f, Extended Data Fig. 3a, 3d, 3j, 3m, 4b, 4d, 4f, 4h, 4i, 5a, 6d, 7a, 7c, 8b, 8e, 8h were performed at least two times independently. Western blots (Extended Data Fig. 9d, 10c, 10f) were performed twice.

Acknowledgements
We thank Adrian P. Bird for sharing *Mecp2* mutant cell lines; Danny Reinberg for sharing purified poly-nucleosomes; Phillip A. Sharp for discussions; Doug Richardson and the Harvard Center for Biological Imaging; Wendy Salmon and the Whitehead W.M. Keck Microscopy Facility; and Ruth Flannery, Jesse Drotar, and Nick Rosenau and the Whitehead Genetically Engineered Models Center for technical support. The work was supported by NIH grant R01 GM123511 (R.A.Y.), NSF grant PHY1743900 (R.A.Y.), NIH grant 2 R01 MH104610-20 (R.J., R.A.Y.), NIH grant R37 CA084198 (R.J.), Hope Funds for Cancer Research Fellowship (A.D.), NIH grant T32 5T32DK007191-45 (J.M.P.), NSF Graduate Research Fellowship (A.V.Z.), and NIH grant K99/R00 MH113813 (X.S.L.).

Author contributions
C.H.L., E.L.C., R.J., and R.A.Y. conceived the project. C.H.L., E.L.C., T.I.L., R.J., and R.A.Y. organized the studies. C.H.L., E.L.C., and R.A.Y wrote the manuscript. E.L.C., J.E.H., O.O., A.V.Z., and J.S. performed in vitro droplet assays. N.M.H. performed protein purification. G.L. performed purification of poly-nucleosomes. C.H.L., E.L.C., and J.E.H. developed and performed computational analyses. E.L.C. and A.D. performed cellular imaging experiments. L.K.A. performed transcriptional reporter assays. C.H.L. generated endogenous-tagged cell lines and gene expression analyses. X.T. and T.L. performed neuronal differentiation. A.D., X.S.L., S.M., D.S.S., and E.W. performed chimeric mouse generation. C.H.L., E.L.C., and J.M.P. generated constructs. R.J. and R.A.Y. supervised the project with help from T.I.L. All authors contributed to editing the manuscript.

Competing interests
R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics, and Dewpoint Therapeutics. R.J. is an advisor/co-founder of Fate Therapeutics, Fulcrum Therapeutics, Omega Therapeutics, and Dewpoint Therapeutics. T.I.L. is a shareholder of Syros Pharmaceuticals and a consultant to Camp4 Therapeutics. All other authors declare no competing interests.

Additional information
Supplementary information is available for this paper. Correspondence and requests for materials should be addressed to R.A.Y. (young@wi.mit.edu) and R.J. (jaenisch@wi.mit.edu). Reprints and permissions information is available at www.nature.com/reprints.

Data availability
Relevant data supporting the findings of this study are available within the paper and Supplementary Information files. RNA-seq datasets generated in this study have been deposited in the Gene Expression Omnibus under accession code GSE139033. Uncropped gel images can be found in Supplementary Figure 1. Additional data are available from the corresponding author upon reasonable request. The following publicly available data were used in this study: GTEx v. 7 RNA-seq Median Gene TPMs by Tissue (www.gtexportal.org), RettBASE *MECP2* Variant List (mecp2.chw.edu.au/mecp2/mecp2_home.php), and UniProt Cluster ID: UniRef50_Q9Z2D6 (www.uniprot.org/uniref/UniRef50_Q9Z2D6).

Code availability

Custom code used for analysis of images from in vitro droplet assays is available at www.github.com/jehenninger/in_vitro_droplet_assay. Custom code used for analysis of heterochromatin condensate volumes is available at www.github.com/jehenninger/MECP2_neuron.

7. REFERENCES

1. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: guardian of the genome. *Annu. Rev. Cell Dev. Biol.* **34**, 265–288 (2018).
2. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
3. Lyst, M. J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* **16**, 261–274 (2015).
4. Ip, J. P. K., Mellios, N. & Sur, M. Rett syndrome: insights into genetic, molecular and circuit mechanisms. *Nat. Rev. Neurosci.* **19**, 368–382 (2018).
5. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).
6. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
7. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
8. Skene, P. J. *et al.* Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* **37**, 457–468 (2010).
9. Shrinivas, K. *et al.* Enhancer features that drive formation of transcriptional condensates. *Mol. Cell* **75**, 549–561 (2019).
10. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
11. Kumar, A. *et al.* Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J. Cell Sci.* **121**, 1128–1137 (2008).
12. Georgel, P. T. *et al.* Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.* **278**, 32181–32188 (2003).
13. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013).
14. Nan, X., Campoy, F. J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471–481 (1997).
15. Baker, S. A. *et al.* An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell* **152**, 984–996 (2013).
16. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
17. Boija, A. *et al.* Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 (2018).
18. van Steensel, B. & Belmont, A. S. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**, 780–791 (2017).
19. Gibson, B. A. *et al.* Organization of chromatin by intrinsic and regulated phase separation. *Cell* **179**, 470–484 (2019).
20. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320**, 1224–1229 (2008).
21. Kruusvee, V. *et al.* Structure of the MeCP2–TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. *Proc. Natl. Acad. Sci.* **114**, E3243–E3250 (2017).
22. Tillotson, R. *et al.* Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. *Nature* **550**, 398–401 (2017).
23. Linhoff, M. W., Garg, S. K. & Mandel, G. A high-resolution imaging approach to investigate chromatin architecture in complex tissues. *Cell* **163**, 246–255 (2015).
24. Li, Y. *et al.* Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons. *Cell Stem Cell* **13**, 446–458 (2013).
25. Wang, L. *et al.* Rett syndrome-causing mutations compromise MeCP2-mediated liquid–liquid phase separation of chromatin. *Cell Res.* **30**, 393–407 (2020).
26. Brown, K. *et al.* The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* **25**, 558–570 (2016).
27. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse

model of Rett syndrome. *Science* **315**, 1143–1147 (2007).
28. Giacometti, E., Luikenhuis, S., Beard, C. & Jaenisch, R. Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. *Proc Natl Acad Sci USA* **104**, 1931–1936 (2007).
29. Wheeler, R. J. *et al.* Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. *bioRxiv* https://doi.org/10.1101/721001 (2019).
30. Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science* **368**, 1386–1392 (2020).
31. Thoma, E. C. *et al.* Ectopic expression of neurogenin 2 alone is sufficient to induce differentiation of embryonic stem cells into mature neurons. *PLoS One* **7**, e38651 (2012).
32. Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
33. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).
34. Jang, S. *et al.* Long-term culture of organotypic hippocampal slice from old 3xTg-AD mouse: an ex vivo model of Alzheimer's disease. *Psychiatry Investig.* **15**, 205–213 (2018).
35. Alberti, S. *et al.* A user's guide for phase separation assays with purified proteins. *J. Mol. Biol.* **420**, 4806–4820 (2018).
36. Wang, J. *et al.* A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699 (2018).
37. Loyola, A., Bonaldi, T., Roche, D., Imhof, A. & Almouzni, G. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol. Cell* **24**, 309–316 (2006).
38. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
39. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
40. Lu, H. *et al.* Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318–323 (2018).
41. Zamudio, A. V. *et al.* Mediator condensates localize signaling factors to key cell kdentity genes. *Mol. Cell* **76**, 753–766 (2019).
42. Pak, C. W. *et al.* Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **63**, 72–85 (2016).
43. Dao, T. P. *et al.* ALS-linked mutations affect UBQLN2 oligomerization and phase separation in a position- and amino acid-dependent manner. *Structure* **27**, 937–951 (2019).
44. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. & Pappu, R. V. CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys. J.* **112**, 16–21 (2017).
45. Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).
46. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
47. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
48. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
49. Christodoulou, J., Grimm, A., Maher, T. & Bennetts, B. RettBASE: the IRSA MECP2 variation database - a new mutation database in evolution. *Hum. Mutat.* **21**, 466–472 (2003).
50. Capra, J. A. & Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **23**, 1875–1882 (2007).

CHAPTER 4 – DISCUSSION AND FUTURE DIRECTIONS

While advances over the last decade have led to a new appreciation of the contribution of genome structure and chromatin condensates in transcriptional regulation, they have also raised several questions that will be important to address. In this chapter, I will discuss some of these questions and initial observations that might prove to be insightful.

1. SPECIFICITY OF ENHANCER-PROMOTER INTERACTIONS

The identification of YY1 as a structuring factor of enhancer-promoter loops[1] (Chapter 2), provides a mechanism by which enhancers and promoters can interact. However, given the large number of enhancers and active gene promoters that exist in each cell type, it remains a mystery how specific interactions occur between enhancers and their target gene promoters. At least two potential models may contribute to the specificity of enhancer-promoter loops, and these models will be discussed below.

One model that could contribute to specificity of enhancer-promoter interactions is constraint imposed by genome structure. As discussed in Chapter 1-3, the genome is organized into a hierarchy of chromosome structures, including individual chromosomes, topologically associating domains (TADs), insulated neighborhoods, and enhancer-promoter loops[2,3]. Enhancer-promoter loops tend to occur within larger chromosome structures[2,4,5], suggesting that these larger structures constrain the set of possible enhancer-promoter interactions, thus contributing to enhancer-promoter specificity. Mutations that disrupt genome structure provide support for this model, as they often result in the formation of aberrant enhancer-promoter interactions and altered gene expression[4–9].

However, two reasons indicate that the specificity of enhancer-promoter interactions is not fully explained by the ability of genome structure to limit the combinatorial complexity of potential enhancer-promoter interactions. First, not all enhancer-promoter interactions are constrained by higher-order genome structure. It has long been observed in *Drosophila* that enhancers can interact with promoters on different chromosomes, a phenomenon termed transvection[10]. These interactions occur despite presence of TADs and insulator proteins in *Drosophila*[11–13], suggesting chromosome structures do not constrain all enhancer-promoter interactions. Evidence of transvection has also been observed in mammalian cells[14,15]. Second, the problem of enhancer-promoter specificity exists even within insulated neighborhoods, which are fundamental units of genome structure and function. Insulated neighborhoods typically contain multiple genes (median of three)[2], raising the question of how specificity of enhancer-promoter interactions occurs within these structures. It is possible that enhancers regulate all genes within the same insulated neighborhood, and evidence from *Drosophila* indicates that this could occur in some instances[16]. However, not all insulated neighborhoods containing multiple genes display concordant gene expression (all genes active or all genes silent)[2], suggesting additional mechanisms contribute to specificity of enhancer-promoter interactions within insulated neighborhood structures.

A second model that could contribute to specificity of enhancer-promoter interactions is the activity of additional enhancer-promoter structuring factors that mediate specific enhancer-promoter interactions (Fig. 1). The discovery that the zinc-finger transcription factor YY1 can structure enhancer-promoter loops[1] (Chapter 2) suggests the possibility that other zinc-finger transcription factors might similarly engage in structuring enhancer-promoter loops. If additional

structuring factors occupy specific enhancers and promoters – as opposed to YY1's more general binding of enhancers and promoters[1] – then these factors could facilitate formation of specific enhancer-promoter interactions. Consistent with this hypothesis, many zinc-finger transcription factors are capable of homo- and hetero-dimerization[17,18], which could allow factors bound at both enhancers and promoters to structure enhancer-promoter interactions via dimerization. Additionally, the zinc-finger transcription factor ZNF143 has been observed to also structure enhancer-promoter loops[19], providing further evidence that additional transcription factors may be engaged in this activity. Further study will be necessary to identify additional structuring factors and to evaluate whether their occupancy of specific enhancers and promoters can contribute to specificity of enhancer-promoter interactions.

Figure 1

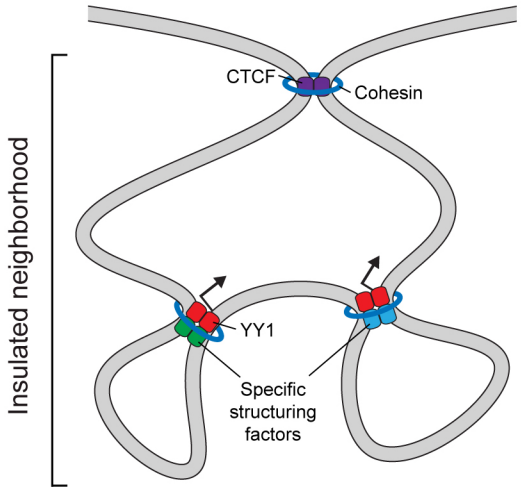


Figure 1. Model of enhancer-promoter specificity by multiple structuring factors
Additional structuring factors could facilitate enhancer-promoter looping specificity, if they bind to and mediate interactions between specific enhancers and promoters. In contrast, YY1 functions as a general enhancer-promoter structuring factor due to its binding to most enhancers and promoters.

2. MECHANISMS OF STRUCTURING FACTOR ACTIVITY

Structuring factors, such as YY1 and CTCF, are thought to structure chromosome looping interactions by forming dimers that bind to the two anchors of each chromosome loop (Fig. 2)[1,20,21]. However, there are at least two problems with this model. First, canonical dimerization interactions occurring between structured protein domains have not been identified for these proteins, suggesting that they engage in other mechanisms of self-association. Second, looping interactions mediated by these proteins are often highly interconnected, involving many DNA loci bound by these structuring factors, and the current model of pairwise interactions formed by dimerization does not adequately account for the existence of these multi-way chromosome interactions in cells[22,23]. These observations indicate that mechanisms other than dimerization likely contribute to the ability of structuring factors to mediate interactions between distinct chromatin loci.

Figure 2

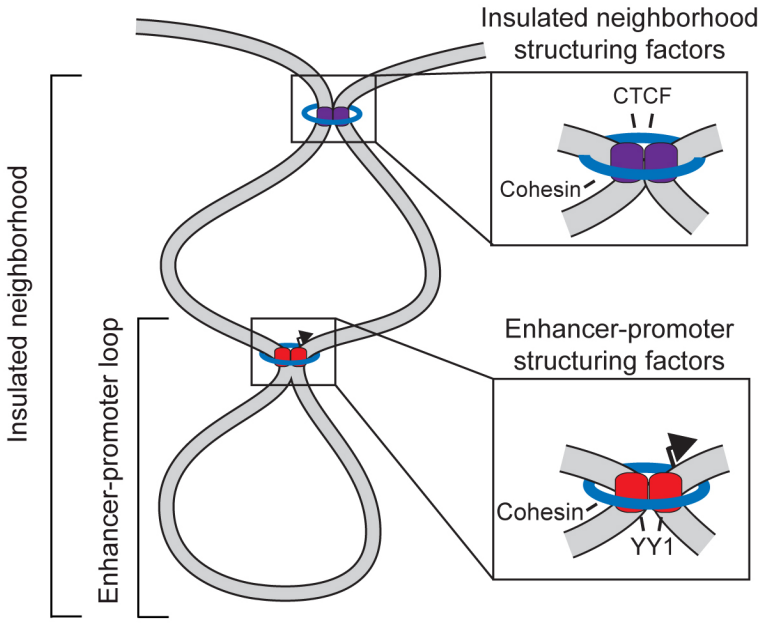


Figure 2. Dimerization model for structuring factors YY1 and CTCF
Transcription factors that regulate genome structure are thought to mediate chromosome looping by forming dimers that bind to two distinct genomic loci. Adapted from Weintraub et al., Cell 2017.

A potential alternative model is that structuring factors self-associate in condensates that mediate chromosome interactions. Several lines of evidence suggest that the structuring factors YY1 and CTCF may engage in condensate interactions. First, both YY1 and CTCF share similar protein domain organizations, with a structured DNA-binding domain and the rest of the protein being predicted to be intrinsically disordered regions (IDRs) (Fig. 3). This domain organization is similar to other transcription factors, which have IDRs that engage in condensate interactions[24,25]. Second, both YY1 and CTCF are capable of forming higher-order multimers[26,27], indicating an ability to engage in multivalent protein-protein interactions, which are often observed for condensate-forming proteins. Self-association of the two proteins is enhanced by RNA[1,27], and RNA can be a key contributing factor for condensate formation[28–31]. Notably, CTCF loops have been observed to depend on RNA-mediated CTCF clustering[32]. Third, both YY1 and CTCF can associate with known nuclear condensates. YY1 is highly enriched at super-enhancers[1], which are associated with active transcriptional condensates[33]. Additionally, YY1 can associate with nuclear speckles containing splicing factors via a poly-histidine stretch in its N-terminal IDR[34]. CTCF interacts with components of the nucleolus and can recruit insulator DNA sequences to the nucleolar periphery[35]. CTCF is also observed to form puncta in cells[32,36], although whether these puncta are associated with better-characterized nuclear condensates or represent distinct condensates of their own remains to be addressed.

Figure 3

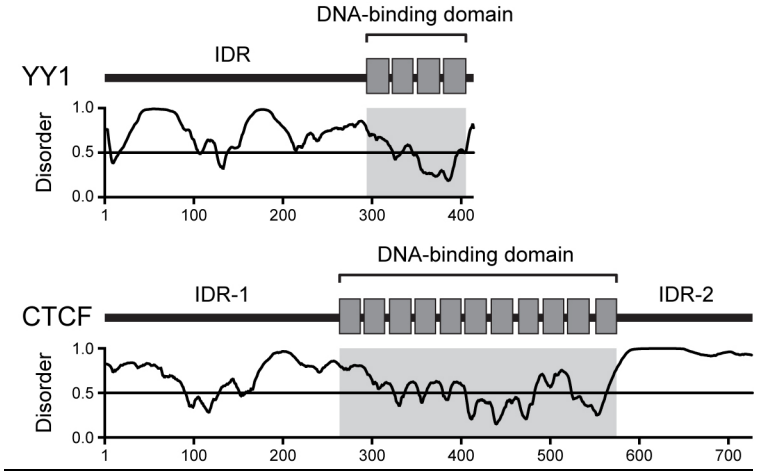


Figure 3. Domain structures of YY1 and CTCF
Plot of the domain structures of YY1 and CTCF with predicted protein disorder (PONDR VSL2) displayed below. Individual zinc fingers within the DNA-binding domains of each protein are displayed as rectangles.

Figure 4

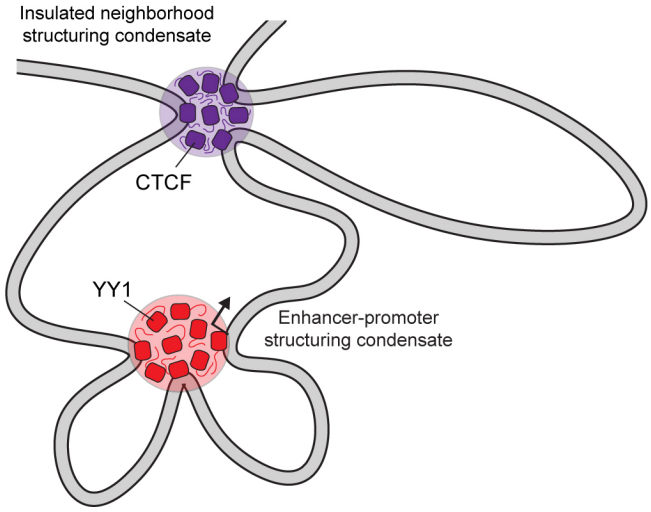


Figure 4. Condensate model for YY1 and CTCF structuring factor activity
A condensate model for structuring factor activity might explain: 1) the absence of strict dimerization domains in YY1 and CTCF, 2) the formation of multi-way chromosome loop interactions, and 3) the ability of enhancers to regulate genes without direct physical interaction of the DNA elements.

A condensate model could help resolve some of the remaining conundrums regarding mechanisms of structuring factor activity (Fig. 4). First, the involvement of YY1 and CTCF in condensates could explain the ability of these structuring factors to self-associate in multimers without having canonical structured dimerization domains. Second, a condensate model would explain how structuring factors can mediate simultaneous interactions between multiple distinct chromatin loci[22,23], as condensates could concentrate many structuring factor molecules, and thus facilitate interactions between multiple chromatin loci that are each bound by these structuring factors. Finally, a condensate model could help explain the confounding observation that, when examined using live-cell imaging techniques, enhancers do not physically touch the promoters they regulate[37]. A condensate of sufficient diameter and containing multiple YY1 molecules could associate with both an enhancer and its target gene promoter to allow their functional interaction, without necessitating the direct physical touching of the two DNA elements. Future studies will be required to experimentally characterize the potential condensate properties of YY1 and CTCF, and the role these properties play in the activity of these proteins as structuring factors.

3. MECHANISMS OF FUNCTIONAL CHROMATIN INSULATION

Despite substantial research into the roles of CTCF and cohesin in genome structure and insulator function, a major gap remains in our understanding of how functional chromatin insulation is mechanistically achieved. The current model for chromatin insulation involves the formation of a chromosome loop structure via interaction between two CTCF-bound insulator DNA elements and reinforcement by the cohesin complex[2,4,5,38]. These loop structures, termed insulated neighborhoods, are thought to mediate insulation by restricting the ability of enhancers located outside the loop from contacting and acting on genes contained within the loop[2,4,5,7]. However, a major limitation of this model is that it does not explain how a chromosome loop can mechanistically prevent chromatin contacts between DNA elements inside and outside the loop (Fig. 5).

Figure 5

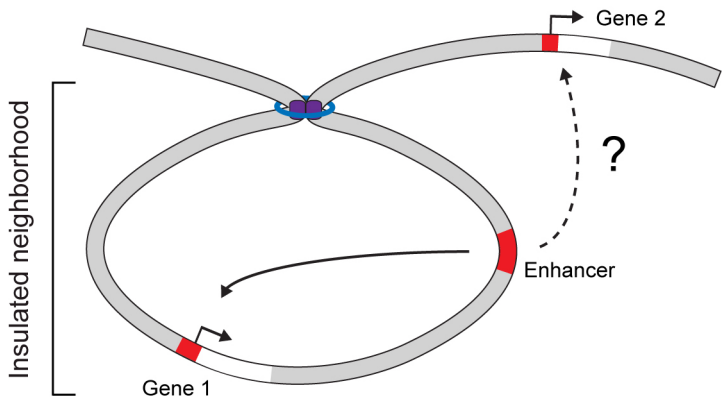


Figure 5. Limitation of the chromosome loop model for insulation
The current model for insulation involves formation of a chromosome loop that constrains enhancers to interact with and regulate genes within the same loop. However, it is not clear how loops can mechanistically prevent interactions that cross loop boundaries.

Several potential mechanisms might be involved in mediating functional chromatin insulation. First, insulated neighborhood loops could alter the propensity of chromatin interactions between DNA elements inside and outside the loop, due to topological constraints arising from the behavior of the chromatin fiber as a polymer. Computational polymer simulations provide support for this model, as simulated loop structures cause reduced frequency of contacts between polymer segments within and outside of the loop[39]. However, the contribution of this mechanism to functional chromatin insulation in vivo remains unclear. In particular, evidence that chromatin loops formed by cohesin and CTCF are dynamic[36,40] suggests that this mechanism alone may not be sufficient to account for insulation in vivo, as in computational simulations, insulation strength depends on the stability of loop structures[39].

A second potential mechanism contributing to chromatin insulation is the formation of insulator condensates. As described above, it is likely that the insulator protein CTCF participates in condensate interactions, and several observations support this view (Chapter 4-2). If insulators are found to behave as condensates, it will be important to evaluate how insulator condensates contribute to functional chromatin insulation. One possibility is that insulator condensates could simply contribute to structuring chromosome looping interactions, and thus mediate functional chromatin insulation. However, since insulated neighborhoods contain genes associated with either super-enhancers or Polycomb domains[5], and both of these behave as condensates[24,33,41], it is intriguing to consider how insulator condensates might interact with transcriptional condensates or heterochromatin condensates in order to mediate functional chromatin insulation. A potential model is that insulator condensates could from a dynamic shell that surrounds a gene and its associated transcriptional or heterochromatin condensate. A similar organization of condensates within condensates has been observed for nucleolar sub-compartments[42], and could provide a mechanism by which insulators act as a physicochemical barrier to interaction with outside DNA regulatory elements. Further experimental investigation will be needed to characterize the relationship between potential insulator condensates and both transcriptional and heterochromatin condensates.

4. ADDITIONAL CONTRIBUTORS TO GENOME STRUCTURE

Recent years have led to the identification and characterization of several key genome structuring factors. However, evidence suggests that additional factors contribute to genome structure. Acute depletion of cohesin, CTCF, and YY1 have been used to examine the contribution of these structuring factors to genome structure[1,43,44]. In each case, the observed change in genome structure after depletion of these factors has been limited. Depletion of YY1 resulted on average in a ~20% reduction in enhancer-promoter interaction frequency[1], suggesting that other factors contribute to structuring enhancer-promoter interactions. The potential for additional enhancer-promoter structuring factors, such as ZNF143 and other zinc-finger transcriptional factors, is discussed above (Chapter 4-1). Depletion of either cohesin or CTCF mostly eliminated the chromosome loop interactions structured by these factors[43,44]. However, higher-order chromosome structures, namely A/B compartments, were either unaffected or even strengthened[43,44], indicating that factors other than cohesin and CTCF are responsible for these higher-order genome structures.

A potential contributor to higher-order genome structure is the organization of chromatin into active and repressive chromatin condensate compartments. Several lines of evidence provide support for this model. First, higher-order genome structure correlates with active and repressive chromatin states. Early studies using Hi-C to map genome structure observed that the genome is organized into two compartments, wherein genomic regions associated with the

same compartment interact with each other, but not with genomic regions associated with the other compartment[45]. The two compartments are associated with transcriptionally active (compartment A) and repressive (compartment B) chromatin states[45], suggesting that these chromatin states could be contributing to genome structure. Second, compartment-level chromosome interactions do not depend on canonical genome structuring factors. Acute degradation of cohesin and CTCF does not disrupt A/B compartment organization[43,44], suggesting that other factors associated with these active and repressive chromatin states are contributing to this level of genome structure. Third, active and repressive chromatin states are associated with condensates that can engage in interactions analogous to those observed for A/B compartments. As described in Chapter 1-4, components of active and repressive chromatin are organized into condensates. Condensates associated with transcriptionally active euchromatin have been observed to interact with each other, both in vivo and in vitro[33,46]. Condensates associated with transcriptionally repressed heterochromatin display a similar ability to interact with each other[47,48] (Chapter 3). However, components of active and repressive condensates do not coalesce in the presence of each other[47,49], suggesting that immiscibility of condensates associated with active and repressive chromatin could contribute to their separation in the nucleus.

The contribution of active and repressive chromatin condensates to genome structure can be examined using at least two approaches. First, acute degradation of key scaffold proteins of active and repressive condensates can be used to disrupt condensates allowing their contribution to genome structure to be examined. By comparison with results from degradation of canonical genome structuring factors like YY1, CTCF, and cohesin, the relative contributions of each factor to genome structure can be assessed. Since degradation of condensate scaffolding proteins may disrupt other functions of these proteins, it will be important to also examine the effects of targeted mutations that specifically disrupt the condensate properties of these proteins. Second, ectopic formation of active and repressive condensates in cells can be used to examine whether these condensates can contribute to structuring the genome. Several engineered systems have been developed to allow ectopic formation of condensates by tethering key condensate scaffolding proteins to specific genomic loci[24,49]. Ectopic formation of active or repressive condensates followed by chromosome structure profiling can be used to test the ability of condensates to contribute to genome structure.

5. CHROMATIN CONDENSATE SPECIFICITY AND DYNAMICS

Recent studies have led to the view that many chromatin-associated processes occur within condensates that concentrate and compartmentalize components required for these processes[50]. These diverse chromatin condensates have distinct identities and are distinguished from each other based on several features, including their spatial localization in the nucleus and on the genome, and the composition of their components.

An emerging model for the formation of condensates at specific genomic locations involves several cooperating mechanisms, including chromatin binding, chromatin-mediated crowding, and dynamic interactions between condensate-forming scaffold proteins and their cofactors. Key condensate-forming proteins recognize and bind to distinct chromatin features present at specific genomic locations. For example, MeCP2 recognizes and binds to methylated DNA present at heterochromatin[51]. Other condensate-forming factors contain domains that recognize additional chromatin features, including histone post-translational modification, specific DNA sequences, and even locally produced RNA species[24,33,48,52–55]. Chromatin-binding facilitates crowding of condensate-forming proteins at specific genomic locations, which in turn promotes

the dynamic interactions between molecules that lead to condensate formation[47,53,54,56]. Finally, interactions between condensate-forming proteins and their cofactors facilitates concentration and compartmentalization of cofactors within condensates[47,52].

Despite this emerging model for the formation of individual condensates at specific genomic locations, how specificity is maintained in the presence of multiple condensates within the nucleus is poorly understood. Below, I discuss two categories of condensate separation – separation between condensates of different types and of the same type – and potential mechanisms that could be contributing to these behaviors.

First, condensates of different types must be separated from each other within the nucleus. The compartmentalization of distinct nuclear processes into separate condensates is important for the function of these processes. Several potential mechanisms could facilitate separation of different types of condensates. Separation could be maintained by constraining the mobility of condensates within the nucleus. This could be accomplished by anchoring condensates to relatively static spatial locations, for example by anchoring condensates to specific genomic loci or to the nuclear lamina[57], thus maintaining separation of different condensates through their spatial isolation. However, condensates of different types often occur in close spatial proximity within the nucleus and on the genome, suggesting other mechanisms must also contribute to their separation.

An emerging mechanism that could contribute to the separation of distinct condensates is immiscibility arising from the physicochemical properties of each condensate. The observation that the heterochromatin-associated protein MeCP2 forms condensates that remain distinct from and immiscible with condensates formed by BRD4[47] (Chapter 3), a component of transcriptional condensates, provides support for condensate immiscibility in the separation of distinct condensates. Several other observations in the literature provide additional support for the relevance of condensate immiscibility in separating different condensate compartments[42,49,54].

Despite observations of condensate immiscibility, the relative contribution of this mechanism to separation among the many different chromatin condensates in the nucleus is not understood. A systematic examination of the contribution of immiscibility to separation can be performed using in vitro droplet assays. Different chromatin condensates can be reconstituted in vitro from their key components. The potential for condensate immiscibility can be examined by combining different reconstituted chromatin condensates in the same reaction. Specific condensate components can be tested to determine which components contribute to observed immiscibility. Immiscibility can also be examined in vivo by ectopically inducing the formation of a condensate where a condensate of a different type is already present[49]. Additionally, loss-of-function studies can be used to determine whether components identified to contribute to immiscibility in vitro are required for separation of condensates of different types in vivo.

Second, condensates of the same type must also be separated from each other within the nucleus. Separation of condensates of the same type is likely important for proper condensate function. For example, many transcriptional condensates are found within the nucleus and are each associated with different genes which they control[24,33]; loss of condensate separation could disrupt the ability of transcriptional condensates to distinctly regulate different genes. In order to maintain separation of like condensates, two processes must be overcome. First is the fusion of two condensates of the same type that are in close spatial proximity. Second is diffusion of molecules from one condensate to another condensate through the dilute phase,

ultimately resulting in the growth of larger condensates at the expense of smaller condensates, in a process termed Ostwald ripening[58–60].

Loss of condensate separation due to fusion events could be mitigated by several potential mechanisms. As described above, separation could be maintained in part by constraining the mobility of condensates within the nucleus such that spatial distance between condensates is maintained and fusion events are avoided. Additionally, the dynamics of condensate dissolution and reformation could also contribute to condensate separation. If the average lifetime of individual condensates were shorter than the average time between fusion events, then temporally transient condensates may be able to avoid fusion events. Transcriptional condensates typically display short average lifetimes around 11 seconds[46], suggesting that they may be sufficiently transient to avoid some fusion events. Heterochromatin condensates are more stable during interphase and fusion events between heterochromatin condensates have been observed[48,61]. However, heterochromatin condensates appear to be dissolved during mitosis[48], perhaps due to an ability of the mitosis-associated H3S10ph chromatin modification to eject heterochromatin condensate-forming HP1 proteins from binding to H3K9me3 modified heterochromatin[62,63]. Dissolution of heterochromatin condensates during mitosis followed by their reformation in interphase could reset the separation of these condensates, thus limiting any potential consequences of loss of separation due to fusion. Notably, evidence that heterochromatin foci tend to cluster together to a greater extent in cells that spend a longer duration in interphase, with the greatest extent of clustering occurring in post-mitotic cells[64], is consistent with this model.

Experimental examination and characterization of the mechanisms that mediate separation of like condensates will be important. The ability to restrict unintended fusion events can be examined by removing constraints on condensate mobility. For example, it is plausible that genome structuring factors such as CTCF and cohesin play roles in constraining condensate mobility, and this model could be tested by degrading these factors and observing whether condensate fusions are increased as a result. Evidence that cohesin degradation results in increased chromosome interactions between super-enhancers[43], which are associated with transcriptional condensates[33], is consistent with this model. The role of condensate dynamics in avoiding unintended fusion events could be examined by modulating the lifetimes of condensates in cells, perhaps by using small molecule inhibitors of enzymes contributing to condensate formation and dissolution.

Suppression of Ostwald ripening is also important in order to mitigate the loss of condensate separation for condensates of the same type. Theoretical non-equilibrium mechanisms have been proposed for the suppression of Ostwald ripening[65]; however, thus far there is limited evidence for these mechanisms in vivo. The dynamics of condensate formation will likely play a role in the suppression of Ostwald ripening for chromatin condensates. Suppression of Ostwald ripening for chromatin condensate could occur if condensate formation and growth at different genomic locations occurs via an autocatalytic mechanism, whereby condensates can actively promote their own growth. Notably, autocatalytic formation of centrosome condensates is thought to explain the ability of centrosomes to suppress of Ostwald ripening[66].

Several mechanisms could contribute to autocatalytic formation of chromatin condensates. First, chromatin condensates likely concentrate and compartmentalize key regulatory enzymes that modify condensate components to further stimulate condensate formation. Notably, formation of chromatin condensates is often stimulated by binding of condensate-forming proteins to specific chromatin modifications, which can facilitate crowding of these proteins to the point of condensate formation[47,52,54]. These condensates concentrate and compartmentalize enzymes

that modify chromatin in a manner that allows further recruitment and binding of the initial condensate-forming protein, leading to a positive feedback mechanism. For example, heterochromatin condensates, formed by binding of condensate-forming HP1 proteins to H3K9me3 modified chromatin, can concentrate SUV39H1[56], an enzyme that catalyzes further generation of H3K9me3 modified chromatin[67]. Second, RNA produced within a condensate might further stimulate condensate formation. Active transcriptional condensates and heterochromatin condensates are associated with genomic loci that are transcribed to produced RNA. Since RNA can contribute to formation of many condensates[28–31], if RNA produced within chromatin condensates act locally to further stimulate condensate formation, then this could provide an additional positive feedback mechanism. Together these positive feedback mechanisms could contribute to the autocatalytic formation of chromatin condensates, thus facilitating the suppression of Ostwald ripening and maintenance of chromatin condensate separation.

The ability of autocatalytic formation of chromatin condensates to suppress Ostwald ripening could be examined by inhibiting positive feedback mechanisms involved in condensate formation. Small molecule drugs targeting chromatin modifying enzymes and RNA polymerase could be used to inhibit the positive feedback mechanisms described above. The effect on suppression of Ostwald ripening could be determined by quantifying changes in condensate numbers and sizes. Interestingly, in a study where treatment of cells with DRB, an inhibitor of transcriptional elongation, was found to reduce splicing factor condensates at active genes[68], fewer but larger splicing factor condensates can also be observed elsewhere in the nucleus of the treated cells. This observation is consistent with the idea that either phosphorylation of the C-terminal domain of RNA polymerase II or the production of RNA during transcriptional elongation – both of which are inhibited by DRB treatment – could stimulate formation splicing factor condensates that suppress the effects of Ostwald ripening.

6. CONDENSATES IN HUMAN DISEASE AND THERAPEUTICS

It is now clear that condensate dysregulation is an important cause of human disease. However, the extent to which condensate dysregulation is involved across the spectrum of human disease remains unknown. Several recent studies have shown that condensate dysregulation is a consequence of disease-causing patient mutations that alter the protein sequence of key condensate-forming proteins[47,69–79]. These studies have thus far been limited to a handful of condensate-forming proteins and diseases. However, given the emerging view that condensates are broadly involved in many key cellular processes[50,80], which are often dysregulated in disease, it is likely that condensate dysregulation is a far more general mechanism contributing to human disease. A broad survey, across the spectrum of human disease, identifying candidate patient mutations that likely cause condensate dysregulation will be instrumental in guiding the biomedical community in leveraging knowledge of condensates to better understand and treat human disease.

Several principles have emerged from our study of MeCP2 in Rett syndrome[47] (Chapter 3) and from other recent studies of condensate dysregulation in disease. These principles will be important to consider when examining new instances of condensate dysregulation in disease and are discussed below:

First, mutations that cause condensate dysregulation can occur in both structured domains and intrinsically disordered regions (IDRs). Studies of condensate dysregulation in disease have largely focused on mutations in IDRs. This is expected as IDRs are prevalent features among

condensate-forming proteins and can engage in the multi-valent interactions that give rise to condensates. However, the finding that mutations in MeCP2's structured methyl-DNA binding domain (MBD) can disrupt condensate formation[47] (Chapter 3) demonstrates that mutations in structured domains can also cause condensate dysregulation. This observation is expected since structured domains can engage in interactions that contribute to valency and condensate formation[81,82]. Mutations disrupting MeCP2's MBD likely cause condensate dysregulation by reducing the ability of methylated DNA sequences containing multiple MeCP2 binding sites to crowd MeCP2 to the point of condensate formation.

Second, multiple types of protein-coding mutations can cause condensate dysregulation. Missense mutations were an initial type of mutation found to affect the IDRs of several proteins involved in condensate dysregulation in amyotrophic lateral sclerosis (ALS)[69,72,74,75]. Recently, missense mutations in structured protein domains have also been implicated in causing condensate dysregulation[47,77,78]. Our finding that truncation mutations that frequently disrupt MeCP2's C-terminal IDR can cause condensate disruption in Rett syndrome provides another type of mutation involved in condensate dysregulation[47] (Chapter 3). This finding is consistent with the idea that truncation mutations could severely abrogate the ability of an IDR to engage in multivalent interactions required for condensate formation. Repeat expansion in IDRs have also been shown to cause condensate dysregulation[71,79]. Finally, gene rearrangements resulting in the production of a fusion protein might lead to condensate dysregulation. This could occur by the loss of key condensate-promoting domains, or by the aberrant gain of a condensate-promoting domain. While this mutation type has yet to be observed to cause condensate dysregulation, gene rearrangements involving proteins with IDRs are implicated in multiple cancers[83,84], suggesting future studies may yet provide support for this mechanism.

Third, varying aberrant condensate phenotypes are associated with condensate dysregulation. Initial studies examining ALS and frontotemporal dementia (FTD), primarily examined mutations that resulted in condensate aggregation[69,72,74,75], where condensates transition from dynamic liquid-like states to more solid-like states, such as gels and amyloid fibrils[85]. In some cases, aberrant formation of liquid-like condensates was observed, and was found to be a precursor to condensate aggregation[75,77]. Condensate dysregulation can also occur via aberrant disruption of condensates. Mutations that affect MeCP2 tend to cause disruption of heterochromatin condensates[47] (Chapter 3). Condensate disruption has also been observed for other proteins[78]. Aberrant changes in condensate composition can also manifest as a consequence of condensate dysregulation. Diseases involving repeat expansions in transcription factors can perturb the composition of condensates containing the affected transcription factors[71]. Notably, there appears to be some correlation between mutation type, affected protein domain, and aberrant condensate phenotypes. Mutations that truncate IDRs tend to cause condensate disruption, while those that increase IDR length tend to cause aberrant condensate formation. Missense mutations in IDRs tend to be associated with condensate aggregation. Additional studies will be needed to further explore these relationships, and to understand how these diverse condensate phenotypes contribute to disease pathogenesis.

Finally, non-coding alterations can also cause condensate dysregulation in disease. Aberrant post-translational modifications (PTMs) of key condensate-forming proteins have been demonstrated to cause condensate dysregulation in multiple neurological diseases[86–88]. Aberrant PTMs likely contribute to condensate dysregulation as PTMs can modulate the ability of condensate-forming proteins to engage in multivalent interactions involved in condensation[89,90]. Increased or decreased protein abundance might also contribute to condensate dysregulation, since condensation can be highly dependent on the concentration of condensate-forming components[91]. Notably, such a mechanism could be contributing to

MeCP2-related diseases, since reduced MeCP2 abundance is a common consequence Rett syndrome patient mutations[92,93] and increased MeCP2 abundance results in *MECP2* duplication syndrome[94,95]. Other biomolecules beyond proteins can also contribute to condensate dysregulation. Aberrant condensates formed by repeat containing RNAs are observed for a number of repeat expansion diseases[96]. It is likely that DNA alterations can also result in condensate dysregulation, since DNA contributes to the recruitment and crowding of other condensate-forming components[47,53]. Notably, the observation that small mutations can result in the formation of an oncogenic super-enhancer in T-cell acute lymphoblastic leukemia[97], suggests the potential for disease mutations to nucleate formation of aberrant transcriptional condensates[98]. Additionally, environmental changes within the cell might contribute to condensate dysregulation. Condensates are sensitive to local chemical environments, such as ATP concentration[99], and environmental factors including ATP concentration can be altered in disease[100]. Further studies will be needed to examine the effects of altered ATP concentration and other environmental factors within the cell on causing condensate dysregulation in disease.

New understanding of the importance of biomolecular condensates not only furthers our understanding of disease mechanisms, but also provides opportunities for new therapeutic approaches. Recent studies have shown that existing small molecule therapeutics can selectively partition into condensates where their targets are concentrated[101]. This suggests that further understanding of the physicochemical properties that govern small molecule partitioning into condensates could facilitate improvements in drug design that enhance the targeting ability of therapeutics. Aberrant condensates can also be directly modulated using small molecules for therapeutic benefit[102]. Further research and development in these areas could provide new avenues for therapeutic intervention for patients with Rett syndrome and other diseases involving condensate dysregulation.

7. REFERENCES

1. Weintraub, A. S. *et al.* YY1 is a structural regulator of enhancer-promoter loops. *Cell* **172**, 1573–1588 (2017).
2. Hnisz, D., Day, D. S. & Young, R. A. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* **167**, 1188–1200 (2016).
3. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
4. Ji, X. *et al.* 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
5. Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
6. Gostissa, M. *et al.* Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3′ regulatory region. *Nature* **462**, 803–807 (2009).
7. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2015).
8. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 3812–3814 (2016).
9. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
10. Lewis, E. B. The theory and application of a new method of detecting chromosomal rearrangements in Drosophila melanogaster. *Am. Nat.* **88**, 225–239 (1954).
11. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
12. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
13. Gurudatta, B. V. & Corces, V. G. Chromatin insulators: lessons from the fly. *Briefings Funct. Genomics Proteomics* **8**, 276–282 (2009).
14. Rassoulzadegan, M., Magliano, M. & Cuzin, F. Transvection effects involving DNA methylation during meiosis in the mouse. *EMBO J.* **21**, 440–450 (2002).
15. Sandhu, K. S. *et al.* Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev.* **23**, 2598–2603 (2009).
16. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
17. Amoutzias, G. D., Robertson, D. L., Van de Peer, Y. & Oliver, S. G. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* **33**, 220–229 (2008).
18. Lamb, P. & McKnight, S. L. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* **16**, 417–422 (1991).
19. Bailey, S. D. *et al.* ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat. Commun.* **6**, 6186 (2015).
20. Hou, C., Zhao, H., Tanimoto, K. & Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20398–20403 (2008).
21. Pant, V. *et al.* Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol. Cell. Biol.* **24**, 3497–3504 (2004).
22. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
23. Quinodoz, S. A. *et al.* Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757 (2018).
24. Boija, A. *et al.* Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 (2018).
25. Zamudio, A. V. *et al.* Mediator condensates localize signaling factors to key cell identity genes. *Mol. Cell* **76**, 753–766 (2019).
26. López-Perrote, A. *et al.* Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. *J. Biol. Chem.* **289**, 22614–22629 (2014).
27. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–734 (2014).

28. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).
29. Schwartz, J. C., Wang, X., Podell, E. R. & Cech, T. R. RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Rep.* **5**, 918–925 (2013).
30. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
31. Shevtsov, S. P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* **13**, 167–173 (2011).
32. Hansen, A. S. *et al.* Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol. Cell* **76**, 395-411.e13 (2019).
33. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
34. Salichs, E., Ledda, A., Mularoni, L., Albà, M. M. & De La Luna, S. Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet.* **5**, e1000397 (2009).
35. Yusufzai, T. M., Tagami, H., Nakatani, Y. & Felsenfeld, G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **13**, 291–298 (2004).
36. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6**, 1–33 (2017).
37. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* **8**, e41769 (2019).
38. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
39. Doyle, B., Fudenberg, G., Imakaev, M. & Mirny, L. A. Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS Comput. Biol.* **10**, e1003867 (2014).
40. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).
41. Plys, A. J. *et al.* Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813 (2019).
42. Feric, M. *et al.* Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).
43. Rao, S. S. P. *et al.* Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
44. Nora, E. P. *et al.* Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
45. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
46. Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
47. Li, C. H. *et al.* MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* (2020).
48. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
49. Shin, Y. *et al.* Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* **175**, 1481–1491 (2018).
50. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular condensates in the nucleus. *Trends Biochem. Sci.* (2020).
51. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905–914 (1992).
52. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
53. Shrinivas, K. *et al.* Enhancer features that drive formation of transcriptional condensates. *Mol. Cell* **75**, 549–561 (2019).
54. Gibson, B. A. *et al.* Organization of chromatin by intrinsic and regulated phase separation. *Cell* **179**, 470–484 (2019).
55. Huo, X. *et al.* The nuclear matrix protein SAFB cooperates with major satellite RNAs to stabilize heterochromatin architecture partially through phase separation. *Mol. Cell* **77**, 368–383 (2020).

56. Wang, L. *et al.* Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. *Mol. Cell* **76**, 646–659 (2019).
57. van Steensel, B. & Belmont, A. S. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**, 780–791 (2017).
58. Brangwynne, C. P. Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.* **203**, 875–881 (2013).
59. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
60. Voorhees, P. W. Ostwald ripening of two-phase mixtures. *Annu. Rev. Mater. Sci.* **22**, 197–215 (1992).
61. Brero, A. *et al.* Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.* **169**, 733–743 (2005).
62. Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
63. Fischle, W. *et al.* Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116–1122 (2005).
64. Ma, Y. & Buttitta, L. Chromatin organization changes during the establishment and maintenance of the postmitotic state. *Epigenetics and Chromatin* **10**, 1–20 (2017).
65. Zwicker, D., Hyman, A. A. & Jülicher, F. Suppression of Ostwald ripening in active emulsions. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **92**, 012317 (2015).
66. Zwicker, D., Decker, M., Jaensch, S., Hyman, A. A. & Jülicher, F. Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2636-2645 (2014).
67. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
68. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
69. Patel, A. *et al.* A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).
70. Batlle, C. *et al.* hnRNPDL phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation. *Cell Rep.* **30**, 1117–1128 (2020).
71. Basu, S. *et al.* Unblending of transcriptional condensates in human repeat expansion disease. *Cell* **181**, 1062–1079 (2020).
72. Murakami, T. *et al.* ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* **88**, 678–690 (2015).
73. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
74. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **24**, 1537–1549 (2016).
75. Mackenzie, I. R. *et al.* TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* **95**, 808–816 (2017).
76. Ryan, V. H. *et al.* Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol. Cell* **69**, 465–479 (2018).
77. Wegmann, S. *et al.* Tau protein liquid–liquid phase separation can initiate tau aggregation. *EMBO J.* **37**, e98049 (2018).
78. Bouchard, J. J. *et al.* Cancer mutations of the tumor suppressor SPOP disrupt the formation of active, phase-separated compartments. *Mol. Cell* **72**, 19–36 (2018).
79. Peskett, T. R. *et al.* A liquid to solid phase transition underlying pathological Huntingtin exon1 aggregation. *Mol. Cell* **70**, 588–601 (2018).
80. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
81. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).

82. Banani, S. F. *et al.* Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
83. Boulay, G. *et al.* Cancer-specific retargeting of BAF complexes by a prion-like domain. *Cell* **171**, 163–178 (2017).
84. Crozat, A., Aman, P., Mandahl, N. & Ron, D. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature* **363**, 640–644 (1993).
85. Alberti, S. & Dormann, D. Liquid–liquid phase separation in disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).
86. Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.* **8**, 275 (2017).
87. Hofweber, M. *et al.* Phase separation of FUS Is suppressed by its nuclear import receptor and arginine methylation. *Cell* **173**, 706–719 (2018).
88. Qamar, S. *et al.* FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. *Cell* **173**, 720–734 (2018).
89. Owen, I. & Shewmaker, F. The role of post-translational modifications in the phase transitions of intrinsically disordered proteins. *Int. J. Mol. Sci.* **20**, 5501 (2019).
90. Boeynaems, S. *et al.* Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
91. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
92. Brown, K. *et al.* The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* **25**, 558–570 (2016).
93. Guy, J. *et al.* A mutation-led search for novel functional domains in MeCP2. *Hum. Mol. Genet.* **27**, 2531–2545 (2018).
94. Van Esch, H. *et al.* Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* **77**, 442–453 (2005).
95. Del Gaudio, D. *et al.* Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **8**, 784–792 (2006).
96. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
97. Mansour, M. R. *et al.* An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–7 (2014).
98. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
99. Patel, A. *et al.* ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017).
100. Johnson, T. A., Jinnah, H. A. & Kamatani, N. Shortage of cellular ATP as a cause of diseases and strategies to enhance ATP. *Front. Pharmacol.* **10**, 1–19 (2019).
101. Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science* **368**, 1386–1392 (2020).
102. Wheeler, R. J. *et al.* Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. *bioRxiv* https://doi.org/10.1101/721001 (2019).