
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


The impact of toggling 3D genome structure on Polycomb-mediated repression, interchromosomal interactions, and X-inactivation

Citation

Kriz, Andrea. 2021. The impact of toggling 3D genome structure on Polycomb-mediated repression, interchromosomal interactions, and X-inactivation. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37370237

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

in the subject of Biological and Biomedical Sciences

have examined a dissertation entitled

The impact of toggling 3D genome structure on Polycomb-mediated repression, interchromosomal interactions, and X-inactivation

presented by Andrea June Kriz

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: Robert Kingston
Robert Kingston (Aug 10, 2021 08:42 EDT)

Typed Name: Dr. Robert Kingston

Signature: Michael Blower
Michael Blower (Aug 10, 2021 09:09 EDT)

Typed Name: Dr. Michael Blower

Signature: Chao-ting Wu
Chao-ting Wu (Aug 18, 2021 13:51 EDT)

Typed Name: Dr. Chao-ting Wu

Signature:

Typed Name: Dr. Erez Lieberman Aiden

Date: July 14, 2021

The impact of toggling 3D genome structure on Polycomb-mediated repression, interchromosomal interactions, and X-inactivation

A dissertation presented

by

Andrea June Kriz

to

The Division of Medical Sciences

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

July 2021

© 2021 by Andrea June Kriz

All rights reserved.

Dissertation Advisor: Dr. Jeannie T. Lee Andrea June Kriz

The impact of toggling 3D genome structure on Polycomb-mediated repression, interchromosomal interactions, and X-inactivation

Abstract

Microscopy and chromosome conformation capture (3C) techniques revealed that mammalian genomes fold into chromosome territories, compartmental structures such as active/inactive (A/B) compartments, and dynamic chromatin loops, mainly formed by the ring-shaped complex cohesin and stabilized by the zinc-finger protein CTCF. Interestingly, previous CTCF and cohesin loss-of-function studies disrupted 3D genome structure but had relatively small impact on gene expression. In this thesis, I revisit this structure-function relationship by first examining the role of a post-translational modification of CTCF, Serine 224 phosphorylation, on gene expression and domain formation. I then examine the role of 3D structure in X chromosome inactivation (XCI), the process which balances dosage of X-linked genes between male and female mammals. XCI is mediated by Xist RNA, which spreads across the inactive X chromosome (Xi), triggering recruitment of repressive factors, such as the Polycomb Repressive Complexes. Simultaneously, the Xi folds into a unique 3D structure not seen elsewhere in the genome. Therefore, XCI provided an ideal system in which to examine the role of 3D structure in epigenetic spreading and gene regulation.

I generated degron lines in which architectural factors CTCF, RAD21, an essential component of the cohesin complex, or WAPL, a cohesin release factor, could be rapidly and efficiently degraded. Surprisingly, CTCF degradation had little impact on 3D genome structure or XCI. Toggling cohesin, in contrast, had striking results. Genome wide, a proper balance of cohesin

was needed to maintain appropriate interchromosomal interactions as well as Polycomb-mediated repression. On the Xi, cohesin degradation ablated megadomain superstructures. However, the *Dxz4-Firre* superloop persisted upon cohesin degradation. In contrast, WAPL degradation did not affect the megadomain border but weakened long-range intra-megadomain interactions. Intriguingly, cohesin imbalance also created Xi-like superstructures on the active X. Finally, I examined how these structural changes impacted XCI. WAPL degradation led to both decreased Xist and Polycomb spreading and disrupted gene silencing, especially to the ends of the chromosome where the most contacts with the Xist transcription locus were lost. Thus, a balance of cohesin is needed for proper Polycomb-mediated repression, interchromosomal interactions, and XCI.

Table of Contents

Abstract……iii
Table of Contents..…….. ……..…v
List of Figures and Tables……..……..….vii
Acknowledgements……..…….……x

Chapter 1. Introduction…...……..……..……..……..……..……..1

Folding the mammalian genome……2
The existence of “TADs” in single cells……6
Genome-wide disruption of CTCF and cohesin-dependent structure has relatively little impact on gene expression……6
Post-translational modifications of architectural proteins……8
X chromosome inactivation: a ‘model’ system of epigenetic silencing……..9
X chromosome inactivation as a model of three-dimensional structural changes…10
Formation of Xi Superstructures……13
Transition of A/B compartments into S1/S2 compartments……..14
Merging of S1/S2 compartments……15
Maintaining Xi structure post-XCI……...16
The role of structural changes in XCI……...19
Other Xi structures during mammalian development……21
X chromosome structures in different systems of dosage compensation……...22

Chapter 2. Exploration of CTCF post-translation modifications uncovers Serine-224 phosphorylation by PLK1 at pericentric regions during the G2/M transition..……25

Abstract…….....26
Introduction……...27
Results……..29
Discussion……58
Methods……61

Chapter 3. Impact of CTCF degradation on three-dimensional genome structure and X-inactivation……76

Introduction……...77
Results……..78
Discussion and Future Directions……85
Methods……86

Chapter 4. Balancing cohesin eviction and retention prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-inactivation……92

Summary……...93
Introduction……...94
Results……..96
Discussion……..125
STAR Methods……...130

Chapter 5. Conclusion and Future Directions……149

Revisiting the relationship between 3D genome structure and gene expression…150
Toggling CTCF and cohesin binding via the dTAG degron system……150
Future development of degron systems……151
Even a small amount of CTCF can maintain 3D genome structure and X-inactivation
……152
Future directions for elucidating the architectural role of CTCF……154
Toggling chromosome architecture via degradation of RAD21 or WAPL……155
Cohesin regulates interchromosomal Hox cluster interactions……157
Cohesin shapes the inactive X chromosome……158
A balance of cohesin is needed for proper X-inactivation……160
Concluding Remarks……161

Appendix……163
References……190

List of Figures and Tables

Chapter 1

Figure 1.1 Folding the mammalian genome……3
Figure 1.2 Folding the Xa into the Xi……11
Figure 1.3 Representation of X chromosome structural changes in population-averaged Hi-C interaction maps throughout XCI……18

Chapter 2

Figure 2.1 Murine CTCF is phosphorylated at Ser^{224}……31
Figure 2.2 CTCF Ser224-P is conserved and accumulates at G2/M……34
Figure 2.3 CTCF Ser^{224} is phosphorylated by PLK1 and prominently labels pericentric chromatin……40
Figure 2.4 CTCF Ser224-P occupies a fraction of CTCF sites outside of pericentric chromatin in interphase……44
Figure 2.5 CTCF S224E phosphomimic mutation is poorly tolerated by dividing cells……47
Figure 2.6 CTCF Ser^{224} is nonessential to nuclear import and DNA binding……50
Figure 2.7 Overexpression of CTCF, S224A or S224E does not impact cell cycle progression or ploidy……52
Figure 2.8 Impact of overexpression of CTCF, S224A and S224E on three-dimensional chromatin structure and gene expression……55

Chapter 3

Figure 3.1 Degradation of CTCF during differentiation……80
Figure 3.2 Impact of CTCF degradation on X chromosome inactivation……83

Chapter 4

Figure 4.1 Toggling cohesin binding and 3D genome structure during differentiation……98
Figure 4.2 Polycomb loops protect against gene dysregulation upon cohesin loss, but increase sensitivity to cohesin gain during differentiation……105
Figure 4.3 Cohesin is required for proper folding of the inactive X chromosome at day 7 of differentiation……112
Figure 4.4 Cohesin is required for proper folding of the inactive X chromosome at day 10 of differentiation……115
Figure 4.5 Cohesin retention results in increase of interactions spanning TAD borders and weakening of megadomains and compartments on the Xi……121
Figure 4.6 Retaining cohesin on the Xi early in differentiation leads to XCI defects, especially at the ends of the chromosome……125
Figure 4.7 Selective cohesin eviction and retention are required to maintain Polycomb targets, fold the Xi, and properly silence Xi genes……128

Appendix

Figure S1.1 Casein kinase 2 (CK2) can phosphorylate CTCF Ser^{224}……165

Figure S1.2 CTCF Ser224-P accumulates in G2 and has a subnuclear distribution comparable to total CTCF……...167
Figure S1.3 CTCF Ser224-P is enriched at pericentric regions of metaphase chromosomes……..169
Figure S1.4 CTCF Ser224-P occupies known CTCF binding sites……171
Supplementary Table S1.1 Antibodies and reagents used in this study……...172
Supplementary Table S1.2 Oligonucleotides used in this study……174
Supplementary Table S1.3 UniProt accession numbers used in Clustal alignments……..176
Figure S2.1 Generation of FKBP^{degron} lines……..177
Figure S2.2 Cell cycle profiling of FKBP^{degron} lines……179
Figure S2.3 Replication of architectural changes across FKBP^{degron} clones……181
Figure S2.4 Additional analysis of cChIP after short term WAPL degradation and gene expression changes after 3 day degradation……183
Figure S2.5 Cohesin significantly contributes to proper upregulation of proximal super-enhancer targets without affecting long-range super-enhancer interactions……..185
Figure S2.6 Additional biological replicate for Xist CHART-seq and H3K27me3 cChIP-seq to examine spreading defects……...186
Supplemental Table S2.1 Statistics for Hi-C datasets……..189

Acknowledgements

My PhD, like all PhDs, was a journey and it would probably be impossible to thank everybody who has helped me. I want to start out my thanking my family who encouraged my academic and scientific work at every step of the way. My parents had my brother and I early on in their own academic careers and worked tirelessly to ensure that we had the best lives and could focus on our studies; I definitely wouldn't be where I am now without all their hard work. I especially want to thank my family in Boston, including my aunt and my grandma, for supporting me when I first moved to the area and giving me a welcoming second home. Also thank you to my aunt for introducing me to lab work in her own lab! Above all I want to thank my mom who provided me a lot of support in the challenging year of 2020.

Next I want to thank Phil Sharp who took a chance on me and allowed me to do undergraduate research in his lab. Thank you to Xuebing Wu for being my mentor and patient with all of my mistakes as I transitioned from 'dry' work to 'wet' work. And thank you to everybody in the Sharp lab, especially the other undergraduate students Victoria Lu, Laura Lu, and Prakriti Paul—the wonderful research environment you welcomed me into was instrumental to me deciding to pursue a PhD.

Finally I of course want to thank everybody in the Lee lab for all of your help, advice (scientific and otherwise), feedback, reagents and just conversations overall (scientific and otherwise). In particular, I first want to thank my advisor Jeannie Lee for letting me do my PhD in her lab. Jeannie, you really let me do the PhD of my dreams and taught me so much about science. Even though my project took many twists and turns you were always there with an optimistic view and ideas for moving forward. I next want to thank the other former graduate students, Lin Yang, Chen-Yu Wang, David Colognori, and John Froberg. You gave me so much advice for getting through graduate school and also became great friends over many dinners and discussions late nights in lab. Thank you especially to John for being my mentor when I was

rotating and first joined the lab. I also wanted to thank Brian Del Rosario for being a great collaborator and postdoc mentor. Thank you also to Bob Kingston, Stirling Churchman and Steve Buratowski for giving their time to be on my dissertation committee and all the useful feedback they gave me.

Chapter 1

Introduction

Attribution

This chapter was modified from the draft of a review article co-written with Chen-Yu Wang.

Folding the mammalian genome

Mammalian genomes consist of meters of DNA which must be packed in a nucleus only micrometers in diameter. Historically, this organization has been studied through microscopy. Such approaches revealed that interphase chromosomes fold into various large scale structures in order to fit in the nucleus (Figure 1.1). At the highest level, each chromosome occupies a distinct "territory", which can be further organized into "compartments" defined by gene-dense, early-replicating chromatin (A compartment) and gene-poor, later-replicating chromatin (B compartment) (Cremer and Cremer, 2001). The advent of chromosome conformation capture (3C) technologies, particularly Hi-C, enabled further dissection of these structures (Figure 1.1). Most initially exciting was the discovery of "Topologically Associating Domains" (TADs), megabase-sized "squares" prominently visible along the diagonals of Hi-C interaction maps (Dixon et al., 2012; Nora et al., 2012). These domains were found to be enriched at their borders for binding of CTCF, an 11-zinc finger DNA-binding protein, and the Structural Maintenance of Chromosomes (SMC) cohesin complex. "TADs" were also highly consistent between cell lines, stages of differentiation, and even between different species.

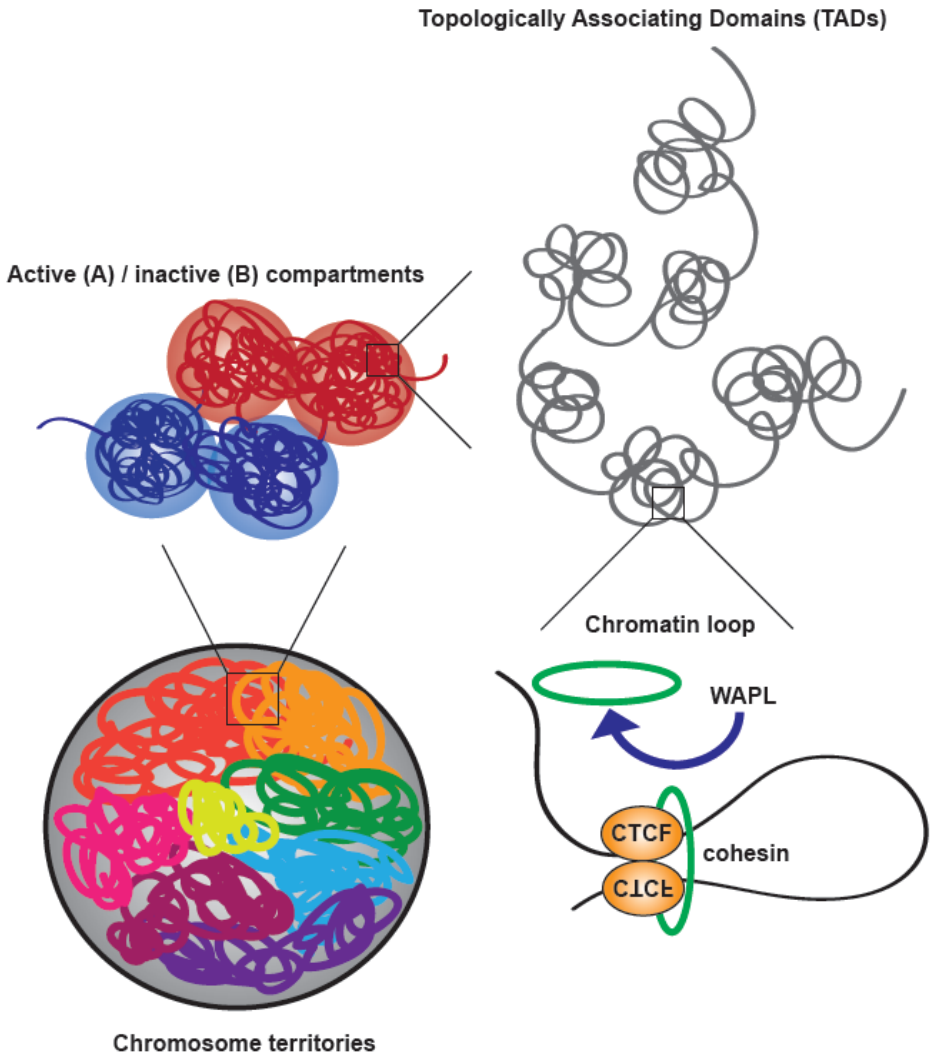


Figure 1.1. Folding the mammalian genome

Model of different layers of chromatin structure in the mammalian nucleus. Chromosome territories (scale: entire chromosomes), A/B compartments (roughly half of each chromosome), TADs (on average 100kb-1Mb in size, depending on definition), and chromatin loops (10-100kb in size on average) are shown.

As “TADs” were first identified from patterns in Hi-C interaction maps, through purely computational methods, their biological significance was unclear. Adding to the confusion, technological improvements in the Hi-C method generated higher resolution interaction maps

which revealed substructures within many “TADs”. For instance, chromatin loops anchored by CTCF, particularly in the convergent orientation, formed the border of many domains (Rao et al., 2014) (Figure 1.1, chromatin loops). This led to a terminological divide in the field, with some groups continuing to refer to all domains as “TADs” and some groups proposing more specialized terminology such as “loop domain” and “sub-TAD”. This terminological confusion still remains to be fully resolved, with the specific terms used to refer to domains varying between studies. This highlights the need to identify biologically functional units of organization, rather than rely purely on patterns visible in Hi-C interaction maps to define domains. For clarity, I will use more precise terms such as “cohesin-dependent loop” or “loop domain” when possible and more neutral terms such as “domains” to encompass the variety of domains called computationally from Hi-C interaction maps.

The next wave of studies sought to determine which architectural factors were required for folding the mammalian genome. Due to their enrichment at loop domain borders, CTCF and cohesin were the top candidates for forming this layer of organization. CTCF, an 11-zinc finger DNA-binding protein, has long been known to have insulatory activity, consistent with a role in border formation (Phillips and Corces, 2009). Cohesin is a Structural Maintenance of Chromosomes (SMC) complex consisting of RAD21, the essential kleisin subunit, the SMC proteins SMC1, SMC3 and either SA1 or SA2 (Nasmyth and Haering, 2009; Peters et al., 2008). These subunits are thought to form a ring-like complex around chromatin and extrude chromatin loops in an ATP-dependent manner (Davidson et al., 2019; Kim et al., 2019). Cohesin binding peaks tend to overlap with those of CTCF; therefore, the vast majority of loops anchored by CTCF, including those at domains borders, are anchored by cohesin as well.

Cohesin is also regulated by various loading factors, including NIPBL, and unloading factors including WAPL. WAPL, identified as a cohesin evictor during prophase (Gandhi et al., 2006; Kueng et al., 2006), also removes cohesin during interphase (Tedeschi et al., 2013) and therefore will be of particular interest in my thesis (Figure 1.1., chromatin loops). Studies suggest

that WAPL-mediated release occurs via opening of a DNA “exit gate” (Buheitel and Stemmann, 2013; Chan et al., 2012; Huis in 't Veld et al., 2014). Importantly, cohesin is not degraded after WAPL-mediated removal.

CTCF binding site deletion studies were the first to confirm that binding of this factor was required for domain border formation, as detected by Hi-C (Sanborn et al., 2015). Disruption of CTCF binding via methylation of CTCF binding motifs also resulted in loss of insulation (Flavahan et al., 2016). Troublingly, CTCF and cohesin knockdown studies only resulted in small changes to insulation genome-wide (Zuin et al., 2014). The development of degron systems (explained below), enabled much more efficient ablation of CTCF, showing genome-wide that CTCF binding was indeed required for most domain borders (Nora et al., 2017). Simultaneously, studies ablating cohesin binding via deletion of the cohesin loader NIPBL or via cohesin degrons showed that this complex was indeed required for formation of cohesin-anchored loops and domains (Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). In contrast, ‘overloading’ cohesin on chromatin by deletion or knockdown of WAPL resulted in the former of more and longer cohesin-anchored loops (Haarhuis et al., 2017; Wutz et al., 2017).

These experiments also had intriguing impacts on compartmental organization. Strikingly, A/B compartmentalization remained intact after CTCF or cohesin ablation (Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). This showed for the first time that compartmentalization and CTCF/cohesin mediated domains were independent layers of organization. Compartmentalization in fact appeared to increase in strength, with finer compartmentalization emerging after ablation of cohesin binding. Thus, the processes mediated by cohesin, possibly loop extrusion, antagonized compartmentalization. Supporting this, increasing cohesin binding via ablation of cohesin unloader WAPL or PDS5A/B resulted in a weakening of compartmentalization (Haarhuis et al., 2017; Wutz et al., 2017). This weakening of long-range interactions also agreed with the phenotype of more linearized ‘vermicelli’ chromosomes observable via microscopy upon WAPL ablation (Tedeschi et al., 2013).

The existence of "TADs" in single cells

The initial Hi-C experiments which detected TADs were population-based; therefore, it was unclear whether these domains represented population averages over millions of input cells or whether genomes in single cells indeed folded into domains. The development of single cell Hi-C, as well as super resolution microscopy, finally addressed this question. Both approaches revealed that individual genomes indeed fold into domains; however, these domains are not always consistent between single cells. Surprisingly, some of these domains are even cohesin-independent (Beagrie et al., 2017; Bintu et al., 2018; Nagano et al., 2013; Nagano et al., 2017; Stevens et al., 2017). These findings are consistent with domains being dynamic rather than static structures. In other words, though domains do appear in single cells, their exact location in population-averaged Hi-C maps reflects a probability rather than a strict delineation in every single cell. This would also be consistent with the loop extrusion model, which posits cohesin constantly moves along chromatin, forming dynamic chromatin loops. According to this model, while chromatin loops anchored at convergent CTCF sites are the most stable, loops anchored at other locations would be present in single cells; the averaging of contacts over all these loop locations would then produce the domain patterns seen in population Hi-C maps. Consistent with this, a recent microscopy study showed that depleting cohesin, CTCF, or WAPL merely modulates frequency of interactions across domain borders in single cells rather than creating or destroying barriers (Luppino et al., 2020). All in all, these studies indicate that the processes mediated by cohesin existing in single cells, such as loop extrusion, may be more biologically relevant than the concept of domains in itself.

Genome-wide disruption of CTCF and cohesin-dependent structure has relatively little impact on gene expression

Initially, 3D genome structure, especially the insulation of domain borders, was predicted to play a global role in gene expression (de Wit et al., 2015; Guo et al., 2015; Merkenschlager and Nora, 2016). This hypothesis was supported by initial studies, in which deletion of domain borders led to ectopic targeting of enhancers to genes across the deleted border (Lupianez et al., 2015). Thus domains bound by convergent CTCFs were hypothesized to encompass insulatory neighborhoods that would restrict enhancer targeting, preventing of inappropriate genes. However, these initial studies made large deletions on the scale of hundreds of kilobases; therefore, it was not clear whether elements in addition to insulatory CTCF sites had been deleted. More precise disruptions of borders, via deletions of specific CTCF binding sites for example, followed. Surprisingly, many such deletions had little impact on gene expression or enhancer targeting (Despang et al., 2019; Paliou et al., 2019; Williamson et al., 2019).

Historically, experiments which might disrupt 3D genome structure more globally were hindered by the fact that the CTCF and cohesin were essential for cell survival; therefore, knockouts of these factors were not viable. Experiments which knocked down CTCF and cohesin by up to 90%, however, only resulted in subtle changes to Hi-C interaction maps (Zuin et al., 2014). It was unclear where the residual 10% remaining was sufficient to mediate most 3D genome structure.

The advent of degron systems enabled more complete ablation of architectural factors than previous knockdown approaches. This strategy relies on a so-called ‘degron tag’, which is knocked into the protein of interest. Upon addition of a specific small molecule, the degron tag becomes tethered to a ubiquitin ligase; this triggers rapid and efficient degradation of the tagged protein. The first widely-adopted degron strategy was the auxin-induced degradation (AID) system derived from plants, which required introduction of the *Tir1* ubiquitin ligase into mammalian cells (Nishimura et al., 2009). The later developed dTAG degron system relied on an engineered bifunctional small molecule, known as dTAG-13, which could tether the $FKBP12^{F36V}$ tag to the Cereblon ligase, already endogenously expressed in most mammalian cells (Nabet et al., 2018).

Auxin degrons were quickly created for CTCF and essential components of the cohesin complex (Nora et al., 2017; Rao et al., 2017; Wutz et al., 2017). The short time frame of degradation allowed for experiments to be completed before cell death, in the case of cohesin, circumventing the need for this complex in cell division. In contrast to the previous knockdown experiments, degradation of either CTCF or cohesin globally disrupted 3D genome structure, particularly at the domain level whil reinforcing compartmentalization. Increasing cohesin binding by ablating the cohesin release factor WAPL conversely increased the number of cohesin-anchored loops while weakening long-range compartmental interactions. Surprisingly, despite these large changes in chromosome folding, the effects on gene regulation, enhancer targeting, or spreading of chromatin marks was relatively small (Haarhuis et al., 2017; Nora et al., 2017; Rao et al., 2017; Wutz et al., 2017).

These layers of chromatin architecture, therefore, largely did not appear to be necessary for baseline gene expression the assayed cell lines. This may not be surprising especially in the case of domains initially defined in Hi-C experiments, which are conserved even in different cell types with widely varying gene expression programs (Dixon et al., 2012). An alternative hypothesis is that 3D genome structure could be more important for gene regulation in dynamic contexts, for example when certain genes must be upregulated or downregulated during development. In my thesis, I examine this possibility using the model system of differentiating mouse embryonic stem cells (mESCs).

Post-translational modifications of architectural proteins

Architectural proteins such as cohesin and CTCF have myriad roles beyond mediating 3D genome structure. In particular, CTCF was originally identified as a transcription regulator (Heger et al., 2012; Lobanenkov et al., 1990), and has since been revealed to have insulatory activity (Merkenschlager and Odom, 2013; Ong and Corces, 2014). Although CTCF's loop-forming capacity has been proposed to underly all of these functions (Phillips and Corces, 2009), tens of

thousands of CTCF binding sites can be detected in mammalian genomes, many of which do not anchor chromatin loops (Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Xie et al., 2007). Many post-translational modifications (PTMs) of CTCF had previously been identified (El-Kady and Klenova, 2005; MacPherson et al., 2009; Sekiya et al., 2017; Torrano et al., 2006; Yu et al., 2004), strongly supporting a role for this layer of regulation in modulating CTCF functions. Before globally disrupting CTCF, we therefore examined the role of PTMs, in particular phosphorylation of CTCF, in regulating gene expression and 3D genome structure. We were curious whether CTCF mutations that disrupted certain PTMs could disrupt its loop-forming role while leaving its other functions intact.

X chromosome inactivation: a 'model' system of epigenetic silencing

Perhaps the most notable example of dynamic gene expression that occurs during differentiation is X chromosome inactivation. Female mammals carry two X chromosomes (XX) whereas males only carry one (XY). As the X chromosome is large (>160Mb) and carries many developmental important genes, this poses an issue of dosage imbalance. Female mammals solve this problem by randomly silencing one of the X chromosomes early in development in a process known as X chromosome inactivation (XCI) (Lyon, 1961). Remarkably, the master regulator of XCI is a noncoding RNA known as Xist (Brockdorff et al., 1991; Brown et al., 1991). Xist, a 17kb-long transcript, is transcribed from the future inactive X chromosome (Xi), spreading and eventually coating the entire chromosome.

Simultaneously, Xist RNA triggers a cascade of chromatin changes, including exclusion of RNA Polymerase II and transcription factors (Chaumeil et al., 2006), deacetylation of histone H3/H4 (Belyaev et al., 1996; Boggs et al., 1996; Jeppesen and Turner, 1993), gene silencing, and recruitment of histone modifying Polycomb complexes, Polycomb repressive complexes 1 and 2 (PRC1 & PRC2), to place ubiquitylation of histone H2A at lysine 119 (H2AK119ub) and trimethylation of histone H3 at lysine 27 (H3K27me3) on the Xi (de Napoles et al., 2004; Fang et

al., 2004; Plath et al., 2003; Plath et al., 2004). After this so-called “initiation phase,” the Xi undergoes further reorganization, including incorporation of histone variant macroH2A (Costanzi and Pehrson, 1998; Mermoud et al., 1999) and DNA methylation of CpG islands (Liskay and Evans, 1980; Lock et al., 1986; Lock et al., 1987; Mohandas et al., 1981; Norris et al., 1991), to establish stable gene silencing (the “establishment phase”).

Initial gene silencing induced by Xist RNA is reversible (Wutz and Jaenisch, 2000). However, sustained Xist coating eventually leads to a “lock-in” state, where multiple repressive mechanisms, including Xist RNA, DNA methylation, and histone deacetylation, operate synergistically to maintain stable silencing (Csankovszki et al., 2001; Mohandas et al., 1981; Wutz and Jaenisch, 2000). In this so-called “maintenance phase,” gene initially persists even after Xist deletion (Csankovszki et al., 2001; Csankovszki et al., 1999; Splinter et al., 2011), although partial reactivation can be seen in the long term (Yildirim et al., 2013; Zhang et al., 2007). Xist deletion also renders the Xi more suspectible to reactivation after drug treatment (Carrette et al., 2018). *De novo* XCI and maintenance of XCI therefore represent two different biological contexts to study the function of Xist. Establishment of XCI, or *de novo* XCI, can be modeled in differentiating mouse embryonic stem cells (mESCs) while the maintenance phase can be examined in fully differentiated cells such as fibroblasts or neural progenitor cells (NPCs).

XCI provides an example of simultaneously one of the most powerful examples of gene silencing, as well as a paradigm of coordination of diverse epigenetics pathways. Therefore, XCI provides an ideal system in which to study epigenetic regulation. In particular, female mESCs undergoing differentiation also undergo XCI. However, it is important to note that XCI does not silence all genes on the Xi. So-called “escapee” genes remain transcribed even in the maintenance phase of XCI. These encompass about 3% of X chromosome genes in mice and at least 15% in humans (Balaton and Brown, 2016; Berletch et al., 2011).

X chromosome inactivation as a model of three-dimensional structural changes

Previous observations hinted at an intriguing role for chromosome architecture in XCI. The structural changes the Xi undergoes during XCI have long-been observed via microscopy. In particular, the Xi condenses into a structure known as a “Barr body” (Barr and Bertram, 1949). Chromosome conformation capture techniques, such as Hi-C, enabled further dissection of how the Xa folds into the more condensed Xi structure (Figure 1.2).

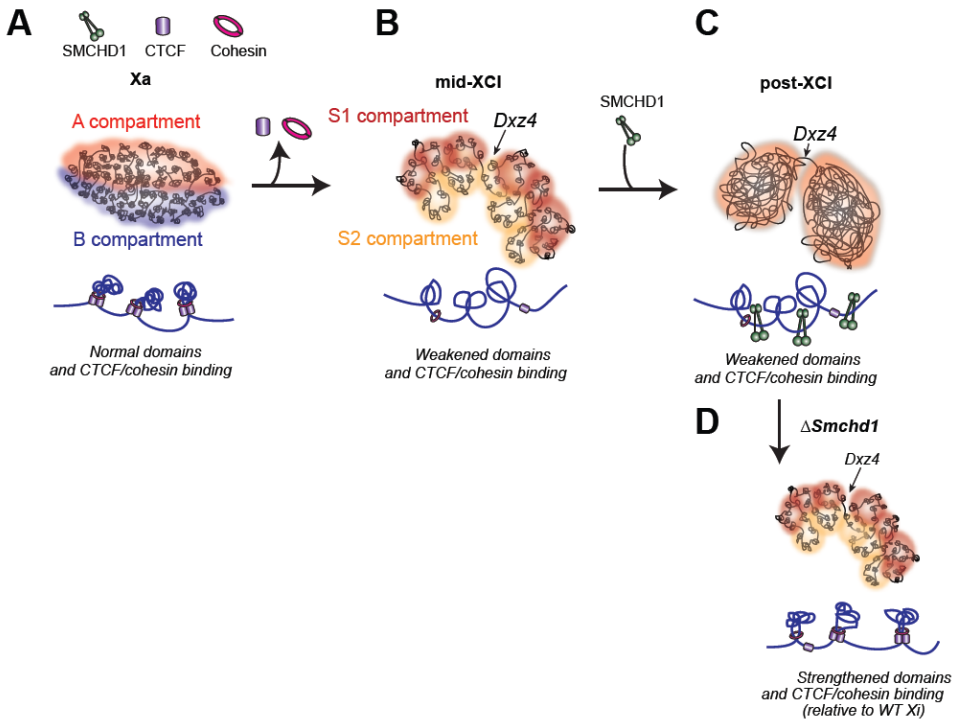


Figure 1.2. Folding the Xa into the Xi

(A) Much like autosomes, the active X (Xa) folds into A/B compartments and CTCF/cohesin-dependent domains (likely loop domains).

(B) As the inactive X (Xi) structure is established, CTCF and cohesin become depleted, weakening domains. Xist and Polycomb spreading mediate transformation of A/B compartments into S1/S2 compartments. Remaining cohesin is hypothesized to further fold the Xi into so-called “superstructures”, such as megadomains.

(C) Late in XCI, SMCHD1 is recruited, merging S1/S2 compartments to fold the Xi into its final state, which lacks strong compartmentalization. CTCF and cohesin become further depleted, along with domains.

(D) When SMCHD1 is deleted, the Xi misfolds and regains strong S1/S2 compartmentalization. In addition, cohesin/CTCF binding as well as domains are strengthened.

Examination of any Hi-C dataset reveals that the active X is organized similarly to autosomes; indeed, the term “TAD” was originally used to describe domains on the Xa, which displays A/B compartments as well (Figure 1.3A) (Nora et al., 2012). Further studies revealed that domain strength was weaker on the Xi than the Xa, with domains being strongest around escapee regions on the Xi (Giorgetti et al., 2016). Binding of CTCF and cohesin is attenuated on the Xi, likely explaining the weakening of domains (Figure 1.3B) (Berletch et al., 2015; Calabrese et al., 2012; Minajigi et al., 2015b).

One Xist interactome study paradoxically identified cohesin as an Xist-binding candidate (Minajigi et al., 2015b), proposing that Xist attenuates domains by directly “repelling” cohesin. As this study was performed in post-XCI cells, further studies will be needed to determine if Xist also binds cohesin during the initiation of XCI. Interestingly, domain strengthening is observed in *de novo* XCI when the Repeat B motif of Xist is deleted (Colognori et al., 2019), suggesting that Xist indeed aids in cohesin eviction during XCI establishment. However, deletion of Repeat B or of Xist entirely does not restore cohesin chromosome-wide in fully differentiated cells, indicating that other factors likely aid Xist in excluding cohesin from the Xi (Colognori et al., 2019; Minajigi et al., 2015b). For example, cohesin release factors like WAPL could also contribute to unloading cohesin from the Xi. The changing epigenetic landscape of the Xi may contribute to repelling

architectural factors as well. For example, CpG methylation can block CTCF binding (Bell and Felsenfeld, 2000).

While domains are weakened on the Xi, how individual cohesin-dependent loops in general are impacted is less clear. Sequencing-based chromosome conformation capture techniques, such as Hi-C, depend on single nucleotide polymorphisms (SNPs) in order to distinguish the Xi from the Xa. Due to not every DNA fragment containing SNPs, allelic Xi and Xa Hi-C maps are of lower resolution than their autosomal counterparts. This makes it difficult to achieve the high resolution needed to reliably detect loops chromosome-wide (Rao et al., 2014). While chromosome-wide time courses of chromatin loops on the Xi have not been generated, the stable CTCF/cohesin loops most readily detected by Hi-C would be predicted to weaken as CTCF levels on the Xi are reduced. However, it is important to note that cohesin and CTCF are not completely removed from the Xi, even in fully differentiated cells. Cohesin and CTCF peaks remain as strong or stronger on the Xi than on the Xa at a few regions, including the repeat elements *Dxz4* and *Firre* (Deng et al., 2015; Giorgetti et al., 2016; Minajigi et al., 2015b; Rao et al., 2014). Cohesin may also be present transiently outside of peaks, as any cohesin loaded onto the Xi would presumably still translocate across the chromosome, forming transient loops via loop extrusion. Such transient loops would be difficult to detect in population-averaged experiments, such as Hi-C. Therefore, improved single-cell Hi-C or super resolution microscopy will be needed to survey the evolution of loops on the Xi throughout XCI.

Formation of Xi Superstructures

As XCI continues, the Xi folds into Xi-specific structures, known as “superstructures”. The entire chromosome folds into two “megadomains”, one of 73Mb and the other of 93Mb—an order of magnitude larger than domains found on autosomes—with a border at a macrosatellite repeat element known as *Dxz4* (Figure 1.2C) (Deng et al., 2015; Giorgetti et al., 2016; Minajigi et al., 2015b; Rao et al., 2014). A “superloop” also forms between *Dxz4* and *Firre* another repeat

element ~30Mb away, a span again exceeding any normally observed on autosomes (Froberg et al., 2018; Rao et al., 2014). Deletion studies showed that *Dxz4* deletion, and to a lesser extent deletion of *Firre*, ablates megadomains, and that deletion of either ablates the superloop (Bonora et al., 2018; Darrow et al., 2016; Froberg et al., 2018; Giorgetti et al., 2016). However, it was not clear what architectural factor mediated these superstructures. While CTCF and cohesin binding are depleted across the Xi, they remain bound to a few regions, including *Dxz4* and *Firre*. Thus, the prevailing hypothesis was that CTCF and cohesin folded the Xi into the megadomain and superloop via loop extrusion (Bonora et al., 2018; Darrow et al., 2016). As single cell studies revealed that domains are dynamic rather than stable structures, it is also important to distinguish whether the same is true of megadomains. DNA FISH verified that the border at *Dxz4* separates megadomain regions in single cells (Giorgetti et al., 2016), suggesting that they do indeed exist on an individual level.

Transition of A/B compartments into S1/S2 compartments

In addition to the ablation of CTCF/cohesin dependent structures and formation of superstructures, compartmental structures are also dramatically change on the Xi. The Xa, much like autosomes, is organized into A/B compartments (Figure 1.2A). However, during differentiation, compartments on the Xi transform from A/B partitions into what are known as S1/S2 compartments, characterized by Xist-rich/Xist-poor rather than active/inactive chromatin (Figure 1.2B) (Wang et al., 2018). S1/S2 compartments are also Polycomb-rich/Polycomb-poor; in particular, PRC1 is important for S1/S2 compartmentalization. The requirement of PRC1 for S1/S2 compartmentalization is consistent with previous observations that PRC1 has phase-separation capacity (Plys et al., 2018; Tatavosian et al., 2018). Polycomb also forms self-associated domains and small compartments on autosomes (Kundu et al., 2017; Rhodes et al., 2020). However, further studies will be needed to determine by what exact mechanism PRC1 drives S1/S2 compartmentalization on the Xi.

An alternative viewpoint proposes that S1/S2 compartments are the same A/B compartments already present on the Xa (Gdula et al., 2019). The disagreement is mainly terminological, as studies across labs agree that the Xi is compartmentalized during the establishment phase of XCI and these compartments encompass much of the same partitions as A/B compartments on the Xa (Gdula et al., 2019; Wang et al., 2018). However, important differences between Xi compartments and A/B compartments merit the differing nomenclature. First, the Xi loses active chromatin; therefore, the main driving force behind Xi compartmentalization cannot be active/inactive states, as for A/B compartments. Secondly, S1/S2 partitions are not exactly equivalent to A/B partitions. In particular, A/B partitions are merged to make S1 or S2 partitions across the Xi (Wang et al., 2018). As these differences likely correspond to differing biological functions of compartments on the Xa versus the Xi, we argue that these merit the S1/S2 nomenclature for compartments on the Xi.

Merging of S1/S2 compartments

Finally, compartments are merged in the late stages of XCI, leading to the final "compartment-less" Xi structure (Figure 1.2D). Such compartment-merging is another example of an Xi-specific process, as no similar process is known to occur on autosomes. Correspondingly, this process is mediated by an Xi-enriched factor, the SMC-like *s*tructural *m*aintenance of *c*hromosomes *h*inge *d*omain containing *1* (SMCHD1) (Blewitt et al., 2008; Gdula et al., 2019; Sakakibara et al., 2018; Wang et al., 2018). SMCHD1 is recruited late in differentiation, after initial Xist spreading, indicating this factor is important for the later folding steps (Figure 1.2C) (Gendrel et al., 2012). Deletion analysis of SMCHD1 during establishment of XCI showed that this factor is required for merging S1/S2 compartments (Figure 1.2D) (Wang et al., 2018). Even when SMCHD1 was deleted after XCI was established, strong compartments reemerged. Therefore, SMCHD1 is needed just to establish, but also to maintain the Xi's final structure. Deletion of SMCHD1 also resulted in strengthening of domains, indicating this factor participates in

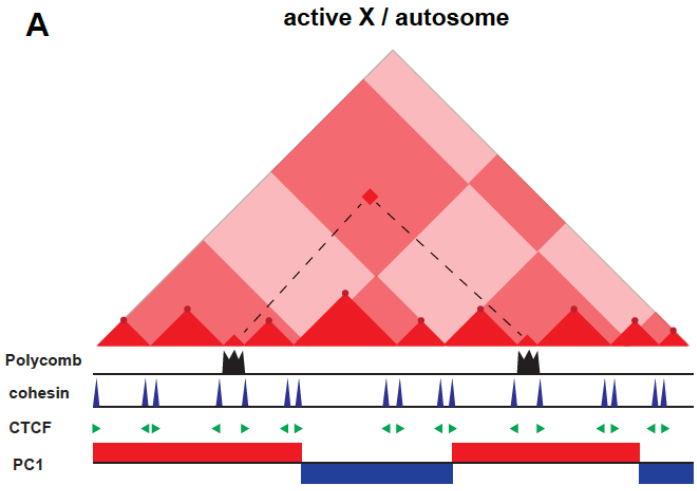
suppression of domains on the Xi as well (Gdula et al., 2019; Jansz et al., 2018a; Jansz et al., 2018b; Wang et al., 2019; Wang et al., 2018).

However, the exact biological threshold at which A/B compartments transition to S1/S2 compartments is unclear. How much active chromatin needs to be lost and Polycomb gained in order for the forces behind compartmentalization to switch from active/inactive chromatin to PRC1-rich/PRC1-poor regions? In addition, S1/S2 compartments have mainly been observed in population averaged Hi-C interaction maps. As with other structures detected via Hi-C, it will be important to distinguish whether and how S1/S2 compartments are present in single cells, via single cell Hi-C or microscopy, for example. These approaches will also be needed to determine whether S1/S2 and A/B compartments could co-exist at intermediate stages in single cells, as Polycomb-enriched compartments and A/B compartments appear to on autosomes.

Maintaining Xi structure post-XCI

In order to gain a full understanding of Xi folding, it is also important to consider the interplay between superstructures and compartments. During *de novo* XCI, Hi-C maps of the Xi display both the “checkerboard” pattern of S1/S2 compartments, as well as a prominent border at *Dxz4*; therefore, at least in population-averaged Hi-C maps, S1/S2 compartments and megadomains co-exist (Figure 1.3B). Importantly, megadomains remain intact after *Smchd1* deletion, indicating that SMCHD1 is not required for megadomain formation. This agrees with the late stage at which SMCHD1 is recruited during *de novo* XCI, i.e., after the megadomain border at *Dxz4* has been established (Froberg et al., 2018; Wang et al., 2018). On the other hand, preventing PRC1 recruitment to the Xi by deleting the Repeat B motif weakens compartments without affecting the strong megadomain border at *Dxz4* (Colognori et al., 2019). Therefore, merging of compartments and formation of megadomains represent different, independent layers of Xi folding. What happens to the *Dxz4-Firre* superloop after *Smchd1* deletion is less clear; higher-resolution Hi-C or microscopy studies will be needed to address this question.

Despite their formation by independent mechanisms, compartments and megadomains still affect each other's strength. In fully-differentiated Xi Hi-C maps, where compartments are lacking, megadomains appear more prominently (Figure 1.3C) (Colognori et al., 2019; Darrow et al., 2016; Deng et al., 2015; Minajigi et al., 2015b; Wang et al., 2018). This may in part be due to how Hi-C interaction maps are presented in 2D, with compartment and superstructure patterns overlaid on top of each other in an interaction matrix: namely, once the compartment "checkerboard" pattern is removed, the megadomain pattern will necessarily become the dominant visual feature. But the antagonism of compartments and megadomains is not merely visual. Depletion of Xist, HNRNPK, or PRC1 also makes megadomains more prominent at the expense of S1/S2 compartments (Colognori et al., 2019; Wang et al., 2019). This suggests that normally these factors mediate interactions across the *Dxz4* border, weakening megadomains. Therefore, S1/S2 compartments and megadomains are formed by independent factors, but are antagonistic to each other. In concept, this could be similar to how cohesin-dependent domains and A/B compartments represent independent but antagonistic layers of organization on autosomes.



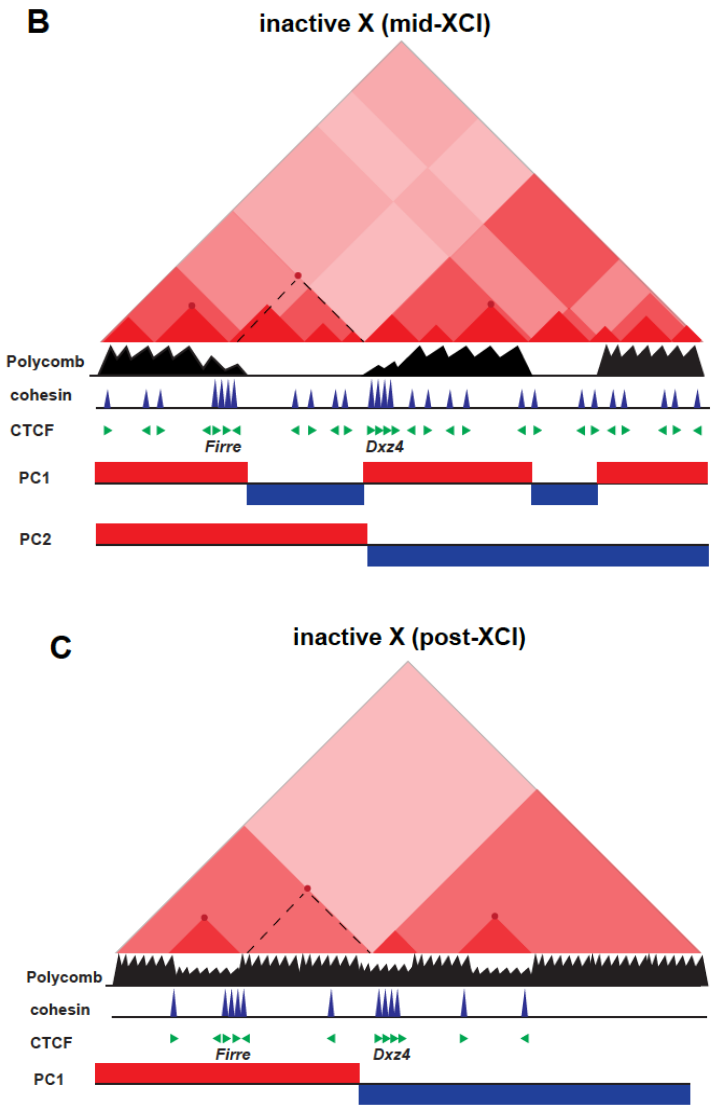


Figure 1.3. Representation of X chromosome structural changes in population-averaged Hi-C interaction maps throughout XCI

(A) Schematic Hi-C interaction map of a typical autosome, which the active X resembles. Darker red indicates greater interaction frequency. Schematic aligned tracks of Polycomb, cohesin, and CTCF binding (arrow indicating motif orientation) and informative Principal Components (PC) shown for comparison.

(B) Schematic Hi-C interaction map of the inactive X during *de novo* XCI, for example in differentiating mouse ES cells.

(C) Schematic Hi-C interaction map of the inactive X during post-XCI, for example in fully differentiated mouse ES cells.

The role of structural changes in XCI

While X chromosome folding has been extensively surveyed throughout XCI, the exact role of these structural changes was unclear. The existing structure of the active X has long been hypothesized to play a role in Xist spreading. Specifically, it has been proposed that Xist spreads via proximity transfer, or the transfer of *Xist* from its transcription locus to its closest existing contacts in 3D space (Simon et al., 2013; Wang et al., 2018). The overall structure of the X chromosome, in particular the location of the *Xist* locus in the A compartment in proximity with active chromatin, thus correlates well with the initial spreading of Xist to active genes (Simon et al., 2013). Though there is a strong correlation between X chromosome structure, particularly A/B compartments, and Xist spreading, it remained to be seen how spreading would be impacted if 3D structure were altered.

Initial studies disrupted Xi superstructure by deleting the border-defining repeat element *Dxz4* and its interacting partner *Firre*. As expected, *Dxz4* deletion, and to a lesser *Firre* deletion. disrupted megadomains (Bonora et al., 2018; Darrow et al., 2016; Froberg et al., 2018; Giorgetti et al., 2016). The *Dxz4-Firre* superloop was also disrupted by either *Dxz4* or *Firre* deletion. Surprisingly, deleting *Dxz4*, *Firre*, or both elements had little effect on gene silencing, with one study additionally not detecting effects on nuclear localization, chromatin accessibility, or H3K27me3 (Froberg et al., 2018). One study found that *Dxz4* deletion reduced escape of certain ‘facultative’ escapee genes, but did not affect constitutive escapees or silenced genes (Giorgetti et al., 2016). Other studies detected slight changes in compartmentalization (Darrow et al., 2016) and increases in chromatin accessibility (Bonora et al., 2018) upon *Dxz4* deletion, although again these were not sufficient to affect gene silencing. Overall, deletion of *Dxz4* and *Firre* greatly disrupted Xi structure but had surprisingly little impact on XCI. While surprising, this result may

be consistent with *Dxz4* and *Firre* not greatly disrupting X chromosome compartment structure in mouse, where the main negative results were observed. As A/B compartments, and in turn S1/S2 compartments, have the greatest correlation with Xist and Polycomb spreading, this layer of organization may be more important to XCI than superstructures. Both Xist and Polycomb are easily able to spread across the megadomain border at *Dxz4*, for example.

However, the conservation of *Dxz4, Firre*, megadomains and superloops from mouse to human hinted at an important role for these Xi superstructures (Darrow et al., 2016). One possibility is that Xi superstructures are important for XCI in a context that has not yet been investigated—in animals living in wild conditions rather than the lab, for example. A recent study proposed a role for domain borders in DNA damage (Arnould et al., 2021). Further studies will be needed to determine if the megadomain border at *Dxz4* could have a similar function for DNA damage repair on the Xi.

Another intriguing possibility is that the function of *Dxz4* and *Firre* is unrelated to XCI. Indeed, several recent studies supported a role for these loci outside of the Xi. One study found that *Dxz4* and *Firre* deletion in mice again did not result in defects in XCI but led to dysregulation of various autosomal genes, with little overlap between genes dysregulated in different organs (Andergassen et al., 2019). *Firre* deletion disrupted the greatest number of genes. Another study showed *Firre* deletion on the Xa affected H3K27me3 levels on the Xi (Fang et al., 2020). Again, *Firre* deletion did not greatly impact gene silencing on the Xi, but did lead to dysregulation of genes on autosomes. Therefore, *Firre* RNA could act both *in cis* on the X chromosome as well as *in trans* genome-wide.

Another possibility is that cohesin levels themselves may be more important to XCI than specific structures formed by cohesin. Even after *Dxz4* and *Firre* deletion, cohesin presumably continues to be loaded onto the Xi. This cohesin may translocate along the Xi and extrude transient loops, even if it no longer forms binding peaks at *Dxz4* and *Firre*. These loops could be

more important to XCI than specific CTCF-mediated borders at *Dxz4* and *Firre.* This could explain why deletion of *Dxz4* and *Firre* has a relatively small impact on the Xi.

Another avenue of disrupting Xi structure is deleting SMCHD1. Deleting *Smchd1* during *de novo* XCI leads to persistent S1/S2 compartments. Such a failure to merge compartments coincides with defective Xist and Polycomb spreading. These defects on spreading have disastrous effects on gene silencing, with ~40% of Xi genes failing to be silenced (Gdula et al., 2019; Sakakibara et al., 2018; Wang et al., 2018). Therefore, disruption of compartment-merging via deletion of *Smchd1* has much stronger consequences for *de novo* XCI than ablation of superstructures via *Dxz4* and *Firre* deletion. Even in the maintenance phase of XCI, deletion of *Smchd1* reverses Xi structure, causing S1/S2 compartments to re-emerge and aberrant accumulation of H3K27me3 in S1 compartment partitions (Wang et al., 2019). While this did not result in immediate reactivation of genes on the Xi, it rendered Xi genes more susceptible to reactivation after treatment with 5-aza-2′-deoxycytidine (Aza). Therefore, merging of compartments via SMCHD1 is important for XCI stability.

Other Xi structures during mammalian development

The 3D structure of the X chromosome has mainly been studied in differentiating mESCs or fully differentiated mouse cell lines. However, it is informative to look at X chromosome structure in other contexts as well. As noted, Xi superstructures such as megadomains and superloops are conserved from mouse, rhesus, to human (Darrow et al., 2016). In contrast, the compartment structure of the Xi varies between mouse and human. While in mouse, the fully differentiated Xi is mostly “compartment-less”, the human Xi has strong compartments. This may be due to the human Xi’s distinct chromatin organization from the mouse Xi. While the fully differentiated mouse Xi displays Polycomb-rich and poor partitions, the human Xi additionally has H3K9me3/HP1 rich partitions, which are lacking in mouse (Darrow et al., 2016). Despite this, SMCHD1 is enriched on the human Xi, much as it is on the mouse Xi (Nozawa et al., 2013). Why

SMCHD1 fails to merge compartments on the human Xi, as it does on the mouse Xi, is not clear. It is possible that the mechanisms that form H3K9me3/HP1-rich and poor compartments are not amenable to mixing. The phase separation qualities of H3K9me3/HP1 could be stronger than that of Polycomb, for example. In addition, SMCHD1 is only enriched on Polycomb-rich partitions on the human Xi, not the H3K9me3/HP1-rich partitions (Nozawa et al., 2013). Such restriction could prevent SMCHD1 from carrying out its merging function. This could provide an intriguing clue as to requirements for gene silencing in H3K9me3/HP1-rich partitions versus Polycomb-rich partitions.

X chromosome structures in different systems of dosage compensation

To fully understand the significance of structural changes in XCI, it is informative to look beyond mammals. Indeed, dosage compensation occurs in every organism that reproduces sexually. Intriguingly, mechanisms of dosage compensation vary drastically.

In the nematode *Caenorhabditis elegans*, both X chromosomes in hermaphrodites (XX) are downregulated in order to balance dosage of X-linked genes with males (XO) (Meyer, 2010). To appreciate the how X chromosome folding changes during dosage compensation, it is informative to examine the general differences between 3D genome structure in *C. elegans* and mammals. Although chromosomes display compartments, similar to A/B compartments in mammals, domains are much less prominent, mainly appearing on the X chromosome (Crane et al., 2015). Intriguingly, the *C. elegans* Dosage Compensation Complex (DCC) belongs to the family of condensins. Condensins, which are SMC complexes like cohesin, play essential and globally conserved roles in chromosome compaction and segregation during mitosis and meiosis (Hirano, 2005). In particular, DCC appears to form domains on the *C. elegans* X chromosome by binding domain boundaries (Crane et al., 2015). Such domains are much less prominent in Dosage Compensation mutants and XO males. Therefore, what occurs on the *C. elegans* X chromosomes contrasts with what occurs on the mammalian Xi, where cohesin binding becomes

weaker. The mechanism by which DCC creates domains on the X chromosome is not clear. Recent studies showed that yeast condensin has loop extrusion activity (Ganji et al., 2018). Higher resolution Hi-C or super resolution microscopy would reveal to what extent DCC extrudes loops on the *C. elegans* hermaphrodite X chromosomes versus autosomes and male X chromosomes. If such loops are detected, it would be interesting to determine whether factors bound to domain borders could insulate such loops, as CTCF does at loop domain borders in mammals. The *C. elegans* observations may also indicate it could be interesting to investigate the role of condensin in shaping the mammalian X chromosome.

In *Drosophila*, instead of females (XX) silencing one X chromosome, males (XY) upregulate the one X chromosome. Hi-C studies showed that the *Drosophila* chromosomes, like mammalian chromosomes, are organized into compartments as well as domains. However, unlike mammals, CTCF does not typically bind the borders of such domains. Instead *Drosophila* domains correspond with transcriptional state; that is, they reflect compartmentalization, rather than loop extrusion by SMC complexes (Rowley et al., 2017). Surprisingly, conformational changes between the male and female *Drosophila* X chromosomes as detected by Hi-C are subtle (Pal et al., 2019). Changes may be more apparent by microscopy, as a recent study suggested that roX RNAs and the male-specific lethal (MSL) complex could induce male X chromosome specific compartments in *Drosophila* (Valsecchi et al., 2021). Another interesting observation is that superstructures are not detected on dosage-compensated X chromosomes in *C. elegans* or *Drosophila*. Therefore, structures such as megadomains and superloops may be specific to the X-inactivation mechanism in mammals.

In my thesis, I revisit the structure-function question for 3D genome structure using XCI as a model. In Chapter 2, we identify phosphorylation sites of CTCF and determine whether disrupting one of these, Serine 224 phosphorylation, impacts gene expression or domain formation. We find only subtle effects of point mutations that disrupt this PTM, supporting the global CTCF ablation approach. In Chapter 3, I investigate the impact of targeted degradation of

CTCF on chromosome structure and XCI. I find that 3D genome structure such as domains is remarkably resistant to loss of CTCF. Following the same trend, CTCF ablation also largely did not impact folding of the Xi or gene silencing. These results suggest that alternative mechanisms exist which could establish borders, or that small amounts of residual CTCF are sufficient for this layer of organization. Despite this, the CTCF degron cell line provides an ideal system for follow-up experiments for CTCF function.

In Chapter 4, I examine how toggling cohesin levels up or down via targeted degradation of RAD21 or WAPL impacts chromosomal contacts, gene expression, and XCI. I found that cohesin degradation ablates cohesin-dependent structures such as chromatin loops without affecting compartments. WAPL degradation, in contrast, increases cohesin binding on chromatin and cohesin-anchored loops while weakening long-range compartmental interactions. Surprisingly, gene expression was fairly resistant to cohesin degradation, although downregulated genes were enriched for Polycomb targets. Interchromosomal contacts between the highly Polycomb-enriched *Hox* clusters also increased. In contrast, WAPL degradation led to upregulation of Polycomb targets and loss of interchromsomal *Hox* interactions. On the Xi, cohesin loss disrupted megadomains; however the *Dxz4-Firre* superloop actually become more prominent. Like autosomes, gene silencing on the Xi was remarkably resistant to cohesin degradation. WAPL degradation did not affect the megadomain border but weakened long-range compartmental interactions on the Xi. Surprisingly, removing cohesin also resulted in an Xi-like “superloop” on the Xa, while removing WAPL resulted in a “megadomain”-like structure on the Xi. Forced cohesin retention via WAPL degradation also compromised spreading of Xist and Polycomb and gene silencing, especially at the ends of the Xi. Therefore, a balance of cohesin is needed for proper interchromosomal contacts, Polycomb repression and XCI.

Chapter 2

Exploration of CTCF post-translation modifications uncovers Serine-224 phosphorylation by PLK1 at pericentric regions during the G2/M transition

Attribution

The content of this chapter has been published in Del Rosario and Kriz *et al.*, Elife, 2019.

ABSTRACT

The zinc finger CCCTC-binding protein (CTCF) carries out many functions in the cell. Although previous studies sought to explain CTCF multivalency based on sequence composition of binding sites, few examined how CTCF post-translational modification (PTM) could contribute to function. Here, we performed CTCF mass spectrometry, identified a novel phosphorylation site at Serine 224 (Ser^{224}-P), and demonstrate that phosphorylation is carried out by Polo kinase 1 (PLK1). CTCF Ser^{224}-P is chromatin-associated, mapping to at least a subset of known CTCF sites. CTCF Ser^{224}-P accumulates during the G2/M transition of the cell cycle and is enriched at pericentric regions. The phospho-obviation mutant, S224A, appeared normal. However, the phospho-mimic mutant, S224E, is detrimental to mouse embryonic stem cell colonies. While ploidy and chromatin architecture appear unaffected, S224E mutants differentially express of hundreds of genes, including p53 and p21. We have thus identified a new CTCF PTM and provided evidence of biological function.

INTRODUCTION

CCCTC-binding protein (CTCF) has been studied in many capacities since its discovery over twenty years ago. Originally identified as a candidate transcription regulator of *c-myc*, this multi zinc finger protein is highly conserved amongst Nephrozoa (Heger et al., 2012; Lobanenkov et al., 1990). Subsequent studies further revealed that CTCF has insulatory activity, specifically in blocking enhancer-promoter interactions, and in separating transcriptionally active genomic regions from heterochromatic domains (Merkenschlager and Odom, 2013; Ong and Corces, 2014). Later genomic studies revealed tens of thousands of CTCF binding sites in mammalian genomes (Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Xie et al., 2007), with >5,000 of these being conserved (Schmidt et al., 2012). In addition to CTCF's role in blocking enhancer-promoter loops, recent chromatin conformation capture (3C) based assays revealed that CTCF paradoxically also plays an architectural role in shaping the genome as well, helping to mediate three-dimensional chromatin loops in some cases. In addition, 3C-based assays also revealed that CTCF binds to the borders of many Topologically Associating Domains (TADs), megabase-sized regions which function to insulate chromatin interactions such as promoter-enhancer loops (Dixon et al., 2012; Nora et al., 2012; Ong and Corces, 2014). Disruption of domains, for example through deletion of CTCF sites, can lead to genetic disease through ectopic enhancer-mediated upregulation of genes (Lupianez et al., 2015; Nora et al., 2017; Rao et al., 2014; Sanborn et al., 2015).

How can a single protein carry out so many different functions in the cell? In addition, how does the cell specify which function CTCF carries out at a given binding site, especially when some of these functions, for example mediating versus blocking chromatin loops, directly contradict each other? Numerous studies have attempted to explain the multivalency of CTCF on the basis of factors such as motif composition of binding site, chromatin state, DNA methylation, and CTCF's role in mediating three-dimensional chromatin interactions (Lu et al.,

2016). For example, it has been proposed that CTCF's insulatory activity can be explain in context of its mediation of three-dimensional chromatin loops.

In addition to these models, which predict the function of CTCF based on the chromatin state around a given binding site, other studies have sought to define CTCF function based on post-translation modifications (PTMs) of the protein itself. Poly(ADP-ribosyl)ation of CTCF has been found to play roles in imprinting and nucleolar transcription (Torrano et al., 2006; Yu et al., 2004) while SUMOylation of CTCF appears to enhance its repressive function (MacPherson et al., 2009). Finally, phosphorylation of CTCF has been proposed to turn CTCF into a transcription activator (El-Kady and Klenova, 2005) or reduce its DNA binding activity (Sekiya et al., 2017), depending on site. These prior studies demonstrate that CTCF PTMs, possibly in combination, could influence the myriad of CTCF functions. The extent to which CTCF is post-translationally modified is currently no known. Many prior PTMs were identified indirectly by candidate-based approaches. Several proteomic screens have been carried out using mass spectrometry (Dephoure et al., 2008; Kettenbach et al., 2011; Olsen et al., 2010; Rigbolt et al., 2011), though phosphorylation sites were called by confidence scores made by probabilistic scoring algorithms without further validation or characterization. Here, we endeavored to screen for new CTCF PTMs by taking an unbiased approach. We identify a novel phosphorylation site at Serine 224 (Ser^{224}-P), perform an extensive characterization of its cellular functions, and uncover an effect on growth of embryonic stem cell colonies and gene regulation.

RESULTS

Murine CTCF is phosphorylated at a highly conserved position, Ser^{224}

To explore the possibility that PTMs regulate CTCF, we took an unbiased approach to both confirm known and identify novel PTMs, specifically S/T/Y phosphorylation using immunoprecipitation and mass spectrometry (Dephoure et al., 2008; Kettenbach et al., 2011; Klenova et al., 2001; Olsen et al., 2010; Rigbolt et al., 2011; Sekiya et al., 2017). We utilized a doxycycline-inducible system to express murine CTCF-3xFLAG (Sun et al., 2013). This allowed us to purify CTCF from cells using a FLAG epitope rather than a CTCF antibody, as the latter could select against PTMs present within the antigenic sequence. The CTCF-3xFLAG transgene was stably introduced into female immortalized rtTA MEFs and doxycycline induction was verified by GFP microscopy (Figure 2.1A) (Jeon and Lee, 2011). Anti-FLAG immunofluorescence confirmed that CTCF-3xFLAG co-stained the nucleus in the same pattern as total CTCF (Figure 2.1B). Exogenous expression of CTCF-3xFLAG in this system only occurs at a modest level, which we confirmed by Western blot (Figure 2.1C) (Sun et al., 2013). Previously, we also found that even slight perturbation of CTCF levels affects differentiation of female ES cells (Sun et al., 2013). However, induction of the transgene was well-tolerated in SV40T immortalized MEFs as they were viable and stably expressing CTCF-3xFLAG even after 9 days (Figure 2.1C). We thus concluded that exogenously expressed CTCF-3xFLAG in immortalized MEFs could be a tractable system for analyzing CTCF PTMs.

We expanded the CTCF-3xFLAG MEFs with or without induction for 3 days and nuclei were isolated. CTCF is tightly associated with chromatin through the central zinc fingers binding its core DNA motif (Nakahashi et al., 2013; Yusufzai and Felsenfeld, 2004). However homogenizing nuclei under mild conditions with nucleases was not preferable as endogenous phosphatase activity would have been present. Thus we generated nuclear extracts under stringent conditions. Following dialysis of the nuclear extracts, anti-FLAG immunoprecipitation was performed and eluted material was resolved by SDS-PAGE and Coomassie Blue stained. A

band corresponding to CTCF-3xFLAG was excised for mass spectrometry analysis (arrow, Figure 2.1D). Following trypsin digestion and extraction, phosphorylated peptides were enriched using immobilized metal affinity chromatography (IMAC) with Fe-NTA resin prior to LC MS/MS. A unique murine site of serine phosphorylation was identified at Ser^{224}, which is ~50aa N-terminal of the CTCF zinc finger domain (Figure 2.1E). Notably, in a phosphoproteome screen of HeLa cells, a single peptide containing human CTCF Ser^{224}-P was identified (Kettenbach et al., 2011). However, the significance of this PTM in human cells was neither validated nor explored. While it is also reported that CTCF is phosphorylated at Ser^{604}, Ser^{609}, Ser^{610} and Ser^{612} (Dephoure et al., 2008; El-Kady and Klenova, 2005; Klenova et al., 2001; Olsen et al., 2010; Rigbolt et al., 2011); we were unable to identify peptides containing these residues. As these phospho-sites lie in close proximity in a region of CTCF poorly cut by trypsin and chymotrypsin, it is possible that resolution of these phospho-sites was infeasible with our approach. Previous mass spectrometry studies which identified these residues were also global studies that did not provide spectra or additional validation of sites, making direct comparison with our study difficult. Likewise, we were also unable to confirm if the linker sequences between the zinc fingers were phosphorylated as previously reported (Sekiya et al., 2017).

Interestingly, CTCF orthologs have only been found in Nephrozoa (Heger et al., 2012). To explore the potential evolutionary significance of CTCF Ser^{224}-P, we compared CTCF amino acid sequences from several Nephrozoa species with Clustal (Figure 2.1F). The CTCF amino acid sequences from mouse, rat, human, chimpanzee, opossum, chicken, frog, zebrafish, sea urchin, fruit fly, and water flea were aligned. A 25 amino acid window centered on mouse Ser^{224} remarkably revealed striking conservation proximal to this position amongst vertebrata, but not echinodermata or arthropoda (arrow, Figure 2.1F). This conservation suggests that Ser^{224} phosphorylation is potentially catalyzed by a conserved vertebrate kinase. Ergo, this PTM may contribute distinctly to the regulatory complexity of vertebrate genomes relative to other deuterostomes or protostomes.

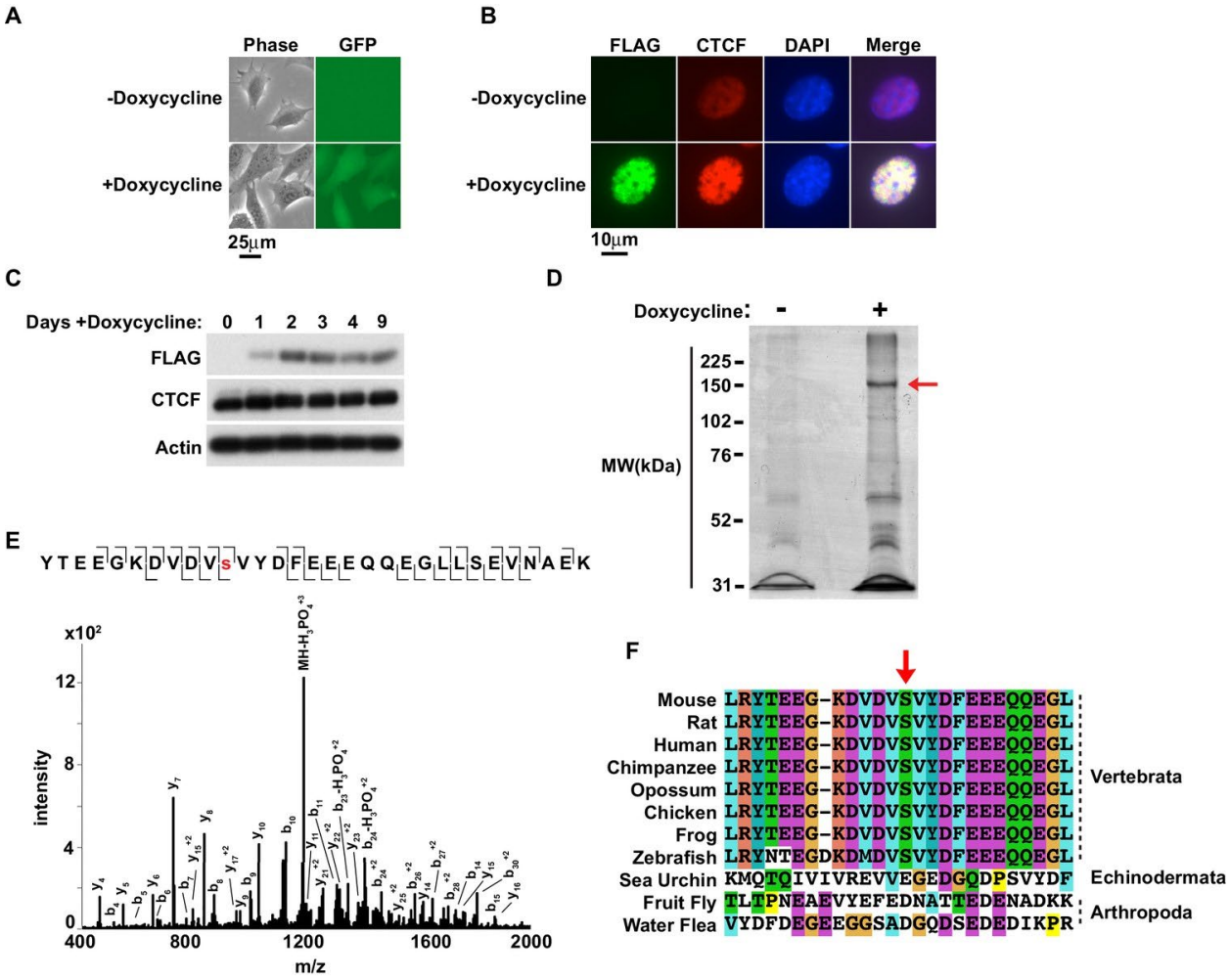


Figure 2.1. Murine CTCF is phosphorylated at Ser^{224}.

(A) GFP microscopy of co-inducible CTCF-3xFLAG, EGFP MEFs +/- 1μg/mL doxycycline for 48hrs. Bar, 25μm.

(B) FLAG and CTCF immunofluorescence accompanying (A). Bar, 10μm.

(C) FLAG, CTCF, and β-actin Western blot of whole cell extracts from inducible CTCF-3XFLAG MEFs treated up to 9 days with 1μg/mL doxycycline.

(D) Coomassie stained gel of anti-FLAG immunoprecipitate from CTCF-3xFLAG MEFs treated 72hrs +/- 1μg/mL doxycycline. Red arrow, band analyzed by mass spectrometry.

(E) Manually validated mass spectra of CTCF peptide Tyr^{214}-Lys^{244} with y and b ions identified. Phosphorylation event at Ser^{224} indicated by red s.

(F) ClustalX alignment of CTCF sequences from the indicated species. Shown is a 25 amino acid window centered on mouse Ser^{224} (red arrow).

CTCF Ser^{224} can be phosphorylated by CK2

Discovery of a novel phosphorylation site at Ser^{224} raised the question as to which kinase could modify this position. It was previously reported that casein kinase 2 (CK2) could phosphorylate CTCF in vitro at Ser^{604}, Ser^{609}, Ser^{610} and Ser^{612} (Klenova et al., 2001). Interestingly, the highly conserved amino acid sequence surrounding Ser^{224} fits the known CK2 recognition site (S-x-x-D/E) (Figure 2.1F) (Meggio and Pinna, 2003). Thus, we decided to test if CK2 also modifies Ser^{224}. To do this, we first generated recombinant CTCF and in vitro phosphorylated it with CK2 and [γ-^{32}P]-ATP (Sun et al., 2013). Following SDS-PAGE, autoradiography revealed that recombinant CTCF can be phosphorylated by CK2 (Figure S1.1A). To determine if Ser^{224} specifically is phosphorylated by CK2, we performed that assay with non-radioactive ATP and excised a Coomassie stained band corresponding to full length CTCF (red box, Figure S1.1B). Mass spectrometry analysis of the band, revealed that CKII phosphorylated CTCF Ser^{224} (Figure S1.1C). As with our analysis of immunoprecipitated CTCF from MEFs, we did not observe phosphorylation at Ser^{604}, Ser^{609}, Ser^{610} or Ser^{612}. However, again, technical hurdles (i.e., the close proximity of these sites in a region poorly cut by trypsin and chymotrypsin) may preclude their identification by mass spectrometry. Furthermore, while we discovered that CK2 could phosphorylate Ser^{224} in vitro, it remained to be seen if this occurs in vivo. Finally, it is also a distinct possibility that CTCF Ser^{224} could be recognized by additional kinases pertinent to specific functions.

Generation of a CTCF Ser^{224}-P antibody

To explore the significance of this PTM in cells, we generated an antibody to CTCF Ser^{224}-P in collaboration with Cell Signaling Technology. Following screening of prospective affinity

purified antibodies, we further tested if the candidate CTCF Ser^{224}-P antibody recognized a phosphorylated protein by Western blot. We generated cell extract from immortalized rtTA MEFs and treated the extract with Lambda Protein Phosphatase. Western blot revealed that the CTCF Ser^{224}-P antibody had reduced affinity for the phosphatase treated samples (Figure 2.2A). The CTCF Ser^{224}-P antibody also recognizes a smaller (~125kDa) band that is not detected by a CTCF monoclonal antibody (Figure 2.2A). Notably CTCF migrates at a higher molecular weight (150kDa) than predicted (84kDa) in SDS-PAGE gels. One possibility is that CTCF Ser^{224}-P occurs on two different CTCF isoforms in vivo, one of which is not recognized by the monoclonal antibody.

However, to further confirm that the affinity purified antibody preferred CTCF Ser^{224}-P, we phosphorylated bacterially expressed recombinant CTCF and tested the antibody by Western blot. We phosphorylated wild type CTCF, S224A, and S224E with CK2. While the Western blot revealed some affinity of the CTCF Ser^{224}-P antibody for unphosphorylated CTCF; the antibody preferred the CK2 phosphorylated protein (2nd column, Figure 2.2B). Moreover, the antibody is likely reactive to Ser^{224}-P rather than other potential CK2 phosphorylation sites as there appeared to be no bias to CK2 phosphorylated S224A or S224E (Figure 2.2B).

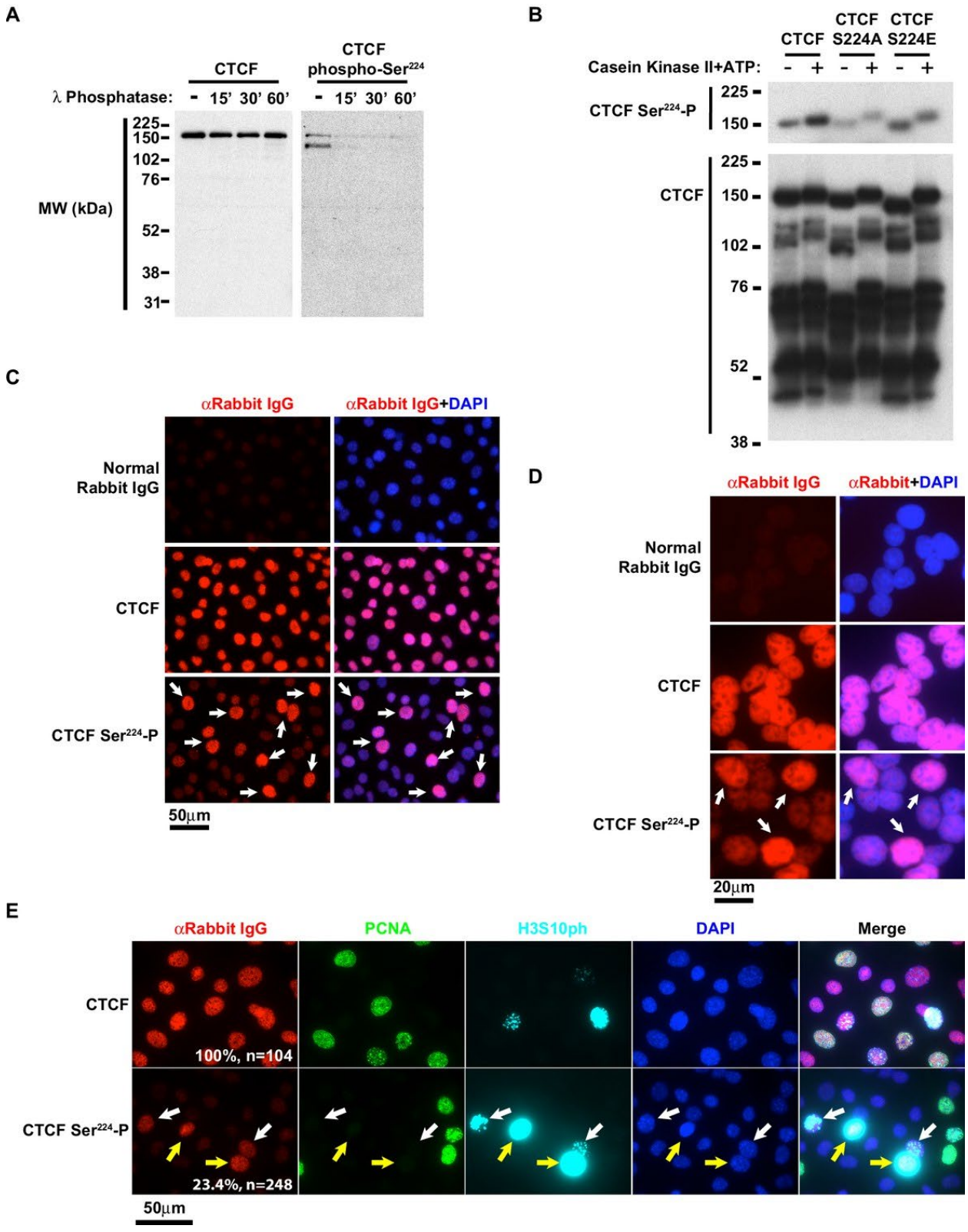


Figure 2.2. CTCF Ser224-P is conserved and accumulates at G2/M.

(A) CTCF and CTCF Ser224-P Western blots of MEF lysates treated with Lambda Protein Phosphatase for the indicated times.

(B) CTCF and CTCF Ser224-P Western blots of recombinant CTCF, CTCF S224A, and CTCF S224E +/- Casein Kinase II in vitro phosphorylation.

(C) Immunofluorescence performed on asynchronous MEFs with the indicated antibodies. Nuclei counterstained with DAPI. Bar, 50μm. White arrows, prominent CTCF Ser224-P cells.

(D) Immunofluorescence performed on asynchronous HEK293 with the indicated antibodies. Nuclei counterstained with DAPI. Bar, 20μm. White arrows, prominent CTCF Ser224-P cells.

(E) Immunofluorescence performed on asynchronous MEFs with the indicated antibodies. Nuclei counterstained with DAPI. Bar, 50μm. % of n cells labeled with CTCF or CTCF Ser224-P antibodies indicated. White arrows, early G2. Yellow arrows, late G2.

Cytological distribution of CTCF Ser^{224}-P is comparable to total CTCF

Using immunofluorescence, we next asked if the CTCF Ser^{224}-P antibody recognized its epitope in situ and if CTCF Ser^{224}-P localization was comparable to total CTCF. Using immortalized tetraploid female MEFs, we found that the CTCF Ser^{224}-P antibody labeled a nuclear antigen in a pattern reminiscent of CTCF (Figure 2.2C, Figure S1.2A). Furthermore, CTCF was visually excluded during interphase from nucleoli and constitutive heterochromatic regions (Burke et al., 2005). Therefore we also scrutinized CTCF Ser^{224}-P relative to the nucleolus, constitutive heterochromatin, and facultative heterochromatin using antibodies against B23, H3K9me3 and H3K27me3 respectively (Bannister and Kouzarides, 2011; Spector et al., 1984). In wildtype female cells, H3K27me3 is known to be prominently enriched on the inactive X chromosome (Xi). In tetraploid MEFs, it usually appears as two nuclear bodies. CTCF Ser^{224}-P was not enriched on

the Xi, and we found no evident visual distinction between CTCF Ser^{224}-P and total CTCF nuclear distribution regardless of co-staining nucleoli and heterochromatin (Figure S1.2B-D). Thus, at a cytological level, CTCF Ser^{224}-P appears to distribute similarly as total CTCF in the nucleus.

CTCF Ser^{224}-P accumulates during the G2/M transition of the cell cycle

However, a remarkable observation from our immunofluorescence assays was that CTCF Ser^{224}-P labeling was not homogeneous in an asynchronous population of MEFs. Some of the nuclei were more intensely labeled by the CTCF Ser^{224}-P antibody whereas nuclei stained with total CTCF antibody were uniformly labeled (arrows, Figure 2.2C). While it also appears that there were less intensely CTCF Ser^{224}-P stained nuclei that were observable particularly when compared to normal IgG staining, this may be attributable to antibody cross reactivity to unmodified CTCF Ser^{224} (Figure 2.2B, C). Since CTCF Ser^{224}-P potentially occurs in human cells (Kettenbach et al., 2011) and vertebrates share a high degree of sequence identity surrounding Ser^{224} (Figure 2.1E), we next asked if the mottled CTCF Ser^{224}-P pattern is discernable in asynchronous human cells. As predicted, CTCF Ser^{224}-P labeling of asynchronous HEK293 revealed that a subset of cells was intensely labeled by the antibody (arrows, Figure 2.2D). This intimates that this PTM may be associated with a vertebrate-conserved CTCF regulatory mechanism.

Given the irregular CTCF Ser^{224}-P staining of asynchronous cells, we posited that perhaps this PTM is regulated by the cell cycle. To explore this possibility, we performed CTCF Ser^{224}-P immunofluorescence on asynchronous MEFs with additional antibodies to cell cycle markers. To identify cells in S phase and G2/M, we labeled cells with antibodies against PCNA and phosphorylated histone H3 Ser^{10} (H3S10ph) respectively (Bravo and Macdonald-Bravo, 1985; Hendzel et al., 1997). While a total CTCF antibody labeled 100% of the cells (n=104), again only a fraction of the cells (23.4%, n=248) were CTCF Ser^{224}-P positive (Figure 2.2E). Comparing this fraction to the cell cycle markers revealed that a striking majority (94.8%) of the CTCF Ser^{224}-

P positive cells were co-labeled with H3S10ph antibody (Figure 2.2E, Figure S1.2E). The CTCF Ser^{224}-P labeling was most prominent in cells in both early G2 and late G2/prophase, as evidenced from the H3S10ph staining pattern (white arrows, early G2; yellow arrows, late G2/prophase. Figure 2.2E, Figure S1.2E). There were also CTCF Ser^{224}-P positive cells co-stained with PCNA antibody (27.6%) (Orange and magenta arrows, Figure S1.2E). However, the majority of these CTCF Ser^{224}-P-PCNA double positive cells (81.3%) were also faintly positive for H3S10ph (magenta arrows, S2.10E). Taken together, these findings indicate that phosphorylation of CTCF Ser^{224} likely initiates at the end of S phase and peaks at G2/M. To further evaluate if CTCF Ser^{224}-P is enriched at G2/M, we arrested TST-1 mESCs 20hrs with the CDK1 inhibitor RO-3306 (Vassilev et al., 2006). RO-3306 treatment resulted in an accumulation of CTCF Ser^{224}-P positive cells blocked at G2/M as evident from the concomitant increase in H3S10ph co-staining (Figure S1.2F). These data pointed to an association of CTCF Ser^{224}-P with the G2/M transition of the cell cycle.

PLK1 phosphorylates CTCF Ser^{224}

To explore how CTCF is specifically phosphorylated at the G2/M transition, we first sought to identify which kinase(s) could be phosphorylating CTCF Ser^{224} at the end of S and in G2/M. Naturally, our first search criterion was that a prospective kinase also had to be expressed at these stages of the cell cycle. We also presumed that the kinase has a bona fide role limited to G2/M. Using these guidelines, Polo-like kinase I (PLK1) was the most attractive candidate. PLK1 is a regulator of the G2/M transition and phosphorylates mitosis-associated substrates (Barr et al., 2004). Importantly, PLK1 expression increases from S phase and peaks during G2 (Golsteyn et al., 1994; Lake and Jelinek, 1993). Comparing the consensus PLK1 substrate site to the amino acid sequence flanking CTCF Ser^{224} revealed that Ser^{224} is a potential target of PLK1 (Figure 2.3A) (Nakajima et al., 2003). Of note, CTCF was not identified as a PLK1 target in HeLa cells (Kettenbach et al., 2011). However, the designation of PLK1 targets in that study was not resultant

from direct assay. Similarly, we did not find PLK1 in our CTCF-3xFLAG IP-MS/MS (Figure 2.1D) likely due to stringent nuclear extraction conditions used to extract CTCF from the chromatin fraction that disrupted native protein-protein interactions (see Methods). We also examined PLK1 conservation across Nephrozoa by clustering the amino acid sequence identity of orthologs using Clustal. For reference, orthologs of the kinases ATM and GSK3B were also compared. We found that PLK1 conservation mirrors that of its potential CTCF Ser^{224} substrate, where the kinase is most conserved amongst vertebrates (Figure 2.1F, 2.3B). This finding may further hint at a homologous CTCF Ser^{224}-P function particular to vertebrates.

To test if CTCF can be phosphorylated by PLK1, we first treated asynchronous immortalized MEFs with the PLK1-specific inhibitor BI 6727 (Rudolph et al., 2009). We performed CTCF and CTCF Ser^{224}-P immunofluorescence after 12hrs of treatment with 100nM or 1000nM BI 6727 (EC_{50} ~10-150nM) (Gorlick et al., 2014; Rudolph et al., 2015; Rudolph et al., 2009). Consistent with inhibition of PLK1, we observed a likely G2/M defect evident in a higher percent of H3S10ph positive cells in BI 6727 treated cells than control (DMSO, 5.6%; 100nM BI 6727, 11.4%; 1000nM BI 6727, 29.2%; t=12hrs; n>100 cells). While 30.8% (n=214) of untreated cells were CTCF Ser^{224}-P positive, only 18.3% (n=175) were positive following treatment with 100nM BI 6727 (Figure 2.3C). The range of EC_{50} values for BI 6727 (10-150nM) suggests that remaining CTCF Ser^{224}-P was likely due to incomplete inhibition of PLK1 at this concentration. Supporting this, 1000nM BI 6727 completely ablated CTCF Ser^{224}-P phosphorylation (n=101) (Figure 2.3C). Total CTCF expression was not affected by 12hr treatment with BI 6726 at either concentration (Figure 2.3C). While this indicates that PLK1 is essential for CTCF Ser^{224}-P, treatment with BI 6727 does not exclude the possibility that our observations were an indirect outcome. Therefore, we also performed an in vitro kinase assay with purified recombinant CTCF and PLK1, using both Casein and PLK1 autophosphorylation as positive controls (*** and ** respectively, Figure 2.3D) (Golsteyn et al., 1995). In support of our hypothesis, we found that PLK1 directly phosphorylated CTCF (*, Figure 2.3D). However, this result did not indicate if CTCF Ser^{224}-P had occurred. To

test this, we performed the PLK1 in vitro kinase assay without radioactive isotope and we excised the band corresponding to full-length CTCF from the Coomassie stained gel (red box, Figure 2.3E). Analysis by mass spectrometry revealed phosphorylation at Ser^{224}, arguing that indeed PLK1 is the kinase for CTCF Ser^{224} (Figure 2.3F). However, as we did not capture *in vivo* interactions of CTCF with PLK1 (or CK2), we could not confirm either as the CTCF kinase for certain. Therefore, both PLK and CK2 remain potential CTCF kinases.

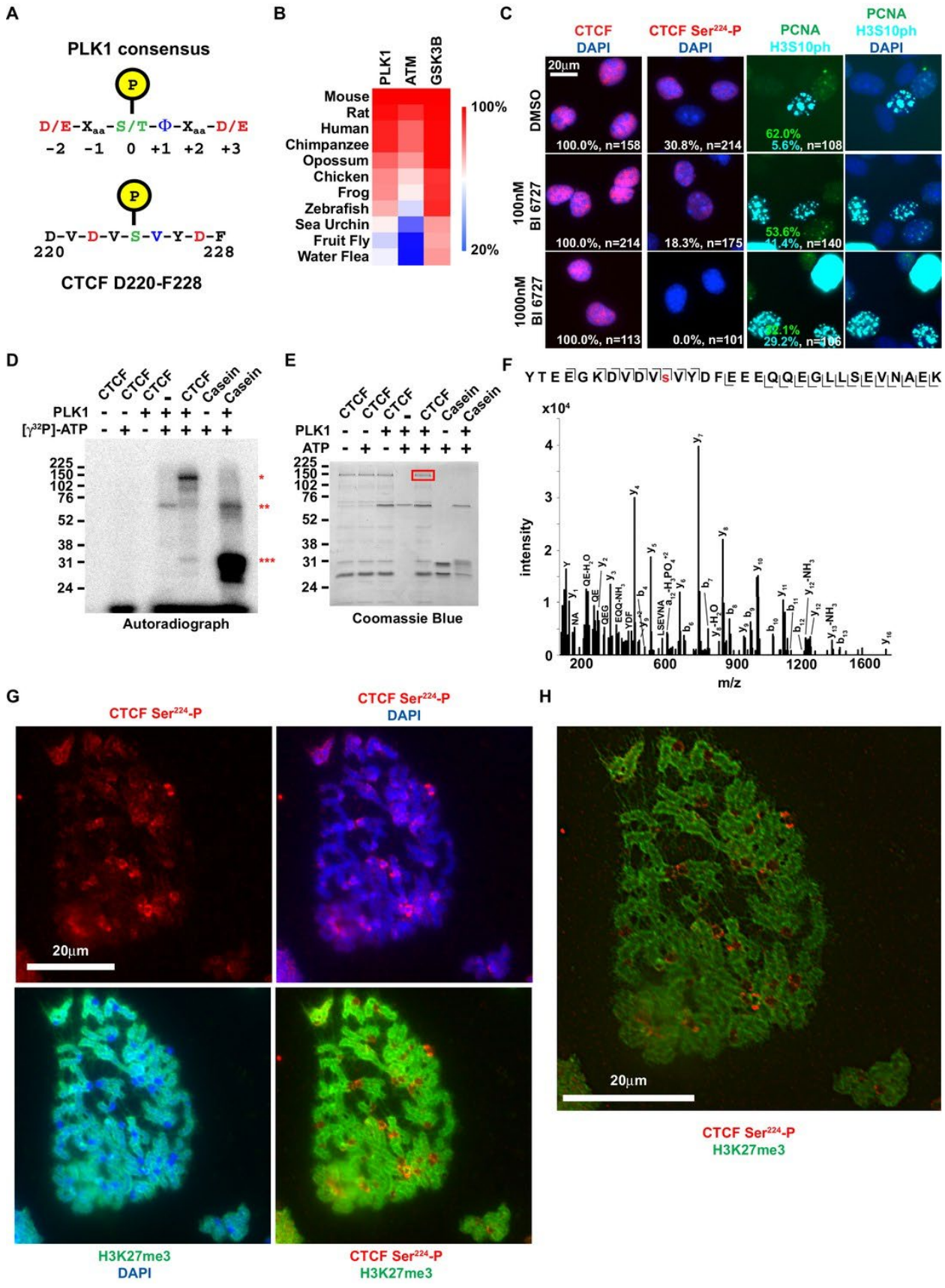

A
PLK1 consensus
D/E-X$_{aa}$-S/T-Φ-X$_{aa}$-D/E
-2 -1 0 +1 +2 +3
D-V-D-V-S-V-Y-D-F
220 228
CTCF D220-F228
B
PLK1
ATM
GSK3B
Mouse
Rat
Human
Chimpanzee
Opossum
Chicken
Frog
Zebrafish
Sea Urchin
Fruit Fly
Water Flea
100%
20%
C
CTCF DAPI
CTCF Ser224-P DAPI
PCNA H3S10ph
PCNA H3S10ph DAPI
20μm
DMSO
100nM BI 6727
1000nM BI 6727
100.0%, n=158
30.8%, n=214
62.0%
5.6%, n=108
100.0%, n=214
18.3%, n=175
53.6%
11.4%, n=140
100.0%, n=113
0.0%, n=101
62.1%
29.2%, n=106
D
CTCF
Casein
PLK1
[γ32P]-ATP
225
150
102
76
52
38
31
24
Autoradiograph
E
ATP
Coomassie Blue
F
Y T E E G K D V D V s V Y D F E E E Q Q E G L L S E V N A E K
x10^4
intensity
m/z
200
600
900
1200
1600
G
CTCF Ser224-P
CTCF Ser224-P DAPI
20μm
H3K27me3 DAPI
CTCF Ser224-P H3K27me3
H
20μm
CTCF Ser224-P H3K27me3

Figure 2.3. CTCF Ser^{224} is phosphorylated by PLK1 and prominently labels pericentric chromatin.

(A) Graphic comparison of PLK1 consensus substrate sequence with CTCF D220-F228. Green S/T with yellow encircled P, phosphorylation site at position 0. Red D/E, aspartic or glutamic acid. Blue Φ, hydrophobic amino acid.

(B) Amino acid sequence identity heat map for the conserved kinases PLK1, ATM, and GSK3B. Nephrozoa species aligned in Figure 1F shown.

(C) CTCF, CTCF Ser224-P, PCNA, and H3S10ph immunofluorescence performed on MEFs treated with DMSO or BI 6727 at the indicated concentrations for 12hrs. Nuclei counterstained with DAPI. Bar, 20μm. % of n cells labeled with CTCF, CTCF Ser224-P, PCNA, or H3S10ph antibodies indicated.

(D) PLK1 in vitro kinase assay with CTCF or dephosphorylated Casein substrates. Red *, phosphorylated CTCF. Red **, autophosphorylated PLK1. Red ***, phosphorylated casein.

(E) In vitro kinase assay performed in parallel to *D* without radioactive isotope. SDS-PAGE gel Coomassie stained. Red box, CTCF band excised for mass spectrometry analysis.

(F) Manually validated mass spectra of CTCF peptide Tyr^{214}-Lys^{244} with y and b ions identified. Phosphorylation event at Ser^{224} indicated by red s.

(G) Immunofluorescence performed on TST-1 mESC metaphase chromosomes with the indicated antibodies. DNA stained with DAPI.

(H) CTCF Ser224-P and H3K27me3 co-stain from (G) deconvolved.

CTCF Ser224-P localizes to pericentric regions during mitosis

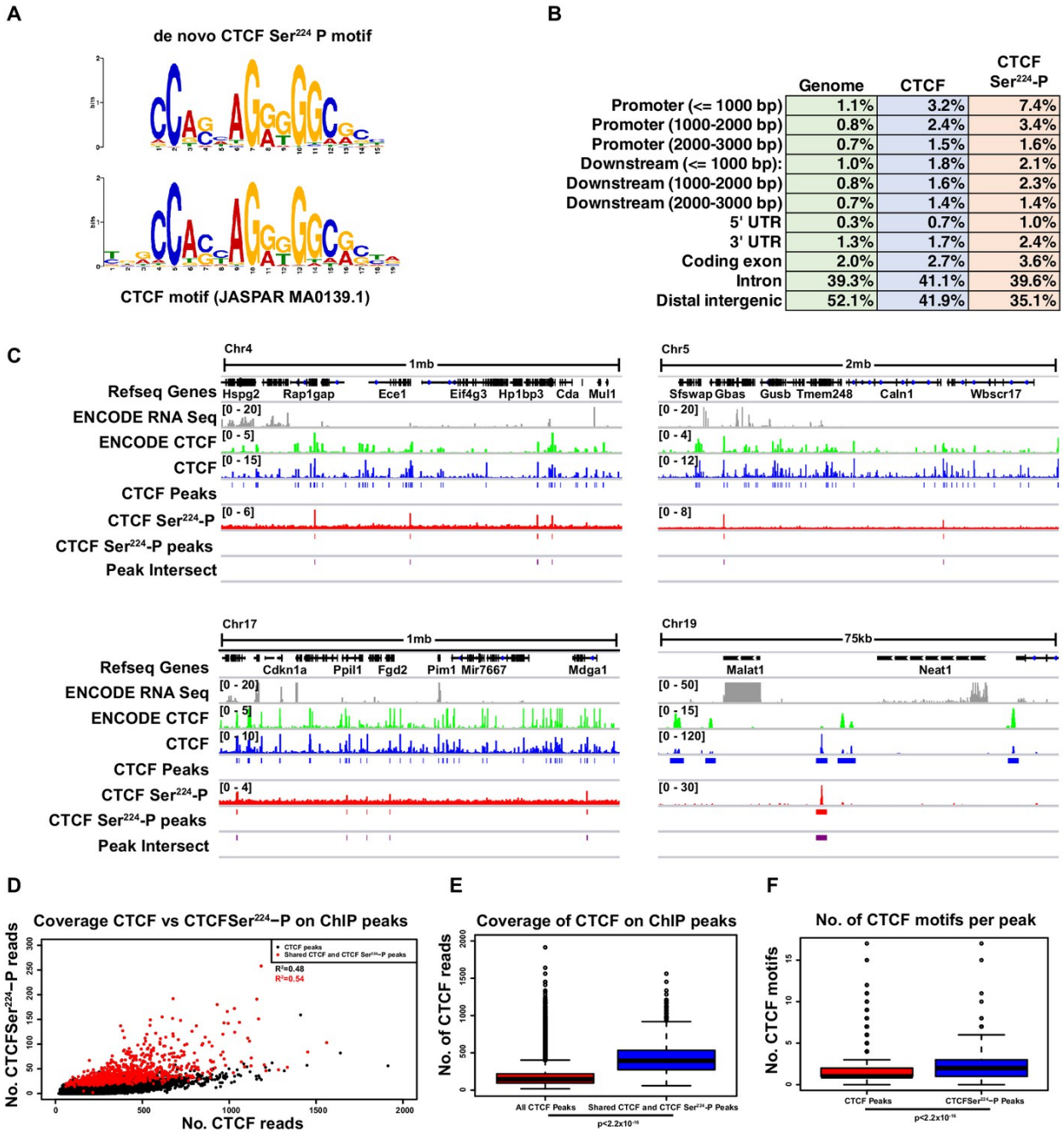
We posited that this modification could be observed on condensed chromatids during metaphase. It was previously reported that CTCF remains bound to mitotic chromosomes – both

along the dyad arms and at centromeres (Burke et al., 2005; Rubio et al., 2008). However it was also reported that CTCF is phosphorylated in its zinc finger domain during mitosis which results in its dissociation from chromatids (Sekiya et al., 2017). Notably, the latter study did not directly identify phosphorylation sites or visualize the mitotic chromatids. Accordingly, to resolve this discord and test our hypothesis, we performed immunofluorescence on mESC metaphase spreads alongside antibody to the facultative heterochromatin mark H3K27me3 that is found on chromatid arms but is excluded from the centromere (Terrenoire et al., 2010). Much like CTCF, CTCF Ser^{224}-P could be observed on chromatid arms, albeit not uniformly like CTCF (Figure 2.3G, Figure S1.3A-B). Remarkably, CTCF Ser^{224}-P appeared distinctly enriched in between the centromeres of the murine acrocentric chromosomes and the H3K27me3 positive dyad arms. We therefore concluded that the metaphase localization of CTCF Ser^{224}-P is pericentric (Figure 2.3G-H, Figure S1.3A-B). As PLK1 is associated with kinetochores, it is possible that PLK1 phosphorylates CTCF in the vicinity of the centromere for a mitosis function (Barr et al., 2004).

CTCF Ser224-P binds to a subset of CTCF binding sites during interphase

As the phospho-specific antibody labeled metaphase chromatin and thus demonstrated in vivo DNA association of CTCF Ser^{224}-P, we next determined its precise genome-wide distribution. To this end, we performed both CTCF Ser^{224}-P and CTCF chromatin immunoprecipitation and sequencing (ChIP-seq) on asynchronous TST-1 mESCs. From our CTCF ChIP-seq, we detected ~50,000 CTCF peaks genome-wide (z=6). MEME-ChIP analysis tellingly revealed significant enrichment of CTCF motifs (JASPAR MA0139.1) centered in these peaks (54% of peaks, p=3.0e-10948) (Figure S1.4A) (Bailey et al., 2009). In contrast, only ~900 CTCF Ser^{224}-P peaks were detected in our CTCF Ser^{224}-P ChIP-seq (z=6). Using MEME-ChIP, we de novo identified a significantly enriched motif in the CTCF Ser^{224}-P peaks (48% of peaks, p=2.7e-88) that closely matched the CTCF motif (p=1.8e-86) and was likewise centered in the peaks (Figure 2.4A, Figure S1.4B).

Further CEAS analysis of the CTCF Ser^{224}-P peaks revealed a genomic feature distribution that was also broadly similar to that of CTCF peaks (Figure 2.4B) (Shin et al., 2009). The distributions of CTCF and CTCF Ser^{224}-P peaks were also spread across all chromosomes (Figure S1.4C,D). Notably, the vast majority (95.9%) of the CTCF Ser^{224}-P peaks intersected with our CTCF peaks (Figure 2.4C). And comparison to mESC RNA-seq data (GSM723776) revealed that CTCF Ser^{224}-P peaks were also proximal to both transcribed and non-transcribed regions (Figure 2.4C) (Shen et al., 2012). Binding to pericentric sequences was difficult to detect in deep-sequencing based assays because of their highly repetitive nature. Knowing that our CTCF Ser^{224}-P antibody showed a low level of cross-reactivity to unmodified CTCF (Figure 2.2A,B), we next wondered whether our CTCF Ser^{224}-P binding could be explained by this cross-reactivity. Accordingly, we examined the correlation between CTCF Ser^{224}-P and CTCF binding as detected by our ChIP-seq (Figure 2.4D). While there was a detectable linear relationship between CTCF Ser^{224}-P and CTCF binding signal, this was only sufficient to explain about half of all CTCF Ser^{224}-P binding (R^2, fraction of variation in CTCF Ser^{224}-P binding explained by CTCF binding, for all peaks = 0.48, R^2 for shared peaks = 0.54). Therefore, at least a subset of CTCF Ser^{224}-P ChIP-seq peaks likely represented real binding sites outside of pericentric regions. In addition, as our ChIP-seq was performed in unsynchronized ES cells, which are mostly in interphase (S phase), this suggests that of CTCF Ser^{224}-P is bound at these regions outside of G2/M as well.



	Genome	CTCF	CTCF Ser224-P
Promoter (<= 1000 bp)	1.1%	3.2%	7.4%
Promoter (1000-2000 bp)	0.8%	2.4%	3.4%
Promoter (2000-3000 bp)	0.7%	1.5%	1.6%
Downstream (<= 1000 bp):	1.0%	1.8%	2.1%
Downstream (1000-2000 bp)	0.8%	1.6%	2.3%
Downstream (2000-3000 bp)	0.7%	1.4%	1.4%
5' UTR	0.3%	0.7%	1.0%
3' UTR	1.3%	1.7%	2.4%
Coding exon	2.0%	2.7%	3.6%
Intron	39.3%	41.1%	39.6%
Distal intergenic	52.1%	41.9%	35.1%

Figure 2.4. CTCF Ser224-P occupies a fraction of CTCF sites outside of pericentric chromatin in interphase.

(A) De novo CTCF Ser224-P motif logo determined with MEME-ChIP (top). JASPAR indexed CTCF motif logo (bottom).

(B) % distribution of CTCF (blue) and CTCF Ser224-P (orange) ChIP-seq peaks across genomic features. % distribution of these features in the genome is shown for comparison (green).

(C) Four example screenshots showing CTCF (blue) and CTCF Ser224-P (red) ChIP-seq coverage tracks and called peaks. Intersected CTCF and CTCF Ser224-P peaks (purple) are also shown. ENCODE CTCF ChIP-seq coverage (green), Refseq Genes (black) and ENCODE RNA-seq (gray) are shown for reference. Chromosome number and window scale are indicated.

(D) CTCF versus CTCF Ser224-P ChIP-seq coverage on CTCF (black) and shared CTCF and CTCF Ser224-P ChIP-seq peaks. R-squared values for both sets of peaks are shown.

(E) CTCF ChIP-seq coverage on CTCF versus shared CTCF and CTCF Ser224-P ChIP-seq peaks ($p < 2.2x10^{-16}$, Wilcoxon rank sum test).

(F) Number of CTCF motifs found in CTCF versus shared CTCF and CTCF Ser224-P ChIP-seq peaks ($p < 2.2x10^{-16}$, Wilcoxon rank sum test).

We next sought to examine features which differentiated regions bound by CTCF Ser^{224}-P from regions bound by unmodified CTCF. Shared CTCF and CTCF Ser^{224}-P ChIP peaks (representing 95.9% of all CTCF Ser^{224}-P peaks) had significantly more CTCF binding signal than CTCF peaks in general (Figure 4E, p < 2.2e-16, Wilcoxon rank sum test). While CTCF motifs found in CTCF Ser^{224}-P peaks did not tend to be more conserved than motifs found in CTCF ChIP peaks (Figure S1.4E, p = 0.1623, Wilcoxon rank sum test), CTCFSer^{224}-P ChIP peaks did tend to contain more motifs than CTCF ChIP peaks (Figure 4F, p < 2.2e-16, Wilcoxon rank sum test). In other words, CTCFSer^{224}-P ChIP peaks tend to be found at higher affinity sites with greater numbers of CTCF binding sites than CTCF ChIP peaks in general. We also found that shared CTCF and CTCF Ser^{224}-P ChIP peaks tended to be larger than CTCF peaks in general (Figure S1.4F, p < 2.2e-16, Wilcoxon rank sum test), which may in part explain these trends.

Finally, we compared our CTCF and CTCFSer^{224}-P ChIP-seq with a previously published cohesin ChIP-seq done in mESCs (Kagey et al., 2010). While 22.8% and 33.3% of our CTCF ChIP-seq peaks overlapped with SMC1 and SMC3 peaks, respectively, 51.6% and 85.1% of our CTCFSer^{224}-P ChIP peaks overlapped with SMC1 and SMC3 peaks, respectively. This suggests that CTCFSer^{224}-P ChIP tends to co-bind more frequently with cohesin as well.

Mutational analysis of CTCF Ser^{224}-P reveals an effect on cell growth

We next decided to examine the effect of expressing either Ser^{224} mutants, S224A or S224E, on cells. The former mutation obviates phosphorylation and the latter glutamic acid substitution is a so-called phosphomimetic that mimics negatively charged phosphorylation by presenting an acidic side chain at that position. We generated doxycycline-inducible S224A- and S224E-3xFLAG mESCs and overexpressed wild type CTCF, S224A or S224E over six days of growth. Ectopic expression of S224E but not wild type or S224A affected growth of mESCs (Figure 2.5A). Namely, mESC colonies overexpressing S224E for 6 days had significantly smaller diameters than uninduced cells, while overexpressing wild type or S224A did not result in significantly smaller colonies (Figure 2.5B). Expression levels of the induced FLAG-tagged proteins were comparable in all clones by Western blot (Figure 2.5C). Taken together, our data so far suggests that the phosphorylated form of CTCF may have a specific function during the cell cycle, as forced constitutive expression of the phosphomimetic form results in a cell growth defect. However, we also note that the continued presence of endogenous wild-type CTCF in our overexpression system may be obscuring detection of further phenotypes of S224A and S224E CTCF.

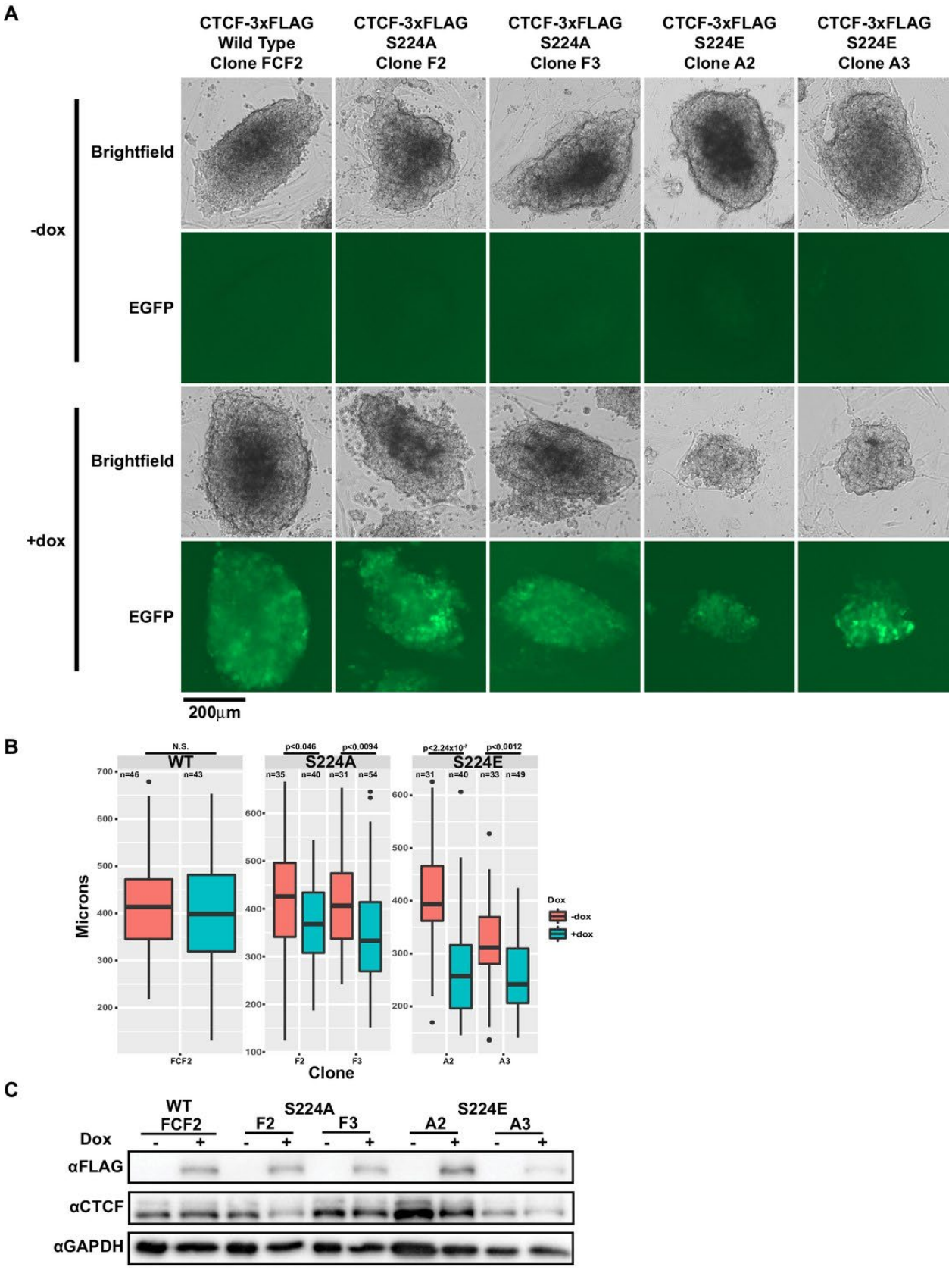

A
CTCF-3xFLAG Wild Type Clone FCF2
CTCF-3xFLAG S224A Clone F2
CTCF-3xFLAG S224A Clone F3
CTCF-3xFLAG S224E Clone A2
CTCF-3xFLAG S224E Clone A3
-dox
Brightfield
EGFP
+dox
Brightfield
EGFP
200μm
B
WT
S224A
S224E
N.S.
p<0.046
p<0.0094
p<2.24x10⁻⁷
p<0.0012
n=46
n=43
n=35
n=40
n=31
n=54
n=31
n=40
n=33
n=49
Microns
Clone
FCF2
F2
F3
A2
A3
Dox
-dox
+dox
C
WT
S224A
S224E
FCF2
F2
F3
A2
A3
Dox
αFLAG
αCTCF
αGAPDH

Figure 2.5. CTCF S224E phosphomimic mutation is poorly tolerated by dividing cells.

(A) Representative brightfield and EGFP images of F1-2.1 mESCs carrying a dox-inducible wild type CTCF, S224A or S224E transgene grown for six days with (bottom) or without (top) doxycycline. Two independent S224A and S224E clones are shown.

(B) Quantification of colony diameters in microns of cell lines shown in (A), grown for six days with (blue) or without (red) doxycycline. Student's t-test was used to calculate p-values between indicated samples, with not significant (N.S.) p-values being > 0.05.

(C) Western blot measuring FLAG and CTCF protein levels of cell lines shown in (A). GAPDH is shown as a loading control.

CTCF Ser^{224} mutations have no effect on nuclear import, DNA binding, cell cycle, or ploidy

Given the phenotype, we further explored S224 mutants to understand the normal function of CTCF Ser^{224}-P. We initially posited that CTCF Ser^{224}-P may be critical to regulating some chromatin function of CTCF. As CTCF Ser^{224}-P was observable in the nucleus in only a fraction of asynchronous cells, it was a remote possibility that Ser^{224} is also critical for CTCF nuclear import. Thus, we examined localization of inducible wild type CTCF, S224A, and S224E in MEFs by immunofluorescence. Regardless of either amino acid substitution at Ser^{224}, nuclear localization of CTCF was not affected. (Figure 2.6A). As we observed CTCF Ser^{224}-P bound to only a fraction of CTCF sites genome-wide, it was also a possibility that this amino acid position is critical for DNA binding. However, since Ser^{224} is not located in the zinc finger domain, we predicted that Ser^{224} would not directly influence sequence-specific DNA binding. To test this, we performed EMSA using a dsDNA probe with a known CTCF motif as well as a probe with mutations in the motif (Spencer et al., 2011). We made recombinant FLAG-tagged S224A and S224E as well as wild type CTCF and GFP as positive and negative controls respectively (Figure 2.6B). As expected, regardless of amino acid polarity or charge, both mutants bound the known

CTCF motif and not the mutated motif (Figure 2.6C). However, it is still a distinct possibility that, in vivo, CTCF Ser^{224}-P may signify a regulatory event that determines which of the thousands of genomic CTCF sites are bound.

To test whether CTCF Ser224-P may play a role on mitotic chromosomes, we asked if could overexpression of S224A or S224E leads to defects in either cell cycle progression or segregation of chromatids. We generated DNA content profiles of mESCs with inducible wild-type CTCF, S224A, or S224E transgenes after six days of doxycycline induction and did not observe any blockages in the cell cycle (Figure 2.7A). We also investigated changes in ploidy by kayotyping these cells, specifically examining metaphase spreads for polyploidy and translocations. However, we did not find significantly higher numbers of polyploid spreads in mESCs overexpressing wild-type or mutant CTCF versus uninduced mESCs (Figure 2.7B, C).

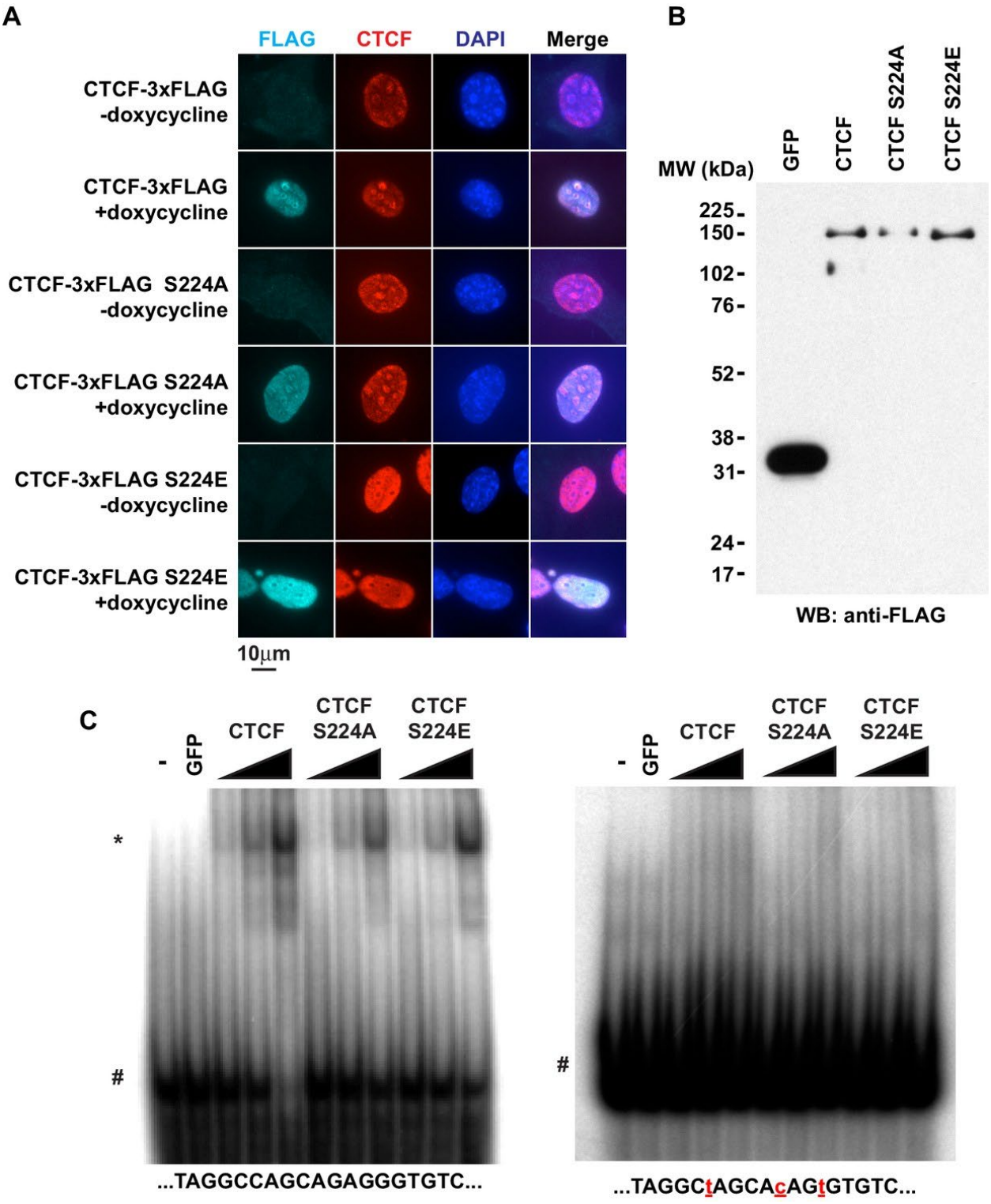


Figure 2.6. CTCF Ser^{224} is nonessential to nuclear import and DNA binding.

(A) FLAG and CTCF immunofluorescence performed on rtTA MEFs with inducible CTCF-3xFLAG transgenes (wild type, S224A, or S224E). Nuclei counterstained with DAPI. Bar, 10μm.

(B) FLAG Western blot of recombinant FLAG-tagged GFP and CTCF (wild type, S224A, or S224E).

(C) DNA EMSA using RS14C (left) and RS14C mutant (right) probes. Red lowercase letters, mutated positions. 2 pmole GFP and 0.5, 1, or 2 pmole CTCF (wild type, S224A, or S224E) were used. *, CTCF shifted probe. #, free probe.

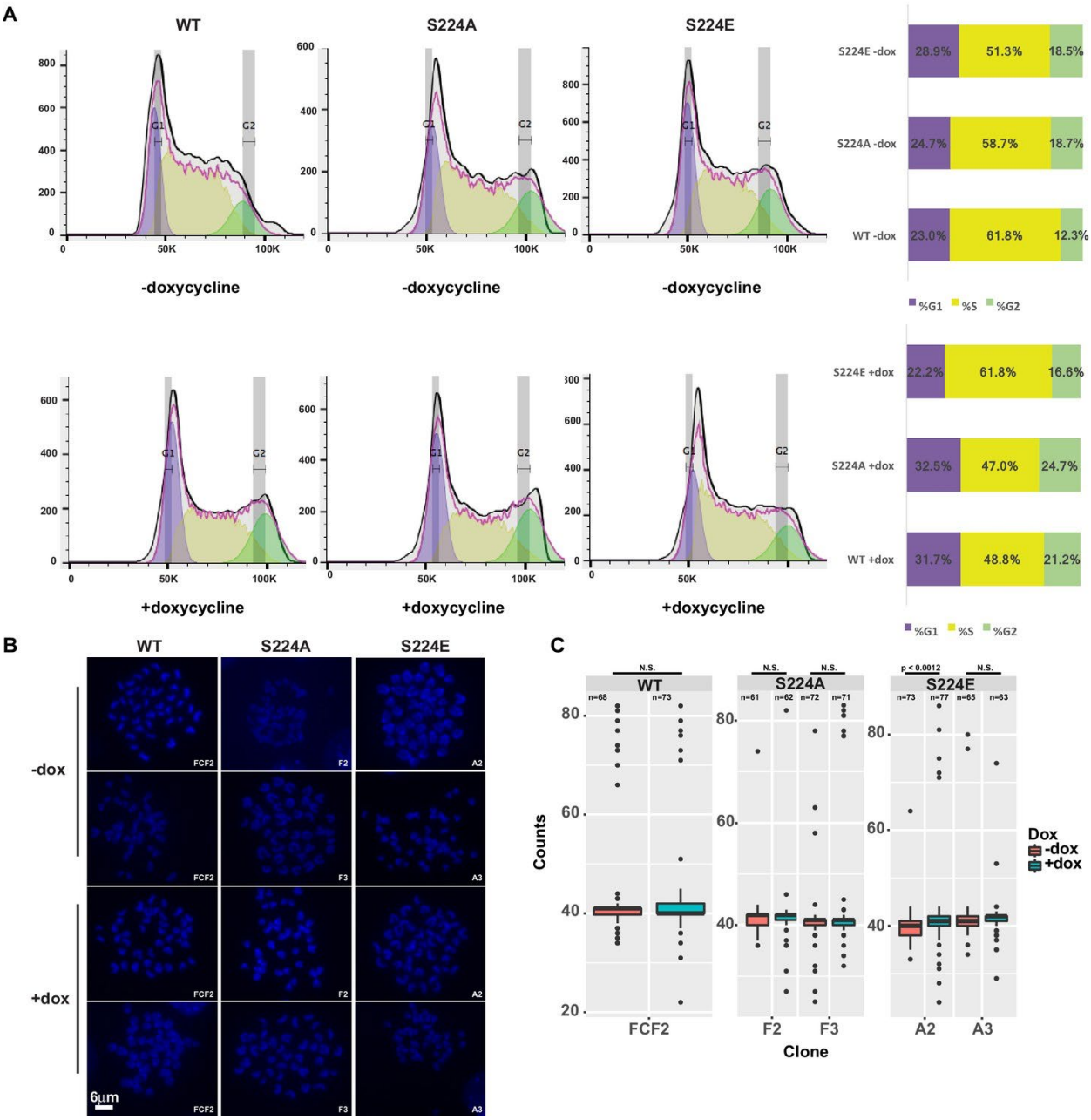


Figure 2.7. Overexpression of CTCF, S224A or S224E does not impact cell cycle progression or ploidy.

(A) Cell cycle profiles of F1-2.1 mESCs carrying a dox inducible CTCF, S224A or S224E transgene grown for six days with (bottom) or without (top) dox induction. Quantification of percent of cells in each stage of the cell cycle indicated in bar graph to the right.

(B) Representative metaphase spreads of cells profiled in *A* grown for six days with (bottom) or without (top) dox induction.

(C) Quantification of chromosome counts for cells profiled in *A* grown for six days with (bottom) or without (top) dox induction. Wilcoxon rank sum test was used to calculate p-values between indicated samples, with not significant (N.S.) p-values being > 0.05.

CTCF Ser^{224}-P and the borders of topologically associating domains (TADs)

Finding that overexpression of wild type, S224A or S224E had little obvious impact on mitotic chromosomes, we investigated whether it could interfere with CTCF function in interphase. We first noted that many of our CTCF Ser^{224}-P ChIP-seq peaks detected in interphase overlapped CTCF peaks at the borders of Topologically Associating Domains (TADs) (Figure 2.8A), megabase-scale organizational structures on chromosomes within which genetic elements show high frequency of interaction (Dixon et al., 2012; Nora et al., 2012). Mammalian chromosomes are generally organized into hundreds of such TADs, with each TAD separated by genetically defined "borders". As previous studies had shown that CTCF binding is important for formation of TAD borders (Nora et al., 2017; Sanborn et al., 2015), we decided to examine the impact of overexpressing wild-type CTCF, S224A or S224E on nuclear architecture using HYbrid Capture Hi-C (Hi-C^2), a cost-effective alternative to genome-wide Hi-C (Sanborn et al., 2015). The TAD containing the gene *Mecp2* was chosen as the capture region as it contained a sub-TAD domain bound by CTCF at both the left and right borders and CTCF Ser^{224}-P at the left border (Figure 2.8A). To minimize secondary impacts on TAD structure, Hi-C was performed after 2 days of wild type CTCF, S224A or S224E overexpression in F1-2.1 mESCs, a time point at which no cell colony defects were observed in any of the three cell lines. By eye, interaction matrices of the *Mecp2* region appeared similar with or without overexpression of wild type CTCF, S224A and S224E (Figure 2.8A). To analyze impact on the interactions within the *Mecp2* TAD quantitatively, we additionally calculated insulation scores across the Hi-C^2 region and used this to calculate a

TAD score for the *Mecp2* TAD in each condition (Crane et al., 2015). The *Mecp2* TAD score was similar with or without overexpression of wild type CTCF, S224A and S224E. Thus, overexpression of CTCF, including CTCF S224E, does not detectably impact three-dimensional chromatin structure (Figure 2.8B).

However, as RNA-seq suggested that levels of S224E CTCF may be modest as compared to wild-type CTCF (see below), we assayed to what extent the mutant S224E CTCF was bound to the Mecp2 TAD. We did CTCF and FLAG ChIP-qPCR in doxycycline inducible S224E-3xFLAG mESCs after 48h of dox induction, the same conditions under which the Hi-C^2 experiment was performed (Figure 8C). While FLAG binding was enriched at positive control region Cirbp over IgG, the –dox sample and a negative control region Oct4 (Figure 8C), the level of binding was much lower than that of CTCF. Similarly, FLAG binding was modestly enriched over IgG, the –dox sample, and Oct4 at at least the two CTCF sites bordering the Mecp2 TAD (Figure 8C, Irak1, Ikbkg), as well as at the site bound by phospho-CTCF in the Mecp2 TAD in one replicate (Figure 2.8C, Flna), although to an extent much less than that of CTCF. This modest binding of S224E CTCF to the Mecp2 TAD, likely due to endogenous wild-type CTCF remaining in the system, may possibly explain why we were unable to detect changes in chromatin architecture in this region.

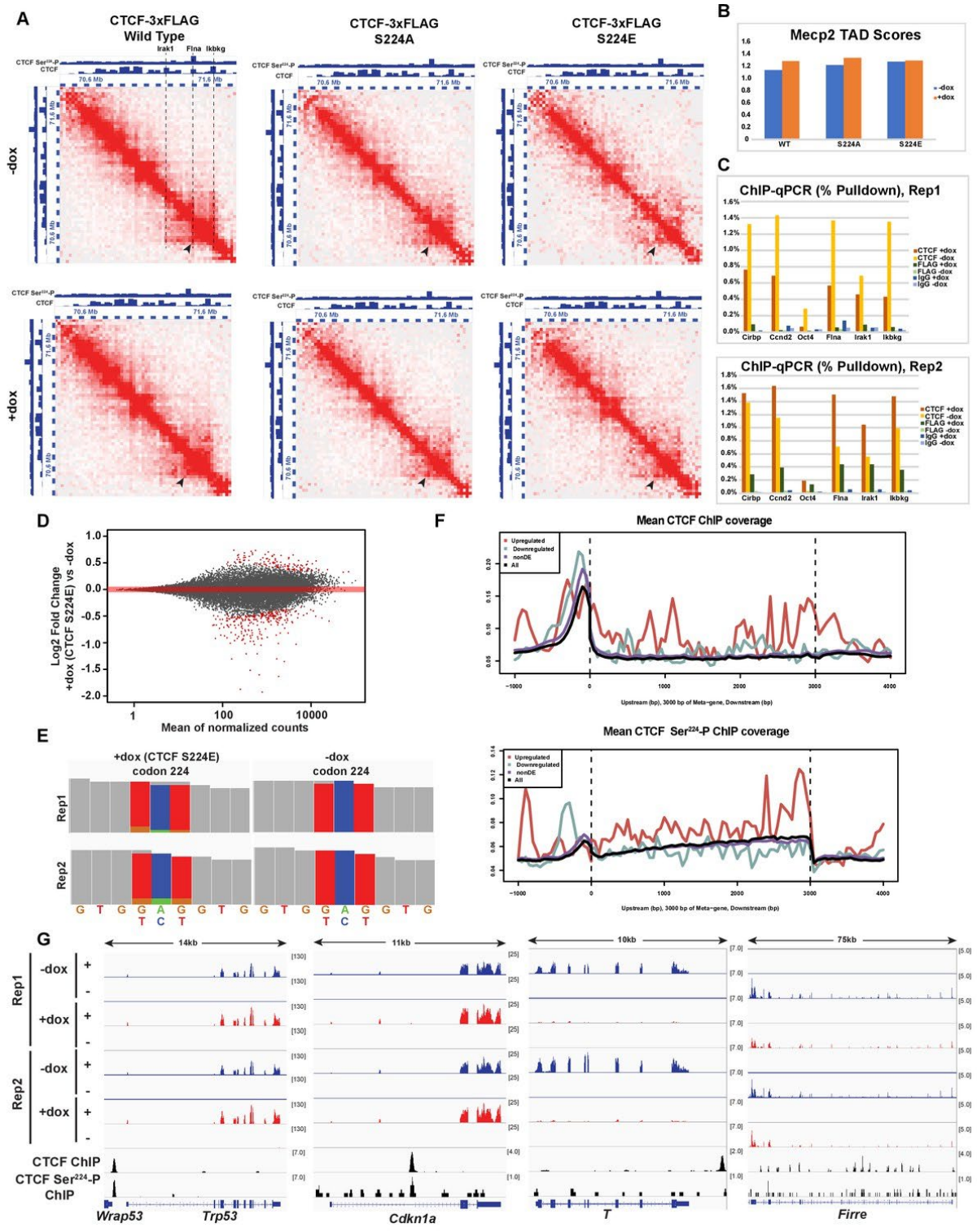


Figure 2.8. Impact of overexpression of CTCF, S224A and S224E on three-dimensional chromatin structure and gene expression.

(A) Hi-C^2 interaction maps at 25kb resolution of the Mecp2 TAD in F1-2.1 mESCs carrying dox-inducible wild type, S224A or S224E CTCF-3xFLAG transgenes grown for 2 days with (bottom) and without (top) doxycycline. CTCF and CTCF Ser224-P ChIP-seq tracks are shown for comparison. Black arrows indicate the left border of a sub-TAD domain bound at both borders by CTCF and at one border by CTCF Ser224-P. In addition, dotted lines and text in the WT -dox Hi-C^2 interaction map indicate locations of ChIP-qPCR primers used in (C), with the Irak1 and Ikbkg lines also indicating the borders of the Mecp2 TAD scored in (B).

(B) TAD scores for the Mecp2 TAD for the Hi-C^2 interaction maps in *A*, with higher TAD score indicating a stronger TAD.

(C) CTCF and FLAG ChIP-qPCR (% Pulldown) in F1-2.1 mESCs carrying dox-inducible S224E CTCF-3xFLAG grown for 2 days with or without doxycycline. Cirbp and Ccnd indicate positive control regions for CTCF binding, while Oct4 indicates a negative control region. Flna, Irak1, and Ikbkg regions are as indicated in (A).

(D) MA plot of RNA-seq expression changes in F1-2.1 mESCs carrying dox-inducible CTCF S224E transgene after 6 days of dox induction. Points in red are DE genes (adjusted p-value < 0.01).

(E) Coverage of RNA-seq reads over codon 224 of CTCF with (left) and without (right) 6 days of dox induction. The number of reads with A (green), C (blue), G (gold), T (red) or N (grey) at positions 1, 2, and 3 of codon 224 are shown.

(F) Metagene coverage of CTCF ChIP-seq reads (top) and CTCF Ser224-P ChIP-seq reads (bottom) over upregulated (red), downregulated (green), non differentially expressed (purple) and all (black) genes.

(G) RNA-seq, CTCF and CTCF Ser224-P ChIP-seq coverage over two representative upregulated (left) and downregulated (right) genes.

Overexpression of CTCF S224E upregulates the p53 signaling pathway and globally plasma membrane proteins at the RNA level

As we could not attribute the mESC growth phenotype to changes in cell cycle progression, CTCF binding or TAD structure, we performed RNA sequencing (RNA-seq) to identify genes that were differentially expressed (DE) when CTCF S224E was induced in mESC's. We found 375 genes that were DE between two replicates (adjusted p-value < 0.01), with 118 (31%) being upregulated and 257 (69%) being downregulated upon overexpression of CTCF S224E (Figure 2.8C). Surprisingly, CTCF was not a DE gene. To confirm that CTCF S224E was expressed, we examined RNA-seq reads overlapping CTCF codon 224 and found that 12-16% of reads had the glutamate codon at position 224 (Figure 2.8D). This indicates that while CTCF S224E was present, it was not grossly overexpressed at the RNA level; therefore, our observations from modest exogenous expression may reflect the importance of this site. We next looked at coverage of CTCF and CTCF Ser^{224}-P ChIP-seq over DE genes in an attempt to determine whether DE genes were directly regulated by CTCF. On average, CTCF binding was enrich=ed along both downregulated and upregulated DE genes, albeit at different locations. In particular, downregulated genes were more enriched in CTCF binding upstream of the transcription start site, while upregulated genes were more enriched in CTCF binding along the gene body. CTCF Ser^{224}-P enrichment followed a similar pattern (Figure 2.8E). Next, in order to determine if specific pathways were enriched among our DE genes, we performed functional annotation analysis of upregulated and downregulated DE genes using the PANTHER Overrepresentation Test (Mi et al., 2017; Nikolsky and Bryant, 2009), with the list of all expressed genes as the background. Among upregulated genes, the only significantly overrepresented PANTHER pathway was the p53 pathway (FDR = 0.0495), most notably including *Trp53* (upregulated 1.3-fold, adjusted p-value < 0.00075) and *Cdkn1a* (upregulated 1.3-fold, adjusted p-value < 0.00043), genes which were also bound by CTCF and CTCF Ser^{224}-P (Figure 2.8F). Meanwhile, among downregulated genes, the most significant overrepresented cellular

component categories were ‘cell periphery’ (FDR = $2.28x10^{-6}$) and ‘plasma membrane’ (FDR = $3.45x10^{-6}$), although, many downregulated genes did not fall into these categories (Figure 2.8F). These results suggest that modest CTCF S224E overexpression could lead to the slight activation of p53 signaling, and also global downregulation of proteins on the plasma membrane. The effect may be direct, as CTCF binding was found to be enriched at DE genes or indirect, as RNA-seq was assayed on cells after six days of S224E overexpression in mESCs, when a colony grown phenotype was already apparent. We conclude that the serine 224 of CTCF plays a critical role in regulating gene expression and suggest that the effect on the p53 signaling pathway may partly explain the growth phenotype in mESCs.

DISCUSSION

In the quarter century since its discovery, CTCF has been ascribed a multitude of chromatin functions. While DNA sequence composition and CpG methylation are determinants of CTCF binding, a regulatory system independent of DNA binding has remained elusive. Here we discovered and confirmed a CTCF PTM conserved among vertebrates. We observe for the first time that CTCF is differentially regulated during the cell cycle, with CTCF Ser^{224}-P being enriched at G2/M and specifically on pericentric regions during metaphase. Notably, CTCF Ser^{224} phosphorylation is regulated by the kinase PLK1 whose expression profile also peaks in G2/M. As exogenously expressing S224E leads to colony formation defects and gene expression changes, our data hint at the importance of CTCF Ser^{224}-P for proper cell function. We were further able exclude several potential functions of CTCF Ser^{224}-P. First, Ser^{224} does not directly affect CTCF binding to DNA. However, while CTCF Ser^{224}-P occupies a subset of all CTCF sites genome wide, how this PTM is limited to a fraction of sites on the chromatids and pericentric satellites is still unclear. Second, while overexpression of a S224E leads to colony formation defects, we did not observe any changes in a TAD’s structure, suggesting that CTCF Ser^{224}-P may not impact the architectural role of CTCF, at least around the gene-rich *Mecp2* TAD. Third,

cell cycle progression and ploidy of cells overexpressing the phospho-mimic were not obviously affected. We conclude that regulation of CTCF Ser^{224}-P is needed to prevent dysregulation of hundreds of genes, directly or indirectly through mechanisms that likely do not involve CTCF interphase binding, architectural function, cell cycle progression or ploidy. However, an intriguing possibility which we did not investigate is that the negative charge introduced by phosphorylation at Ser224 could alter the conformation of the CTCF protein, altering the manner in which CTCF regulates genes or interacts with other proteins. We hope these observations will provide starting points for further elucidating the natural functions of this CTCF PTM and valuable resources for future study.

With our observation of CTCF Ser^{224}-P enrichment proximal to the centromere, several avenues of investigation become conspicuous. First, what could be the pericentric function of CTCF Ser^{224}-P? For example, CTCF interacts with cohesins and the latter remain at the centromere to facilitate cohesion during metaphase (Morales and Losada, 2018; Rubio et al., 2008; Stedman et al., 2008; Xiao et al., 2011). CTCF has also been found to impede 'loop extrusion' activity of the cohesin complex during interphase (Vian et al., 2018a). Could a similar capacity be expected for pericentric CTCF Ser^{224}-P and the cohesin complex? Likewise, the presence of pericentric CTCF Ser^{224}-P may demarcate a topological boundary between the chromatid arms and satellite DNA. Notably, PLK1 also phosphorylates cohesins, promoting their prophase dissociation from the chromatid arms (Gimenez-Abian et al., 2004; Hauf et al., 2005; Losada et al., 2002; Sumara et al., 2002). PLK1 phosphorylation of CTCF may also be linked to cohesin dissociation from the chromatids. Another prospective undertaking would be to determine if CTCF Ser^{224}-P is present at all pericentromeres. Are all pericentromeres identical? Thus, the identification of this PTM and development of a specific antibody may prove to be an invaluable resource to the cell cycle and mitosis communities.

PLK1 mediated phosphorylation of CTCF Ser^{224} may also not be a finite event. Coordinated signaling could occur. The Polo Box Domain (PBD) of PLK1 targets the kinase

activity by tethering PLK1 to a phosphopeptide (Lowery et al., 2004). While CTCF appears to lack an ideal PBD binding site, perhaps PLK1 is recruited to CTCF Ser^{224} by a proximal phosphoprotein (Elia et al., 2003). Conversely, what triggers dephosphorylation of CTCF Ser^{224}-P? Thus exploring the vicinal events surrounding Ser^{224} phosphorylation could elucidate the G2/M function of this PTM. Furthermore, as CTCF contains 100 S/T/Y amino acids, there exists the possibility that other bona fide phosphorylation events exist. Similarly, with 65 lysines, CTCF acetylation and/or methylation may also occur and orchestrate a much more complex symphony of insulator regulation. Hence by confirming Ser^{224} phosphorylation, we are likely only previewing what could be an expansive CTCF signaling network. While our assays did not elucidate a conspicuous mechanism or functional consequence of CTCF Ser^{224}-P, nonetheless its identification and generation of a specific antibody to it can be a useful resource. From a disease perspective, this PTM may also serve as a useful biomarker of dividing cells or prove to be a therapeutic target for competitive or allosteric inhibition, if targeting PLK1 directly proves adverse. Lastly, at least 5000 CTCF binding locations are conserved amongst mammalian genomes and these are often associated with syntenic chromatin domain boundaries (Dixon et al., 2012; Schmidt et al., 2012; Vietri Rudan et al., 2015). As CTCF Ser^{224}-P also appears conserved amongst vertebrates, it is an attractive possibility that Ser^{224} phosphorylation along with yet confirmed CTCF PTMs may constitute a conserved code that underlies chromatin organization and gene regulation akin to the histone code (Jenuwein and Allis, 2001). Elucidation of such a CTCF code may also reveal insights into how interphase contacts are remembered and restored before and after mitosis. This prospective code may also be bifurcated species-specifically by the degree of conservation of particular CTCF PTMs, potentially exposing mechanisms of speciation. Thus, further unmasking CTCF Ser^{224}-P and thoroughly validating the entirety of CTCF PTMs would be invaluable to understanding how the ubiquitous insulator's functions are specified.

METHODS

Plasmids and Cells

The *Ctcf*^{S224A} and *Ctcf*^{S224E} mutations were generated by site-directed mutagenesis with Phusion polymerase (NEB) on pcDNA3-CTCF (Chao et al., 2002). Mutations were confirmed by DNA sequencing. Recombinant FLAG-CTCF-6xHis and FLAG-GFP-6xHis were made using plasmids pBD39 and pBD40 (Sun et al., 2013). The *Ctcf*^{S224A} and *Ctcf*^{S224E} mutations were cloned into pFLAG2 with 6xHis C-terminal tags (pBD116 and pBD117) to produce recombinant FLAG-CTCF(S224A)-6xHis and FLAG-CTCF(S224E)-6xHis in *E. coli*. Tet-inducible co-expression of CTCF-3xFLAG and EGFP in mammalian cells was achieved with pBD102 (Sun et al., 2013). *Ctcf*^{S224A} and *Ctcf*^{S224E} mutations were incorporated into this vector to generate pBD118 and pBD119. To generate Tet-on CTCF, CTCF(S224A) and CTCF(S224E) cells; AseI linearized pBD102, pBD118 and pBD119 were electroporated into SV40T immortalized tetraploid female Rosa26-M2rtTA MEFs or F1-2.1 female mouse ESCs (Jeon and Lee, 2011; Luikenhuis et al., 2001). Stable transformants were selected with hygromycin. Female mouse ESCs with a monoallelic mutation in *Tsix* (TST-1) were previously described (Ogawa et al., 2008). HEK 293 and the rtTA MEFs were passaged in DMEM GlutaMAX 4.5g/L D-glucose, 110mg/L sodium pyruvate, 15% FBS, 25mM HEPES pH 7.5, 0.1mM MEM non-essential amino acids, 0.1mM 2-mercaptoethanol, 100U/mL penicillin, and 100ug/mL streptomycin at 37C 5% CO_2. TST-1 and F1-2.1 ES cells were propagated in the aforementioned media supplemented with LIF to 500U/mL. Chemical reagents utilized in cell culture conditions are listed in Supplementary File 1. All mESCs and MEFs originated from our lab. HEK293 cells were purchased from ATCC. Identity of all lines was confirmed via examination of deep-sequencing datasets of genomic DNA generated in our lab for relevant features (e.g., species, SNPs, presence of transgenes). All cells tested negative for mycoplasma contamination. Quantification of colony size shown in Figure 6B was done blinded to cell line identity and doxycycline induction status.

CTCF Immunoprecipitation

Female immortalized rtTA MEFs stably maintaining pBD102 were grown for 72hrs in MEF media supplemented with 100μg/mL hygromycin with or without 1μg/mL doxycycline. ~5x10^7 cells from both treatments were manually collected and washed in cold 25mM HEPES pH 7.5, 150mM NaCl, 1mM DTT, 1mM EDTA, phosphatase inhibitors (10mM sodium pyrophosphate, 10mM B-glycerophosphate, 5mM NaF, 1mM Na_3VO_4) and protease inhibitors (1mM PMSF, 1μg/mL pepstatin, and 1μg/mL leupeptin). Cell pellets were incubated on ice in 5 cell pellet volumes of swelling buffer (10mM Tris-HCl pH 7.5, 10mM KCl, 1.5mM $MgCl_2$, 1mM DTT, phosphatase and protease inhibitors) and incubated for 10minutes. Digitonin was added to the suspension to 20μg/mL. Nuclei were pelleted at 2000g 4C. Nuclei were then washed twice in swelling buffer 0.1% Triton X-100 and pelleted. After aspirating the supernatant, the nuclei were frozen at -80C. Frozen nuclei were resuspended in lysis buffer (25mM HEPES pH 7.5, 2M NaCl, 10mM EDTA, 0.5% Sarkosyl, 5mM DTT, phosphatase and protease inhibitors) and kept on ice for one hour. 150-212μm acid-washed glass beads (Sigma) were washed in 95% EtOH and then lysis buffer. The washed glass beads were then mixed with the cold nuclear lysate by gentle inversion. The glass beads and the viscous phase of the nuclear lysate were pelleted by centrifugation at 2000g 5 minutes 4C. The supernatant was passed through a 25G syringe to further reduce the lysate's viscosity and insoluble material was removed by centrifugation at 16000g 4C. It was then dialyzed 1:250 overnight (3000 mwco) at 4C in dialysis buffer (25mM HEPES pH 7.5, 300mM NaCl, 0.1mM $ZnSO_4$, 1mM sodium pyrophosphate, 1mM β-glycerophosphate, 0.1mM Na_3VO_4, 10% glycerol). It was further dialyzed 1:250 another 5 hours at 4C in IP buffer (dialysis buffer with 0.1% Triton X-100). Sepharose 4B (Sigma) was washed with IP buffer with protease inhibitors and used to preclear the dialyzed nuclear extract. Anti-FLAG M2 agarose (Sigma) was washed and blocked with 25mM HEPES pH 7.5, 150mM NaCl, 10% Gly-Gly, and protease inhibitors at 4C. The

blocked anti-FLAG agarose was washed with IP buffer with protease inhibitors. Nuclear lysate protein concentrations from uninduced and induced cells were equilibrated by Bradford assay. Equal concentrations of lysates were combined with the blocked anti-FLAG agarose and incubated for 5 hours at 4C with rotation. Immunoprecipitations were copiously washed with IP buffer with protease inhibitors. Bound protein was eluted with 25mM HEPES pH 7.5, 1% SDS and then TCA precipitated. Precipitates were resuspended and heated in SDS-PAGE sample buffer. Samples were resolved by 7.5% SDS-PAGE followed by Coomassie Blue R-250 staining.

Mass Spectrometry

Immunoprecipitated CTCF was extracted from a Coomassie stained gel. After destaining with 40% ethanol/10% acetic acid, CTCF were reduced with 20mM DTT (Sigma) for 1h at 56°C and then alkylated with 60mM iodoacetamide (Sigma) for 1h at 25°C in the dark. Proteins were then digested with 12.5ng/μL modified trypsin (Promega) in 50μL 100mM ammonium bicarbonate, pH 8.9 at 25°C overnight. Peptides were extracted by incubating the gel pieces with 50% acetonitrile/5%formic acid then 100mM ammonium bicarbonate, repeated twice followed by incubating the gel pieces with 100% acetonitrile then 100mM ammonium bicarbonate, repeated twice. Each fraction was collected, combined, and reduced to near dryness in a vacuum centrifuge. For the CTCF-CK2 *in vitro* kinase reaction, CTCF was extracted from a Coomassie stained gel and peptides were extracted in the same manner as described above except 125 ng/μL chymotrypsin (Roche) in 100mM Tris, 10mM calcium chloride, pH7.8 was used. Phosphorylated peptides were enriched using NTA-Fe resin as previously described (Ficarro et al., 2009). Peptides were eluted with 250mM sodium phosphate pH 8.9 and then loaded on a precolumn and separated by reverse phase HPLC (Agilent) over a 120 minute gradient before nanoelectrospray using an Orbitrap XL mass spectrometer (Thermo). The mass spectrometer was operated in a data-dependent mode. A full scan MS spectrum (injection time of 1000ms,

resolution of 60,000 across 400-2000 *m/z*) was followed by MS/MS for the top 10 precursor ions in each cycle. For the MS/MS on the LTQ (collision-induced dissociation (CID)) was set at 35% energy, injection time of 100ms, and isolation width 3 *m/z*. Dynamic exclusion for the data-dependent scans was set to 30s. Raw mass spectral data files (.raw files) were converted into .mgf file format using Proteome Discoverer, and then searched against a SwissProt 2015 database containing *Mus musculus* protein sequences (16,724 entries) using Mascot version 2.4.1 (Matrix Science). Mascot search parameters were: mass tolerance for precursor ions was 10 ppm; the fragment ion mass tolerance was 0.8 Da; 2 missed cleavages of trypsin or chymotrypsin; fixed modification was carbamidomethylation of cysteine; variable modifications were methionine oxidation, tyrosine phosphorylation, and serine/threonine phosphorylation. Mascot peptide identifications and phosphorylation site assignment were verified manually with the assistance of CAMV (http://web.mit.edu/fwhitelab/software.html) (Curran et al., 2013).

Recombinant proteins & In vitro phosphorylation

FLAG-CTCF-6xHis (wild type, S224A, and S224E) and FLAG-GFP-6xHis were made as previously described in Rosetta-Gami B (DE3) *E. coli* (EMD Millipore) using pBD39, pBD40, pBD116, and pBD117 (Sun et al., 2013). 2.5μg FLAG-CTCF-6xHis was *in vitro* phosphorylated with 250IU Casein Kinase 2 (CK2) (NEB, P6010) and 200μM ATP (1μCi [γ^{32}P] ATP) for 30 minutes at 30C. 1μg FLAG-CTCF-6xHis and 0.5μg dephosphorylated casein (SignalChem C03-54BN) were also *in vitro* phosphorylated with 0.1μg Polo Like Kinase 1 (PLK1) (Sigma, P-0060) in 5mM MOPS pH7.2, 2.5mM β-glycerophosphate, 4mM $MgCl_2$, 1mM EGTA, 0.05mM DTT, and 200μM ATP (1μCi [γ^{32}P] ATP) for 60 minutes at 30C. CK2 and PLK1 labeling was detected by SDS-PAGE and autoradiography. For mass spectrometry, FLAG-CTCF-6xHis was phosphorylated with non-radioactive ATP.

EMSA

DNA EMSA was performed as previously described (Sun et al., 2013). dsDNA oligonucleotides to RS14C and RS14C with a mutated CTCF motif were used (Spencer et al., 2011). Sequences are listed in Supplementary File 2.

Antibody generation

The phospho-CTCF Ser224 rabbit polyclonal antibody was developed in collaboration with Cell Signaling Technology. This antibody was produced by immunizing rabbits with a synthetic phosphopeptide spanning amino acids 215 to 229 of human CTCF. The antibody was purified by protein A and peptide affinity chromatography.

Western blot

SDS-PAGE and Western blot transfer to PVDF membranes were carried out under standard conditions. Membranes were blocked and blotted in PBS 0.2% Tween-20 4% nonfat dry milk except for blots performed with the CTCF Ser224-P antibody where 4% BSA was used in lieu of milk. Membranes were washed with PBS 0.2% Tween-20 and proteins were detected by chemiluminescence using Western Lightning Plus-ECL (Perkin Elmer) and film. Antibodies used are listed in Supplementary File 1.

Lambda Protein Phosphatase assay

SV40T immortalized tetraploid female Rosa26-M2rtTA MEFs were passaged asynchronously, harvested, PBS washed and aliquoted 4x10^6 cells/tube. Cell pellets were then flash frozen in liquid nitrogen. Pellets were thawed and then sonicated on ice in 1x NEBuffer Pack for Protein MetalloPhosphatases with 1mM $MnCl_2$ (New England Biolabs) and freshly added 1mM PMSF, 1μg/mL pepstatin and 1μg/mL leupeptin. 1x10^6 cell equivalents were treated with 1000U of Lambda Protein Phosphatase (New England Biolabs) at 30C for 15, 30, or 60 minutes. Lambda

Protein Phosphatase was excluded from the negative control sample which was incubated at 30C for 60minutes in buffer supplemented with 10mM sodium pyrophosphate, 10mM B-glycerophosphate, 5mM NaF, 1mM Na_3VO_4. Reactions were stopped with an equal volume of 2x Laemmli sample buffer and heated at 95C for 4 minutes. Samples were analyzed by SDS-PAGE and Western blot.

Immunofluorescence

Cells on glass slides were first permeabilized 4 minutes on ice with CSKT (100mM NaCl, 300mM Sucrose, 10mM PIPES pH 6.8, 3mM $MgCl_2$, 0.5% Triton X-100) and then fixed in PBS 4% paraformaldehyde for 10 minutes at room temperature followed by a PBS wash step. Fixed cells were blocked (PBS, 0.2% Tween-20, 4% bovine serum albumin (BSA), 2% normal goat serum (NGS) (Sigma)) for 30 minutes at room temperature. Antibodies were diluted in blocking buffer. Samples were washed with PBS300.2T (PBS supplemented to 300mM NaCl, 0.2% Tween-20). Slides were mounted using Vectashield with DAPI (Vector Labs). Slides were imaged with a Nikon Eclipse 90i microscope with Volocity software (Perkin Elmer). Antibodies used are listed in Supplementary File 1.

Mitotic chromosome immunofluoresence

TST-1 mESCs grown under feeder-free conditions were treated for 3hrs in media supplemented with 1μg/mL colcemid (Thermo Fisher). The mitotic cells were collected by shake off. Cells were then washed in ice cold PBS twice. Then the cells were resuspended in cold cell swelling buffer (CSB) (10mM Tris-HCl pH 7.5, 10mM KCl, 1.5mM $MgCl_2$, 0.1% Tween-20) and incubated on ice for 10 minutes at an approximate density of 1-2x10^5 cells/mL. Cell suspension was applied to slides (2-4x10^4 cells/slide) by centrifugation at 350g for 10 minutes using a Cytospin 3 (Shandon). Cells were extracted by treating the slides for 10 minutes at 4C in Potassium Chromosome Media (KCM) (120mM KCl, 20mM NaCl, 10mM Tris-HCl pH 7.5, 0.5mM EDTA,

0.1% Triton X-100). Mitotic chromosomes were blocked with cold KCM 1% bovine serum albumin 1% normal goat serum for 5 minutes at 4C. Samples were then incubated with primary antibodies diluted in the blocking buffer for one hour at 4C. The mitotic chromosomes were washed for 5 minutes with cold KCM 3 times. Secondary antibodies were similarly applied and washed from the samples. Slides were then fixed with PBS 4% paraformaldehyde for 5 minutes at room temperature. Slides were washed once in PBS then sequentially dehydrated in 70%, 80%, 90%, and 100% ethanol. After briefly drying the slides, coverslips were mounted using Vectashield with DAPI (Vector Labs). Slides were imaged with a Nikon Eclipse 90i microscope with Volocity software (Perkin Elmer). Antibodies used are listed in Supplementary File 1.

Mitotic spread karyotyping

F1-2.1 mESCs overexpressing CTCF, S224A or S224E for 6 days were arrested at metaphase with colcemid for 3 hours and harvested by shake off. Harvested cells were then spun down for 5 minutes at 300g and resuspended in 0.056M KCl. Cells were incubated in 0.056M KCl for 30min and spun down at 300g. Supernatant was removed, leaving 200μl, and cells gently resuspended. Cells were crosslinked by adding 1mL of 3:1 Methanol:Glacial acetic acid solution, spun down at 160g for 10min at RT. The methanol:acetic acid treatment and centrifugation steps were repeated three times more for a total of four times. Cells were then resuspended in 200ul 3:1 methanol:acetic acid solution and incubated at -20C overnight. The next day, cells were dropped from one meter onto glass slides frozen at -80C overnight. Slides were mounted using Vectashield with DAPI (Vector Labs) and imaged with a Nikon Eclipse 90i microscope with Volocity software (Perkin Elmer).

Cell-cycle analysis by flow cytometry

F1-2.1 mESCs overexpressing WT, S224A or S224E CTCF for 6 days were washed with PBS and fixed by slowly mixing in 70% EtOH chilled at -20C. Fixed mESCs were stored at least

overnight in -20C, washed once with PBS, resuspended in PBS and treated with 100μg/ml of RNase A for 5-10 minutes at room temperature. Triton X-100 was added to 0.1% and Propidium Iodide added to 20μg/ml. The sample was then mixed and incubated in the dark for 15 minutes at room temperature. DNA content profiles were generated via flow cytometry on a FACSAria 3 instrument at the HSCI CRM Flow Cytometry Core. Cell cycle profiles were generated using FlowJo10 (FlowJo, LLC) using the Watson (Pragmatic) model. G1 and G2 coefficients of variation (CVs) were set to 10 and peak locations were adjusted manually to result in the lowest possible root-mean-square deviation (RMSD).

ChIP-seq

ChIP-seq was performed using undifferentiated 2x10^7 feeder-free TST-1 cells. Harvested cells were crosslinked for 5 minutes in 1% formaldehyde. Crosslinking was quenched by the addition of glycine to 125mM. Fixed cells were PBS washed and cell pellets were frozen at -80C. Approximately 2x10^7 cells were used per ChIP. Pellets were resuspended in Lysis buffer 1 (25mM HEPES pH 7.5, 25mM EDTA, 0.5% SDS). Chromatin was sheared in a Qsonica Q800R for 20 minutes at 40% power 30sec on 30sec off at 4C to ~150bp. After centrifuging away debris, 4 volumes of Lysis buffer 2 (25mM HEPES pH 7.5, 187.5mM NaCl, 1.25% Triton X-100, 0.625% sodium deoxycholate, 6.25% glycerol, 12.5mM sodium pyrophosphate, 12.5mM β-glycerophosphate, 6.25mM NaF, 1.25mM Na_3VO_4) were added with protease inhibitors (1mM PMSF, 1μg/mL pepstatin, and 1μg/mL leupeptin). Sepharose 4B (Sigma-Aldrich) was washed with ChIP Wash Buffer 1 (25mM HEPES pH 7.5, 150mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 5% glycerol, 5mM EDTA, 10mM sodium pyrophosphate, 10mM β-glycerophosphate, 5mM NaF, 1mM Na_3VO_4, protease inhibitors) and used to preclear the chromatin samples. 25μL/IP Protein G Dynabeads (ThermoFisher) were blocked in ChIP Wash Buffer 1 100μg/mL Acyclovir, 100μg/mL Zalcitabine, 5% Gly-Gly (Sigma-Aldrich). Beads were

then washed with ChIP Wash Buffer 1. Precleared chromatin, 2μg of anti-CTCF or anti-CTCF Ser224-P antibodies (Supplementary File 1), and blocked Protein G Dynabeads were incubated together overnight at 4C with rotation. 10% of the precleared chromatin input was also saved. IP samples were washed with ChIP Wash Buffer 1, then with ChIP Wash Buffer 2 (25mM HEPES pH 7.5, 300mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 5% glycerol, 5mM EDTA, 10mM sodium pyrophosphate, 10mM β-glycerophosphate, 5mM NaF, 1mM Na_3VO_4, protease inhibitors) and then with TENG (50mM Tris-HCl pH 8.0, 1mM EDTA, 50mM NaCl, 5% glycerol, protease inhibitors). IPs were incubated for 30minutes at 37C in TENG with 200μg/mL RNase A. 10% Input samples were also treated with 200μg/mL RNase A. IPs were washed with ChIP Wash Buffer 2 and then TENG. IPs were then eluted at 65C in Elution Buffer (50mM Tris-HCl pH7.5, 10mM EDTA, 1% SDS). An equal volume of TE (Tris-HCl pH 8.0, 1mM EDTA) was added to the eluate with 40μg Proteinase K and incubated for 2hrs at 45C. The input samples were similarly protease treated in parallel. Both IP and input cross-links were reversed at 65C overnight and then phenol/chloroform/isoamyl alcohol extracted. The aqueous phase was further chloroform extracted and then alcohol precipitated. The DNA was suspended in TE. ChIP-seq libraries were generated from both IP and Input DNA samples using the NEBNext ChIP-seq Library Prep Master Mix for Illumina with NEBNext Muliplex Oligos for Illumina (NEB). DNA was quantified with both a Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher) and a KAPA Library Quantification Kit (KAPA Biosystems). Libraries were size selected using Agencourt AMPure XP resin (Beckman Coulter) and sizes confirmed on a 2100 Bioanalyzer (Agilent). Sequencing was performed on Illumina HiSeq 2000 instrument (MGH Bioinformatics Core), resulting in approximately 20-25 million paired-end 50 bp reads per sample.

ChIP-seq analysis

BWA was used to align reads against the mm9 reference genome (Ballas et al., 2009). Alignments were filtered for uniquely mapped reads and duplicates were removed. Input-normalized coverage tracks were generated using SPP (Kharchenko et al., 2008). To identify prospective CTCF binding sites, we resolved regions of ChIP-Seq tag enrichment. Tag counts were analyzed in a 1 Kb window over the chromosome length with a 200bp step and estimated statistical significance of enrichment of ChIP vs input using negative binomial distribution, with the estimate of the mean based on the tag counts in input, and the size parameter(s) selected based on manual inspection of resulting peak calls. Regions of significant enrichment were generated by merging adjacent significantly enriched windows separated by 1 Kb or less. Broad regions of enrichment were called using SPP (Kharchenko et al., 2008). Genomic location of the resultant 50,749 CTCF and 912 CTCF Ser224-P peaks (z=6) was categorized using CEAS according to RefGene annotation of introns, exons, and UTRs (Shin et al., 2009). CTCF and CTCF Ser224-P data was visualized using IGV with ENCODE mESC CTCF ChIP-seq (GSM723015) and RNA-seq (GSM723776) data plotted for reference. Motif search of the ChIP-seq peak datasets (z=6) and logo generation was performed using MEME-ChIP with default parameters (http://meme-suite.org/) (Bailey et al., 2009). To calculate CTCF and CTCF Ser224-P coverage over ChIP peaks, bam files of CTCF and CTCF Ser224-P ChIP-seq read alignments were converted to bed files. Bedtools coverage was then used to calculate coverage of CTCF and CTCF Ser224-P over all CTCF peaks or shared CTCF and CTCF Ser224-P peaks (defined as CTCF peaks that overlap with CTCF Ser224-P peaks). Linear correlation between CTCF and CTCF Ser224-P coverage was calculated using the lm() function in R with default parameters. To calculate number of CTCF motifs falling in CTCF and CTCF Ser224-P peaks, bed files of CTCF motifs were downloaded from PWMscan (http://ccg.vital-it.ch/pwmtools) using the mm9 genome and p-value cutoff of $3x10^{-5}$. Number of CTCF motifs falling within each CTCF or CTCF Ser224-P peak was then calculated with bedtools coverage using the -counts option. To calculate average conservation of CTCF motifs falling within CTCF and CTCF Ser224-P peaks, the mm9 phyloP 30-way conservation tracks (Pollard et

al., 2010; Siepel et al., 2005) were downloaded from the UCSC genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/mm9/phyloP30way) and converted to one genome-wide bigwig track. Average conservation over CTCF motifs intersecting with CTCF and CTCF Ser224-P peaks were then calculated using bigWigAverageOverBed. All Wilcoxon rank sum and Student's t-tests indicated in the text were done using the wilcox.test() and t.test() functions in R, respectively, with default parameters. ChIP-seq data have been deposited in GEO (GSE119697).

Previously published cohesin (SMC1 and SMC3) ChIP-seq data was downloaded from GEO (GSE22562) (Kagey et al., 2010) and mapped to the mm9 reference genome using bowtie2 with default parameters. Peaks were called using macs2 with default parameters. SMC1 and SMC3 ChIP-seq peaks were then intersected with CTCF and CTCF Ser224-P ChIP-seq peaks called above using bedtools with the -u option.

HYbrid Capture Hi-C (Hi-C^2)

HYbrid Capture Hi-C (Hi-C^2) probes were designed and hybridization to *in-situ* Hi-C libraries carried out as described previously (Sanborn et al., 2015) with 3-5 million F1-2.1 mESCs carrying a WT (clone FCF2), S224A (clone F3), or S224E (clone A2) transgene grown for 2 days with or without 1μg/mL doxycycline. Probe sets were designed to enrich interactions in the region of interest: chrX:70,370,161-71,832,975 (mm9). Briefly, 120 bp probes were designed around the MboI restriction sites in the regions of interest as previously described (Sanborn et al., 2015) and custom synthesized pools of single stranded oligodeoxynucleotides ordered from CustomArray, Inc. (Bothell, WA). Single stranded DNA oligos were amplified and biotinylated in a MAXIScript T7 transcription reaction (Ambion). The resulting biotinylated RNA probes were hybridized to 250-300 ng of *in situ* Hi-C libraries for 24 hours at 65C. DNA hybridized to the RNA probes was pulled down by streptavidin beads (Dynabeads MyOne Streptavidin C1, Life Technologies), washed,

and eluted as described (Sanborn et al., 2015). The resulting DNA was desalted using a 1X SPRI cleanup and amplified with Illumina primers for 18 cycles to prepare for sequencing.

Hi-C^2 libraries were sequenced to a depth of 8-15 million 50 bp paired-end reads. Reads were trimmed using cutadapt with the options --adapter=GATCGATC (MboI ligation junction) and --minimum-length=20. Reads of each pair were individually mapped to the mus and cas reference genomes using novoalign and merged into non-allelic Hi-C summary files and filtered using HOMER as previously described (Minajigi et al., 2015a). To avoid computational complexities arising from normalization of sparse, non-enriched regions in the Hi-C contact map, only Hi-C interactions falling within the capture region were analyzed further. For each capture, awk was used to pull out the filtered Hi-C interactions falling within the target region from the HOMER tag directories. Hi-C contact maps of the capture regions were then generated from these HOMER tags using the 'pre' command of Juicer tools (Durand et al., 2016b). The resulting Hi-C contact maps in .hic format were visualized and normalized with the 'Coverage (Sqrt)' option in Juicebox (Durand et al., 2016a).

Insulation and TAD score analysis of Hi-C^2 data

Insulation scores were computed for Hi-C^2 datasets as previously described (Crane et al. 2015, Giorgetti et al. 2016) with a few modifications. Briefly, 25kb resolution 'Coverage (Sqrt)' normalized Hi-C^2 interaction matrices were extracted from Juicebox using the 'dump' function of Juicer tools and converted to the cworld matrix format using custom shell and R scripts. Insulation scores were then computed from the cworld matrices using the 'matrix2insulation.pl' cworld script and the parameters '-v --is 125000 --ids 75000 –im sum' and manually normalized using the total number of reads falling in the capture region, rather than total number of reads on the chromosome. TAD scores were then calculated from the insulation scores using the 'insulation2tads.pl' cworld script with default parameters and the Mecp2 TAD boundaries

manually defined as the bins chrX: 71,225,000-71,250,000 and chrX: 71,675,000-71,700,000, based on these bins containing CTCF peaks with convergent motifs that aligned visually with the 'box' of the TAD region.

RNA-seq

RNA from F1-2.1 mESCs carrying a dox-inducible CTCF S224E transgene (clone A2) grown for six days with or without 1μg/mL doxycycline for 6 days was purified by TRIzol (Thermo Fisher) extraction via manufacturer's instructions. Polyadenylated mRNA was the selected from total cell RNA using oligo(dT) beads (New England BioLabs). Libraries were constructed using the NEBNext Ultra Direction RNA Second Strand Synthesis Module and the NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (New England BioLabs) and sequenced on a HiSeq 2000, resulting in 50-60 million 50-nt paired end reads per sample. Two biological replicates were performed.

RNA-seq analysis

RNA-seq reads were trimmed using cutadapt with the options -e 0.2 -q 20 -a AGATCGGAAGAGC -m 20 --overlap 12. Trimmed reads were aligned to the mm9 reference genome using TopHat2 as previously described. PCR duplicates were removed and HOMER used to tabulate unique reads mapping to exons of each gene. DESeq2 was then used to identify differentially expressed (DE) genes. DE genes were counted as those with adjusted p-value < 0.01. Only genes with at least one count in one of the four samples were included in the analysis. MA plot was generated using the plotMA() function of DESeq with LFC estimates generated with the lfcShrink() function of DESeq2 with default parameters.

For visualizing RNA-seq reads overlapping codon 224 of CTCF, RNA-seq reads were aligned to the CTCF S224E transgene sequence using the same parameters as above and an index of the

CTCF S224E transgene sequence generated using bowtie2 with the default parameters. The resulting bam file of aligned reads was then visualized on IGV.

Overrepresentation analysis of upregulated and downregulated DE genes was done using the PANTHER Classification System (http://www.pantherdb.org). Gene symbols for downregulated or upregulated DE genes used as input lists and the gene symbols of all expressed genes (i.e., all genes used in DESeq2 analysis) were used as the reference list.

Metagene plots of ChIP-seq over DE upregulated, downregulated and all genes were generated via CEAS using bigwig files of reads from CTCF or CTCF Ser224-P ChIP-seq (see above).

Clustal Alignment

UniProt indexed CTCF, PLK1, ATM and GSK3B amino acid sequences (Supplementary File 3) from mouse (*Mus musculus*), rat (*Rattus norvegicus*), human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), frog (*Xenopus laevis*), zebrafish (*Danio rerio*), sea urchin (*Strongylocentrotus purpuratus*), fruit fly (*Drosophila melanogaster*), and water flea (*Daphnia pulex*) were aligned using Clustal (http://www.clustal.org/clustal2/). The multiple alignments were iterated at each step using the BLOSUM62 substitution matrix with gap opening, extension, and distance penalties set to 10, 0.2, and 4 respectively. Percent identity matrices and heat maps were generated for the PLK1, ATM and GSK3β alignments.

Author Contributions

B.C.D., A.J.K. and J.T.L. developed the project. B.C.D. and A.J.K. performed the experiments. A.M.D. performed the mass spectrometry and analyzed the data with F.M.W. A.J.K. A.A. and

R.I.S. analyzed sequencing data. C.F. generated the CTCF Ser224-P antibody. B.C.D., A.J.K. and J.T.L wrote the manuscript.

Acknowledgements

This work was supported by NSF GRFP grants and a Herchel Smith Fellowship to A.J.K., and grants from the NIH (R01-GM058839 and the Howard Hughes Medical Institute to J.T.L.

Declaration of interests

J.T.L. is a cofounder and member of the Scientific Advisory Boards of Translate Bio and Fulcrum Therapeutics.

Data availability: All sequencing data has been deposited in the National Center for Biotechnology Information GEO repository under accession GSE119697.

Chapter 3

Impact of CTCF degradation on three-dimensional genome structure and X-inactivation

Introduction

Many different functions for CCCTC-binding protein (CTCF) have been discovered since its discovery over twenty years ago. Originally identified as a candidate transcription regulator of *c-myc*, this multi zinc finger protein is highly conserved amongst Nephrozoa (Heger et al., 2012; Lobanenkov et al., 1990). Subsequent studies further revealed that CTCF has insulatory activity, specifically in blocking enhancer-promoter interactions, and in separating transcriptionally active genomic regions from heterochromatic domains (Merkenschlager and Odom, 2013; Ong and Corces, 2014). Later genomic studies revealed tens of thousands of CTCF binding sites in mammalian genomes (Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Xie et al., 2007), with >5,000 of these being conserved (Schmidt et al., 2012).

The advent of chromosome conformation capture based techniques revealed another dimension of CTCF function in the cell. It was long known that the mammalian genome folds into large-scale structures in the nucleus, such as chromosome territories (Cremer and Cremer, 2001; Fraser et al., 2015; Paulson and Laemmli, 1977). Chromosome conformation capture (3C) based techniques enabled further dissection of these structures. Importantly, chromosomes further fold into compartments, partitioned on active/inactive chromatin (Rowley and Corces, 2018). Within these compartments, chromatin folds in domains, initially identified as Topologically Associating Domains (TADs) (Dixon et al., 2012; Nora et al., 2012). The majority of these, known as loop domains, are bordered by loops anchored by CTCF, in particular in the convergent orientation (Rao et al., 2014). Therefore, it was proposed that the myriad functions of CTCF could be mediated through its role in structuring the 3D genome.

Indeed, initial studies suggested that deletion of domain borders containing CTCF sites could induce ectopic enhancer targeting (Lupianez et al., 2015). Therefore, it was hypothesized that such domains in particular played an important role in gene regulation (de Wit et al., 2015; Guo et al., 2015; Merkenschlager and Nora, 2016). However, deletions made in the initial studies tended to be large, spanning hundreds of kilobases and containing many elements in addition to

CTCF binding sites. Therefore, more precise CTCF binding site deletions were made for domains containing well-established enhancer-promoter regulatory elements (Despang et al., 2019; Paliou et al., 2019; Williamson et al., 2019). Surprisingly, these deletions resulted in only small impacts on gene expression or enhancer targeting within their borders.

Such results indicate that not all CTCF sites at domain borders function to regulate genes at all times. A more nuanced model is that certain CTCF-mediated structures are important for gene regulation in certain contexts. Therefore, global CTCF ablation experiments may be more informative that deletion of specific CTCF sites. A global CTCF ablation study addressed this mainly in undifferentiated mouse embryonic stem cells (mESCs) (Nora et al., 2017); surprisingly, relatively little impact on gene expression genome-wide was detected after several days of CTCF ablation. Reasoning that CTCF may play a greater role in gene regulation during development, I revisited this question usual the model system of differentiating mESCs. In particular, the 'model' epigenetic system, X chromosome inactivation (XCI) is recapitulated in mESCs.

XCI is the process in which female mammals silence or 'inactive' one of their X chromosomes in order to balance dosage of X-linked genes with mammals. A noncoding RNA, Xist, mediates the process by being transcribed from spreading across the inactive X (Xi), leading to recruitment of repressive factors, heterochromatin formation, and gene silencing (Disteche, 2016; Lee, 2011; Starmer and Magnuson, 2009; Wutz, 2011). Several lines of evidence indicate that changes in the 3D structure of the Xi are correlated with these processes. In particular, domains are weakened on the Xi. Instead, the Xi folds into superstructures, including two (>50Mb) megadomains (Deng et al., 2015; Minajigi et al., 2015b; Rao et al., 2014). The border of these megadomains, a repeat element known as *Dxz4*, is highly bound by CTCF. Therefore, XCI provides the ideal model system to test the role of CTCF-mediated structures in gene regulation, such as gene silencing.

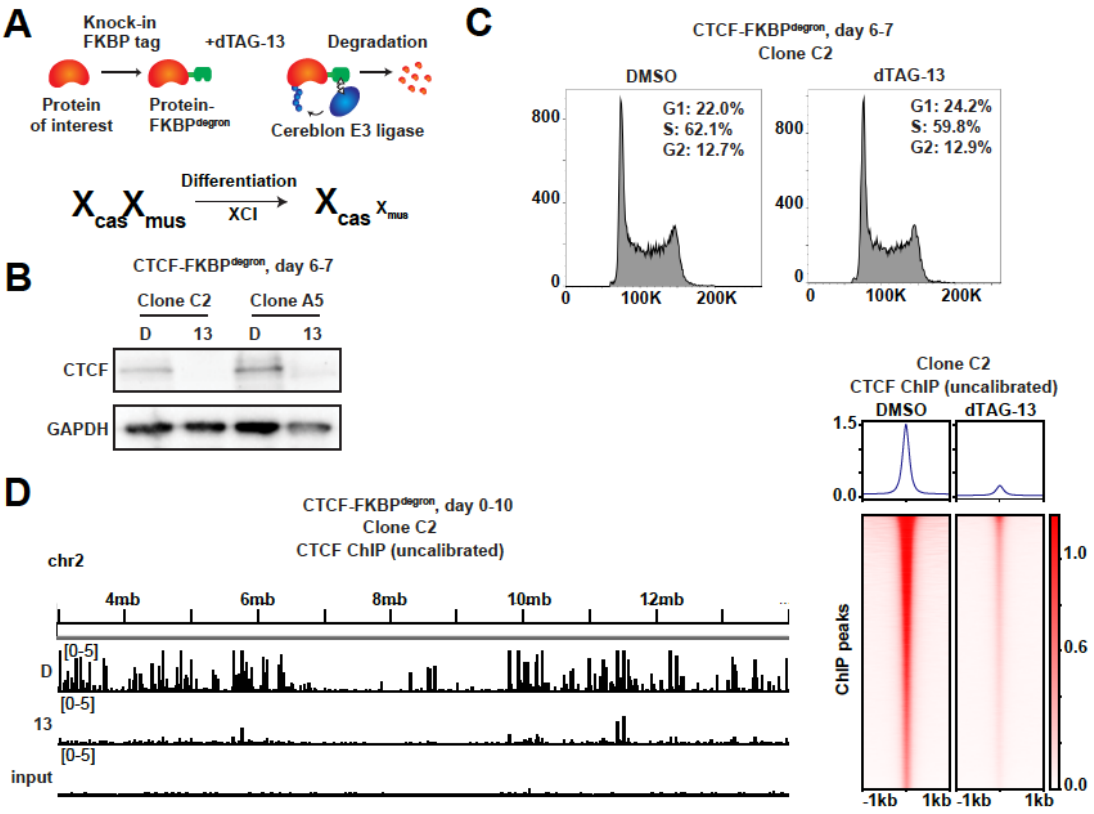
RESULTS

Generation of CTCF dTAG degron cell lines

Previous studies indicating that deletion of CTCF is lethal to developing embryos and CTCF ablation causes cell growth defects after several days (Moore et al., 2012; Nora et al., 2017). In order to circumvent these issues, I generated dTAG degron (Nabet et al., 2018; Weintraub et al., 2017) mESC lines to acutely and selectively degrade CTCF in female TST ES cells (Figure 3.1A). In the dTAG system, the heterobifunctional dTAG molecule (dTAG-13) recruits an endogenous E3 ubiquitin ligase, Cereblon, to degrade $FKBP12^{F36V}$ (herein referred to as $FKBP^{degron}$)-tagged proteins. By using an endogenous E3 ubiquitin ligase, dTAG also circumvents issues that arose in the previous CTCF degron study, which precluded degradation timecourses during differentiation; namely, difficulties in maintaining Tir1 expression (Nora et al., 2017). I knocked in the $FKBP^{degron}$ tag at the C-terminal end of both alleles of the *Ctcf* gene using CRISPR/Cas9 and a homology-directed repair template. Two CTCF-$FKBP^{degron}$ mESC lines were isolated. Western blots confirmed that CTCF could be rapidly degraded in differentiating mESCs (Figure 3.1A). In addition, cell cycle profiles were not impacted by up to 24 hours of CTCF degradation (Figure 3.1B). This lack of impact on the cell cycle agrees with our previous CTCF overexpression experiments (Chapter 2), as well as the previous CTCF degron study (Nora et al., 2017), indicating that imbalance of CTCF in either direction does not impede cell cycle progression.

Surprisingly, cells tolerated up to 10 days of degradation during mESC differentiation into embryoid bodies without visible cell death compared to the DMSO vehicle control. I therefore performed Chromatin Immunoprecipitation and sequencing (ChIP-seq) of CTCF in order to determine if CTCF binding was completely ablated on chromatin. Visual inspection of CTCF binding coverage tracks indicated that there was indeed residual binding of CTCF on chromatin after 10 days of dTAG-13 treatment. The ChIP was uncalibrated, preventing exact quantification of residual binding. However, peak calling detected 43,554 CTCF peaks in the DMSO control and only 3,047 peaks in the dTAG-13 treatment. Although this represents a 94% decrease in CTCF

binding peaks, it is possible that the small amount of CTCF remaining bound to chromatin could carry out necessary essential functions in the cell, enabling cell survival despite dTAG-13 treatment. Together, these data demonstrate that CTCF binding can be ablated using the CTCF dTAG degron system, although some residual CTCF binding does remain.



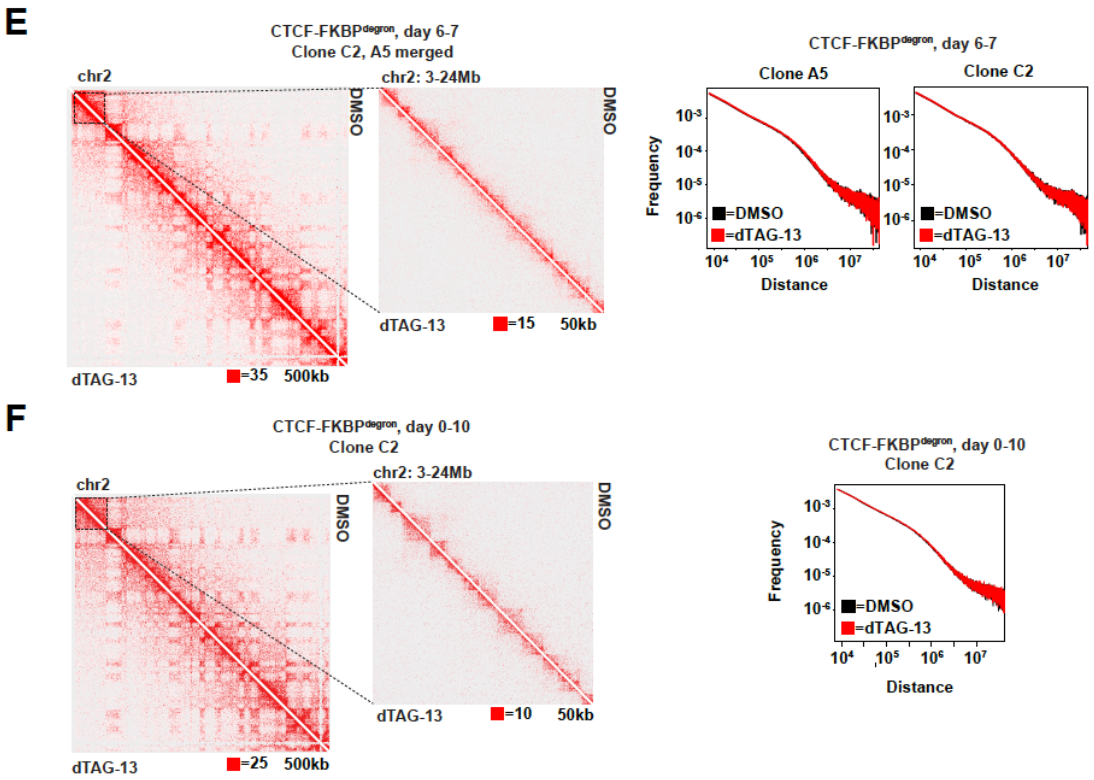


Figure 3.1. Degradation of CTCF during differentiation

(A) Schematic of dTAG degron system (top) and TST mouse embryonic stem cell model of X chromosome inactivation (bottom).

(B) Western blot for CTCF-FKBP^{degron} and GAPDH loading control in two degron clones, C2 and A5, after 24 hours of treatment with DMSO or 500nM dTAG-13 at day 6 of differentiation into embryoid bodies (EBs).

(C) Cell cycle profiling of CTCF-FKBP^{degron} clone C2 treated as in (B).

(D) Uncalibrated CTCF ChIP-seq tracks for a representative region after 10 days of DMSO or 500nM dTAG-13 treatment throughout EB differentiation (left). Genome-wide metagene average and heatmaps of CTCF binding over peaks (right).

(E) Low resolution in-situ Hi-C maps CTCF-FKBP^{degron} clones (merged) treated as in (B). Genome-wide quantification of contact distance versus probability (right).

(F) Low resolution in-situ Hi-C maps CTCF-FKBP^{degron} clone C2 treated as in (D). Genome-wide quantification of contact distance versus probability (right).

CTCF degradation has surprisingly little impact on 3D architecture during differentiation

I next performed *in situ* Hi-C in order to examine the impact of CTCF degradation on 3D genome structure. The *in situ* Hi-C was low-depth (10-30 million read pairs) but enabled assessment of large-scale impact on chromosome folding. Surprisingly, Hi-C interaction maps appeared visually similar to DMSO controls after 24 hours (Figure 3.1D) or even 10 days (Figure 3.1E) of dTAG-13 treatment. In order to assess changes in 3D structure genome-wide, I made plots of interaction frequency versus distance for both conditions (Figure 3.1D-E). All these distributions appeared similar between DMSO control and dTAG-13 treatment. Therefore, CTCF degradation for up to 10 days during differentiation largely did not change distribution of 3D contacts in the genome. The similar distributions of contact frequency versus distance in control and CTCF degradation conditions is consistent with a previous CTCF degron study (Nora et al., 2017). However, Nora et al. noted that domains became weakened upon CTCF degradation. While such changes were not immediately obvious upon examination of my Hi-C interaction maps (Figure 3.1D-E), I note that my the Hi-C experiments were sequenced at very low depth; therefore, assessment of finer structural changes, such as those of CTCF-anchored loops, was not possible. It is possible that deeper sequencing of Hi-C experiments may reveal subtle changes in 3D genome structure after CTCF degradation, similar to the previous CTCF degron study. In addition, another CTCF degron studies also observed resistance of domain structure to CTCF degradation (Kubo et al., 2021), similar to my degron line. Indeed, Nora et al. also observed persistence of some domain borders to CTCF degradation. It appears, therefore, that the impact of CTCF on 3D genome structure varies between cell lines. Further work will be needed to determine whether

this is due to resiliencies of different cell types to CTCF loss, or variances between degron systems, for example in amount of residual CTCF remaining after degradation.

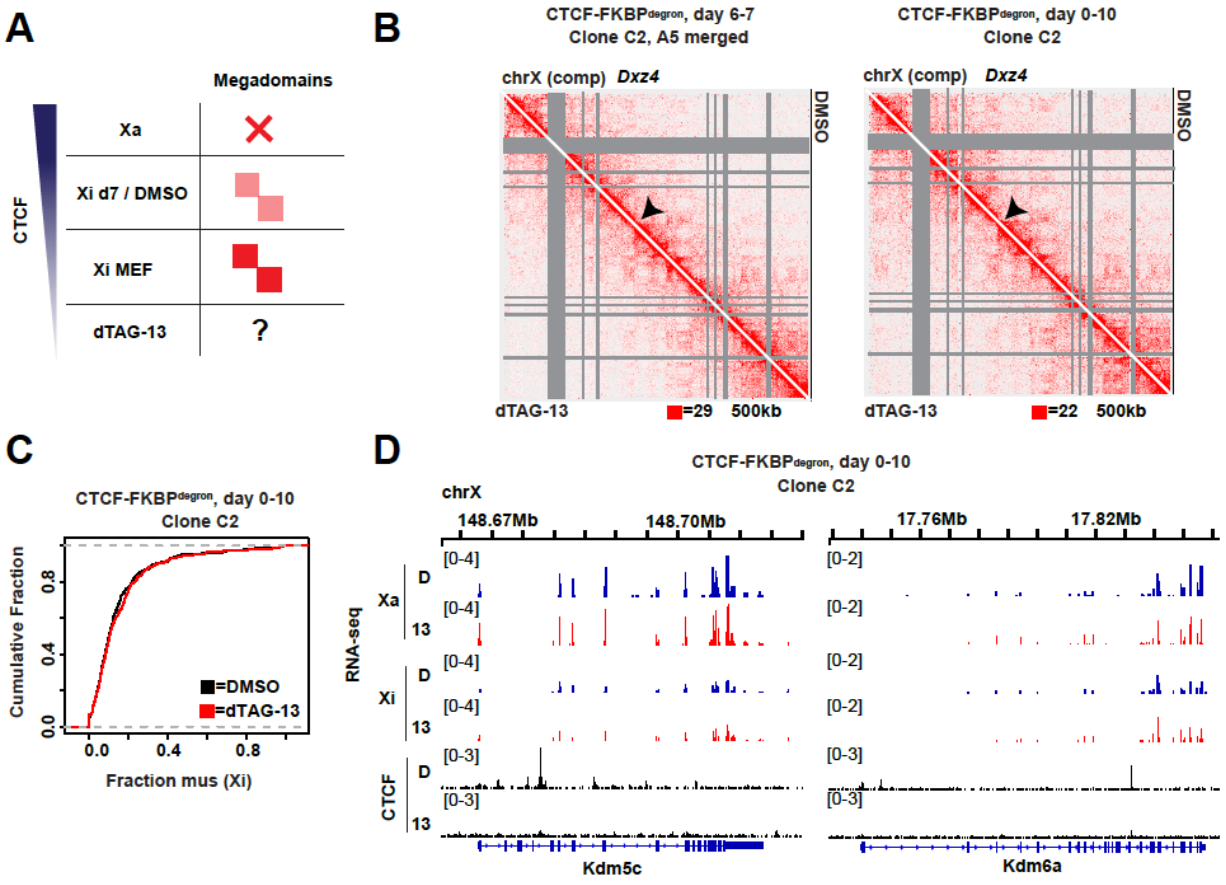


Figure 3.2. Impact of CTCF degradation on X chromosome inactivation

(A) Schematic of expected CTCF levels versus megadomain superstructures detected on the X chromosome at different stage of XCI, based on prior studies. Levels corresponding to day 10 DMSO control and dTAG-13 treated cells indicated.

(B) Low-resolution composite X chromosome Hi-C interaction maps at 500kb resolution after 24 hours of treatment at day 6 (left) or 10 days of treatment at day 10 (right).

(C) Cumulative distribution plots of chromosome-wide Xi expression in CTCF-FKBP^{degron} clone C2 treated from days 0-10 of differentiation with DMSO or 500nM dTAG-13.

(D) Expression of escapee genes *Kdm5c* and *Kdm6a* in CTCF-FKBP^{degron} clone C2 by RNA-seq after treatment from days 0-10 of differentiation with DMSO or 500nM dTAG-13.

Inactive X structure and gene silencing remain largely intact after CTCF degradation

I next examined the impact of CTCF degradation on X chromosome structure. As mentioned, the Xi adopts a unique structure during XCI when binding of cohesin and other architectural factors is depleted. In particular, the Xi folds into “megadomain” super structures with a border at the repeat element *Dxz4*. As *Dxz4* is highly bound by CTCF, it was hypothesized that, much as for domains on autosomes, CTCF was needed for the megadomain border. While my *in situ* Hi-C experiments were not sequenced deeply enough to effectively assess changes on the Xi specifically, they do enable assessment of large-scale structures on the composite X chromosome (Figure 3.2A). In particular, the prominent megadomain border at *Dxz4* can be seen even in composite X chromosome Hi-C interaction maps (Figure 3.2B). This border was apparent in the DMSO controls as well as after 1 day or even 10 days of CTCF degradation (Figure 3.2B). Therefore, the megadomains likely persist even after CTCF degradation.

Finally, I assessed impacts on gene silencing by RNA-seq. I focused on the 10 day degradation time course as the most extreme and therefore most likely to display an effect. Interestingly, no cumulative shift in expression from the Xi was observed (Figure 3.2C). These results indicate that CTCF degradation does not trigger broad changes in gene silencing. Alternatively, the residual amount of CTCF left after degradation could be sufficient to mediate any role CTCF may play in gene silencing.

While there was not a significant overall shift in Xi expression, it is possible that certain types of Xi genes could be impacted by CTCF degradation. For example, CTCF has been proposed to have a role in maintaining expression of genes that “escape” silencing during XCI. For example, CTCF could act as an insulator to block silencing machinery from spreading to escapee genes (Filippova et al., 2005; Goto and Kimura, 2009). I therefore looked at two escapee genes, *Kdm5c* and *Kmd6a*. Intriguingly, expression of these two genes were largely similar after

10 days of CTCF degradation to the DMSO control (Figure 3.2D). In other words, even 10 days of CTCF degradation did not affect the ability of *Kdm5c* and *Kmd6a* to escape from gene silencing.

DISCUSSION AND FUTURE DIRECTIONS

CTCF has long been proposed to carry out a wide range of functions in the cell. Recently, these have been reframed in the context of CTCF's mediation of 3D chromatin structures, such as loops and domains. However, CTCF degron studies observed relatively small impacts on gene expression and Polycomb spreading (Kubo et al., 2021; Nora et al., 2017). To reframe this question in the context of XCI, I generated female mESC CTCF degron lines. Surprisingly, even in the context of differentiation and XCI, CTCF degradation had relatively little impact on autosomal or Xi 3D structure and gene silencing.

While surprising, several technical aspects of the CTCF degron lines should be examined in more detail before making definitive conclusions. First, while extremely efficient at ablation, degron systems are still not equivalent to a complete knockout. Therefore, some residual CTCF remains even after long degradation time courses. Further investigation will be needed to determine whether this residual CTCF could mediate most essential functions in the cells, explaining the relatively lack of phenotype.

Indeed, while all three CTCF degron studies observed residual domain structure after CTCF degradation, they appear to differ in extent which residual structure was observed. Careful assessment of residual CTCF binding to chromatin between the different degron cell lines, for example by calibrated ChIP-seq, will be needed to determine if differing phenotypes could be due to varying efficiencies of degrading CTCF. In addition, most experiments I performed in my study were done with an extreme 10 day time course of degradation. It is possible that such a long time course may select against cells sensitive to CTCF loss. Therefore, ChIP-seq and RNA-seq experiments should be repeated in my degron lines with a more comparable time frame to the previous CTCF degron studies.

Taking the results at face value, however, the CTCF degron studies suggest that mESCs express CTCF orders of magnitude higher than necessary to maintain domain borders and XCI, which in itself is an intriguing observation. What would be the biological purpose of such overexpression? Could it be a mechanism by which mESCs protect against the catastrophic consequences of CTCF loss? Alternatively, cells in culture could simply require much less CTCF than in animals. To address this, knockdown or deletion of one CTCF allele could be paired with degron cell lines to further decrease CTCF levels in cells.

Despite these caveats, CTCF degron lines serve as an exciting system in which to further study aspects of CTCF function. One exciting avenue to explore are the post-translational modifications we and others showed CTCF undergoes (Chapter 2). Intriguingly, one CTCF phosphorylation site, Serine-224, is located in the N-terminal region of CTCF recently identified to interact with and stabilize cohesin (Li et al., 2020; Nora et al., 2020; Pugacheva et al., 2020). In our study, we assayed the impact of disrupting Serine-224 phosphorylation by overexpressing phospho-mutants of CTCF. However, the presence of wild-type CTCF likely mitigated the expression level of the phospho-mutant as well as possible phenotypes. The CTCF degron line provides an ideal system in which to express phospho-mutants while eliminating the majority of wild-type CTCF, thus enabling cleaner assessment of the impact of disrupting Serine-224 phosphorylation. Such experiments paired with biochemistry and cohesin binding assays could determine if this post-translation mark modulates CTCF-cohesin interactions.

METHODS

Cell Lines

CTCF-FKBP^{degron} mESCs were generated by transfecting *Tsix*$^{TST/+}$ mESCs with (1) a plasmid (pSpCas9-(BB) 2A Puro) expressing Cas9 with an sgRNA targeted to the STOP codon (2) a homology repair template plasmid including 900bp homology arms around the STOP codon, the FKBP^{degron} tag in a pUC19 backbone and (3) a homology-directed repair plasmid (pMB 1609-pRR-

EGFP) with the sgRNA target site cloned in. Specifically, 500,000 TST mESCs were transfected using Lipofectamine 3000 1.5ug of pSpCas9-(BB) 2A Puro, 1.5ug of the non-linearized repair plasmid, and 400ng of pMB 1609-pRR-EGFP. Transfected cells were GFP sorted after 48 hours and the entire GFP positive pool plated. After 5 days, the pool was split and plated at low density, and after 5 more days, clones were picked from the low-density plate in a 96-well plate. Clones were grown to confluency, then triplicate plated in 96-well plates. One 96-well plate was used to PCR screen clones. Clones were then thawed from one of the triplicate plates, expanded, and re-screened by PCR and Sanger sequenced to confirm correct knock-in of the $FKBP^{degron}$ tag.

ES cell differentiation and induced degradation

Undifferentiated mESCs were grown on γ-irradiated MEF feeders in medium containing DMEM, high glucose, GlutaMAX supplement, pyruvate, 15% Hyclone FBS (Sigma), 25 mM HEPES pH 7.2-7.5, 1x MEM non-essential amino acids, 1x Pen/Strep, 0.1 mM βME, and 500 U/mL ESGRO recombinant mouse Leukemia Inhibitory Factor (LIF) protein (Sigma, ESG1107) at 37°C with 5% CO2. To induce differentiation in EBs, mESCs were trypsinized and resuspended in the same media without LIF. For the first 3 days of differentiation, EBs were grown in suspension. On day 4 and subsequent days, EBs were grown on gelatinized plates until harvested at day 7 or day 10.

Degradation of RAD21 or WAPL in $FKBP^{degron}$ lines during differentiation was induced by adding a 1:20,000 dilution of 10mM dTAG-13 in DMSO to cell culture media (500nM final concentration) for the indicated time frames. The small molecule dTAG-13 was synthesized as previously described (Erb et al., 2017; Nabet et al., 2018). For DMSO vehicle controls, the same volume of DMSO alone was added to cell culture media.

Western blot

Cells were counted and ~100,000 cells per sample lysed in 100ul 1X Laemmli buffer. Protein extracts were boiled for 5 min at 95C. Equal cell counts (5-15ul) were loaded and resolved by SDS page. Protein was transferred to an Immobilon-P PVDF membrane (EMD Millipore). After transfer, membrane was blocked with 5% milk in PBST (PBS/0.05% Tween-20) for 30 min at room temperature. Membrane was incubated with primary antibody (1:5000 for CTCF, 1:20,000 for GAPDH) in 5% milk in PBST at 4C overnight, then washed three times with PBST. Membrane was then blocked with secondary antibody-HRP conjugate (1:5000) in 5% milk in PBST for 1 hour at room temperature. Membrane was incubated with Western Lightning Plus-ECL (PerkinElmer) and exposed using the Bio-Rad ChemiDoc MP Imaging System.

Antibodies

For ChIP-seq the following antibodies were used: CTCF (CST 3418). For Western blot: CTCF (CST 3418), GAPDH (CST 5174T).

ChIP-seq and calibrated ChIP-seq

ChIP-seq was performed as described previously with minor modifications (Del Rosario et al., 2019). Briefly, cells were trypsinized, resuspended in 1 million cells/mL of media and crosslinked with 1% formaldehyde for 10 min at room temperature with rotation. Cells were then quenched with 0.2 M glycine for 5 min. Crosslinked cells were washed once with cold PBS, pelleted, and snap frozen in an EtOH/dry ice bath. Cell pellets were stored at -80C until further processing. 5-10 million cross-linked cells per IP were thawed on ice and resuspended in 300ul Lysis Buffer 1 (25mM HEPES pH 7.2-7.5, 25mM EDTA, 0.5% SDS). Nuclei were sonicated (Qsonica Q800 Sonicator) in polystyrene tubes at 40% power setting, 30s ON/30s OFF for a total sonication (ON) time of 6 min. After sonication, 1.2 mL of Lysis Buffer 2 (25mM HEPES pH 7.2-7.5, 187.5mM

NaCl, 1.25% Triton X-100, 0.625% Na-deoxycholate, 6.25% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail) was added. 30ul of Dynabeads Protein G (Thermo Fisher Scientific) per IP were washed 3 times with ChIP Wash Buffer 1 (25mM HEPES pH 7.2-7.5, 150mM NaCl, 5mM EDTA, 1% Triton X-100, 0.5% Na-deoxycholate, 5% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail) and resuspended in 60ul ChIP Wash Buffer 1. Each IP was blocked with 20ul washed beads for 2 hours at 4C with rotation. The other 40ul of washed beads were incubated with antibody (5ul of RAD21 antibody or 2ul of H3K27me3 antibody/IP) for 2 hours at 4C. 5% input was taken from blocked lysates. Beads conjugated to antibody were added to blocked lysates and IP allowed to proceed overnight at 4C with rotation. IPs were then washed 3 times with cold ChIP Wash Buffer 1, 3 times with cold ChIP Wash Buffer 2 (25mM HEPES pH 7.2-7.5, 300mM NaCl, 5mM EDTA, 1% Triton X-100, 0.5% Na-deoxycholate, 5% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail). And 1 time with cold TENG (50mM Tris-HCl pH 8, 1mM EDTA, 50mM NaCl, 5% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail). All washes were performed by inverting tubes and setting aside for ~2 minutes per wash. IPs were then resuspended in 500ul TENG with 100ug RNAse A (Thermo Fisher Scientific) and incubated for 30 min at 37C. IPs were then washed 2 times with cold ChIP Wash Buffer 2 and 2 times with cold TENG without protease inhibitors. IPs were resuspended in 200ul elution buffer (50mM Tris-HCl pH 7.5, 10mM EDTA, 1% SDS) and incubated 15 min at 65C. Eluate was then transferred to a new tube and 198ul TE buffer (Ambion) and 40ug Proteinase K (Sigma) added. Inputs were thawed, TE buffer added to equalize volume to IPs and 200ul elution buffer and 40ug Proteinase K added. Inputs and IPs were then incubated for 2h at 45C and crosslinks reversed overnight at 65C. DNA was purified with phenol-chloroform extraction and resuspended in 50ul TE. Efficiency of ChIP was assayed via qPCR, with samples displaying over 10-fold enrichment of positive regions over negative control prepared for sequencing. Libraries were prepared from ~500ng of input or IP DNA using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB) according to kit instructions. Experiments were performed in two biological replicates. Input and

ChIP-seq libraries were sequenced on Illumina HiSeq, generating 20-50 million paired end 50 or 150 nucleotide reads per sample.

ChIP-seq analysis

For uncalibrated ChIP-seq, input and ChIP-seq reads were aligned to the 129S1/SvJm (mus) and CAST/Eih (cas) genome as previously described (Kung et al., 2015; Minajigi et al., 2015b). PCR duplicates were removed and uniquely mapped reads used for further analysis. Fpm-normalized bigwig tracks were generated for all uniquely mapped (comp), mus-specific (mus), and cas-specific (cas) reads.

Heatmaps and metaplots were generated using deepTools functions computeMatrix and plotHeatmap with the reference point as the summit of the called ChIP peaks.

RNA-seq

Total RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific) and mRNA isolated using the NEBNext Poly(A) Magnetic Isolation Module (NEB) according to the kit instructions. Two biological replicates of RNA-seq libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) according the kit instructions. Libraries were sequenced on the Illumina HiSeq2500 (rapid run), generating 30-50 million 50 nt paired end reads per sample or the Illumina HiSeq4000, generating 16-27 million 150 nt paired end reads per sample.

***In situ* Hi-C and visualization**

In situ Hi-C was done as described previously (Rao et al., 2014) using 2 clones for each degron line and 5 million cells per sample.

Reads were trimmed using cutadapt with the options --adapter=GATCGATC (MboI ligation junction) and --minimum-length=20. Reads of each pair were individually mapped to the mus (Xi) and cas (Xa) reference genomes using novoalign and merged into Hi-C summary files and filtered using HOMER as previously described (Minajigi et al., 2015). Hi-C contact maps were generated from reads in the filtered HOMER mus and cas tag directories using the ‘pre’ command of Juicer tools with the genomeID mm9 (Durand et al., 2016b). The resulting Hi-C contact maps in .hic format were visualized and normalized with the ‘Coverage Square Root’ option in Juicebox version 1.8.8 (Durand et al., 2016a).

Chapter 4

Balancing cohesin eviction and retention prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-inactivation

Attribution

The content of this chapter has been published in Kriz *et al.*, Molecular Cell, 2021.

SUMMARY

Depletion of architectural factors globally alters chromatin structure, but only modestly affects gene expression. We revisit the structure-function relationship using the inactive X chromosome (Xi) as a model. We investigate cohesin imbalances by forcing its depletion or retention using degron-tagged RAD21 (cohesin subunit) or WAPL (cohesin release factor). Cohesin loss disrupts Xi superstructure, unveiling superloops between escapee genes, with minimal effect on gene repression. By contrast, forced cohesin retention markedly affects Xi superstructure, compromises spreading of Xist RNA-Polycomb complexes, and attenuates Xi silencing. Effects are greatest at distal chromosomal ends, where looping contacts with the *Xist* locus are weakened. Surprisingly, cohesin loss creates an “Xi superloop” and cohesin retention creates “Xi megadomains” on the active X. Across the genome, a proper cohesin balance protects against aberrant inter-chromosomal interactions and tempers Polycomb-mediated repression. We conclude that a balance of cohesin eviction and retention regulates X-inactivation and inter-chromosomal interactions across the genome.

INTRODUCTION

Classical microscopy studies revealed that mammalian chromosomes fold into large scale structures and occupy discrete nuclear territories in order to fit into this space (Cremer and Cremer, 2001; Fraser et al., 2015; Paulson and Laemmli, 1977). The advent of genome-wide chromosome conformation capture (3C)-based assays revealed how in particular, chromosomes form "loops" at several levels (Paulson and Laemmli, 1977; Rao et al., 2014; Splinter et al., 2006) and partition into "compartments" based on chromatin properties and gene expression states (Rowley and Corces, 2018). Chromosomes have also been proposed to fold into “Topologically associating domains” (TADs) (Dixon et al., 2012; Nora et al., 2012), though single cell Hi-C and microscopy studies reveal that TADs represent population-based averages, rather than stable domains (Beagrie et al., 2017; Bintu et al., 2018; Nagano et al., 2013; Nagano et al., 2017; Stevens et al., 2017). Two factors dynamically regulate chromosome architecture — cohesin and CTCF. Cohesin, a multi-subunit complex, is thought to form a ring-like structure around chromatin and extrudes chromatin loops in an ATP-dependent manner (Davidson et al., 2019; Kim et al., 2019). RAD21 is the essential kleisin subunit of the cohesin complex, which additionally consists of SMC proteins SMC1, SMC3, and either SA1 or SA2 (Nasmyth and Haering, 2009; Peters et al., 2008). CTCF, an 11-zinc finger DNA-binding protein, insulates enhancers from promoters and typically anchors long-range loops (Bell and Felsenfeld, 2000; Rao et al., 2014).

It was hypothesized that 3D genome structures had an important role in gene regulation (de Wit et al., 2015; Guo et al., 2015; Lupianez et al., 2015; Merkenschlager and Nora, 2016). Unexpectedly, a series of loss-of-function studies targeting CTCF, cohesin component RAD21, and cohesin release factor WAPL showed that depletion of each of these factors dramatically alters 3D genome structure, but only modestly impacts gene expression (Haarhuis et al., 2017; Nora et al., 2017; Rao et al., 2017). These results beg the question of what precise role 3D genome organization plays during epigenetic regulation. A caveat is that these studies were

conducted in a limited number of cell types and developmental timepoints. Therefore, 3D genome structure could play a more acute and global role during epigenetic regulation than is currently appreciated.

Here we revisit the role of cohesin-mediated architecture using a model epigenetic phenomenon, X chromosome inactivation (XCI) — the process by which female mammals 'inactivate' one of their X chromosomes to balance dosage of X-linked genes with males. XCI occurs during differentiation of female mouse embryonic stem cells (mESCs). The process is mediated by Xist, a noncoding RNA which is transcribed from and spreads across the inactive X (Xi), triggering recruitment of repressive factors such as Polycomb Repressive Complexes 1 and 2 (PRC1/2), formation of heterochromatin, and gene silencing (Disteche, 2016; Lee, 2011; Starmer and Magnuson, 2009; Wutz, 2011). XCI provides an extreme example of changes in 3D genome structure. While the active X (Xa) folds similarly to autosomes, the Xi shows greatly weakened TAD structure and appears to be devoid of A/B compartmentalization. The Xi is instead organized into two large (>50Mb) megadomains (Deng et al., 2015; Minajigi et al., 2015b; Rao et al., 2014) and the A/B compartments are reorganized into "S1" and "S2" compartments before assuming the super-structure that characterizes the globally silenced chromosome (Wang et al., 2019; Wang et al., 2018). Restructuring from A/B to S1/S2 compartments requires Xist RNA and Polycomb repressive complex 1 (PRC1), and further reorganization of S1/S2 into a single large compartment requires the noncanonical SMC protein, SMCHD1 (Gdula et al., 2019; Wang et al., 2019; Wang et al., 2018), a cohesin and condensin relative (Blewitt et al., 2008).

While SMCHD1 and the S1/S2 structures are required for gene silencing, a remaining pivotal question is whether accompanying structural changes are required for or are merely byproducts of silencing the entire chromosome. Clarifying these questions would impact general understanding of 3D genome regulation, as well as XCI biology. To this aim, here we perturb cohesin levels through targeted degradation of a cohesin component RAD21 and a release factor

of cohesin, WAPL, using the dTAG system. WAPL evicts cohesin from DNA in prophase (Gandhi et al., 2006; Kueng et al., 2006) as well as during interphase (Tedeschi et al., 2013), possibly by opening a DNA 'exit gate' (Buheitel and Stemmann, 2013; Chan et al., 2012; Huis in 't Veld et al., 2014). By degrading either RAD21 or WAPL, we modulate cohesin levels upwards and downwards in a mouse embryonic stem cell (mESCs) model and examine the consequences for loop formation, compartments, and XCI.

RESULTS

Toggling cohesin binding using a RAD21- and WAPL-dTAG degron system

We employed the dTAG degron system (Nabet et al., 2018; Weintraub et al., 2017) to acutely and selectively degrade a cohesin component RAD21 and cohesin release factor, WAPL, in female TST ES cells (Figure 4.1A). The female TST ES line carries a *Tsix* termination cassette on one X chromosome (*Tsix*$^{TST/+}$) to ensure silencing of the *Mus musculus* X chromosome in this hybrid ES line (Ogawa et al., 2008) (Figure 4.1A). In the dTAG system, the heterobifunctional dTAG molecule (dTAG-13) recruits an endogenous E3 ubiquitin ligase, Cereblon, to degrade FKBP12^{F36V} (herein referred to as FKBP^{degron})-tagged proteins. The dTAG degron system thus avoids previously described issues with auxin degrons (Nora et al., 2017), namely difficulties maintaining Tir1 expression and effective degradation during differentiation. The FKBP^{degron} tag was knocked in at the endogenous STOP codons of the *Rad21* and *Wapl* genes using CRISPR/Cas9 and a homology-directed repair template (Figure S2.1A-E). We isolated 2 RAD21-FKBP^{degron} clones with comparable expression levels to endogenous RAD21 (Figure 4.1B, S2.1B-C). We also isolated 2 WAPL- FKBP^{degron} clones, although these had different baseline levels of WAPL expression (Figure 4.1B, S2.1D). We later learned that the higher WAPL expression in clone H7

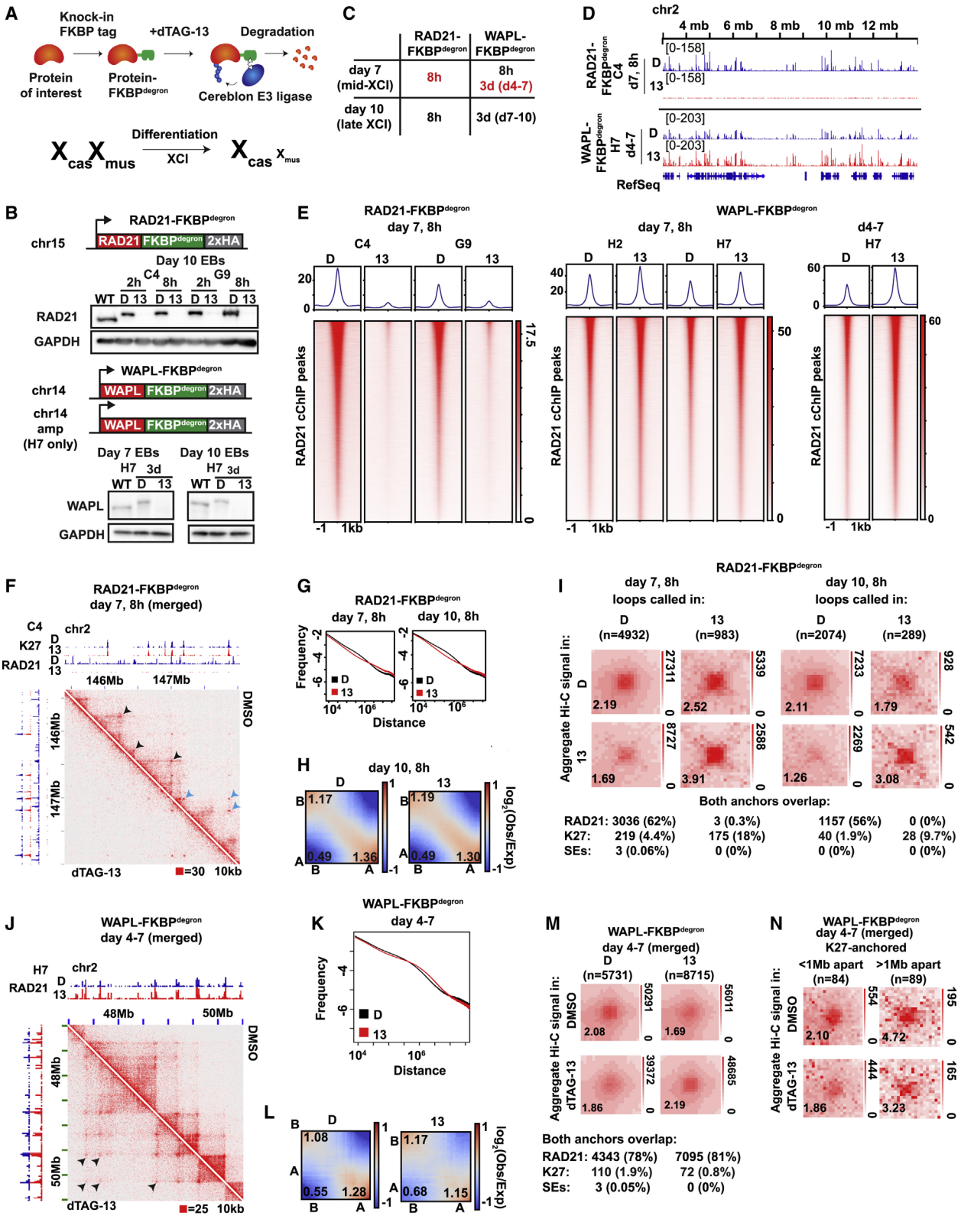


Figure 4.1: Toggling cohesin binding and 3D genome structure during differentiation.

(A) Schematic of dTAG degron system (top) and the TST mouse embryonic stem cell system for X chromosome inactivation (bottom).

(B) Western blots for FKBP^{degron}-tagged proteins and GADPH loading control in unmodified parental TST (WT) and indicated FKBP^{degron} lines. Samples treated with DMSO vehicle control (D) or 500nM dTAG-13 (13).

(C) Table of investigated time frames in each degron line. Red, focuses of this study.

(D) Representative tracks of RAD21 binding by calibrated ChIP-seq (cChIP-seq) in RAD21-FKBP^{degron} lines and WAPL-FKBP^{degron} lines.

(E) Metaplot and heatmaps of RAD21 binding by cChIP-seq in RAD21-FKBP^{degron} and WAPL-FKBP^{degron} lines.

(F) Representative Hi-C interaction maps in RAD21-FKBP^{degron} lines (10kb resolution, clones merged) at day 7 after 8 hours of dTAG-13 or DMSO control treatment. Cohesin-dependent (black arrows) and cohesin-independent (blue arrows) loops indicated.

(G) Frequency versus distance plots of RAD21-FKBP^{degron} Hi-C maps (clones merged). Treatment as in (F).

(H) Saddle plot analysis of compartment strength in RAD21-FKBP^{degron} Hi-C maps (clones merged). Scores represent B-B, B-A, and A-A compartment interaction strength. Treatment as in (F).

(I) Aggregate peak analysis (APA) plots of aggregate Hi-C signal over sets of loops called in dTAG-13 or DMSO RAD21-FKBP^{degron} Hi-C maps (clones merged). Lower left: average loop enrichment (P2LL ratio). Numbers and percent of dTAG-13 and DMSO loops with both anchors overlapping RAD21, H3K27me3 cChIP peaks or mESC super-enhancers (Whyte et al. 2013) are shown.

(J) Representative Hi-C interaction maps in RAD21-FKBP^{degron} lines (15kb resolution, clones merged) at day 7 after 3 days of dTAG-13 or DMSO control treatment. Longer cohesin-anchored chromatin loops appearing upon WAPL degradation indicated (black arrows).

(K) Frequency versus distance plots of WAPL-FKBP^{degron} Hi-C maps (clones merged). Treatment as in (J).

(L) Saddle plot analysis of compartment strength in WAPL-FKBP^{degron} Hi-C maps (clones merged). Scores as in (H). Treatment as in (J).

(M) APA plots of aggregate Hi-C signal over sets of loops called in dTAG-13 or DMSO WAPL-FKBP^{degron} Hi-C maps (clones merged). Lower left: average loop enrichment (P2LL ratio). Numbers and percent of dTAG-13 and DMSO loops with both anchors overlapping RAD21, H3K27me3 cChIP peaks in corresponding samples or mESC super-enhancers (Whyte et al. 2013) are shown.

(N) APA plots of aggregate WAPL-FKBP^{degron} Hi-C signal over sets of cohesin-independent H3K27me3-anchored loops (clones merged). Lower left: average loop enrichment (P2LL ratio).

was due to an amplification of the Wapl-FKBP^{degron} gene. We therefore took extra care to characterize the second clone, H2, which lacked the amplification but had a much lower baseline WAPL level relative to wildtype cells (Figure S2.1F). Because clone H7 had a baseline WAPL level closer to wildtype cells, we considered it important to analyze both H7 and H2. Thus, all analyses were carried out with both clones for both RAD21 and WAPL-FKBP^{degron}, with results agreeing between clones.

Western blots and immunofluorescence confirmed that RAD21 and WAPL were rapidly degraded during ES differentiation upon treatment with dTAG-13 but not with the DMSO vehicle (Figure 4.1B, S2.1C-D, G). To study effects on XCI, we performed analyses at days 7 and 10 of differentiation — timepoints corresponding to the mid and late stages of XCI, respectively (Figure 4.1C). For RAD21, we chose 8 hours of degradation to compare to previous degron studies (Rao et al., 2017; Rhodes et al., 2020). This degradation timeframe resulted in less impact on the cell cycle (Figure S2.2A-B) than experienced in a previous study (Rhodes et al., 2020). For WAPL,

we performed both an 8-hour and 3-day degradation, reasoning that it may take time to build up cohesin levels and observe resulting physiological effects. WAPL degradation did not affect cell cycle profiles (Figure S2.2C-E), consistent with the idea that separase can cleave cohesin at centromeres in the absence of WAPL, enabling cell cycle progression (Tedeschi et al., 2013).

We then performed calibrated ChIP sequencing (cChIP-seq) to examine effects on cohesin chromatin binding. RAD21 binding was reduced dramatically (Figure 4.1D-E). Whereas 29,519 RAD21 significant peaks were observed in DMSO controls, only 351 peaks were present after dTAG-13 treatment in clone C4 (99% decrease). Similarly, whereas 23,659 peaks occurred in DMSO controls, only 780 peaks remained after dTAG-13 treatment in clone G9 (97% decrease), This degree of loss in cohesin binding is comparable to a prior study (Rao et al., 2017). We also performed calibrated ChIP-seq (cChIP-seq) for RAD21 following WAPL degradation. After 8 hours of degradation, we observed increased cohesin binding (Figure 4.1E). Specifically we detected 49,130 (clone H2) and 42,086 (clone H7) peaks in DMSO control conditions and 53,480 (clone H2) and 47,866 (clone H7) peaks after dTAG-13 treatment. After 3 days of degradation, 45,415 peaks were called in DMSO control and 66,488 peaks called after dTAG-13 treatment in clone H7 (Figure 4.1D-E). In clone H2, similar numbers of peaks were called (73,706 in DMSO and 68,881 in dTAG-13), but became wider on average after dTAG-13 treatment (Figure S2.3A). The 3-day increase in cohesin binding was similar to that observed in a previous WAPL knockout study (Haarhuis et al., 2017). Together, these data demonstrate that we can successfully toggle cohesin levels using the RAD21- and WAPL-degron system.

Changes in 3D architecture in response to cohesin eviction and retention

To investigate changes in 3D architecture, we generated *in situ* Hi-C interaction maps, with a minimum of 360 million read pairs for each day 7 sample (totaling 2.8 billion read pairs and

2.7 billion valid contacts across 14 samples, Supplemental Table S1.1) achieving a resolution of 10kb. For RAD21, 8 hours of degradation at day 7 or 10 ablated topological domains evident along diagonals of contact heatmaps (Figure 4.1F-G, S2.3B-D), as expected. By contrast, A/B compartmentalization was not greatly altered (Figure 4.1H, S2.3B), agreeing with previous studies (Rao et al., 2017; Wutz et al., 2017). In contrast, there were intriguing effects on long-range loops, visible in Hi-C matrices as dots of high interaction frequency (Figure 4.1F, arrows). HiCCUPs (Durand et al., 2016b) and metaloop analysis showed a clear decrease in the number of looping interactions following cohesin depletion (Figure 4.1I) — 4,932 (DMSO) versus 983 (dTAG13) at day 7, 2074 versus 289 at day 10.

Previous studies found that ablation of cohesin dysregulates Polycomb target genes in undifferentiated mESCs (Rhodes et al., 2020), with subunits SA1 and SA2 having opposing effects on interactions between Polycomb domains (Cuadrado et al., 2019). We independently observed that a substantial fraction (10-18%) of remaining loops were anchored by H3K27me3 (Figure 4.1I). However, most cohesin-independent loops were not anchored by H3K27me3, indicating chromatin features in addition to cohesins and PRC2 are involved in loop formation (Figure 4.1I, S2.3E). Interestingly, metaloop analysis showed that loops persisting after cohesin degradation were strengthened (Figure 4.1I). No persistent loops occurred between super-enhancers (SE) (Whyte et al., 2013) (Figure 4.1I). SE interactions may occur at lower frequency during cell differentiation than in pluripotent ES cells (Rao et al., 2017; Rhodes et al., 2020). These data indicate that depleting cohesin binding reduces the number of loops across the genome, but persistent loops paradoxically strengthen.

We next examined consequences of increasing cohesin binding by degrading WAPL. When we degraded WAPL for 8 hours (to mirror the timeframe of RAD21 depletion), there was only a modest increase in cohesin binding (Figure 4.1E). Thus, we degraded WAPL for 3 days between days 4-7 or days 7-10. These timeframes resulted in an increase in interactions of 1-

10Mb (Figure 4.1J-K, S2.3B,F) and a decrease in far-cis compartmental interactions of >10Mb (Figure 4.1K-L,S2.3F). Concurrently, there was a large increase in number of chromatin loops, with HiCCUPs calling 5731 loops in controls versus 8715 loops in WAPL-depleted cells, of which the vast majority were anchored by RAD21 (Figure 4.1M). Whereas RAD21-depletion strengthened Polycomb-anchored loops (Figure 4.1I), WAPL degradation weakened them (Figure 4.1N). The effect appeared to be distance-dependent, as loops spanning >1Mb were weakened more than loops spanning <1Mb (Figure 4.1N). Thus, whereas cohesin depletion reduces loop number and strengthens persistent loops, cohesin retention did the opposite.

Polycomb loops protect against effects of cohesin loss

We next investigated effects on gene expression by RNA-seq. After 8-hr RAD21 in day 10 cells, 43 genes were significantly upregulated and 98 were downregulated (adjusted p-value < 0.1), with downregulated genes being enriched for H3K27me3 (Figure 4.2A, top, Supplemental Table S1.2). These numbers are similar to those reported for other cell types (Rao et al., 2017; Rhodes et al., 2020). We wondered if they may be an underestimate because of the median mRNA half-life of ~10 hr (Greenberg, 1972). We therefore performed a transcriptome-wide analysis of nascent transcription using PRO-seq. However, a similar number of genes were revealed to be differentially expressed after 8 hr of RAD21 depletion on day 7 (Figure 4.2A**;** 37 downregulated, 32 upregulated), reinforcing a modest effect of acute cohesin loss on gene expression. Loss of RAD21-loop anchors did not correlate with gene dysregulation at day 10 (Figure S2.4A). At day 7, there was a significant tendency for genes overlapping cohesin-dependent loop anchors to be more upregulated, but again the effect was modest (Figure S2.4A).

However, Polycomb targets inside a H3K27me3-marked persistent loop anchor were less downregulated than Polycomb targets not in such loop anchors in day 7 cells (Figure 4.2B, bottom). This contrasts with day 0 ES cells, in which Polycomb loops were correlated with

increased repression upon cohesin degradation (Rhodes et al., 2020). Further examination revealed that even in the DMSO control, clustered Polycomb targets tended to be more repressed than those not involved in such interactions (Figure 4.2B). This suggests that, normally, interactions between Polycomb domains reinforce repression and cohesin depletion makes repression even more robust. Changes in gene expression upon cohesin removal were not correlated with changes in H3K27me3 by cChIP-seq (Figure 4.2C, top). Thus, Polycomb target genes are strikingly resistant to cohesin degradation in differentiating ES cells.

Cohesin protects against inter-chromosomal *Hox* interactions and tempers Polycomb-mediated repression

Among classic Polycomb targets are the 4 *Hox* clusters, each located on a different chromosome, with each cluster being coordinately regulated in cis to specify anterior-posterior patterning in mammals (Deschamps and Duboule, 2017; Pearson et al., 2005; Wellik, 2007). As *Hox* regulation is recapitulated during ES differentiation (Lanctot et al., 2007), we studied how architectural organization of *Hox* genes is affected by cohesin loss. Examination of Hi-C maps revealed an entire network of inter-*Hox* contacts between chromosomal clusters (Figure 4.2D, S1.3G-I). A small subset of these contacts had been seen previously in ES cells (Lanctot et al., 2007; Schoenfelder et al., 2015). Heretofore unseen, however, are interactions that extend across entire *Hox* clusters and throughout cell differentiation. Most intriguingly, these trans-interactions strengthened without cohesin (Figure 4.2D, S2.3G-I). A previous RAD21-degron study did not report these contacts (Rhodes et al., 2020). Reanalysis of their Hi-C dataset indicated that they indeed exist in day 0 cells—but only after RAD21 degradation (Figure S2.3G)**.** Thus, analysis of our degron-tagged ES cells showed a dynamic and extensive network of *Hox-to-Hox* contacts (Figure 4.2E).

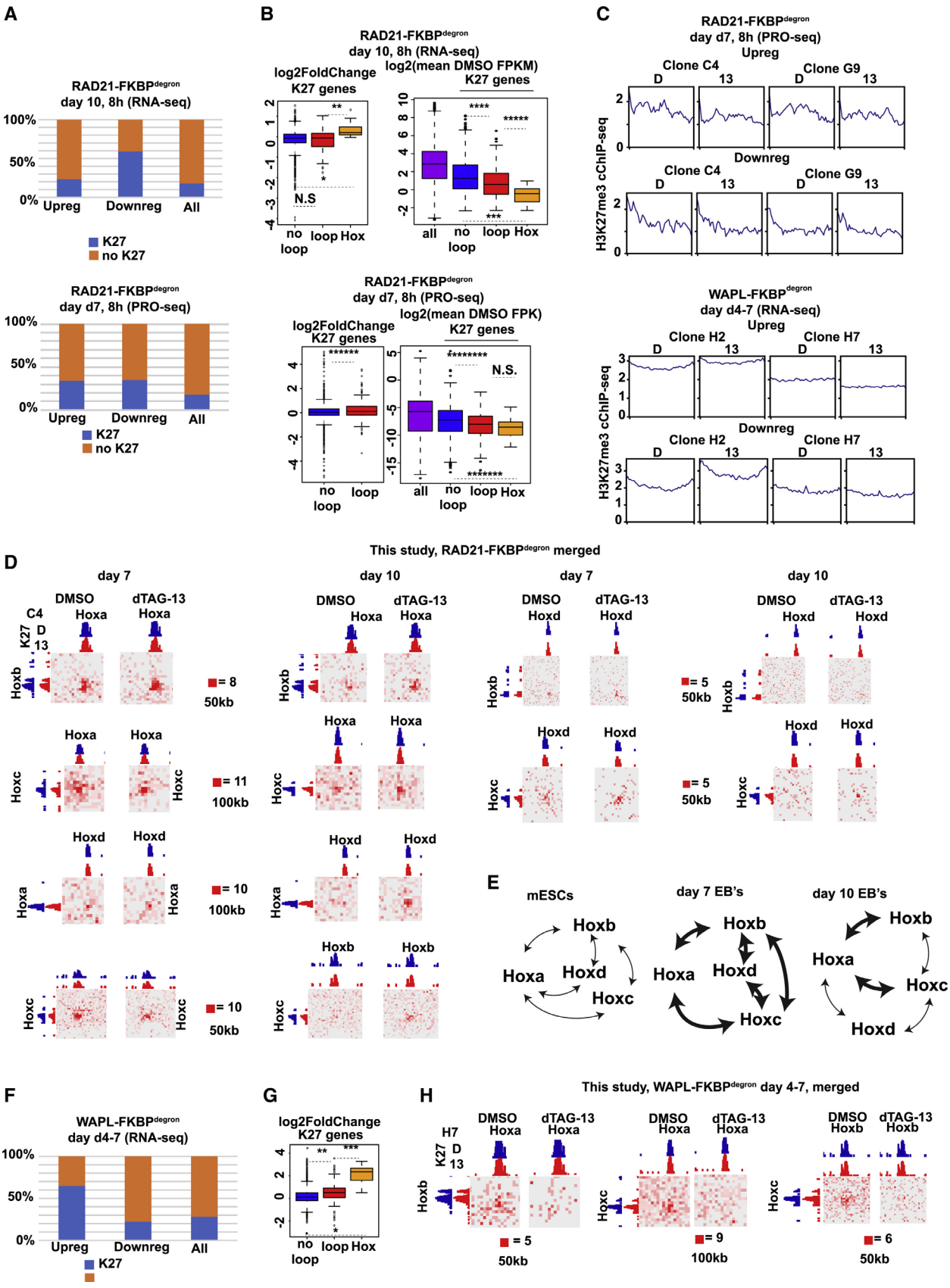

A
RAD21-FKBP^degron
day 10, 8h (RNA-seq)
100%
50%
0%
Upreg
Downreg
All
K27
no K27
RAD21-FKBP^degron
day d7, 8h (PRO-seq)
100%
50%
0%
Upreg
Downreg
All
K27
no K27
B
RAD21-FKBP^degron
day 10, 8h (RNA-seq)
log2FoldChange
K27 genes
N.S
no loop
loop
Hox
log2(mean DMSO FPKM)
K27 genes
all
no loop
loop
Hox
RAD21-FKBP^degron
day d7, 8h (PRO-seq)
log2FoldChange
K27 genes
no loop
loop
log2(mean DMSO FPK)
K27 genes
N.S.
all
no loop
loop
Hox
C
RAD21-FKBP^degron
day d7, 8h (PRO-seq)
Upreg
Clone C4
D
13
Clone G9
D
13
H3K27me3 cChIP-seq
Downreg
Clone C4
D
13
Clone G9
D
13
H3K27me3 cChIP-seq
WAPL-FKBP^degron
day d4-7 (RNA-seq)
Upreg
Clone H2
D
13
Clone H7
D
13
H3K27me3 cChIP-seq
Downreg
Clone H2
D
13
Clone H7
D
13
H3K27me3 cChIP-seq
D
This study, RAD21-FKBP^degron merged
day 7
day 10
day 7
day 10
DMSO
dTAG-13
C4
K27
D
13
Hoxa
Hoxd
Hoxb
Hoxc
= 8
50kb
= 11
100kb
= 10
100kb
= 10
50kb
= 5
50kb
= 5
50kb
E
mESCs
day 7 EB's
day 10 EB's
Hoxa
Hoxb
Hoxc
Hoxd
F
WAPL-FKBP^degron
day d4-7 (RNA-seq)
100%
50%
0%
Upreg
Downreg
All
K27
G
log2FoldChange
K27 genes
no loop
loop
Hox
H
This study, WAPL-FKBP^degron day 4-7, merged
H7
K27
D
13
DMSO
Hoxa
dTAG-13
Hoxa
DMSO
Hoxb
dTAG-13
Hoxb
Hoxb
Hoxc
= 5
50kb
= 9
100kb
= 6
50kb

Figure 4.2. Polycomb loops protect against gene dysregulation upon cohesin loss, but increase sensitivity to cohesin gain during differentiation.

(A) Genes upregulated or downregulated after indicated time courses of RAD21 degradation (relative to DMSO control) overlapping with H3K27me3 cChIP-seq peaks. All genes shown for reference.

(B) RAD21-FKBP^{degron} RNA-seq (day 7: PRO-seq) DESeq2 log2-fold (dTAG-13/DMSO) change (left) or log2 FPKM (day 7: FPK, right, averaged between clones) of different categories of expressed (day 7: all) H3K27me3-marked genes. (No) loop: (not) overlapping anchor of H3K27me3-associated, cohesin-independent loop, Hox: Hox gene located in Hox cluster. All genes shown for reference. Significant p-values by Wilcoxon rank test indicated by * (p<7.344e-05), ** (p<0.0006566), *** (p<2.847e-05), **** (p<0.0005113), ***** (p<0.003082), ****** (p<0.0006716), ******* (p<0.001102), ********* (p<0.000146). N.S., not significant.

(C) Average DMSO and dTAG-13 H3K27me3 cChIP-seq signal over upregulated and downregulated genes in day 7 RAD21-FKBP^{degron} and WAPL-FKBP^{degron} samples.

(D) Inter-chromosomal Hi-C interaction maps between Hox cluster regions in RAD21-FKBP^{degron} lines at day 7 and day 10 at indicated resolutions (clones merged). H3K27me3 cChIP-seq signal in day 7 RAD21- FKBP^{degron} lines shown for reference (scale: 200).

(E) Schematic of inter-chromosomal Hox interactions in mESCs, day 7 and day 10 of differentiation. Bold arrows, interactions detected by Hi-C under control conditions.

(F) Genes upregulated or downregulated upon WAPL degradation from days 7-10 (relative to DMSO control) overlapping with H3K27me3 cChIP-seq peaks. All genes shown for reference.

(G) WAPL-FKBP^{degron} RNA-seq log2-fold (dTAG-13/DMSO) change of different categories of expressed H3K27me3-marked genes in WAPL-FKBP^{degron} lines. Categories as in (B).

Significant p-values by Wilcoxon rank test indicated by * (p<6.703e-08), ** (p<1.462e-07), *** (p<4.638e-05).

(H) Inter-chromosomal Hi-C interaction maps between Hox cluster regions in WAPL-FKBP^{degron} lines (clones merged) at indicated resolutions. H3K27me3 cChIP-seq signal in day 7 WAPL- FKBP^{degron} lines shown for reference (scale: 100).

Next we asked how inappropriate cohesin retention affects these *Hox* dynamics. Because 8-hour WAPL depletion only modestly increased cohesin retention (Figure S2.4B-D), we depleted WAPL from days 4-7 of differentiation, where we observed much more cohesin retention (Figure 4.1E). RNA-seq revealed differential expression of 2386 genes, of which 981 were downregulated and 1405 were upregulated (adjusted p-value < 0.1). WAPL degradation at a later timepoint, days 7-10, resulted in fewer DEGs (39 upregulated and 19 downregulated, Figure S2.4E). We therefore focused further analysis on the day 4-7 timecourse. As might be expected, genes that are normally anchored by cohesin did not change expression patterns when cohesins were more strongly retained (Figure S2.4A). Intriguingly, upregulated DEGs tend to be Polycomb targets, particularly those in H3K27me3-marked loop anchors that persisted after RAD21 degradation (Figure 4.2F,G). This was especially evident at the *Hox* loci, where WAPL depletion and consequent cohesin retention led to a loss of trans-interaction networks and marked increase in gene expression (Figure 4.2G-H, S2.3J). Among the upregulated DEGs were also the *Zscan4* and *Usp17l* families (Figure S2.4F), markers of the two-cell embryonic development stage (Falco et al., 2007; Macfarlan et al., 2012).Thus, disruption of Polycomb interaction networks, both in cis and in trans, through inappropriate cohesin retention can have catastrophic effects on gene repression.

Taken together, our data indicate that cohesins play an important role in Polycomb regulation. As Polycomb domains have an inherent tendency to cluster (Bonev and Cavalli, 2016;

Denholtz et al., 2013; Joshi et al., 2015; Kundu et al., 2017; McLaughlin et al., 2019; Rhodes et al., 2020; Schoenfelder et al., 2015; Vieux-Rochas et al., 2015), we propose that a key role of cohesins is to organize chromatin in two respects: First, to prevent ectopic long-range cis-interactions as well as inappropriate trans-interactions; and second, to temper such interactions when they occur in the normal physiological context. Indeed, when cohesin is ablated, trans-interactions strengthen between chromosomes, as exemplified by inter-*Hox* interactions. When cohesins are aberrantly retained, such interactions are disrupted, with a dramatic increase in expression of certain Polycomb targets.

Long-term WAPL or CTCF degradation in ES cells induces expression of "2C-like" genes

We wondered whether the same Polycomb target genes were dysregulated in the WAPL and RAD21 degron cells. Although one might expect the DEGs to overlap but be dysregulated in opposite direction (e.g., down in RAD21-FKBP^{degron} cells, up in WAPL-FKBP^{degron} cells), there was little overlap between the two sets of dysregulated genes. In our day 7 datasets, 0 genes were DE in opposite directions in the RAD21 and WAPL degrons, while 8 genes were upregulated and 13 genes downregulated in both. On the other hand, this may not be surprising due to our WAPL degradation time course (3 days) being much longer than that of RAD21 (8 hours). We therefore wondered whether WAPL degradation would better phenocopy long term degradation of other architectural factors. Accordingly, we examined a published CTCF auxin-AID degron RNA-seq dataset, in which CTCF had been degraded for 4 days in mESCs (Nora et al., 2017). Interestingly, there was high overlap between genes DE in the same direction after WAPL or CTCF degradation (959 upregulated in both, 467 downregulated in both), with less of an overlap between genes DE in opposite directions (301 upregulated after CTCF degradation and downregulated after WAPL degradation, 74 downregulated after CTCF degradation and upregulated after WAPL degradation).

Among the genes strongly upregulated upon both WAPL and CTCF degradation were the *Zscan4* and *Usp17l* families (Figure S2.3E), markers of the two-cell embryonic development stage (Falco et al., 2007; Macfarlan et al., 2012). Subpopulations of mESCs in cell culture have been found to express such a "2C-like" gene signature, providing an explanation for why these genes could be expressed in the ES cell degron lines of us and others (Kolodziejczyk et al., 2015; Macfarlan et al., 2012). Notably, *Usp17l* is a family of de-ubiquitinating enzymes (Reyes-Turcu et al., 2009). As both dTAG and the auxin-AID degron system rely on tethering degron-tagged proteins to ubiquitin ligases, upregulation or selection of a subpopulation expressing these "2C-like" genes could reflect use of degrons in mESCs for extended time periods. Alternatively, a recent study implicated cohesin in regulation of 2-cell genes (Zhang et al., 2020), supporting that the upregulation we observe in the case of CTCF and WAPL degradation could be a consequence of toggling cohesin localization on chromatin. We emphasize that, although "2C-like" genes were strongly upregulated, their overall expression level remained relatively low and WAPL-FKBP^{degron} ES cells were still able to differentiate and efficiently degrade WAPL upon addition of dTAG-13 (Figure 4.1B). Therefore, the percent of "2C-like" cells in the population likely remained negligible.

Cohesin regulates genes expression associated with super-enhancers in a looping-independent manner

Our RAD21/WAPL degron system enabled us to revisit an important question about dependence of super-enhancers (SEs) on cohesin for which there is disagreement in the field (Rao et al., 2017; Rhodes et al., 2020). To examine proximity of DEGs to SEs, we restricted analysis to expressed genes (FPKM > 0.5), leveraged our PRO-seq data to focus on nascent transcription, and measured distances from SEs to all expressed genes, downregulated DEGs, or upregulated DEGs (Figure S2.5A). Significantly, downregulated DEGs in cohesin-depleted cells were closer to super-enhancers (Figure S2.5A-B). Approximately 54% of downregulated

DEGs were <150kb away, and ~29% were <20kb away (Figure S2.5B). Overall, loss of cohesin resulted in ~2-fold downregulation of these targets. To ask whether this could be due to an associated shift in looping interactions, we examined Hi-C maps for interactions between SEs. Meta-loop analysis using APA plots revealed a slight enrichment of aggregate signal over SE anchors, suggesting low-frequency contacts between SEs (Figure S2.5C). However, the contacts were not appreciably disrupted by cohesin loss in either day 7 or day 10 cells. These data contrast with those in day 0 ES cells, where SE interactions are cohesin-dependent (Rhodes et al., 2020). We conclude that cohesins play role in regulating SE-associated genes independent of looping in differentiating ES cells.

Xi superloops form between escapees in a cohesin- and Polycomb-independent manner

The inactive X chromosome (Xi) is a strong Polycomb target, becoming highly enriched with H3K27me3 in female ES cells undergoing differentiation and XCI (Plath et al., 2003; Silva et al., 2003; Wang et al., 2001; Zhao et al., 2008). XCI brings about major architectural changes to the X chromosome, as Xist RNA evicts up to 90% of cohesins while it spreads along the Xi (Minajigi et al., 2015b) (Figure 4.3A and 4.4A**,** Xi MEF row). Here we followed the Xi from mid (day 7) to late (day 10) XCI and indeed observed 40- 50% loss by days 7 to 10 (Figure 4.3B and 4.4B). Changes in cohesin binding during differentiation are known to correlate with (i) suppression of “TADs” across the Xi (Wang et al., 2018), (ii) appearance of two megadomains with a strong border at the macrosatellite repeat, *Dxz4,* (Deng et al., 2015; Froberg et al., 2018; Giorgetti et al., 2016; Minajigi et al., 2015b; Rao et al., 2014) and (iii) establishment of a ‘superloop’ spanning ~30 Mb between *Dxz4* and another repeat element, *Firre* (Darrow et al., 2016; Hacisuleyman et al., 2014; Horakova et al., 2012; Rao et al., 2014) (Figure 4.3A). The superloops are not present on the Xa. Current models postulate that superloops form through loop extrusion of cohesin (Bansal et al., 2019; Bonora et al., 2018), but cohesin’s exact role in

XCI is unclear. Does cohesin binding antagonize the establishment of the Xi? Is cohesin eviction required for XCI? Does the remaining cohesin binding have a role in folding the Xi?

Our ability to toggle cohesin levels using RAD21/WAPL degrons provided an opportunity to address these questions (Figures 4.3-4.4). Our *in situ* Hi-C was performed in a hybrid female ES line (*Tsix*$^{TST/+}$) carrying a *Mus musculus* (mus) and a *Mus castaneus* (cas) X chromosome, with the mus X chromosome bearing a *Tsix* mutation that ensures inactivation of the X-chromosome in cis (Ogawa et al., 2008).

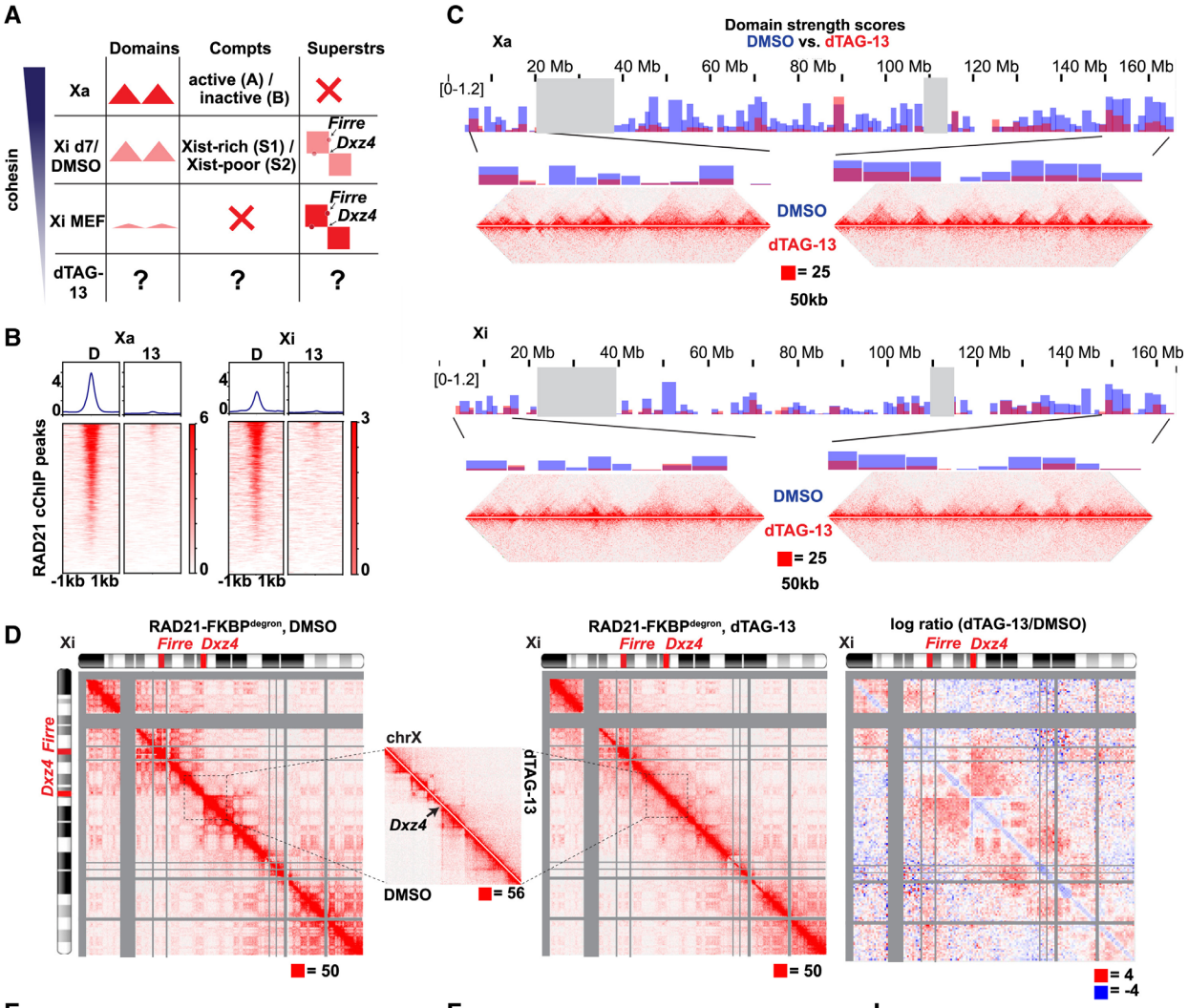

A
Domains
Compts
Superstrs
cohesin
Xa
active (A) / inactive (B)
Xi d7/ DMSO
Xist-rich (S1) / Xist-poor (S2)
Xi MEF
dTAG-13
Firre
Dxz4
B
Xa
Xi
D
13
RAD21 cChIP peaks
-1kb 1kb
C
Domain strength scores
DMSO vs. dTAG-13
Xa
Xi
[0-1.2]
20 Mb
40 Mb
60 Mb
80 Mb
100 Mb
120 Mb
140 Mb
160 Mb
DMSO
dTAG-13
= 25
50kb
D
RAD21-FKBP^degron, DMSO
RAD21-FKBP^degron, dTAG-13
log ratio (dTAG-13/DMSO)
Dxz4 Firre
chrX
= 50
= 56
= 4
= -4
E
C4
chrX:47.9-48.0Mb
chrX:72.9-73.1Mb
RAD21
K27
Firre
Dxz4
= 25
100kb
F
chrX:98.50-99.25Mb
chrX:148.5-148.75Mb
Ogt
Kdm5c
= 131
= 44
250kb
I
log2(Obs/Exp)
G
chrX:45.6-45.8Mb
chrX:11.5-11.8Mb
Utp14a
Bcorl1
Bcor
= 21
H
RAD21-FKBP^degron, PC1
20Mb
80Mb
140Mb
Xi, D
Xi, 13
Xa, D
Xa, 13
J
Cumulative Fraction
counts (norm. to length)
mus (Xi), D
mus (Xi), 13
cas (Xa), D
cas (Xa), 13
chr13
mus, D
mus, 13
cas, D
cas, 13

Figure 4.3. Cohesin is required for proper folding of the inactive X chromosome at day 7 of differentiation.

(A) Schematic of expected cohesin levels versus TAD, compartment, and superstructures detected on the X chromosome at different stages of XCI, based on prior studies. Levels corresponding to day 7 DMSO control and dTAG-13 treated cells indicated.

(B) Metaplot and heatmaps of active (Xa) and inactive (Xi) X chromosome RAD21 binding by cChIP-seq in day 7 RAD21-FKBP^{degron} embryoid bodies treated for 8 hours with 500nM dTAG-13 or DMSO.

(C) TAD scores calculated from day 7 Xa and Xi RAD21-FKBP^{degron} allelic Hi-C interaction maps binned at 100kb. DMSO and dTAG-13 samples as in (B). Representative Hi-C maps shown for the indicated regions (50kb resolution, clones merged).

(D) Chromosome-wide day 7 Xi RAD21-FKBP^{degron} allelic Hi-C maps (250kb resolution, clones merged) and log2-ratio map (1Mb resolution). Treatment as in (B). Inset, 50kb resolution.

(E) Day 7 composite X, Xi, and Xa Hi-C maps at *Dxz4-Firre* superloop (100kb resolution, clones merged). RAD21 and H3K27me3 cChIP-seq at superloop anchors shown for reference. Treatment as in (B).

(F) Day 7 composite X, Xi, and Xa Hi-C maps at *Kdm5c-Ogt* loop (250kb resolution, clones merged). RAD21 and H3K27me3 cChIP-seq at loop anchors shown for reference. Treatment as in (B).

(G) Day 7 composite X, Xi, and Xa Hi-C maps at *Bcor-Bcorl1* loop (100kb resolution, clones merged). RAD21 and H3K27me3 cChIP-seq at loop anchors shown for reference. Treatment as in (B).

(H) First Principal Component (PC1) tracks of day 7 Xa and Xi RAD21-FKBP^{degron} Hi-C maps. Treatment as in (B).

(I) Saddle plot analysis of compartment strength in RAD21-FKBP^{degron} Xi and Xa Hi-C maps (clones merged). Scores represent B-B (Xi: S2-S2), B-A (Xi: S2-S1), and A-A (Xi: S1-S1) compartment interaction strength. Treatment as in (B).

(J) Cumulative distribution plots of allelic chromosome X and chromosome 13 expression by PRO-seq in RAD21-FKBP^{degron} lines. Treatment as in (B).

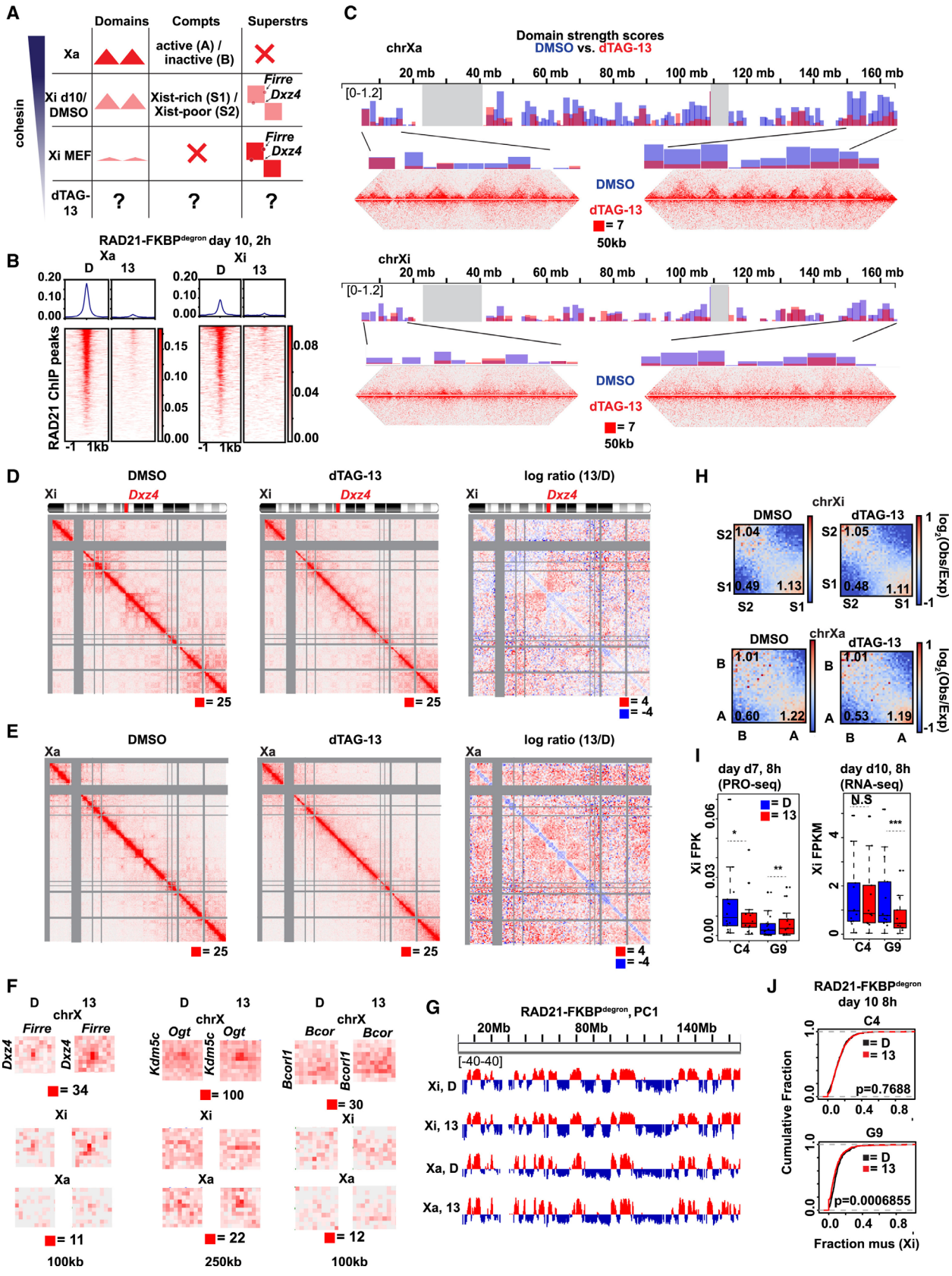

A
Domains
Compts
Superstrs
cohesin
Xa
active (A) / inactive (B)
Xi d10/ DMSO
Xist-rich (S1) / Xist-poor (S2)
Firre
Dxz4
Xi MEF
dTAG-13
B
RAD21-FKBP^degron day 10, 2h
Xa
Xi
D
13
RAD21 ChIP peaks
-1
1kb
C
Domain strength scores
DMSO vs. dTAG-13
chrXa
chrXi
[0-1.2]
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
140 mb
160 mb
DMSO
dTAG-13
= 7
50kb
D
DMSO
dTAG-13
log ratio (13/D)
Xi
Dxz4
= 25
= 4
= -4
E
Xa
F
chrX
Firre
Dxz4
= 34
= 11
100kb
Ogt
Kdm5c
= 100
= 22
250kb
Bcor
Bcorl1
= 30
= 12
100kb
G
RAD21-FKBP^degron, PC1
20Mb
80Mb
140Mb
[-40-40]
Xi, D
Xi, 13
Xa, D
Xa, 13
H
chrXi
chrXa
S2
S1
B
A
1.04
1.13
0.49
1.05
1.11
0.48
1.01
0.60
1.22
0.53
1.19
log2(Obs/Exp)
I
day d7, 8h (PRO-seq)
day d10, 8h (RNA-seq)
Xi FPK
Xi FPKM
N.S
C4
G9
J
RAD21-FKBP^degron day 10 8h
Cumulative Fraction
p=0.7688
p=0.0006855
Fraction mus (Xi)

Figure 4.4. Cohesin is required for proper folding of the inactive X chromosome at day 10 of differentiation

(A) Schematic of expected cohesin levels versus TAD, compartment, and superstructures detected on the X chromosome at different stages of XCI, based on prior studies. Levels corresponding to DMSO control and dTAG-13 treated cells indicated.

(B) Metaplot and heatmaps of active (Xa) and inactive (Xi) X chromosome RAD21 binding by uncalibrated ChIP-seq in day 10 RAD21-FKBP^{degron} embryoid bodies treated for 8 hours with 500nM dTAG-13 or DMSO.

(C) TAD scores calculated from day 10 Xa and Xi RAD21-FKBP^{degron} allelic Hi-C interaction maps binned at 100kb. DMSO and dTAG-13 samples as in (B). Representative Hi-C maps shown for the indicated regions (50kb resolution, clones merged).

(D) Chromosome-wide day 10 Xi RAD21-FKBP^{degron} Hi-C maps (250kb resolution, clones merged) and log2-ratio map (1Mb resolution). Treatment as in (B).

(E) Chromosome-wide day 10 Xa RAD21-FKBP^{degron} Hi-C maps (250kb resolution, clones merged) and log2-ratio map (1Mb resolution). Treatment as in (B).

(F) Day 7 composite X, Xi, and Xa Hi-C maps at *Dxz4-Firre* superloop (100kb resolution, clones merged), *Kdm5c-Ogt* loop (250kb resolution, clones merged) and *Bcor-Bcorl1* loop (100kb resolution, clones merged). RAD21 and H3K27me3 cChIP-seq at superloop anchors shown for reference. Treatment as in (B).

(G) First Principal Component (PC1) tracks of day 7 Xa and Xi RAD21-FKBP^{degron} Hi-C maps. Treatment as in (B).

(H) Saddle plot analysis of compartment strength in RAD21-FKBP^{degron} Xi and Xa Hi-C maps (clones merged). Scores represent B-B (Xi: S2-S2), B-A (Xi: S2-S1), and A-A (Xi: S1-S1) compartment interaction strength. Treatment as in (B).

(I) Xi escapee (including *Ogt, Bcor1, Bcorl1*) gene expression in RAD21-FKBP^{degron} lines. Significant p-values by Wilcoxon signed rank test indicated by * ($p<0.001663$), ** ($p<0.00232$), *** ($p<0.0002441$). N.S., not significant.

(J) Cumulative distribution plots of allelic chromosome X and chromosome 13 expression by RNA-seq in RAD21-FKBP^{degron} lines. Treatment as in (B).

We were therefore able to perform allele-specific analysis to examine the Xa and Xi separately. In day 7 control cells, whereas the Xa showed topological domains, the Xi showed weakened (but not ablated) local domains (Figure 4.3C, 4.4C) and megadomains (Figure 4.3D, 4.4D), as expected. When cohesin was depleted by dTAG-13 treatment, the topological domains were nearly abolished on both Xa and Xi (Figure 4.3C, 4.4C). The loss of interactions observed on the Xa was similar to autosomes (Figure 4.4E,4.1F). On the Xi, the megadomains and *Dxz4* border also disappeared in the absence of cohesin (Figure 4.3D, 4.4D). Thus, although cohesin becomes depleted on the Xi during the normal course of XCI, the remaining cohesin is required for folding into megadomains. Notably, *Dxz4* contains cohesin-binding sites (Figure 4.3E, RAD21 track) that normally persist on the Xi despite Xist-mediated cohesin repulsion (Bonora et al., 2018; Froberg et al., 2018; Giorgetti et al., 2016).

The *Firre* locus also normally retains cohesin binding on the Xi (Figure 4.3E, RAD21 track) (Bonora et al., 2018; Froberg et al., 2018; Giorgetti et al., 2016). It may thereby form a superloop with *Dxz4* through cohesin-mediated loop extrusion across a 30Mb (Bansal et al., 2019; Bonora et al., 2018). Intriguingly, our allelic Hi-C map revealed a cohesin-dependent ‘stripe’ extending from *Dxz4* (Figure 4.3D, inset). This stripe is indicative of a unidirectional loop extrusion (Vian et al., 2018b) whereby cohesin remains bound at the *Dxz4* anchor while extruding chromatin. However, this stripe extended away from *Firre*, implying that *Dxz4* does indeed anchor loops, but loop extrusion may not be the mechanism by which it engages *Firre*. Indeed, the *Dxz4-Firre*

superloop was not perturbed by cohesin degradation, but paradoxically became stronger (Figure 4.3E, 4.4F). Not only did the superloop strengthen on the Xi, it became visible on the Xa as well (Figure 4.3E, 4.4F). Thus, surprisingly, the *Dxz4-Firre* superloop is not cohesin-dependent and is in fact inhibited by cohesin.

Firre and *Dxz4* were depleted for H3K27me3 relative to the rest of the Xi (Figure 4.3E), arguing against Polycomb-dependent looping for these loci. We noted that *Firre* belongs to a group of "escapees", X-linked genes that are not subject to XCI, and surmised that escapee status might render genes resistant to cohesin depletion. Intriguingly, allelic Hi-C analysis revealed a clustering of multiple escapee or Polycomb-depleted genes, including *Kdm5c*, *Ogt*, *Bcor,* and *Bcorl1* (Figure 4.3F-G and 4.4F). All such loops were reinforced by RAD21 degradation. While some escapees were previously observed to associate using 4C (Splinter et al., 2011), our high resolution allelic Hi-C maps enabled us to identify these superloops on a chromosome-wide basis and demonstrate that cohesin antagonizes their formation. No replicable changes to escapee expression was detected by PRO-seq at day 7 (8 hours degradation). A trend for decreased escapee expression could be detected by RNA-seq at day 10, but was only significant in one replicate (Figure 4.4I)**.** We conclude that escapees form long-range loops through mechanisms other than cohesin-mediated loop extrusion and Polycomb-based co-clustering, and may instead self-associate via active chromatin. We propose that escapee clustering may aid in their persistent expression on the Xi, independently of cohesin and Polycomb.

Transitory S1/S2 compartments on the Xi do not require cohesins

We next investigated effects on Xa and Xi compartments. Prior to XCI, the two X-chromosomes are organized into A/B compartments, much as any autosome. At the onset of XCI, Xist RNA — together with PRC1 that it recruits — fuses the A/B compartments into larger intermediate structures called S1 and S2 compartments (Figure 4.3A**,** middle column) (Wang et

al., 2019). Xist then recruits the non-canonical SMC protein, SMCHD1, to merge S1 and S2 compartments into a unique Xi-specific superstructure in which compartments are not visible (Wang et al., 2018). We asked how the transition from A/B to S1/S2 to the Xi superstructure is affected by cohesin perturbations. At days 7 and 10, S1/S2 compartments were visible on the control Xi as “checkerboard” patterns, and also as alternating +/- segments in the 1st Principal Component of the Hi-C map (Figure 4.3D,H,4.4D,G). Classic A/B compartments were evident on the Xa (Figure 4.3D,H,4.4E,G). After 8 hours of cohesin degradation, these patterns were unchanged on both Xa and Xi (Figure 4.3I, 4.4H). There were also no significant effects on Xi silencing at day 7 or 10 (Figure 4.3J, 4.4J). While these findings are in line with the fact that up to 90% of cohesins are normally evicted from the Xi during XCI (Minajigi et al., 2015b), they were somewhat unanticipated, given how much residual cohesin binding contributes to the Xi superstructure. Thus, like some Polycomb targets on autosomes, Xi superstructure and transcriptomics appear to be resistant to cohesin loss. We surmise that the Xi’s enrichment for Polycomb marks at least partially underlies the resistance.

Forced cohesin retention disrupts the 3D structure of Xi and creates megadomains on Xa

What would happen to the Xi if we forced cohesin retention? In our WAPL-FKBP^{degron} lines, 8 hours of WAPL depletion only succeeded in retaining cohesins at the 30Mb centromeric and distal telomeric ends of the Xi (Figure S2.4C). We therefore focused analysis on 3 days of WAPL depletion (days 4-7), when cohesins became enriched 2-fold across the Xi (Figure 4.5B). A 2-fold enrichment also occurred on the Xa. Interestingly, cohesin levels did not become equalized between the Xi and the Xa (Figure 4.5B). This suggests that recruitment of WAPL to the Xi may not be the sole means of cohesin eviction during XCI. It is also possible that degradation from days 4-7 is too late to fully restore cohesin levels. Nevertheless, the increased cohesin levels on the Xi enabled us to examine consequences for Xi architecture and gene expression.

Intriguingly, forced cohesin retention did not substantially change strength of domains on the Xi (so-called “TADs”; Figure 4.5C) — contrasting with the marked attenuation seen when cohesins were depleted (Figure 4.3C,4.4C). Furthermore, the *Dxz4* megadomain border remained strong and the architectural stripe anchored at *Dxz4* appeared to lengthen (Figure 4.5D), consistent with cohesin retention and a unidirectional loop extrusion originating at *Dxz4*. The preservation of shorter-range looping interactions (Figure 4.5C) occurred at the expense of longer-range intra-megadomain interactions (Figure 4.5D, right panel – subtractive map). Longer-range interactions were also affected on the Xa (Figure 4.5E), in line with increased cohesin retention on the Xa. However, PC1 analysis showed that cohesin retention did not prevent the appearance of transitory S1/S2 compartment structures except for disruptions at the distal centromeric end of the Xi (Figure 4.5F, arrow). Saddle plot analysis (Schwarzer et al., 2017) confirmed changes in compartments, with S1-S1 interactions weakening and S1-S2 interactions strengthening (Figure 4.5G), which could potentially explain the disruptions to S1/S2 compartments. Finally, no effects on the *Dxz4-Firre* superloop was observed (Figure 4.5H), further enforcing the idea that the superloop is cohesin-independent. Thus, forced cohesin retention resulted in failure to attenuate local looping interactions and disrupted the Xi superstructure without affecting superloops.

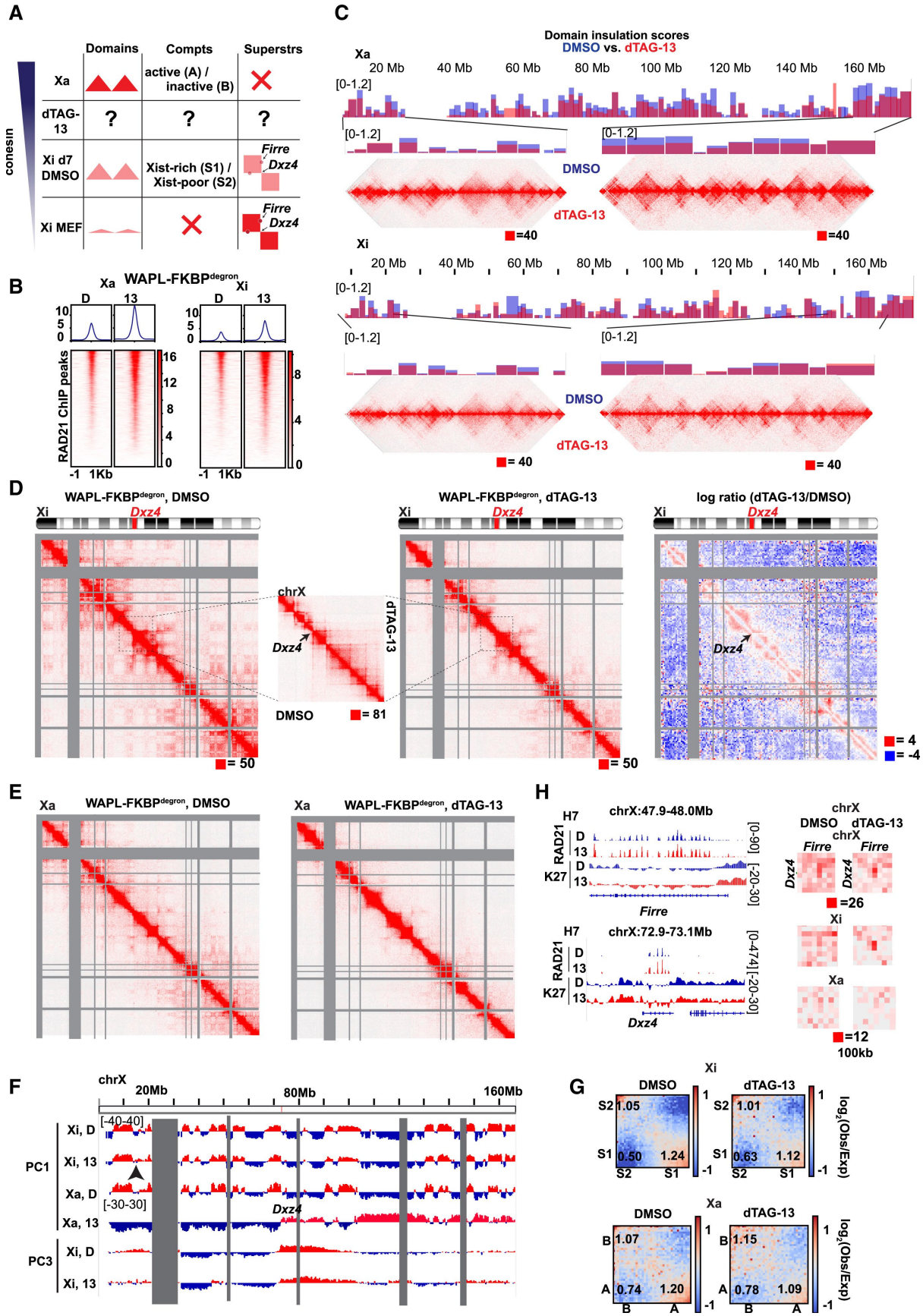

A
Domains
Compts
Superstrs
Xa
active (A) / inactive (B)
dTAG-13
Xi d7 DMSO
Xist-rich (S1) / Xist-poor (S2)
Firre
Dxz4
Xi MEF
consein
B
WAPL-FKBP^degron
Xa
Xi
D
13
RAD21 ChIP peaks
-1
1Kb
C
Domain insulation scores
DMSO vs. dTAG-13
20 Mb
40 Mb
60 Mb
80 Mb
100 Mb
120 Mb
140 Mb
160 Mb
[0-1.2]
DMSO
dTAG-13
=40
= 40
D
WAPL-FKBP^degron, DMSO
WAPL-FKBP^degron, dTAG-13
log ratio (dTAG-13/DMSO)
chrX
Dxz4
= 81
= 50
= 4
= -4
E
H
H7
chrX:47.9-48.0Mb
chrX:72.9-73.1Mb
RAD21
K27
[0-90]
[-20-30]
[0-47.4]
=26
=12
100kb
F
20Mb
80Mb
160Mb
[-40-40]
[-30-30]
PC1
PC3
Xi, D
Xi, 13
Xa, D
Xa, 13
G
S2
S1
B
A
1.05
0.50
1.24
1.01
0.63
1.12
1.07
0.74
1.20
1.15
0.78
1.09
log2(Obs/Exp)

Figure 4.5. Cohesin retention results in increase of interactions spanning TAD borders and weakening of megadomains and compartments on the Xi.

(A) Schematic of expected cohesin levels versus TAD, compartment, and superstructures detected on the X chromosome at different stages of XCI, based on prior studies. Levels corresponding to day 7 DMSO control and dTAG-13 treated cells indicated.

(B) Metaplot and heatmaps of active (Xa) and inactive (Xi) X chromosome RAD21 binding by cChIP-seq in day 7 WAPL-FKBP^{degron} embryoid bodies treated for 3 days with 500nM dTAG-13 or DMSO.

(C) TAD scores calculated from day 7 Xa and Xi WAPL-FKBP^{degron} allelic Hi-C interaction maps binned at 100kb. DMSO and dTAG-13 samples as in (B). Representative Hi-C maps shown for the indicated regions (50kb resolution, clones merged).

(D) Chromosome-wide day 7 Xi WAPLFKBP^{degron} allelic Hi-C maps (250kb resolution, clones merged) and log2-ratio map (1Mb resolution). Treatment as in (B). Inset, 50kb resolution.

(E) Chromosome-wide day 7 Xa WAPL-FKBP^{degron} allelic Hi-C maps (250kb resolution, clones merged). Treatment as in (B).

(F) First and third Principal Component (PC1 and PC3) tracks of day 7 Xa and Xi WAPL-FKBP^{degron} Hi-C maps. Treatment as in (B).

(G) Saddle plot analysis of compartment strength in WAPL-FKBP^{degron} Xi and Xa Hi-C maps (clones merged). Scores represent B-B (Xi: S2-S2), B-A (Xi: S2-S1), and A-A (Xi: S1-S1) compartment interaction strength. Treatment as in (B).

(H) Zoomed-in day 7 composite X, Xi, and Xa Hi-C maps at *Dxz4-Firre* superloop (100kb resolution, clones merged). RAD21 and H3K27me3 cChIP-seq at superloop anchors shown for reference. Treatment as in (B).

Findings for the Xa were equally intriguing. The Xa also enriched for cohesins by 2-fold after WAPL degradation (Figure 4.5B). While this retention did not affect topological domains (Figure 4.5C), A/B compartments could not be detected in PC1 of the Xa Hi-C map (Figure 4.5F). Instead, Xa adopted a megadomain-like pattern, which is typically only seen for the Xi (Figure 4.5F, PC3). Saddle plot analysis showed weakened A-A interactions and strengthened A-B interactions (Figure 3.5G), which may partially account for the sudden change in the overall Xa structure. Thus, unexpectedly, megadomain organization can emerge independently of XCI.

Failure of cohesin eviction compromises Xist spreading and Xi gene silencing

Previous work demonstrated that merging S1/S2 compartments is critical for Xist to spread and silence genes on the Xi (Wang et al., 2018). Given the changes in compartment structure (Figure 4.5), we asked if forced cohesin retention affected Xist spreading and Xi silencing. We performed CHART-seq to map genomic binding sites for Xist RNA in WAPL-depleted cells (dTAG-13 treated between days 4-7). Xist RNA still bound the Xi despite retained cohesins (Figure 4.6A and S1.6A). However, the localization appeared to be compromised, as there was a 2-fold drop in Xist coverage at the centromeric and telomeric ends (Figure 4.6A,S2.6A**,** Xist ∆ track). Notably, these regions are furthest from the *Xist* locus. In the Hi-C contact map, these were also the regions that tend to lose contact with *Xist* (Figure 4.6A,S2.6A**,** top track). Statistical analysis confirmed Xist RNA depletion across the entire chromosome, with greatest effect size at the 30 Mb centromeric and telomeric ends (Figure 4.6B, S2.6B). Concomitantly, H3K27me3 levels decreased up to 4-fold at the Xi ends (Figure 4.6A,S2.6A**,** H3K27me3 tracks). Statistical analysis confirmed this decrease (Figure 4.6C,S2.6C). These effect sizes are Xi-specific in both biological replicates. While Xist and H3K27me3 coverages decreased significantly in both S1 and S2 compartments, the S1 compartment near the chromosomal ends were most affected (Figure 4.6A,S2.6A, yellow highlights) and the effect size was largest in S1 compartments overall (Figure S1.6F-G). These

data support the notion that weakened compartmentalization negatively impacts Xist and H3K27me3 spreading.

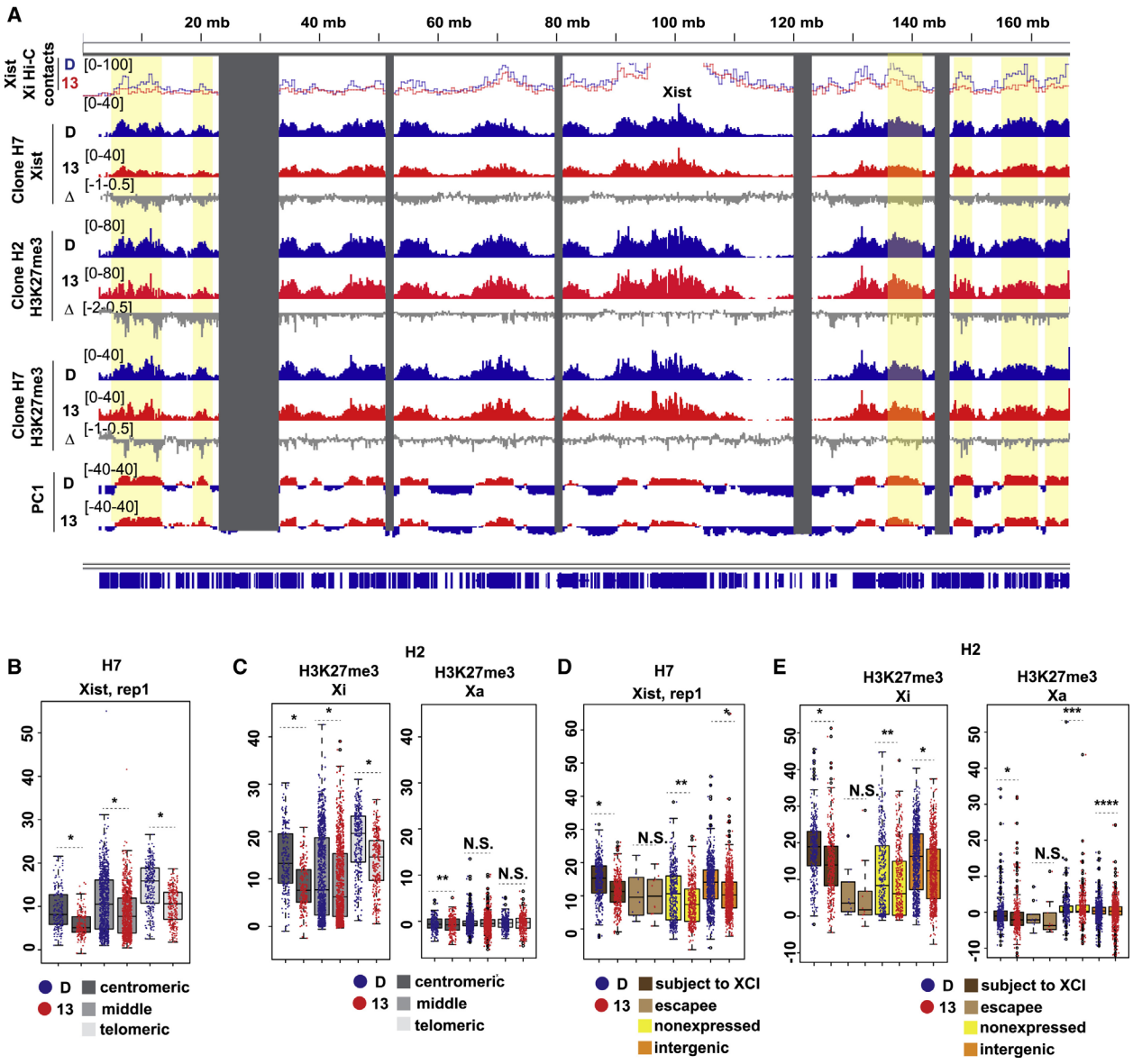


Figure 4.6. Retaining cohesin on the Xi early in differentiation leads to XCI defects, especially at the ends of the chromosome.

(A) Tracks of Xi Hi-C contacts to the Xist transcription locus, Xist binding (CHART-seq) and H3K27me3 (cChIP-seq) in WAPL-FKBP^{degron} lines treated from days 4-7 of differentiation with DMSO (D) or 500nM dTAG-13 (13). Log fold ratio tracks (dTAG-13/DMSO) are also shown (Δ). Principal component corresponding to compartments on the Xi (PC1, positive S1, negative S2) shown for reference.

(B) Boxplot showing Xist coverage over 100kb bins in the centromeric most 30Mb, middle 100Mb, and telomeric most 30Mb of the X chromosome in WAPL-FKBP^{degron}, clone H7 treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * ($p<2.2e-16$).

(C) Boxplots showing allelic H3K27me3 coverage over 100kb bins in the centromeric most 30Mb, middle 100Mb, and telomeric most 30Mb of the X chromosome in WAPL-FKBP^{degron}, clone H2 treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * ($p<2.2e-16$) and ** ($p<0.001187$). N.S., not significant.

(D) Boxplot showing average Xist density over different classes of genes on the X chromosome in WAPL-FKBP^{degron}, clone H7 treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * ($p<2.2e-16$). N.S., not significant.

(E) Boxplot showing average allelic H3K27me3 density over different classes of genes on the X chromosome in WAPL-FKBP^{degron}, clone H2 treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * ($p<2.2e-16$), ** ($p<8.094e-16$), *** ($p<0.00206$), **** ($p< 0.0001193$). N.S., not significant.

The increased number of longer chromatin loops (Figure 3.1K,M) could also play a role in weakened Xist and PRC2 spreading. As looping contacts mainly increased in the 1-10Mb range (Figure 3.1K), we examined how spreading was affected within 10Mb around *Xist*. While Xist coverage decreased within 10Mb, H3K27me3 was not replicably changed (Figure S1.6H).

Therefore, while additional chromatin loops may inhibit Xist spreading, this does not seem to be sufficient to impede H3K27me3 deposition at short ranges.

Finally, we investigated effects on Xi gene silencing. Xist and H3K27me3 coverages were significantly decreased over genes subject to XCI (Figure 3.6D-E,S1.6D-E). They were also reduced over intergenic regions, but not over escapees which normally lacked Xist and H3K27me3. We performed allele-specific RNA-seq analysis in WAPL-depleted cells (dTAG-13 treated days 4-7) and observed a significant cumulative shift in expression from the Xi (Figure 3.7A) — indicating either a failure of inactivation during the establishment phase, or an initial inactivation and then a partial reactivation in later stages. Consistent with compromised spreading of the Xist-PRC2 complex, failure of silencing was greater at the centromeric and telomeric 30Mb ends (Figure 3.7B). Examination of specific loci revealed strong (>1.7-fold) upregulation from various genes, including *Clcn5*, *Apex2*, *Tspan7*, *Spin2c*, and *Ap1s2* (Figure 3.7C) — all of where were located at the chromosome ends. We conclude that aberrant cohesin retention disrupts the ability of the Xist-PRC2 complex to spread at long range to the ends of the Xi, with consequences for gene silencing.

DISCUSSION

Large and finer scale 3D genome structures have long been hypothesized to play a pivotal role in gene regulation (de Wit et al., 2015; Guo et al., 2015; Merkenschlager and Nora, 2016). However, recent degron studies targeting architectural factors and their genomic binding sites have challenged this model. Here, we revisited the genomic structure-function relationship through the lens of XCI and demonstrate that a fine balance between cohesin eviction and retention is necessary to shape the Xi. Additionally, this balance is required to protect against inappropriate inter-chromosomal contacts and to temper Polycomb-mediated gene repression on autosomes.

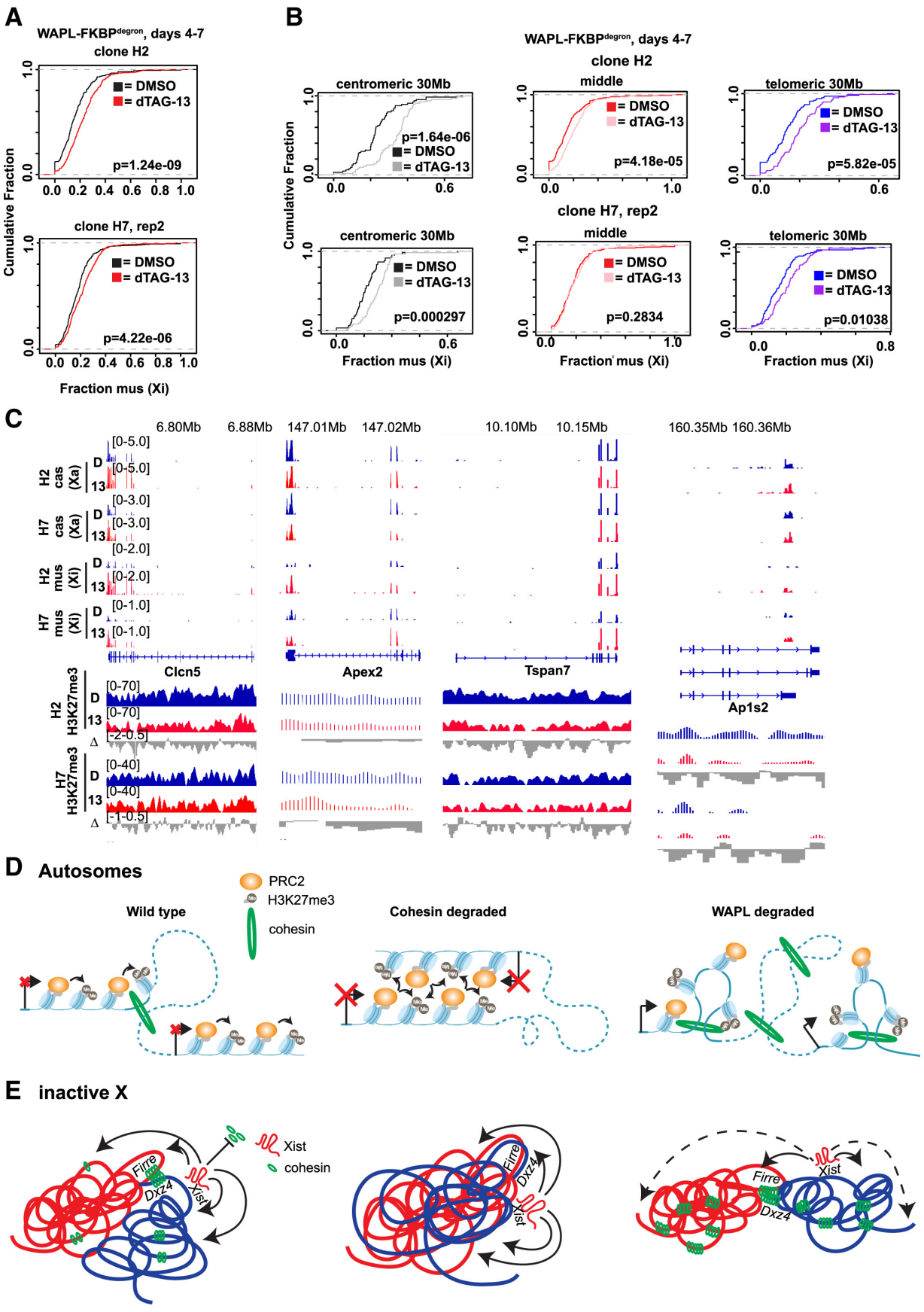

A
WAPL-FKBP^degron, days 4-7
clone H2
= DMSO
= dTAG-13
p=1.24e-09
clone H7, rep2
p=4.22e-06
Cumulative Fraction
Fraction mus (Xi)
B
WAPL-FKBP^degron, days 4-7
clone H2
centromeric 30Mb
middle
telomeric 30Mb
p=1.64e-06
p=4.18e-05
p=5.82e-05
clone H7, rep2
p=0.000297
p=0.2834
p=0.01038
C
6.80Mb
6.88Mb
147.01Mb
147.02Mb
10.10Mb
10.15Mb
160.35Mb
160.36Mb
H2 cas (Xa)
H7 cas (Xa)
H2 mus (Xi)
H7 mus (Xi)
Clcn5
Apex2
Tspan7
Ap1s2
H3K27me3
D Autosomes
PRC2
H3K27me3
cohesin
Wild type
Cohesin degraded
WAPL degraded
E inactive X
Xist
cohesin
Firre
Dxz4

Figure 3.7. Selective cohesin eviction and retention are required to maintain Polycomb targets, fold the Xi, and properly silence Xi genes.

(A) Cumulative distribution plots of chromosome-wide Xi expression in 2 WAPL-FKBP^{degron} clones treated from days 4-7 of differentiation with DMSO or 500nM dTAG-13. P-value calculated by the KS test.

(B) Cumulative distribution plots of Xi expression from the centromeric most 30Mb, middle 100Mb, and telomeric most 30Mb of the chromosome in 2 in 2 WAPL-FKBP^{degron} clones treated as in (A). P-value calculated by the KS test.

(C) Examples of genes at the far centromeric (Clcn5,Tspan7) and telomeric (Apex2,Ap1s2) ends of the X chromosome. Allelic RNA-seq and H3K27me3 cChIP-seq shown for 2 WAPL-FKBP^{degron} clones treated as in (A).

(D) Model: cohesin tempering of Polycomb targets on autosomes.

(E) Model: cohesin tempering of Xist and Polycomb spreading on the Xi.

While it is now established that Xist RNA evicts cohesins and the Xi is generally depleted of architectural factors, it was previously not known whether some cohesins must be retained for proper architectural transformation into a silent chromosome. By toggling cohesin levels at will, we demonstrate that XCI requires that cohesins be evicted from some regions and retained on others. Depleting cohesin ablates the sharp megadomain border, whereas forcing cohesin retention does not compromise the sharp border but reduces intra-megadomain interactions. Likewise, loop domains or “TADs” on the Xi require cohesin. When cohesin is depleted, these structures disappear almost entirely. We also demonstrate that Xi compartments — specifically the S1/S2 compartments — are weakened by stabilization of cohesin. Superloops, such as the those between *Dxz4-Firre* and between escapee *Kdm5c* and Polycomb-depleted gene *Ogt*, are also antagonized by cohesins and strengthened when cohesins are degraded. Thus, selective

regions on the Xi indeed must retain cohesin binding in order to construct the Xi superstructure. An interesting question for future consideration is how Xist evicts one set of cohesins while retaining others on the Xi.

The present work also reveals that Xist and Polycomb spreading and gene silencing all depend on cohesin eviction across the Xi. Forced cohesin retention results in reduced spreading of Xist RNA and H3K27me3 across the whole Xi, especially at the distal ends. Gene silencing was also blunted to varying degrees along the whole chromosome, with the strongest effects again at the Xist-distal ends. Expression of escapee genes, by contrast, is not affected by cohesin retention. However, cohesin depletion led to a surprising reinforcement of clustering of various escapees. The *Dxz4-Firre* superloop also strengthened upon cohesin depletion. It is possible that the depletion of cohesin that normally occurs on the Xi reinforces clustering of escapees to ensure their continued expression. Intriguingly, a recent study implicated *Firre* RNA as a regulator of the Xi and autosomes *in trans* (Fang et al., 2020). Future work will be needed to determine if interactions of the *Firre* genomic locus could be involved as well.

Our data show that cohesin imbalances also affect the Xa, causing the Xa to adopt some Xi-like structures. Indeed, a megadomain-like pattern appears mysteriously on the Xa when cohesins are aberrantly enriched, suggesting that enhanced local contacts may spontaneously drive partitioning of an X-chromosome into two large domains. Importantly, this means that megadomain organization can emerge independently of Xist and XCI. A *Dxz4-Firre* interaction also appeared on the Xa after cohesin degradation, also suggesting that the superloop could be a consequence of cohesin depletion rather than XCI.

Finally, our study revealed unexpected effects on autosomes as well. For instance, cohesins antagonize interactions between Polycomb targets and temper Polycomb-mediated repression (Rhodes et al., 2020). Moreover, the effects on Polycomb domains are not limited in cis, as loss of cohesin strengthens clustering of Polycomb targets *in trans,* as exemplified by

trans-interactions among the four *Hox* gene clusters. Inter-chromosomal bridges between the *Hox* loci are especially prominent during differentiation, in contrast to mESCs, where such interactions are not detected in control conditions. When cohesins are aberrantly retained, such interactions are disrupted and expression of the Polycomb targets are dramatically upregulated.

Altogether, the relationship between cohesin and gene expression on autosomes is analogous to that observed on the Xi (Figure 3.7D-E). On autosomes under wild type conditions, current models postulate that cohesin translocates along chromatin during the process of loop extrusion, passing through Polycomb domains (Figure 3.7D, left). Polycomb targets participating in interactions with each other tend to be highly repressed and are therefore indifferent to the presence of cohesin. However, certain Polycomb targets which are not engaged in interactions can become further repressed upon loss of cohesin. As Polycomb-anchored loops appear to be reinforced through cohesin loss, it is possible that this occurs due to increased interactions between previously non-interacting regions (Figure 3.7D, middle). As the Xi as a whole is highly repressed and covered with Polycomb that forms self-associating compartments, it is logical that Xi expression would also be indifferent to the loss of cohesin, similar to interacting Polycomb targets on autosomes (Figure 3.7E, middle). When WAPL is degraded, cohesin levels increase genome-wide, weakening interactions between Polycomb domains. Such disruptions lead to upregulation of Polycomb targets previously engaged in such interactions (Figure 3.7D, right). We also observe upregulation across the Xi. In contrast to autosomes, repression on the Xi is mediated by Xist (Figure 3.7E, left). Therefore, we observe the largest impact on gene silencing at the distal ends of the chromosome furthest from Xist, where the most Hi-C contacts are lost as well (Figure 3.7E, right). In conclusion, we propose that a key role of cohesins is to organize local chromatin and prevent ectopic long-range cis- and trans-interactions. In doing so, cohesins tempers repression of Polycomb targets as well as X chromosome inactivation.

ACKNOWLEDGEMENTS

We thank J. Froberg and C.-Y. Wang for critical reading of the manuscript, and we thank all lab members for support. We thank N. Gray for support and the generous gift of dTAG-13. The authors would like to thank the Nascent Transcriptomics Core at Harvard Medical School, Boston, MA for performing PRO-seq library construction. Nucleosomes in Figure 3.7 and Graphical Abstract created with BioRender.com. A.J.K. is supported by 1F31HD100109-01 and J.T.L. by grants from the National Institutes of Health (9R01HD097665-09) and the Howard Hughes Medical Institute. B.N. is supported by an American Cancer Society Postdoctoral Fellowship (PF-17-010-01-CDD) and the Katherine L. and Steven C. Pinard Research Fund.

AUTHOR CONTRIBUTIONS

A.J.K. and J.T.L. conceived the project, analyzed data, and wrote the paper. A.J.K. performed all bioinformatics analyses. D.C. and H.S. performed ChIP-seq in parental TST mESCs. A.J.K. performed all other experiments. B.N. generated dTAG-13 and provided advice in generation of $FKBP^{degron}$ lines.

DECLARATION OF INTERESTS

J.T.L. is a cofounder of Translate Bio and Fulcrum Therapeutics, and is also an advisor to Skyhawk Therapeutics. B.N. is an inventor on patent applications related to the dTAG system (WO/2017/024318, WO/2017/024319, WO/2018/148440 and WO/2018/148443).

STAR★Methods

RESOURCE AVAILABILITY

Contact for Reagent and Resource Sharing

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Jeannie T. Lee (lee@molbio.mgh.harvard.edu).

Materials Availability

Requests for materials generated in this study should be directed to and will be fulfilled by the Lead Contact upon completion of a Material Transfer Agreement.

Data and Code Availability

Original unprocessed blot and microscope images in this manuscript have been deposited at Mendeley Data and are available at: http://dx.doi.org/10.17632/9b83hvkgm5.1

Raw high-throughput sequencing data and processed files for RNA-seq, PRO-seq, cChIP-seq, CHART-seq, and Hi-C reported in this paper have been deposited at GEO under accession number: GSE144834

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

RAD21-FKBP^{degron} and WAPL-FKBP^{degron} mESCs were generated by transfecting *Tsix*$^{TST/+}$ mESCs with (1) a plasmid (pSpCas9-(BB) 2A Puro) expressing Cas9 with an sgRNA targeted to the STOP codon (in the case of Rad21) or to a region in the last codon near the STOP codon (in the case of Wapl) (2) a homology repair template plasmid including 900bp homology arms around

the STOP codon, the FKBP^{degron} tag in a pUC19 backbone and (3) a homology-directed repair plasmid (pMB 1609-pRR-EGFP) with the sgRNA target site cloned in. Specifically, 500,000 TST mESCs were transfected using Lipofectamine 3000 1.5ug of pSpCas9-(BB) 2A Puro, 1.5ug of the non-linearized repair plasmid, and 400ng of pMB 1609-pRR-EGFP. Transfected cells were GFP sorted after 48 hours and the entire GFP positive pool plated. After 5 days, the pool was split and plated at low density, and after 5 more days, clones were picked from the low-density plate in a 96-well plate. Clones were grown to confluency, then triplicate plated in 96-well plates. One 96-well plate was used to PCR screen clones. Clones were then thawed from one of the triplicate plates, expanded, and re-screened by PCR and Sanger sequenced to confirm correct knock-in of the FKBP^{degron} tag.

METHOD DETAILS

ES cell differentiation and induced degradation

Undifferentiated mESCs were grown on γ-irradiated MEF feeders in medium containing DMEM, high glucose, GlutaMAX supplement, pyruvate, 15% Hyclone FBS (Sigma), 25 mM HEPES pH 7.2-7.5, 1x MEM non-essential amino acids, 1x Pen/Strep, 0.1 mM βME, and 500 U/mL ESGRO recombinant mouse Leukemia Inhibitory Factor (LIF) protein (Sigma, ESG1107) at 37°C with 5% CO2. To induce differentiation in EBs, mESCs were trypsinized and resuspended in the same media without LIF. For the first 3 days of differentiation, EBs were grown in suspension. On day 4 and subsequent days, EBs were grown on gelatinized plates until harvested at day 7 or day 10.

Degradation of RAD21 or WAPL in FKBP^{degron} lines during differentiation was induced by adding a 1:20,000 dilution of 10mM dTAG-13 in DMSO to cell culture media (500nM final concentration) for the indicated time frames. The small molecule dTAG-13 was synthesized as previously

described (Erb et al., 2017; Nabet et al., 2018). For DMSO vehicle controls, the same volume of DMSO alone was added to cell culture media.

Western blot

Cells were counted and ~100,000 cells per sample lysed in 100ul 1X Laemmli buffer. Protein extracts were boiled for 5 min at 95C. Equal cell counts (5-15ul) were loaded and resolved by SDS page. Protein was transferred to an Immobilon-P PVDF membrane (EMD Millipore). After transfer, membrane was blocked with 5% milk in PBST (PBS/0.05% Tween-20) for 30 min at room temperature. Membrane was incubated with primary antibody (1:1000 for RAD21 and WAPL, 1:20,000 for GAPDH) in 5% milk in PBST at 4C overnight, then washed three times with PBST. Membrane was then blocked with secondary antibody-HRP conjugate (1:5000) in 5% milk in PBST for 1 hour at room temperature. Membrane was incubated with Western Lightning Plus-ECL (PerkinElmer) and exposed using the Bio-Rad ChemiDoc MP Imaging System.

Immunofluorescence

Immunofluorescence was performed as described previously with minor modifications (Del Rosario et al., 2019). Briefly, cells on glass slides were first permeabilized 2 min on ice with CSKT (100 mM NaCl, 300 mM Sucrose, 10 mM PIPES pH 6.8, 3 mM MgCl2, 0.5% Triton X-100) and then fixed in PBS 4% paraformaldehyde for 10 min at room temperature followed by a PBS wash step. Fixed cells were blocked (PBS, 0.2% Tween-20, 4% bovine serum albumin (BSA), 2% normal goat serum (NGS) (Sigma)) for 30 min at room temperature and incubated with primary antibody in a humid chamber at 4C overnight. Antibodies were diluted in blocking buffer. Samples were washed with PBS300.2T (PBS supplemented to 300 mM NaCl, 0.2% Tween-20) and incubated with secondary antibody for 1 hour at room temperature. Slides were mounted using

Vectashield with DAPI (Vector Labs). Slides were imaged with a Nikon Eclipse 90i microscope with Volocity software (Perkin Elmer).

Antibodies

For ChIP-seq the following antibodies were used: H3K27me3 (GeneTex GTX60892), RAD21 (Abcam ab992). For Western blot: RAD21 (Millipore 05-908), WAPL (Abcam ab70741), GAPDH (CST 5174T). For immunofluorescence: HA (Roche 11867423001).

ChIP-seq and calibrated ChIP-seq

ChIP-seq was performed as described previously with minor modifications (Del Rosario et al., 2019). Briefly, cells were trypsinized, resuspended in 1 million cells/mL of media and crosslinked with 1% formaldehyde for 10 min at room temperature with rotation. Cells were then quenched with 0.2 M glycine for 5 min. Crosslinked cells were washed once with cold PBS, pelleted, and snap frozen in an EtOH/dry ice bath. Cell pellets were stored at -80C until further processing. 5-10 million cross-linked cells per IP were thawed on ice and resuspended in 300ul Lysis Buffer 1 (25mM HEPES pH 7.2-7.5, 25mM EDTA, 0.5% SDS). Nuclei were sonicated (Qsonica Q800 Sonicator) in polystyrene tubes at 40% power setting, 30s ON/30s OFF for a total sonication (ON) time of 6 min. After sonication, 1.2 mL of Lysis Buffer 2 (25mM HEPES pH 7.2-7.5, 187.5mM NaCl, 1.25% Triton X-100, 0.625% Na-deoxycholate, 6.25% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail) was added. 30ul of Dynabeads Protein G (Thermo Fisher Scientific) per IP were washed 3 times with ChIP Wash Buffer 1 (25mM HEPES pH 7.2-7.5, 150mM NaCl, 5mM EDTA, 1% Triton X-100, 0.5% Na-deoxycholate, 5% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail) and resuspended in 60ul ChIP Wash Buffer 1. Each IP was blocked

with 20ul washed beads for 2 hours at 4C with rotation. The other 40ul of washed beads were incubated with antibody (5ul of RAD21 antibody or 2ul of H3K27me3 antibody/IP) for 2 hours at 4C. 5% input was taken from blocked lysates. Beads conjugated to antibody were added to blocked lysates and IP allowed to proceed overnight at 4C with rotation. IPs were then washed 3 times with cold ChIP Wash Buffer 1, 3 times with cold ChIP Wash Buffer 2 (25mM HEPES pH 7.2-7.5, 300mM NaCl, 5mM EDTA, 1% Triton X-100, 0.5% Na-deoxycholate, 5% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail). And 1 time with cold TENG (50mM Tris-HCl pH 8, 1mM EDTA, 50mM NaCl, 5% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail). All washes were performed by inverting tubes and setting aside for ~2 minutes per wash. IPs were then resuspended in 500ul TENG with 100ug RNAse A (Thermo Fisher Scientific) and incubated for 30 min at 37C. IPs were then washed 2 times with cold ChIP Wash Buffer 2 and 2 times with cold TENG without protease inhibitors. IPs were resuspended in 200ul elution buffer (50mM Tris-HCl pH 7.5, 10mM EDTA, 1% SDS) and incubated 15 min at 65C. Eluate was then transferred to a new tube and 198ul TE buffer (Ambion) and 40ug Proteinase K (Sigma) added. Inputs were thawed, TE buffer added to equalize volume to IPs and 200ul elution buffer and 40ug Proteinase K added. Inputs and IPs were then incubated for 2h at 45C and crosslinks reversed overnight at 65C. DNA was purified with phenol-chloroform extraction and resuspended in 50ul TE. Efficiency of ChIP was assayed via qPCR, with samples displaying over 10-fold enrichment of positive regions over negative control prepared for sequencing. Libraries were prepared from ~500ng of input or IP DNA using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB) according to kit instructions. Experiments were performed in two biological replicates. Input and ChIP-seq libraries were sequenced on Illumina HiSeq, generating 20-50 million paired end 50 or 150 nucleotide reads per sample.

Calibrated ChIP-seq was performed using the same protocol as ChIP-seq except that 50,000 HEK293 cells per 5 million mouse cells were mixed in at the initial lysis step. Cells were counted and equal numbers of cells used per IP.

ChIP-seq and calibrated ChIP-seq analysis

For uncalibrated ChIP-seq, input and ChIP-seq reads were aligned to the 129S1/SvJm (mus) and CAST/Eih (cas) genome as previously described (Kung et al., 2015; Minajigi et al., 2015b). PCR duplicates were removed and uniquely mapped reads used for further analysis. Fpm-normalized bigwig tracks were generated for all uniquely mapped (comp), mus-specific (mus), and cas-specific (cas) reads.

Heatmaps and metaplots were generated using deepTools functions computeMatrix and plotHeatmap with the reference point as the summit of the called ChIP peaks.

Calibrated ChIP-seq analysis was performed as previously described (Hu et al., 2015). Briefly, paired-end reads were aligned to a concatenated mouse and human genome (mm9 + hg19) using Bowtie2 with "--no-mixed" and "--no-discordant" options. PCR duplicates were filtered out with SAMBAMBA (Tarasov et al., 2015). Uniquely aligned mouse reads were normalized to human spike-in by subsampling mm9 reads so that total number of hg19 reads was matched between samples and bigWig tracks generated from normalized samples. For calibration of RAD21 cChIP between RAD21-FKBP^{degron} clones, only reads within hg19 peaks (hg19 tracks combined) were counted for normalization purposes. Subsampling factors were further adjusted based on ratio of hg19 to mm9 reads in matched input samples. After normalization factor was determined, paired end reads were aligned to mus and cas genomes as previously described, except using Bowtie2 with the same parameters as above (with "--no-mixed" and "--no-discordant") instead of

Novoalign. PCR duplicates were filtered out with SAMBAMBA and allelic reads normalized using the same normalization factor as above.

Input subtracted H3K27me3 calibrated ChIP coverage profiles were generated with SPP (Kharchenko et al., 2008) with smoothing using 1kb windows recorded every 500 bp, as previously described (Simon et al., 2013). For input subtraction, inputs from DMSO and dTAG-13 treated samples were combined to ensure equal profiles were subtracted from each calibrated IP.

Heatmaps and metaplots were generated using deepTools functions computeMatrix and plotHeatmap with the reference point as the summit of the called ChIP peaks. Plots of calibrated H3K27me3 signal over upregulated or downregulated genes were calculated using "scale-regions" with the entire gene scaled to 3kb.

To compare the allelic densities of H3K27me3 over different classes of X-linked genes, we computed average densities over gene bodies using multiBigwigSummary from deepTools with default parameters. Non-expressed genes were defined as X-linked genes have 0 RPKM in at least one RNA-seq replicate of DMSO control. Genes subject to XCI were identified as the set of genes have allelic information (see RNA-seq analysis above), with escapees, Xist and Tsix excluded. Escapees were as defined previously (Wang et al., 2018). To compare allelic densities of H3K27me3 over S1 and S2 compartments, we defined S1 compartments as regions having positive value in the first Principal Component calculated in the WAPL-FKBP^{degron} DMSO control days 4-7 sample and S2 compartments as regions have negative value. S1 and S2 compartments were then binned into 100kb windows using bedtools makewindows, excluding windows that overlapped with unmappable regions as defined previously (Wang et al., 2018). We then we computed average densities over these 100kb bins using multiBigwigSummary from deepTools with default parameters.

RNA-seq

Total RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific) and mRNA isolated using the NEBNext Poly(A) Magnetic Isolation Module (NEB) according to the kit instructions. Two biological replicates of RNA-seq libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) according the kit instructions. Libraries were sequenced on the Illumina HiSeq2500 (rapid run), generating 30-50 million 50 nt paired end reads per sample or the Illumina HiSeq4000, generating 16-27 million 150 nt paired end reads per sample.

RNA-seq analysis

RNA-seq reads were aligned allele-specifically to the 129S1/SvJm (mus) and CAST/Eih (cas) genomes using TopHat2 (Kim et al., 2013) as previously described (Kung et al., 2015; Minajigi et al., 2015b). PCR duplicates were removed and unique reads mapping to exons of each gene quantified by Homer (Heinz et al., 2010). DESeq 2 was used to call differentially expressed genes with a cut off of adjusted p-value < 0.1. For analyses on all genes, no FPKM threshold was applied, but genes were required to have at least 1 count in one sample. An expression threshold of FKPM > 0.5 was applied for analyses on expressed genes. Expressed genes with a minimum of 13 allelic-specific reads in at least one sample were counted as genes with allelic information (Colognori et al., 2019). For a given gene, fraction mus was calculated as the fraction of allelic reads that were mus-specific (mus / (mus+cas) reads).

Log2FoldChange and log2baseMean from DESeq2 were plotted for indicated classes of genes. H3K27me3 genes were defined as genes overlapping H3K27me3 cChIP-seq peaks in their respective samples (DMSO and dTAG-13 peaks combined). H3K27me3 loop genes were defined as H3K27me3 genes overlapping H3K27me3-bound loop anchors called in RAD21-FKBP^{degron}

dTAG-13 Hi-C maps (i.e., cohesin-independent H3K27me3 anchored loops). For overlap with RAD21 loop anchors, loops were called in respective DMSO Hi-C maps.

PRO-seq library construction

PRO-Seq library construction was performed by the Nascent Transcriptomics Core at Harvard Medical School, Boston, MA. Aliquots of frozen (-80°C) permeabilized cells were thawed on ice and pipetted gently to fully resuspend. An aliquot was removed and diluted 1:10 in 1xPBS for cell counting using an automatic counter (Biorad TC-20) with counts spot checked by manual recount using a hemocytometer as needed. An aliquot of each dilution was also stained with Trypan Blue and counting repeated to determine permeabilization efficiency. For each sample, 1 million permeabilized cells were used for nuclear run-on. For normalization, 50,000 permeabilized *Drosophila* S2 cells were added to each sample of 1 million cells. Nuclear run on assays and library preparation were then performed essentially as described [Elrod, N.D., et al. (2019) *Molecular Cell* **76**: 738-752.e7] with modifications noted below.

The nuclear run-on buffer was prepared as 4X stock (20mM Tris (pH 8), 20mM MgCl2, 2mM DTT, 600mM KCl, 40uM/ea biotin-11-NTPs (Perkin Elmer), 20U SuperaseIN (Thermo)). The 4X stock was mixed 1:1 with 2% sarkosyl (Sigma) to yield 2X complete run-on mix. The 3' adapter (RNA oligo: 5'P-GAUCGUCGGACUGUAGAACUCUGAAC-3'InvdT) was pre-adenylated prior to use (5' DNA adenylation kit, NEB, according to the manufacturer's instructions). Adenylated oligo was purified by ethanol precipitation, resuspended in water and quality was modification confirmed by electrophoresis on a 15% TBE-Urea gel (Novex) next to appropriate controls. The adapter concentration was adjusted to 10uM and 1uL was used for each ligation which used T4 RNA ligase 2, truncated KQ (NEB) with 15% PEG-8000 in the reaction mix. Reactions were incubated at 16°C overnight. Each ligation reaction was then diluted with 180uL of betaine binding buffer

(1.42g of betaine brought to 10mL with binding buffer and sterile filtered) containing 1uL of 100uM blocking oligo (TCCGACGATCCCACGTTCCCGTGG/3InvdT/). The blocking oligo is complementary to and overhangs the 3' end of the 3' adapter and reduces 5'/3' adapter dimer formation during the 5' adapter ligation step. The blocking oligo has modifications at its 3' end, to prevent extension by reverse transcriptase. Blocking oligo was also included in final wash solution prior to the 5' adapter ligation and in the ligation reaction itself, both at 3-fold molar excess over the initial amount of 3' adapter (i.e. 1uL of 30uM per reaction). The 5' adapter ligation was performed as in Elrod except PEG-8000 was increased to 15%.

After reverse transcription, cDNA was immediately amplified for 5-cycles ("preCR" step). A PCR cocktail consisting of NEBNext Ultra II Q5 master mix (NEB) and Illumina TruSeq PCR primers (RP-1, common; and RPI-X, indexing) was added to each cDNA and amplification was done according to manufacturer's suggested 2-step cycling conditions for NGS applications. To determine optimal library amplification, the preCRs were thawed on ice and an aliquot serially diluted. Each dilution was mixed with a PCR cocktail (Q5 DNA polymerase with optional high GC enhancer) containing 10uM/ea of Illumina universal P5 and P7 primers and amplified 15-cycles. The reactions were electrophoresed on a 2.2% agarose gel (SeaKem) cast with SYBR gold (Thermo) and analyzed to determine the optimal number of cycles for final amplification. preCRs were transferred from ice to a pre-heated thermal cycler block and amplified for the appropriate number of additional cycles to reach the total determined in the test amplification.

Pooled libraries were sequenced using Illumina NovaSeq S1 (PE 50) platform.

PRO-seq analysis

Paired-end reads were aligned to a concatenated mouse and fly genome (mm9 + dm6) using Bowtie2 with "--no-mixed" and "--no-discordant" options. DMSO and dTAG-13 samples for each clone were normalized to drosophila spike-in by subsampling mm9 reads so that total number of dm6 reads was matched between samples. For RAD21-FKBP^{degron} clone G9, downsampling factors corresponded to library size (i.e., were equivalent to downsampling based on total number of reads). For RAD21-FKBP^{degron} clone C4, downsampling factors were 1 for DMSO and 0.73 for dTAG-13 although library sizes were similar, indicating that there were overall lower levels of nascent transcription in the dTAG-13 sample. Duplicates were not removed to preserve reads corresponding to TSS and promoter proximal PolII positions. Reads mapping to genes (full gene bodies) were quantified by Homer (Heinz et al., 2010), removing miRNA, snoRNA and scaRNAs. DESeq 2 was used to call differentially expressed genes with a cut off of adjusted p-value < 0.1. For analyses on all genes, no FPKM threshold was applied, but genes were required to have at least 1 count in one sample. An expression threshold of FKPM > 0.5 was applied for analyses on expressed genes.

Log2FoldChange and log2baseMean from DESeq2 were plotted for indicated classes of genes. H3K27me3 genes were defined as genes overlapping H3K27me3 cChIP-seq peaks in their respective samples (DMSO and dTAG-13 peaks combined). H3K27me3 loop genes were defined as H3K27me3 genes overlapping H3K27me3-bound loop anchors called in RAD21-FKBP^{degron} dTAG-13 Hi-C maps (i.e., cohesin-independent H3K27me3 anchored loops). For overlap with RAD21 loop anchors, loops were called in respective DMSO Hi-C maps.

For allelic analysis, PRO-seq reads not mapping to the fly genome (dm6) were aligned to mus and cas genomes using Bowtie2 with the same parameters as above ("--no-mixed" and "--no-discordant") and downsampled using the same factors as calculated above. Expressed genes with a minimum of 13 allelic-specific reads in at least one sample were counted as genes with

allelic information (Colognori et al., 2019). For CDPs, normalized mus and cas counts divided by gene length were plotted.

In situ Hi-C and visualization

In situ Hi-C was done as described previously (Rao et al., 2014) using 2 clones for each degron line and 5 million cells per sample. See Supplemental Table 1 for Hi-C statistics.

Reads were trimmed using cutadapt with the options --adapter=GATCGATC (MboI ligation junction) and --minimum-length=20. Reads of each pair were individually mapped to the mus (Xi) and cas (Xa) reference genomes using novoalign and merged into Hi-C summary files and filtered using HOMER as previously described (Minajigi et al., 2015). Hi-C contact maps were generated from reads in the filtered HOMER mus and cas tag directories using the 'pre' command of Juicer tools with the genomeID mm9 (Durand et al., 2016b). The resulting Hi-C contact maps in .hic format were visualized and normalized with the 'Coverage Square Root' option in Juicebox version 1.8.8 (Durand et al., 2016a).

In situ Hi-C analysis

Principal component analysis (PCA) was done using the runHiCpca.pl command of HOMER with the filtered HOMER Hi-C tag directory as input and the options -res 100000 -superRes 100000 -genome mm9. Pearson correlation maps were also computed by HOMER using the analyzeHiC command with the filtered HOMER Hi-C tag directory as input and the options -chr chrX -res 200000 –corr.

Loops were called from Hi-C interaction maps using the GPU version of HiCCUPS (Durand et al., 2016b) with the following parameters: 'java -jar juicer_tools_1.13.02.jar hiccups -m 512 -r 25000,50000 -k KR -f .1,.1 -p 1,1 -i 3,3 -t 0.02,1.5,1.75,2 -d 50000,100000', adding the --

ignore_sparsity option when needed. Loops were called from merged Hi-C interaction maps for each condition (DMSO and dTAG-13). Loops were additionally filtered to require at least 27 contacts in the peak center, 2.25 enrichment over the donut neighborhood and an FDR < 0.05 in the donut neighborhood. APA was performed on loops called by HiCCUPS using Juicer (Durand et al., 2016b) at 5kb resolution.

For inter-chromosomal Hox interactions, APA was performed at 100kb resolution with the -e option using an input loop list of inter-chromosomal interactions anchored at 500kb intervals centered at Hox clusters.

For super-enhancer interactions, APA was performed at 100kb resolution using an input loop list of interactions anchored at 500kb intervals centered at super-enhancers (Whyte et al., 2013). Super-enhancer intervals were split into 5 quantiles based on distance between anchors.

TAD scores were calculated as described previously. Specifically, 100-kb resolution 'Balanced' normalized chromosome Hi-C interaction matrices were extracted from Juicebox using the 'dump' function of Juicer tools (juicer dump observed KR name.hic X X BP 100000 name.KR.100kb.dump). Custom shell and R scripts were used to convert this output matrix into the cworld matrix format. Insulation scores were then calculated for the X chromosome using the resulting cworld matrix and the matrix2insulation.pl cworld script with parameters –is 500000 –ids 400000 –im iqrMean–nt 0 –ss 200000 –yb 1.5 –nt 0 –bmoe 0. As described previously, regions of insulation score minima represent regions of locally high insulation and are considered potential TAD boundaries (Crane et al., 2015). TAD scores were then calculated using insulation scores from given samples and TAD borders called previously on the Xa (Colognori et al., 2019) with the cworld script insulation2tads.pl with default parameters.

For saddle plot compartmental analysis, .hic format Hi-C maps were first converted to .mcool format using 'hic2cool convert' with default options (Abdennur and Mirny, 2020). Publicly available

code from https://cooltools.readthedocs.io/en/latest/notebooks/04_saddle-plots.html was then used to generate saddle plots from .mcool files.

CHART-seq

Xist CHART-seq was performed as previously described (Wang et al., 2018). Briefly, cells were trypsinized, resuspended in 1 million cells/mL of media and crosslinked with 1% formaldehyde for 10 min at room temperature with rotation. Cells were then quenched with 0.2 M glycine for 5 min. Crosslinked cells were washed once with cold PBS, pelleted, and snap frozen in an EtOH/dry ice bath. Cell pellets were stored at -80C until further processing. 25 million cross-linked cells per CHART were thawed on ice and resuspended in 1 mL cold sucrose buffer. (10 mM HEPES pH 7.5, 0.3 M sucrose, 1% Triton X-100, 100 mM potassium acetate, 0.1 mM EGTA) supplemented with 0.5 mM spermidine, 0.15 mM spermine, 1 mM DTT, 1x protease inhibitor cocktail (Sigma, P8340), 10 U/mL RNase inhibitor (Roche). The cell suspension was rotated for 10 min at 4°C, after which it was diluted with additional 2 mL cold sucrose buffer. This was added to a pre-chilled 15-mL glass Wheaton dounce tissue grinder (Thermo Fisher Scientific) that had been pre-rinsed once with RNaseZAP RNase decontamination solution (Thermo Fisher Scientific), three times with DEPC-treated water, and once with 2 mL sucrose buffer. Release of nuclei was accomplished by douncing cells 20 times with a tight pestle, and verified by staining with Trypan Blue Solution (Thermo Fisher Scientific). The nuclear suspension was then layered atop a cushion of 7.5 mL cold glycerol buffer (10 mM HEPES pH 7.5, 25% glycerol, 1 mM EDTA, 0.1 mM EGTA, 100 mM potassium acetate) supplemented with 0.5 mM spermidine, 0.15 mM spermine, 1 mM DTT, 1x cOmplete EDTA-free protease inhibitor cocktail (Roche), and 5 U/mL RNase inhibitor, and centrifuged at 1500 g for 10 min at 4°C. The purified nuclear pellet was resuspended in 6 mL PBS and further cross-linked with 3% formaldehyde for 30 min at room temp with rotation. Afterward, nuclei were pelleted by centrifugation at 1,000 g for 5 min at 4°C and washed three

times with cold PBS. Nuclei were resuspended in in 1 mL cold nuclear extraction buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 0.1 mM EGTA, 0.5% sarkosyl, 0.1% sodium deoxycholate, 5 mM DTT, 10 U/mL RNase inhibitor) and incubated for 10 min at 4°C with rotation. Nuclei were then centrifuged at 400 g for 5 min at 4°C and resuspended in 230 µL cold sonication buffer (50 mM HEPES pH 7.5, 75 mM NaCl, 0.1 mM EGTA, 0.5% sarkosyl, 0.1% sodium deoxycholate, 0.1% SDS, 5 mM DTT, 10 U/mL RNase inhibitor) to a final volume of ~270 µL. Nuclei were sonicated for 5 min at 4°C using a Covaris E220e (140W peak incident power, 10% duty factor, 200 cycles/burst). Sonicated chromatin was centrifuged at 16,000 g for 20 min at 4°C to remove debris. The supernatant (~220 µL) was mixed with 110 µL sonication buffer to a final volume of ~330 µL, which was split into two CHART reactions (Xist capture using antisense probes and control capture using sense probes).

For two CHART reactions, 600 µL MyOne Streptavidin C1 beads (Thermo Fisher Scientific) were used, which were pre-washed with 1 mL DEPC-treated water, blocked with 600 µL blocking buffer (33% sonication buffer, 67% 2x hybridization buffer, 500 ng/µL yeast RNA [Thermo Fisher Scientific], 1% BSA) at room temp for 1 hr, washed with 1 mL DEPC-treated water again, and resuspended in 610 µL 1x hybridization buffer (33% sonication buffer, 67% 2x hybridization buffer [see below for composition]). For each CHART, 160 µL chromatin was mixed with 320 µL 2x hybridization buffer (50 mM Tris pH 7.0, 750 mM NaCl, 1% SDS, 1 mM EDTA, 15% formamide, 1 mM DTT, 1 mM PMSF, 1x cOmplete EDTA-free protease inhibitor cocktail, 100 U/mL RNase inhibitor) and pre-cleared with 60 µL pre-blocked MyOne Streptavidin C1 beads at room temp for 1 hr with rotation. Pre-cleared chromatin was separated from beads, with 1% saved as “input” sample. The remaining pre-cleared chromatin (~500 µL) was mixed with 36 pmol of either antisense (Xist-targeting) or sense (control) biotinylated capture probes. We used a pool of 9 capture probes, which were selected from the 11 probes described previously (Simon et al.,

2013), excluding probes X.3 and X.8. Hybridization was carried out at 37°C for 4 hr followed by incubation with 240 µL blocked MyOne Streptavidin C1 beads at 37°C for 1 hr. The beads were washed once with 1x hybridization buffer (33% sonication buffer, 67% 2x hybridization buffer) at 37°C for 10 min, five times with wash buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 2% SDS, 2 mM EDTA, 2 mM EGTA, 1 mM DTT) at 37°C for 5 min, and twice with elution buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 0.5% NP-40, 3 mM MgCl2, 10 mM DTT) at 37°C for 5 min. 10% of the final wash was saved as an "RNA capture" sample to estimate the efficiency of Xist RNA capture afterward by qRT-PCR. CHART-enriched DNA was eluted twice in 200 µL elution buffer supplemented with 5 U/µL RNase H (New England BioLabs) at 37°C for 30 min. An "input" sample and CHART DNA were treated with 0.5 mg/mL RNase A (Thermo Fisher Scientific) at 37°C for 1 hr and then with 1% SDS, 10 mM EDTA, and 0.5 mg/mL proteinase K (Sigma) at 55°C for 1 hr. Reversal of crosslinks was performed by addition of 150 mM NaCl (final 300 mM) and incubation at 65°C overnight. DNA was purified by phenol-chloroform extraction and further sheared to < 500-bp fragments using Covaris E220e (140W peak incident power, 10% duty factor, 200 cycles/burst). Sonicated DNA was purified using 1.8x Agencourt AMPure XP beads (Beckman Coulter). Input and CHART-seq libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB) according to kit instructions. Libraries were sequenced on Illumina, generating ~50 million 150-nt paired-end reads per sample.

CHART-seq analysis

Input and ChIP-seq reads were aligned to the 129S1/SvJm (mus) and CAST/Eih (cas) genome as previously described (Kung et al., 2015). PCR duplicates were removed and uniquely mapped reads used for further analysis. Fpm-normalized bigwig tracks were generated for all uniquely mapped (comp), mus-specific (mus), and cas-specific (cas) reads. CHART libraries were sampled to equal read depth and input subtracted CHART coverage profiles were generated with SPP

(Kharchenko et al., 2008) with smoothing using 1kb windows recorded every 500 bp, as previously described (Simon et al., 2013). After scaling, we computed base 2 logarithm of the ratio of the coverage of DMSO control divided by that of dTAG-13 using bigWigCompare from deepTools with --binSize 1000 and default parameters otherwise. To compare Xist densities over centromeric, middle, and telomeric regions, we split the first 30Mb, middle 100Mb and last 30Mb of the X chromosome into 100kb windows using bedtools makewindows, excluding windows that overlapped with unmappable regions as defined previously (Wang et al., 2018) and computed average Xist densities over these bins using multiBigwigSummary from deepTools with default parameters. To compare the densities of Xist over different classes of X-linked genes, we computed average densities over gene bodies using multiBigwigSummary from deepTools with default parameters. Non-expressed genes were defined as X-linked genes have 0 FPKM in at least one RNA-seq replicate of DMSO control. Genes subject to XCI were identified as the set of genes have allelic information (see RNA-seq analysis above), with escapees, Xist and Tsix excluded. Escapees were as defined previously (Wang et al., 2018). To compare densities of Xist over S1 and S2 compartments, we defined S1 compartments as regions having positive value in the first Principal Component calculated in the WAPL-FKBP^{degron} DMSO control days 4-7 sample and S2 compartments as regions have negative value. S1 and S2 compartments were then binned into 100kb windows using bedtools makewindows, excluding windows that overlapped with unmappable regions as defined previously (Wang et al., 2018). We then we computed average densities over these 100kb bins using multiBigwigSummary from deepTools with default parameters.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters including the statistical tests used and the values of n and p are reported in the figures, figure legends, or associated main texts. Statistical significance is determined by the value of $p < 0.05$ by the indicated tests.

Chapter 5

Conclusion and Future Directions

Revisiting the relationship between 3D genome structure and gene expression

When I started work on my PhD thesis, chromosome conformation capture (3C) based methods, such as Hi-C, had just revealed myriad layers of genome folding in the nucleus, not previously appreciated by microscopy. Most exciting were the “Topologically Associating Domains” (TADs), megabase-sized “square” domains prominently visible along the diagonals of Hi-C interaction maps and enriched for CTCF and cohesin binding at their borders (Dixon et al., 2012; Nora et al., 2012). Initially, this excitement seemed borne out by large deletion studies of “TAD” borders, which caused catastrophic enhancer mistargeting phenotypes (Lupianez et al., 2015). However, studies which disrupted 3D genome structure more precisely had puzzling results. Targeted deletions ablating CTCF binding sites were in general not sufficient to cause enhancer mistargeting or improper gene expression (Despang et al., 2019; Paliou et al., 2019; Williamson et al., 2019). Even more surprisingly, CTCF, cohesin, and WAPL loss-of-function studies globally disrupted 3D genome structure but had relatively small impacts on gene regulation, enhancer targeting, or spreading of chromatin marks (Haarhuis et al., 2017; Nora et al., 2017; Rao et al., 2017; Wutz et al., 2017). Rather than all domain structure being functional at all times, these studies supported a model where only specific structures were needed at certain points of development. I therefore looked at the role of chromosome structure in a prototypical model system of development, differentiating mouse embryonic stem cells (mESCs). In particular, I revisited the relationship between chromosome folding and gene expression in the context of X chromosome inactivation (XCI), the “model” epigenetic process by which Xist RNA and other repressive factors, such as Polycomb Repressive Complexes 1 and 2, silence and entire X chromosome in female mammals.

Toggling CTCF and cohesin binding via the dTAG degron system

The advent of degron technology enabled me to bypass the technical issues with knocking out proteins such as CTCF, RAD21, and WAPL. By knocking-in a degron tag to the C-terminal

end of these proteins of interest, I could trigger their rapid and efficient degradation upon addition of a small molecule. I chose to use the dTAG degron system, which utilizes the endogenous ubiquitin ligase, cereblon (Nabet et al., 2018). This enabled me to further bypass issues with expressing the Tir1 ubiquitin ligase required for the auxin degron system during differentiation. Indeed, all my degron lines were extremely efficient at degradation. This enabled ablation of CTCF binding as well as toggling of cohesin binding down or up depending on whether RAD21 or WAPL was degraded, respectively.

Future development of degron systems

While degron systems have proven to be extremely powerful tools both in my thesis and many other studies, it is worth noting they could still be improved for future use. In particular, an issue with destabilization of tagged proteins is noticeable for both the auxin and dTAG degron tags. While the extent to which proteins are affected varies, CTCF, RAD21, and WAPL all suffer from such destabilization with both the dTAG and auxin degron tags (Kriz et al., 2021; Liu et al., 2020; Nora et al., 2017; Perea-Resa et al., 2020; Rhodes et al., 2020; Wutz et al., 2017). Therefore, it is important to keep in mind that all these degron lines have lower baseline levels of their tagged protein than their respective parental wild-type cells. In other words, even in control conditions they likely display a different phenotype than their wild-type counterparts. Therefore, wild-type cells are not an ideally matched control. In my thesis, I circumvented this problem by using a DMSO treatment control in the matched cell line for each degradation time course. However, a downside to this rigorous control strategy is that control versus degradation conditions may underestimate effect size due to levels of the tagged protein already being lower prior to treatment. Future iterations of degron technology will hopefully address this issue of degron “leakiness”. Indeed, solutions have already been proposed for the auxin system (Yesbolatova et al., 2019).

Another avenue in which it would also be exciting to develop degron technology is to enable degradation of proteins at specific phases of the cell cycle. This would be especially useful for studies of cohesin, which plays an essential role in cell division. A system in which degradation could be limited to interphase would enable separation of phenotypes induced by cell cycle defects from those resulting from cohesin's interphase role in mediating 3D genome structure, especially over longer degradation timecourses. Indeed, with current technology it is not possible to ablate cohesin in dividing cells for longer than the length of a cell cycle without massive cell death.

Already, by engineering new small molecules, the use of degron tags has been expanded. A recent study developed dTAG^{V}-1, a new small molecule that can be used with the FKBP12^{F36V} (here called FKBP^{degron}) tag to recruit von Hippel-Lindau (VHL) E3 ligase complex to degrade proteins instead of Cereblon (Jaeger et al., 2020; Nabet et al., 2020). In this way, dTAG^{V}-1 could induce degradation of FKBP^{degron} tagged proteins that had resisted Cereblon-mediated degradation. In concept, the FKBP12^{F36V} tag could be used with other bifunctional small molecules to tether tagged proteins to proteins not involved in degradation pathways, such as dCas9-fusions. Such approaches could allow for inducible recruitment, for example (Jaeger and Winter, 2021). A particularly intriguing application would be to trigger recruitment of cohesin to specific regions of the genome, enabling more specialized dissection than WAPL degradation, which increases cohesin occupancy genome-wide.

Even a small amount of CTCF can maintain 3D genome structure and X-inactivation

The initial studies that identified TADs found that CTCF binding was highly enriched at certain domain borders, especially in the convergent orientation. This correlation, along with CTCF's known role as an insulator, lead to the prevailing model of CTCF establishing borders through its insulatory activity. Degron technology enabled acute loss-of-function studies to test this model in undifferentiated mESCs (Kubo et al., 2021; Nora et al., 2017) and cancer cells (Wutz

et al., 2017). In all of these studies, CTCF degradation indeed disrupted borders. Surprisingly this only had a small impact on gene expression and epigenetic spreading. Hypothesizing CTCF might be more important during dynamic processes such as development, I revisited CTCF's role in mESCs undergoing differentiation. In particular, I focused on the model epigenetic process, XCI.

I generated female mESC CTCF dTAG-13 degron lines, verifying that CTCF levels could indeed by ablated rapidly upon degradation during differentiation via Western blot and uncalibrated ChIP-seq. Due to the hybrid nature of the X chromosomes in these cell lines, the Xi and Xa could further be specifically detected. Intriguingly, this massive depletion of CTCF had relatively little impact on genome structure, either on autosomes on the X chromosome, by low-resolution Hi-C. While deeper sequencing of Hi-C could reveal subtle impacts, the results are generally in keeping with previous CTCF degron studies, which all detected at least subsets of domain borders that persisted even after CTCF degradation (Kubo et al., 2021; Nora et al., 2017). Some of these borders could plausibly be attributed to elements other than CTCF that confer insulation, pointing to a need for more precise terminology than "TAD" for domain-level structures. In particular, terms for CTCF-dependent versus CTCF-independent domains would help clarify much discussion. It is likely that others resistant borders are due to residual CTCF binding that remains even after degradation.

Furthermore, even 10 days of CTCF degradation did not significantly shift expression of the Xi relative to DMSO controls. As previous degron studies did not or could not assay effects of CTCF degradation on XCI due to using male cells, it was not possible to compare these results to others. However, as with domain borders, it is possible that the residual amounts of CTCF could be sufficient to maintain borders.

Taken together, these results indicate that mESCs may express CTCF orders of magnitude higher than needed for its essential roles in 3D genome structure and XCI. Further investigation will be needed to clarify why this may be the case. CTCF deletion causes embryonic lethality in mice before embryonic day 5.5 (Moore et al., 2012); therefore, one possibility is that

embryonic cells overexpress CTCF in order of protect against catastrophic effects of CTCF loss. Another possibility is that while CTCF degradation may not cause massive disruption of XCI and 3D genome structure in cell lines, the consequence of these changes in animals may be much more severe.

Future directions for elucidating the architectural role of CTCF

CTCF's hypothesized role in mediating chromosome structure continued to evolve during my PhD. Initial models of "TAD" formation could not explain why convergent orientations of CTCF binding at borders were favored above all others. Higher-resolution Hi-C studies revealed that stable CTCF/cohesin loops are present at the majority of domain borders (Rao et al., 2014). Cohesin is thought to extrude loops through its ring-like structure; therefore, it was hypothesized that convergent CTCF sites, but not CTCFs bound in other orientations, had the ability to halt this loop extrusion activity, thus creating stable loops. Only recent studies, however, revealed the mechanism behind this observation. Namely, the N-terminal end of CTCF was found to bind cohesin, specifically at the site where cohesin normally interacts with WAPL, one of its release factors. In this way, CTCF "protects" cohesin from unloading by WAPL, thereby stabilizing CTCF/cohesin loops. An especially important observation is that the C-terminal end of CTCF did not have cohesin-binding activity (Li et al., 2020; Nora et al., 2020; Pugacheva et al., 2020). Therefore, only convergently-oriented CTCF pairs would be predicted to fully stabilize cohesin. These results at last provided a biological explanation for the convergent orientation rule. Intriguingly, the region of CTCF found to bind cohesin includes Serine 224, the residue we found to be phosphorylated in our study of CTCF post-translational modifications (Chapter 2). A future direction would be to investigate whether Serine 224 phosphorylation and other post-translations modifications could play a role in modulating binding to cohesin or protecting cohesin from WAPL-mediated removal. Indeed, recent studies have shown that this process can be regulated by post-

translation modifications (PTMs) of cohesin (Wutz et al., 2020), supporting that modulatory PTMs of CTCF could exist as well.

Another outstanding question is why the majority of CTCF binding peaks do not demarcate borders. One possibility is that elements in addition to CTCF binding are needed to establish border insulation. Indeed, a recent study helped clarify additional requirements for CTCF binding sites to have insulatory activity, using a *Sox2* enhancer reporter assay in mESCs (Huang et al., 2021). This study revealed that multiple CTCF binding sites in tandem are needed to confer insulation between *Sox2* and its super-enhancer. In addition, CTCF binding sites that demarcate domain borders confer the most insulation. Analysis of sequences flanking these insulatory CTCF binding sites revealed an additional facilitating motif that possibly interacts with CTCF zinc fingers 9-11. Another possibility is that additional elements may be able to establish certain borders even when only small amounts of CTCF are present in the cell. Supporting this, all CTCF degron studies reveal subsets of domains which resist CTCF degradation (Kubo et al., 2021; Nora et al., 2017). Future work will be needed to elucidate why these borders persist and what elements are responsible for their insulatory activity in addition to CTCF.

Toggling chromosome architecture via degradation of RAD21 or WAPL

In contrast to the subtler effects of CTCF degradation, degradation of either RAD21 or WAPL dramatically altered chromosome architecture. Degradation of RAD21 completely ablated domain structure as well as cohesin-anchored loops within hours, as previously observed (Rao et al., 2017; Wutz et al., 2017). Compartments, in contrast, persisted, again agreeing with previous studies. WAPL degradation, which resulted in accumulation of cohesin binding, resulted in a large increase of cohesin loops and weakening of compartments. This trend agreed with a WAPL knockout study, although the magnitude of effects was only similar to the knockout after several days of WAPL degradation (Haarhuis et al., 2017). My degrons also revealed classes of loops that resisted cohesin degradation, in fact becoming stronger on average when cohesin was

removed. Many of these were anchored by Polycomb, in my study assayed by the PRC2 mark H3K27me3. Surprisingly, the majority of these loops did not appear to be anchored by H3K27me3. One possibility is that PRC1 may be a better marker of Polycomb-dependent loops. Indeed, a recent study also observed Polycomb-anchored, cohesin-independent loops, further showing that these were ablated by RING1B degradation (Rhodes et al., 2020). However, it is also likely that factors other than Polycomb are responsible for mediating loops in differentiating mESCs. Further work will be needed to elucidate these. Possibilities that have been shown to be mediate 3D contacts in other contexts include YY1 and H3K27ac (Rao et al., 2017; Weintraub et al., 2017). Another noteworthy finding was that these cohesin-independent loops as a whole were on average weakened in strength when WAPL was degraded and cohesin levels increased on chromatin. Therefore, cohesin antagonizes both Polycomb-anchored and loops anchored by unknown factors.

Gene expression was surprisingly resistant to cohesin degradation, with similar numbers becoming dysregulated as in previous cohesin degron studies (Rao et al., 2017; Rhodes et al., 2020); therefore, cohesin-dependent structures as a whole do not appear to be more essential for gene expression overall during differentiation. The majority of significantly downregulated genes were marked by Polycomb, agreeing with the results of a previous cohesin degron study (Rhodes et al., 2020). This suggests that cohesin's antagonism of Polycomb interactions may serve to prevent certain Polycomb targets from becoming repressed beyond appropriate levels. Interesting, Polycomb targets engaged in looping interactions tended to already be more repressed than non-interacting Polycomb targets. This trend may explain why the majority of Polycomb targets were not significantly dysregulated by cohesin degradation.

When WAPL was degraded, increasing cohesin levels on chromatin, significantly upregulated genes were enriched in Polycomb targets, in contrast. Polycomb targets engaged in looping interactions tended to especially highly upregulated, suggesting that balance of cohesin at these loci is especially important for proper expression.

In both cohesin and WAPL degradation conditions, H3K27me3 levels did not significantly change at Polycomb targets. Therefore, the observed change in gene expression were likely due to alterations in proximity, or alternatively in frequency of cohesin-mediated loop extrusion passing through the loci, rather than alteration of Polycomb activity. This model could be tested by altering proximity or cohesin levels at specific loci. For example, dCas9 fused to a cohesin loader, such as *NIPBL* could be targeted to candidate Polycomb targets engaged in looping interactions.

Cohesin regulates interchromosomal *Hox* cluster interactions

Most intriguing was the impact of cohesin toggling on interchromosomal loops. Unexpectedly, interchromosomal "loops" between the developmentally important and heavily Polycomb-marked *Hox* clusters was visible in my high resolution Hi-C interaction maps. All *Hox* clusters participated in these interactions, which were resistant to cohesin degradation and weakened by WAPL degradation. Examination of previously generated data revealed that these interchromosomal interactions were not visible in control conditions in undifferentiated mESCs, only being detected after cohesin was degraded (Rhodes et al., 2020). The dynamic nature of these interchromosomal *Hox* interactions throughout differentiation suggests they may play a role in development, for example by aiding in coordination of *Hox* gene expression. The antagonism of cohesin against these interchromosomal interactions further suggests that cohesin may play a role against preventing inappropriately strong interactions between *Hox* clusters. However, losing these interactions could also cause inappropriate expression of *Hox* clusters. Indeed, when WAPL was degraded and interchromosomal *Hox* interactions weakened, *Hox* genes became among the most highly upregulated of Polycomb targets.

Further investigation will be needed to test this model. A possibility is that extremely strong *Hox* clustering in an "inactive" compartment may prevent activation of *Hox* cluster genes when needed; therefore, cohesin is needed to partially disrupt such interchromosomal interactions. My WAPL degradation experiments, which increase cohesin levels on chromatin, further show that

balance of cohesin is important. Degradation of cohesin in a system that more accurately models *Hox* regulation during development, such as mouse limb buds, would enable testing of these hypothesis. Indeed, dTAG degrons have been demonstrated to work effectively in mice (Nabet et al., 2018). These results also indicate searching systemically for interchromosomal interactions, for example by expansion of existing Hi-C loop-calling algorithms to interchromosomal regions, would likely be fruitful.

Cohesin shapes the inactive X chromosome

The mechanisms behind the unique structural changes the Xi undergoes have long been unclear. It was known that certain architectural factors, like CTCF and cohesin, become depleted from the Xi. The SMC-like SMCHD1, in contrast, becomes recruited to the Xi during the later stages of XCI (Minajigi et al., 2015b; Wang et al., 2018). The Xi therefore loses autosomal structures, such as cohesin-dependent domains and A/B compartments, folding instead into S1/S2 compartments and "superstructures"—the megadomains, with a border at *Dxz4*, and a "superloop" forming between *Dxz4* and *Firre*. As cohesin remains highly bound to both *Dxz4* and *Firre*, it was hypothesized that the remaining cohesin on the Xi was responsible for both superstructures. With my cohesin degron line, I confirmed that the megadomains were indeed cohesin-dependent.

Unexpectedly, the *Dxz4-Firre* superloop not only persisted after cohesin degradation but became more prominent. Further studies will be needed to elucidate which factors mediate superloop formation. However, another striking feature of *Dxz4* and *Firre* is that they, among a few other regions on the Xi including the "escapee" genes, remain active during XCI. Therefore, one possibility is that the active chromatin state of these loci mediated looping. Indeed, I also detected other regions depleted for Polycomb, one including the escapee gene *Kdm5c*, that formed interactions on the Xi. This hypothesis could be tested by deleting the promoter regions of *Dxz4* and *Firre*, in a way that would disrupt transcription without disrupting the majority of

transcription factor binding. Alternatively, dCas9 fused to a silencing complex could be targeted to *Dxz4* and *Firre*. As previous studies have shown that *Dxz4* and *Firre* deletion largely do not disrupt XCI in cell lines, silencing these loci would likely not have secondary effects that would confound any observed changes in superloop structure. Overall, these results suggest that in addition to forming megadomains, the remaining cohesin on the Xi could function to antagonize interactions between still-active regions on the Xi. This could be necessary to enable active regions silenced later in XCI to be inactivated eventually, for example.

Increasing cohesin on the Xi by degrading WAPL, in contrast, did not disrupt the megadomain border at *Dxz4* or the *Dxz4-Firre* superloop. The most salient effect was on compartments, with long-range compartmental interactions weakening and the Xi becoming more "linear" upon WAPL degradation and increased cohesin retetention. This effect mirrored the impact of WAPL degradation on autosomal compartments, observed both in my and a previous WAPL knockout study (Haarhuis et al., 2017). An interesting future direction would be to investigate how increased cohesin levels affect SMCHD1 recruitment and function.

Puzzlingly, toggling of cohesin also resulted in Xi-like superstructures appearing on the Xa. After cohesin degradation, an interaction between *Dxz4* and *Firre* became apparent on the Xa. After WAPL degradation, in contrast, a megadomain-like pattern with a border near *Dxz4* could be detected by Principal Component (PC) analysis of the Xa Hi-C interaction map. These results suggest that the Xi's unique superstructures may be a result of changes in cohesin distribution on the Xi, rather than XCI in itself. Namely, the *Dxz4-Firre* superloop may be a result of overall depletion of cohesin. The megadomain border at *Dxz4* may be a result of increased pile-up of cohesin at *Dxz4*, which retains CTCF binding, relative to surrounding regions which do not. Superstructures reflecting cohesin distribution, rather than gene silencing, would also support previous *Dxz4* and *Firre* deletion studies which showed that superstructures largely are not essential for gene silencing (Bonora et al., 2018; Froberg et al., 2018; Giorgetti et al., 2016). A pivotal question that remains to be addressed, then, is what the function of superstructures is.

Subsequent *Firre* and *Dxz4* deletion studies in mice paradoxically revealed impacts on autosomal gene expression (Andergassen et al., 2019; Fang et al., 2020). Another study revealed a role for cohesin-mediated loop extrusion in the DNA damage response (Arnould et al., 2021). Examining the response of female cells lacking superstructures on the Xi to DNA damage and other stress could thus reveal more insight into the biological function of these unique structures, which are conserved from mouse to human (Darrow et al., 2016).

A balance of cohesin is needed for proper X-inactivation

A longstanding question in the XCI field is whether the Xi's unique structure is required for, or merely a byproduct of gene silencing. Surprisingly, although cohesin degradation dramatically altered Xi structure, expression from the Xi did not shift significantly. These results, however, agree with the relatively small effects observed after cohesin degradation on autosomes. They also support previous *Dxz4* deletion studies showing that the megadomains are not needed for gene silencing (Bonora et al., 2018; Froberg et al., 2018; Giorgetti et al., 2016). However, a caveat is that cohesin degradation timecourses were relatively short, limited by cohesin's necessity in cell division. Future development of degron technology, for example by enabling interphase-specific degradation of cohesin, would enable assessment of longer-term effects of cohesin degradation on XCI.

WAPL could be degraded, and cohesin levels inappropriately retained on the Xi, for up to 3 days without altering cell cycle profiles, enabling examination of XCI in the presence of higher cohesin levels. Xist and Polycomb spreading both decreased, especially at the ends of the chromosome. Interestingly, the *Xist* locus correspondingly lost the most Hi-C contacts to the ends of the chromosome. Expression of X chromosome genes also shifted as a whole to higher expression from the Xi. As with the spreading defects, the shift toward higher fraction Xi expression was especially prominent at the ends of the chromosome, where the most Hi-C contacts to the *Xist* locus were lost. Therefore, a balance of cohesin is needed for appropriate

XCI, with inappropriate retention of cohesin and loss of long-range contacts causing defects, especially to the ends of the X chromosome.

Concluding Remarks

Almost ten years after the discovery of “TADs”, pivotal questions about the functions of these and other chromatin structures in the mammalian genome remain to be fully elucidated. Cell lines are remarkably resilient to genome-wide loss of certain types of chromatin architecture via cohesin or CTCF degradation, suggesting that 3D genome structure is not be required for steady state gene expression, at least in cell culture. The future of the field, then, will likely be investigating specific contexts where chromosome folding is needed to dynamically upregulate or downregulate certain groups of genes. The process of X-inactivation, in which the Xi refolds dramatically while being silenced, provides a proof-of-concept of this hypothesis. Previous studies showed that S1/S2 compartments mediated by PRC1 and merged by SMCHD1 are essential for proper XCI. My thesis further showed that a balance of cohesin is needed as well. Too much cohesin being retained on the Xi antagonizes gene silencing, especially at regions where long-range contacts are lost. However, XCI was resistant to short-term changes in structure caused by cohesin loss. In addition, my thesis revealed an intriguing correlation between Polycomb repression and cohesin balance, especially at the *Hox* cluster genes. Both interaction frequency of certain Polycomb targets as well as expression state correlated with cohesin levels on chromatin.

In my thesis, I modeled extreme loss of CTCF, cohesin, and WAPL using dTAG degron cell lines. However, the importance of cohesin balance may provide valuable clues to human diseases as well. One such disease is Cornelia de Lange syndrome, where mutations in the cohesin loading factor *NIPBL* as well as cohesin components are often mutated. To my knowledge, the impact of such mutations on *Hox* cluster expression or XCI during development has not been studied. Although such mutations would not alter cohesin levels as dramatically as

cohesin or *WAPL* degradation, it is conceivable that even slight imbalances over a lifetime could accumulate into dysregulation of similar magnitude.

Appendix

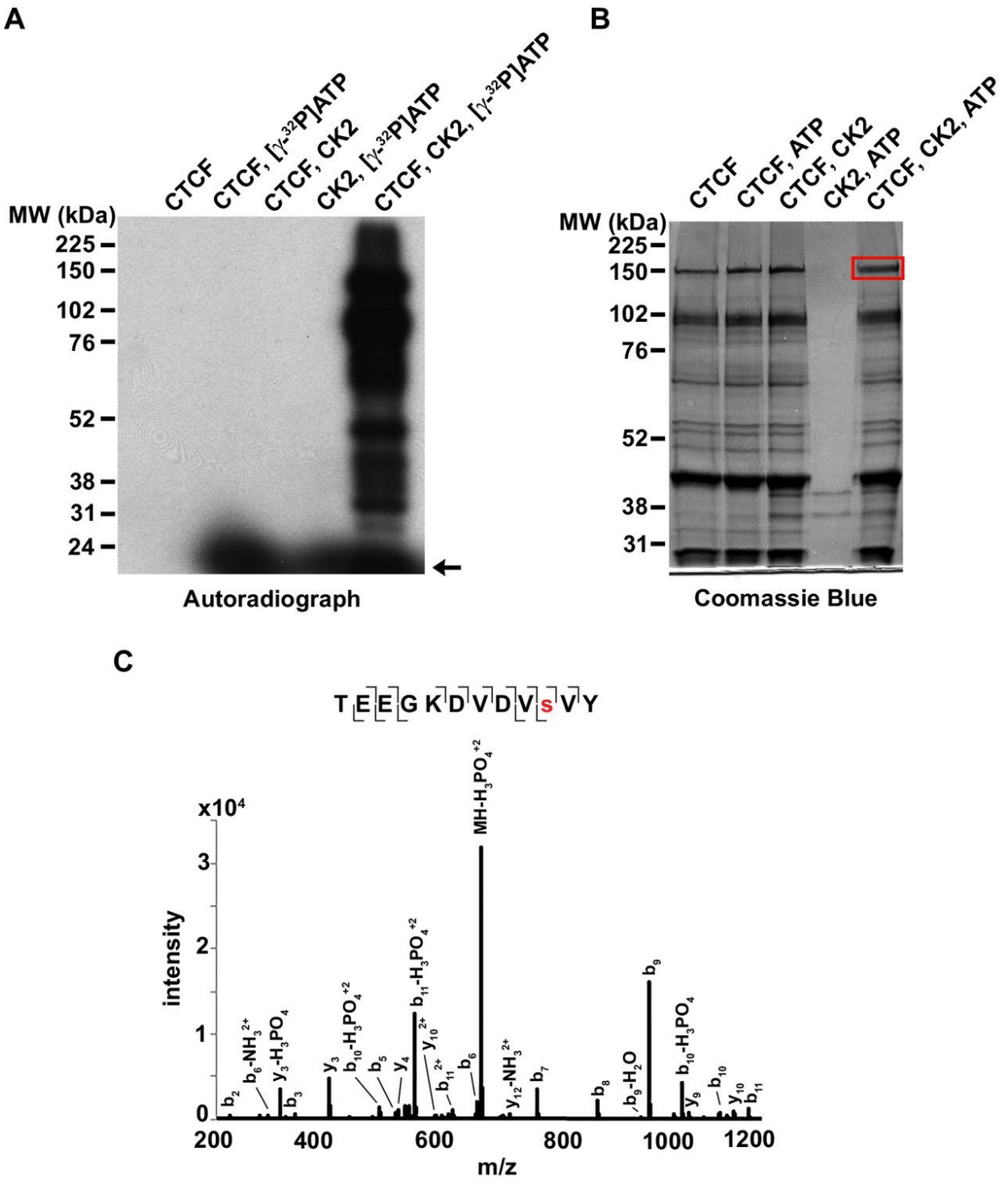


Figure S1.1. Casein kinase 2 (CK2) can phosphorylate CTCF Ser^{224}.

(A) In vitro kinase assay autoradiograph of recombinant FLAG-CTCF-6xHis phosphorylated by CK2 with [γ-^{32}P]ATP. Black arrow, free [γ-^{32}P]ATP.

(B) Coomassie Blue stained gel of FLAG-CTCF-6xHis phosphorylated with CK2 and cold ATP performed in parallel as in A. Red box, Full-length FLAG-CTCF-6xHis band excised for mass spectrometry analysis.

(C) Manually validated mass spectra of CTCF peptide Thr^{215}-Tyr^{226} with y and b ions identified. Phosphorylation at Ser^{224} indicated by red s.

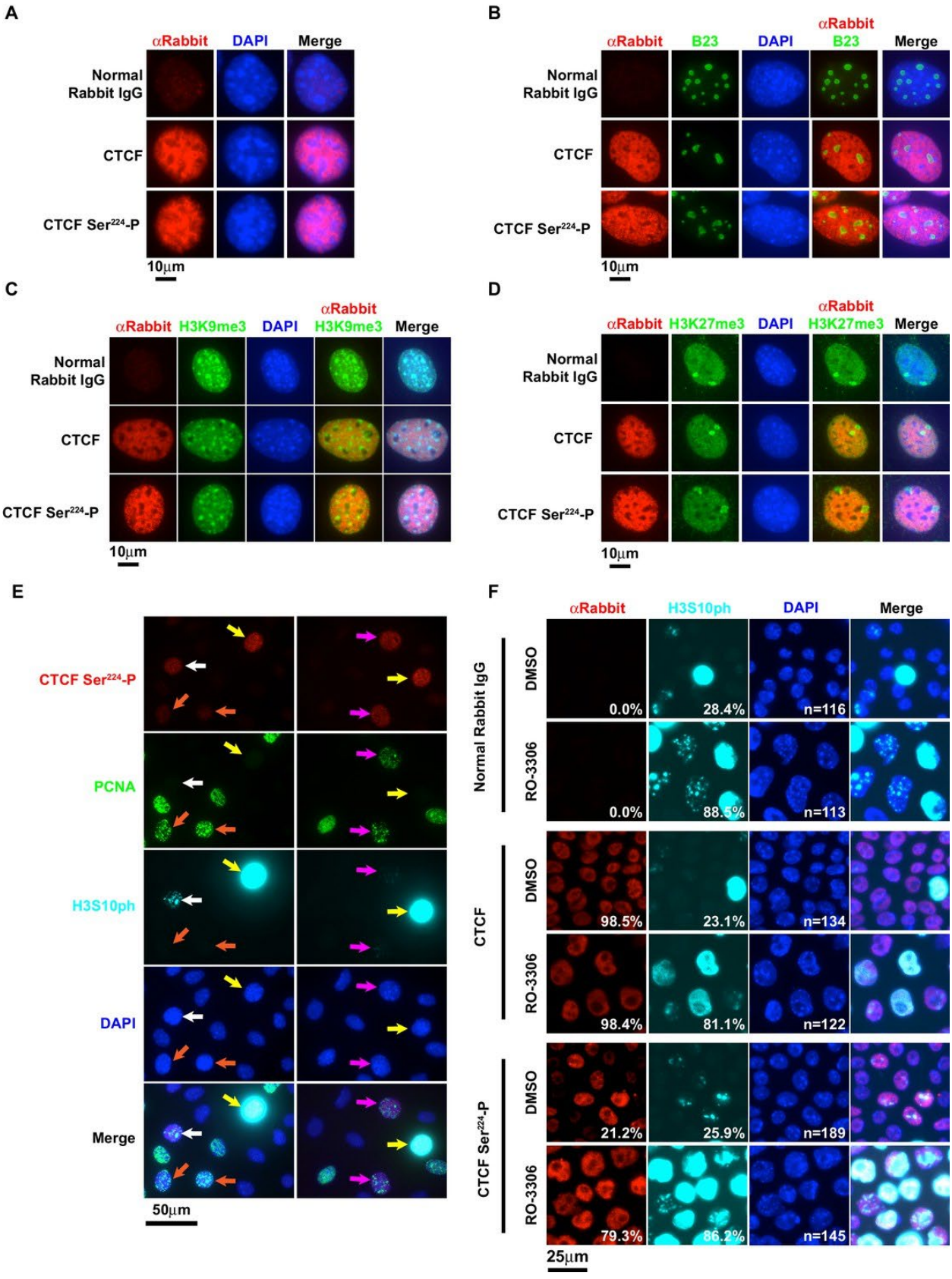

A
αRabbit
DAPI
Merge
Normal Rabbit IgG
CTCF
CTCF Ser224-P
10μm
B
αRabbit
B23
DAPI
αRabbit
B23
Merge
Normal Rabbit IgG
CTCF
CTCF Ser224-P
10μm
C
αRabbit
H3K9me3
DAPI
αRabbit
H3K9me3
Merge
Normal Rabbit IgG
CTCF
CTCF Ser224-P
10μm
D
αRabbit
H3K27me3
DAPI
αRabbit
H3K27me3
Merge
Normal Rabbit IgG
CTCF
CTCF Ser224-P
10μm
E
CTCF Ser224-P
PCNA
H3S10ph
DAPI
Merge
50μm
F
αRabbit
H3S10ph
DAPI
Merge
Normal Rabbit IgG
DMSO
0.0%
28.4%
n=116
RO-3306
0.0%
88.5%
n=113
CTCF
DMSO
98.5%
23.1%
n=134
RO-3306
98.4%
81.1%
n=122
CTCF Ser224-P
DMSO
21.2%
25.9%
n=189
RO-3306
79.3%
86.2%
n=145
25μm

Figure S1.2. CTCF Ser224-P accumulates in G2 and has a subnuclear distribution comparable to total CTCF.

(A-D) Immunofluorescence performed on immortalized MEFs using the indicated antibodies. Nuclei counterstained with DAPI. Bar, 10μm.

(E) Immunofluorescence performed on asynchronous MEFs with the indicated antibodies. Nuclei counterstained with DAPI. Bar, 50μm. White arrows (PCNA negative, speckled H3S10ph), early G2. Yellow arrows (PCNA negative, pronounced H3S10ph), late G2. Orange arrows (Speckled PCNA, no H3S10ph), S. Magenta arrows (Speckled PCNA, faint speckled H3S10ph), late S/early G2.

(F) Immunofluorescence performed on TST-1 mESCs with the indicated antibodies. Cells were treated 20hrs with 10μM RO-3306 or DMSO alone control. Bar, 25μm.

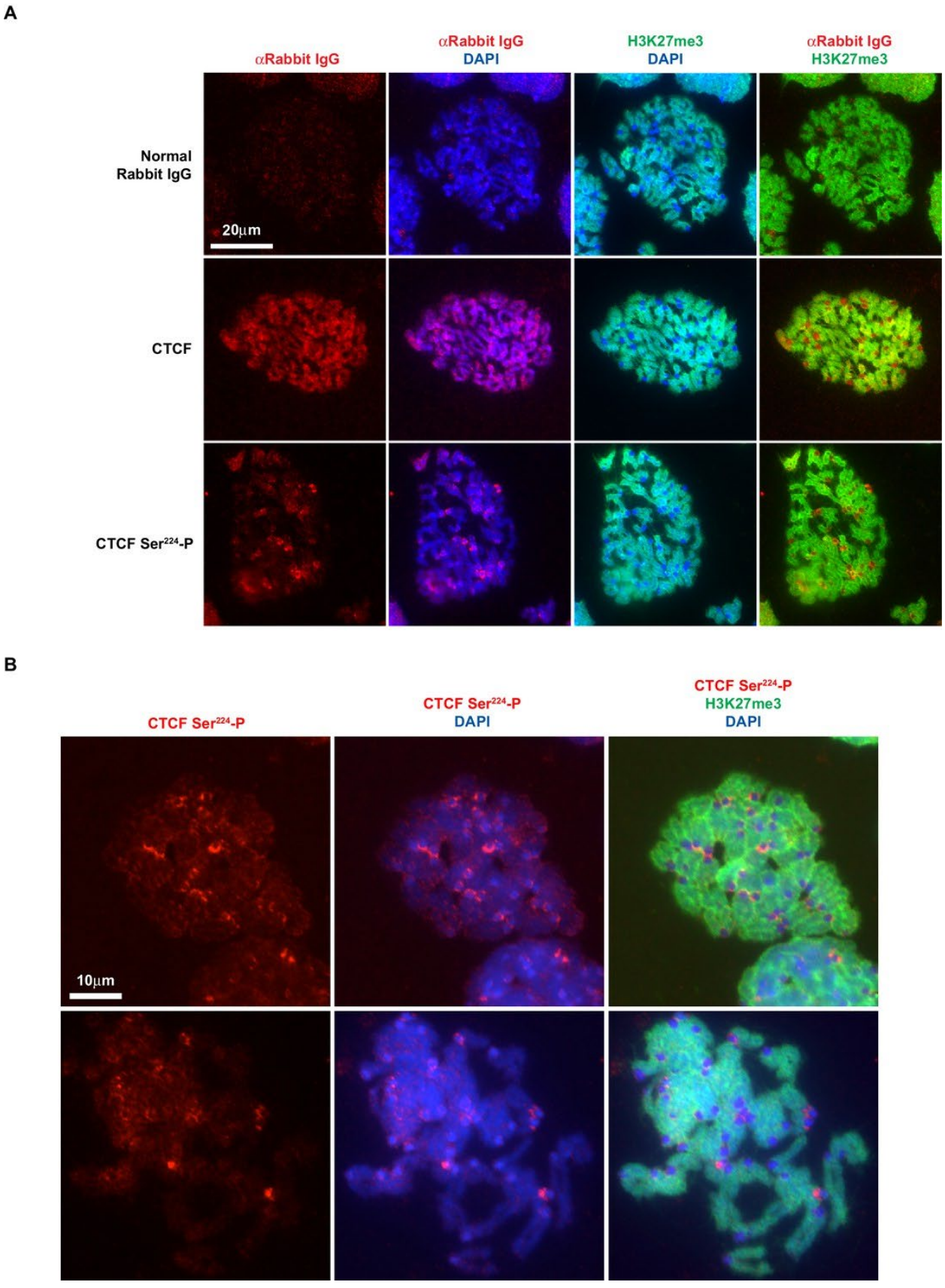

A
αRabbit IgG
αRabbit IgG
DAPI
H3K27me3
DAPI
αRabbit IgG
H3K27me3
Normal
Rabbit IgG
20μm
CTCF
CTCF Ser224-P
B
CTCF Ser224-P
CTCF Ser224-P
DAPI
CTCF Ser224-P
H3K27me3
DAPI
10μm

Figure S1.3. CTCF Ser224-P is enriched at pericentric regions of metaphase chromosomes.

(A) Immunofluorescence performed on TST-1 mESC metaphase chromosomes using the indicated antibodies. DNA stained with DAPI. CTCF Ser224-P images are replicated from Figure 3G and are shown for comparison purposes. Bar, 20μm.

(B) Two additional examples of immunofluorescence on TST-1 mESC metaphase chromosomes using anti-CTCF Ser224-P and anti-H3K27me3 antibodies. DNA stained with DAPI. Bar, 10μm.

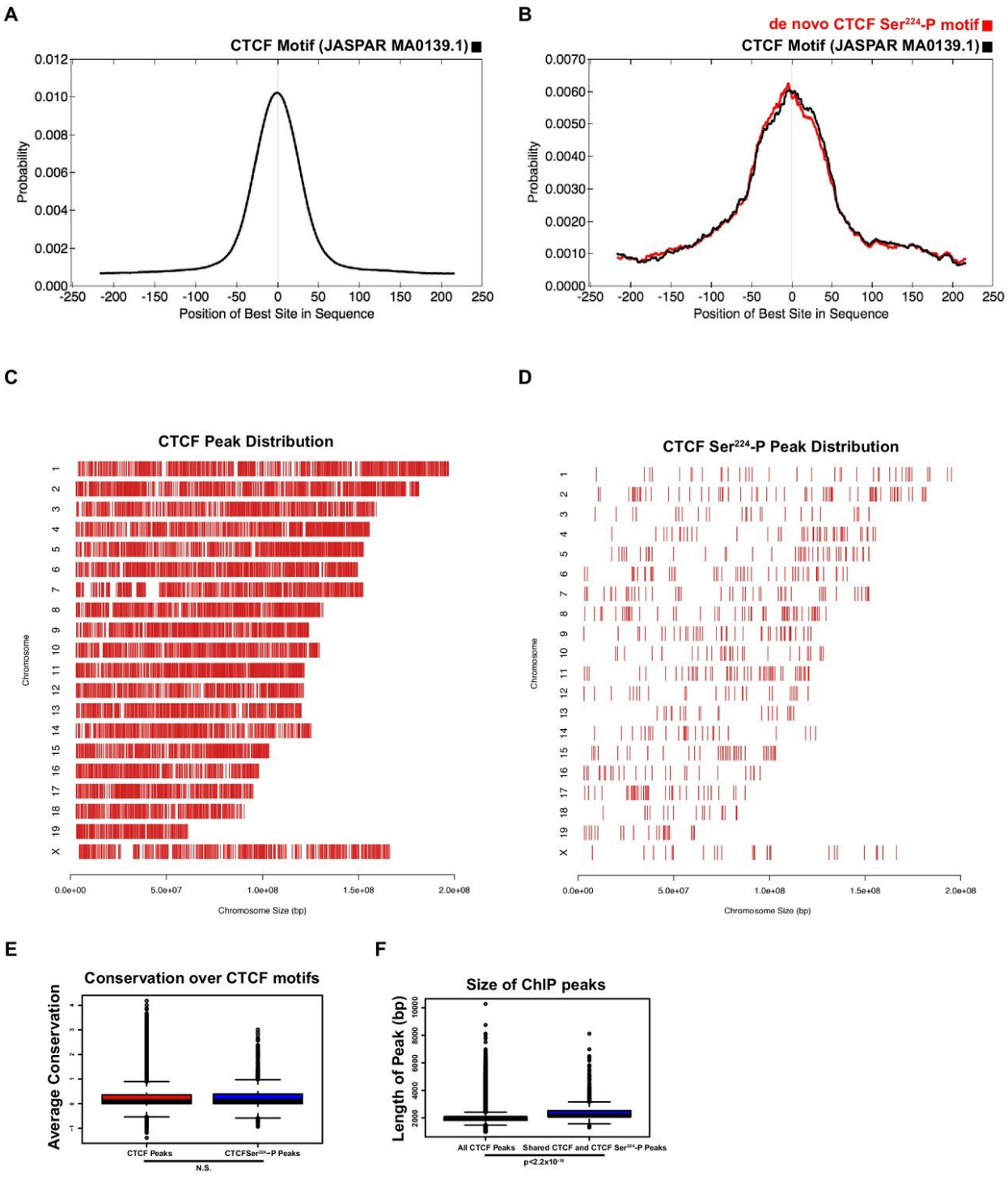


Figure S1.4. CTCF Ser224-P occupies known CTCF binding sites.

(A) CentriMo site probability distribution plot generated from CTCF ChIP-seq called peaks (z=6) with the observed consensus CTCF motif plotted (black trace).

(B) Site probability distribution plot generated from CTCF Ser224-P ChIP-seq called peaks (z=6) with both de novo identified (red) and consensus CTCF (black) motifs plotted.

(C) CTCF and CTCF Ser224-P ChIP-seq peak (z=6) distribution plots across all chromosomes.

(D) CTCF and CTCF Ser224-P ChIP-seq peak (z=6) distribution plots across all chromosomes.

(E) Average conservation of CTCF motifs falling in CTCF versus CTCF Ser224-P ChIP-seq peaks (p-value > 0.05, Wilcoxon rank sum test).

(F) Length of CTCF peaks versus shared CTCF and CTCF Ser224-P peaks (p-value < 2.2×10^{-16}, Wilcoxon rank sum test).

Supplementary Table S1.1. Antibodies and reagents used in this study

Table 1 - Antibodies and reagents used in this study			
Description	**Vendor**	**Catalog #**	**Usage**
Rabbit anti-CTCF, clone D31H2	Cell Signaling Technology	#3418	Western Blot, Immunofluoresence, ChIP-seq
Normal rabbit IgG	Cell Signaling Technology	#2729	Immunofluoresence
Rabbit anti-Phospho-CTCF (Ser224)	Cell Signaling Technology	This study	Western Blot, Immunofluoresence, ChIP-seq
Mouse anti-FLAG clone M2	Sigma-Aldrich	F3165	Western Blot, Immunofluoresence

Mouse anti-PCNA clone PC10	Cell Signaling Technology	#2586	Immunofluoresence
Rat anti-H3S10ph clone 6G8B7	Active Motif	61623	Immunofluoresence
Rabbit anti-GAPDH clone D16H11	Cell Signaling Technology	#5174	Western Blot
Mouse anti-B23 clone FC82291	Sigma-Aldrich	B0556	Immunofluoresence
Mouse anti-H3K27me3 clone 0323	Active Motif	61017	Immunofluoresence
Mouse anti-H3K9me3 clone 6F12-H4	Sigma-Aldrich	05-1242	Immunofluoresence
Mouse anti-B-actin clone AC-15	Sigma-Aldrich	A1978	Western Blot
Goat anti-mouse IgG, HRP conjugate	Promega	W4021	Western Blot
Goat anti-rabbit IgG, HRP conjugate	Promega	W4011	Western Blot
Goat anti-rabbit IgG, Alexa Fluor 555 conjugate	Thermo Fisher	A21428	Immunofluoresence
Goat anti-mouse IgG, Alexa Fluor 488 conjugate	Thermo Fisher	A11029	Immunofluoresence
Goat anti-rat IgG, Alexa Fluor 647 conjugate	Thermo Fisher	A21247	Immunofluoresence

Goat anti-mouse IgG, Alexa Fluor 647 conjugate	Thermo Fisher	A21236	Immunofluoresence
Reagent	**Vendor**	**Catalog #**	
Doxycycline	Sigma-Aldrich	D9891	
Volasertib (BI 6727)	Selleck Chemicals	S2235	
RO-3306	Selleck Chemicals	S7747	
Colcemid	Thermo Fisher	15210040	

Supplementary Table S1.2. Oligonucleotides used in this study

Table 2 - Oligonucleotides used in this study			
Identifier	**Use**	**Description**	**Sequence (5'-->3')**
oBD64	EMSA	RS14C Forward	ctatgcattaggtttgggtgttatacccgtgtaggccagcagagggtgtcggatcccaaggaaaccaagttacagacgcc
oBD65	EMSA	RS14C Reverse	ggcgtctgtaacttggtttccttgggatccgacaccctctgctggcctacacgggtataacacccaaacctaatgcatag

oBD66	EMSA	RS14C Forward mutant	ctatgcattaggtttgggtgttatacccgtgtaggcTagcaCagTgtgtcgga tcccaaggaaaccaagttacagacgcc
oBD67	EMSA	RS14C Reverse mutant	ggcgtctgtaacttggtttccttgggatccgacacActGtgctAgcctacacg ggtataacacccaaacctaatgcatag
AK226	ChIP-qPCR	AK226 Cirbp F	CTCTGGTTTCCGTCAGTGATAA
AK227	ChIP-qPCR	AK227 Cirbp R	GACTCATGTCCCACCTCTTTC
AK230	ChIP-qPCR	AK230 Ccnd2 F	CTC TTT GCG TTC TAG GCA ATT C
AK231	ChIP-qPCR	AK231 Ccnd2 R	TGA ACT GAG CCC TTT GTG AG
AK166	ChIP-qPCR	AK166 Oct4 F	TAGGTGAGCCGTCTTTCCA
AK167	ChIP-qPCR	AK167 Oct4 R	GTGAGAAGGCGAAGTCTGAAG
AK304	ChIP-qPCR	AK304 Flna F	GGGTGTCCTAGCCTTAGAGT
AK305	ChIP-qPCR	AK305 Flna R	CTCATCTGGATCTTTGGGATGG
AK308	ChIP-qPCR	AK308 Irak1 F	GGGAGTCCCTAATGGGAGAA

AK309	ChIP-qPCR	AK309 Irak1 R	CCTGTTCAGCCCTATTCAGATG
AK312	ChIP-qPCR	AK312 Ikbkg F	CAGTCAGCAGTGCTCTCTTAT
AK313	ChIP-qPCR	AK313 Ikbkg R	AATCCTCAGCAGATCCACAAG

Supplementary Table S1.3. UniProt accession numbers used in Clustal alignments

	CTCF	PLK1	ATM	GSK3B
Mouse	Q61164	Q07832	Q62388	Q9WV60
Rat	Q9R1D1	Q62673	D4ACL8	P18266
Human	P49711	P53350	Q13315	P49841
Chimpanzee	H2QBB9	G2HFG8	K7B6I0	K7CFX8
Opossum	F6QDS6	F7F258	F6YNZ6	F6PGQ1
Chicken	Q08705	Q5ZJ36	E1C0Q6	F1NPL8
Frog	Q6PF79	P70032	Q5MPF8	Q91757
Zebrafish	Q6JAA4	Q4KMI8	F1QJB6	Q9YH60
Sea Urchin	W4Y9Y1	W4YQ73	B0FWN6	W4ZIA0
Fruit Fly	Q9VS55	P52304	Q5EAK6	P18431
Water Flea	E9HLQ9	E9G7B9	E9GRN4	E9H6Q6

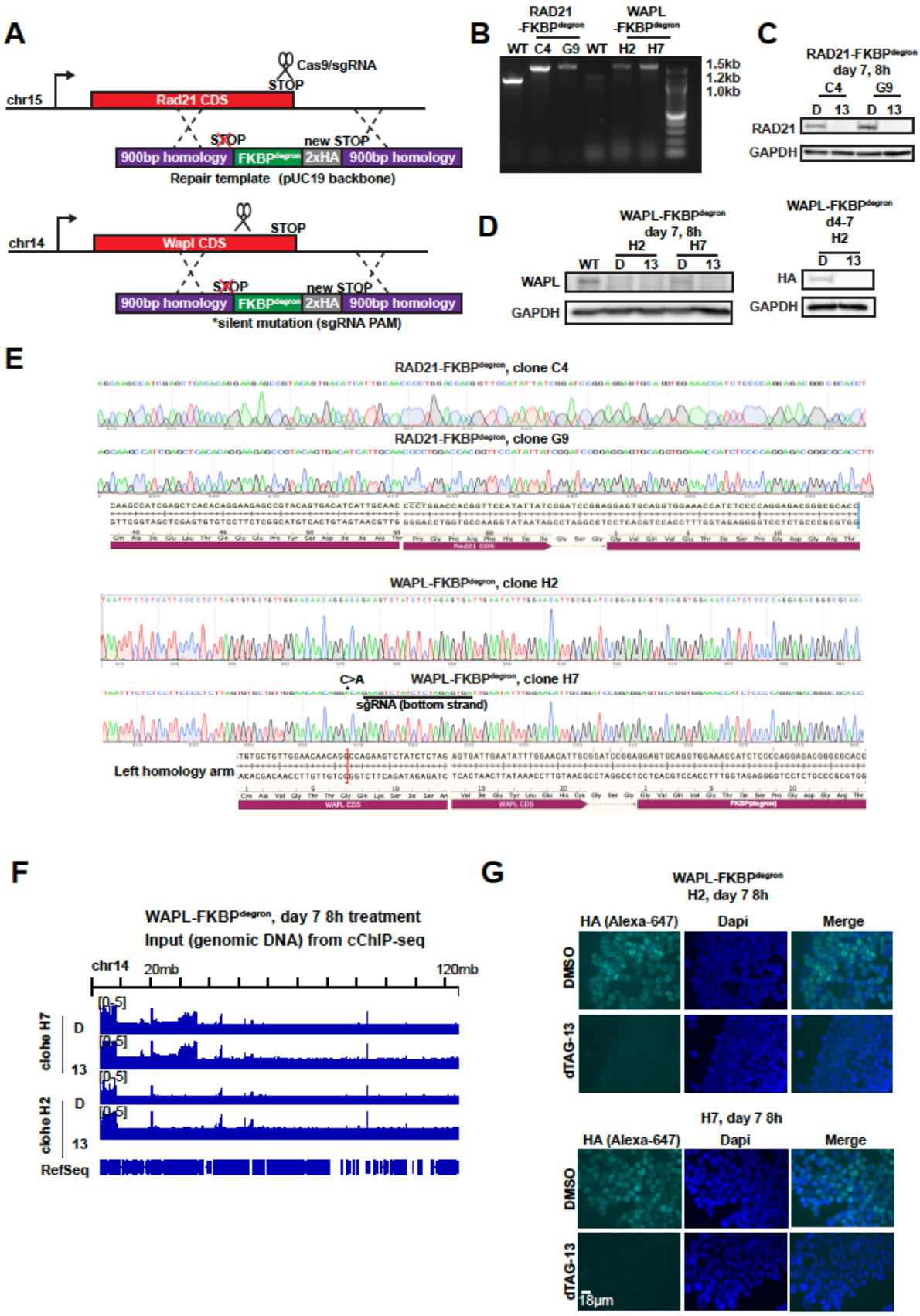

A
Cas9/sgRNA
STOP
chr15
Rad21 CDS
STOP
new STOP
900bp homology
FKBPdegron
2xHA
900bp homology
Repair template (pUC19 backbone)
chr14
Wapl CDS
STOP
STOP
new STOP
900bp homology
FKBPdegron
2xHA
900bp homology
*silent mutation (sgRNA PAM)
B
RAD21 -FKBPdegron
WAPL -FKBPdegron
WT C4 G9 WT H2 H7
1.5kb
1.2kb
1.0kb
C
RAD21-FKBPdegron day 7, 8h
C4
G9
D 13 D 13
RAD21
GAPDH
D
WAPL-FKBPdegron day 7, 8h
H2
H7
WT D 13 D 13
WAPL
GAPDH
WAPL-FKBPdegron d4-7 H2
D 13
HA
GAPDH
E
RAD21-FKBPdegron, clone C4
RAD21-FKBPdegron, clone G9
WAPL-FKBPdegron, clone H2
C>A
WAPL-FKBPdegron, clone H7
sgRNA (bottom strand)
Left homology arm
F
WAPL-FKBPdegron, day 7 8h treatment
Input (genomic DNA) from cChIP-seq
chr14
20mb
120mb
clone H7
D
13
clone H2
D
13
RefSeq
G
WAPL-FKBPdegron H2, day 7 8h
HA (Alexa-647)
Dapi
Merge
DMSO
dTAG-13
H7, day 7 8h
HA (Alexa-647)
Dapi
Merge
DMSO
dTAG-13
18μm

Figure S2.1. Generation of FKBP^{degron} lines.

(A) Outline of protocol used to generate RAD21 and WAPL-FKBP^{degron} mESCs

(B) PCR amplifying knock-in site of parental TST mESC lines (WT) and RAD21 and WAPL-FKBP^{degron} clones. One PCR primer was located outside of the homology arms to ensure that only the endogenous locus was amplified. FKBP^{degron} tag is ~400bp.

(C) Western blots for RAD21 and GADPH loading control in RAD21-FKBP^{degron} lines at day 7 of differentiation. Samples treated with DMSO vehicle control (D) or 500nM dTAG-13 (13).

(D) Western blots for WAPL and GADPH loading control in WAPL-FKBP^{degron} lines at day 7 of differentiation. Samples treated with DMSO vehicle control (D) or 500nM dTAG-13 (13).

(E) Sanger sequencing of RAD21-FKBP^{degron} and WAPL-FKBP^{degron} lines.

(F) Comparison of genomic DNA (input DNA) in WAPL-FKBP^{degron} lines at chromosome 14. Sharp dip in WAPL-FKBP^{degron}, clone H7 input corresponds to the end of the right 900bp homology arm used for knock-in of the tag. Reads mapping to the plasmid backbone used for the homology repair template were also found here, although they did not appear to be incorporated into the main Wapl transcript.

(G) Immunofluorescence of WAPL-HA in WAPL-FKBP^{degron} lines at day 7 of differentiation. Samples treated with DMSO vehicle control (D) or 500nM dTAG-13 (13).

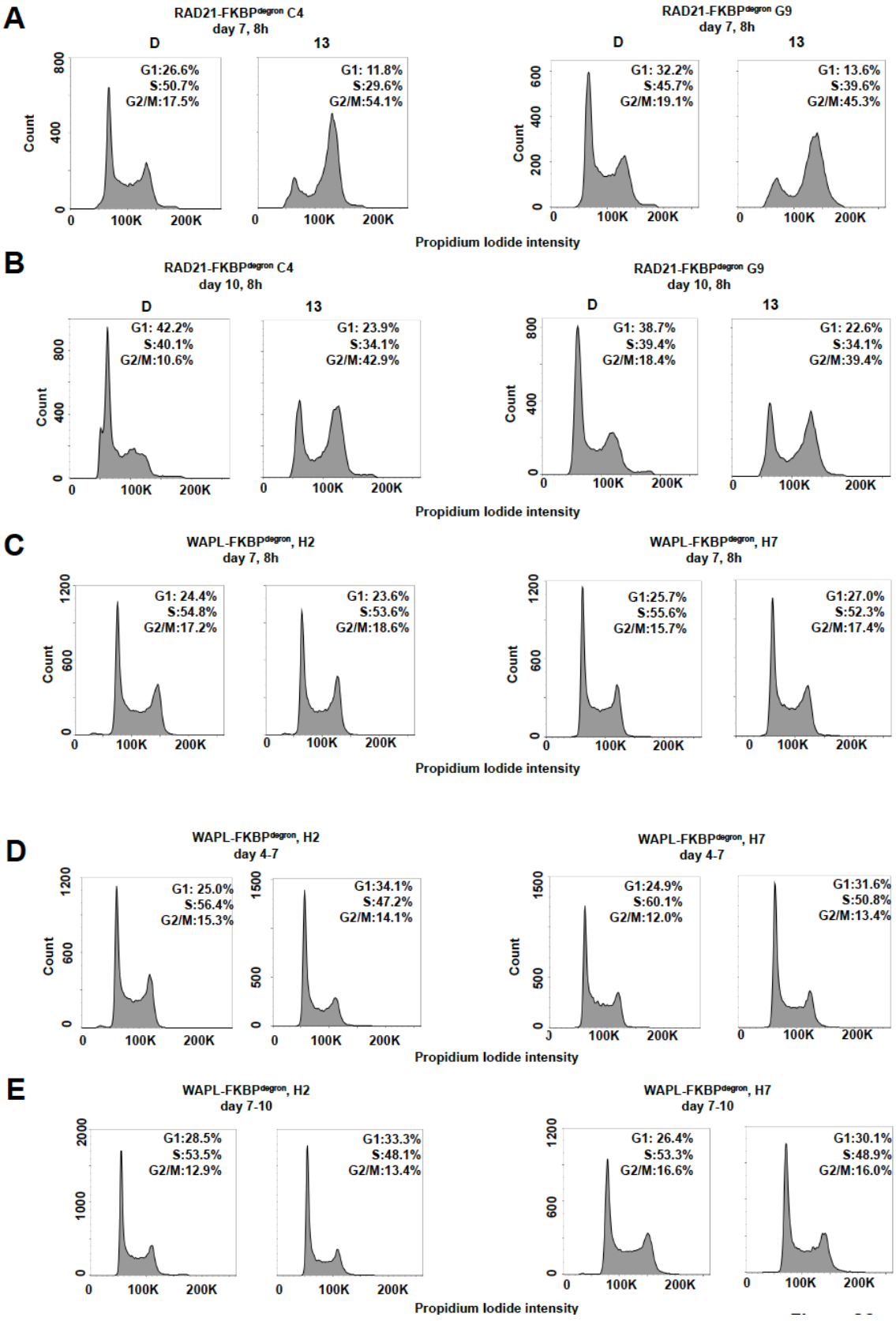

A
RAD21-FKBPdegron C4
day 7, 8h
D
13
G1:26.6%
S:50.7%
G2/M:17.5%
G1: 11.8%
S:29.6%
G2/M:54.1%
RAD21-FKBPdegron G9
day 7, 8h
D
13
G1: 32.2%
S:45.7%
G2/M:19.1%
G1: 13.6%
S:39.6%
G2/M:45.3%
Count
Propidium Iodide intensity
B
RAD21-FKBPdegron C4
day 10, 8h
D
13
G1: 42.2%
S:40.1%
G2/M:10.6%
G1: 23.9%
S:34.1%
G2/M:42.9%
RAD21-FKBPdegron G9
day 10, 8h
D
13
G1: 38.7%
S:39.4%
G2/M:18.4%
G1: 22.6%
S:34.1%
G2/M:39.4%
Count
Propidium Iodide intensity
C
WAPL-FKBPdegron, H2
day 7, 8h
G1: 24.4%
S:54.8%
G2/M:17.2%
G1: 23.6%
S:53.6%
G2/M:18.6%
WAPL-FKBPdegron, H7
day 7, 8h
G1:25.7%
S:55.6%
G2/M:15.7%
G1:27.0%
S:52.3%
G2/M:17.4%
Count
Propidium Iodide intensity
D
WAPL-FKBPdegron, H2
day 4-7
G1: 25.0%
S:56.4%
G2/M:15.3%
G1:34.1%
S:47.2%
G2/M:14.1%
WAPL-FKBPdegron, H7
day 4-7
G1:24.9%
S:60.1%
G2/M:12.0%
G1:31.6%
S:50.8%
G2/M:13.4%
Count
Propidium Iodide intensity
E
WAPL-FKBPdegron, H2
day 7-10
G1:28.5%
S:53.5%
G2/M:12.9%
G1:33.3%
S:48.1%
G2/M:13.4%
WAPL-FKBPdegron, H7
day 7-10
G1: 26.4%
S:53.3%
G2/M:16.6%
G1:30.1%
S:48.9%
G2/M:16.0%
Propidium Iodide intensity

Figure S2.2. Cell cycle profiling of FKBP^{degron} lines.

(A) DNA content profiles via propidium iodide staining of RAD21-FKBP^{degron} lines at day 7 of differentiation. Samples treated with DMSO vehicle control (D) or 500nM dTAG-13 (13) as indicated.

(B) DNA content profiles via propidium iodide staining of RAD21-FKBP^{degron} lines at day 10 of differentiation.

(C) DNA content profiles via propidium iodide staining of WAPL-FKBP^{degron} lines at day 7 of differentiation.

(D) DNA content profiles via propidium iodide staining of WAPL-FKBP^{degron} lines at day 7 of differentiation.

(E) DNA content profiles via propidium iodide staining of WAPL-FKBP^{degron} lines at day 10 of differentiation.

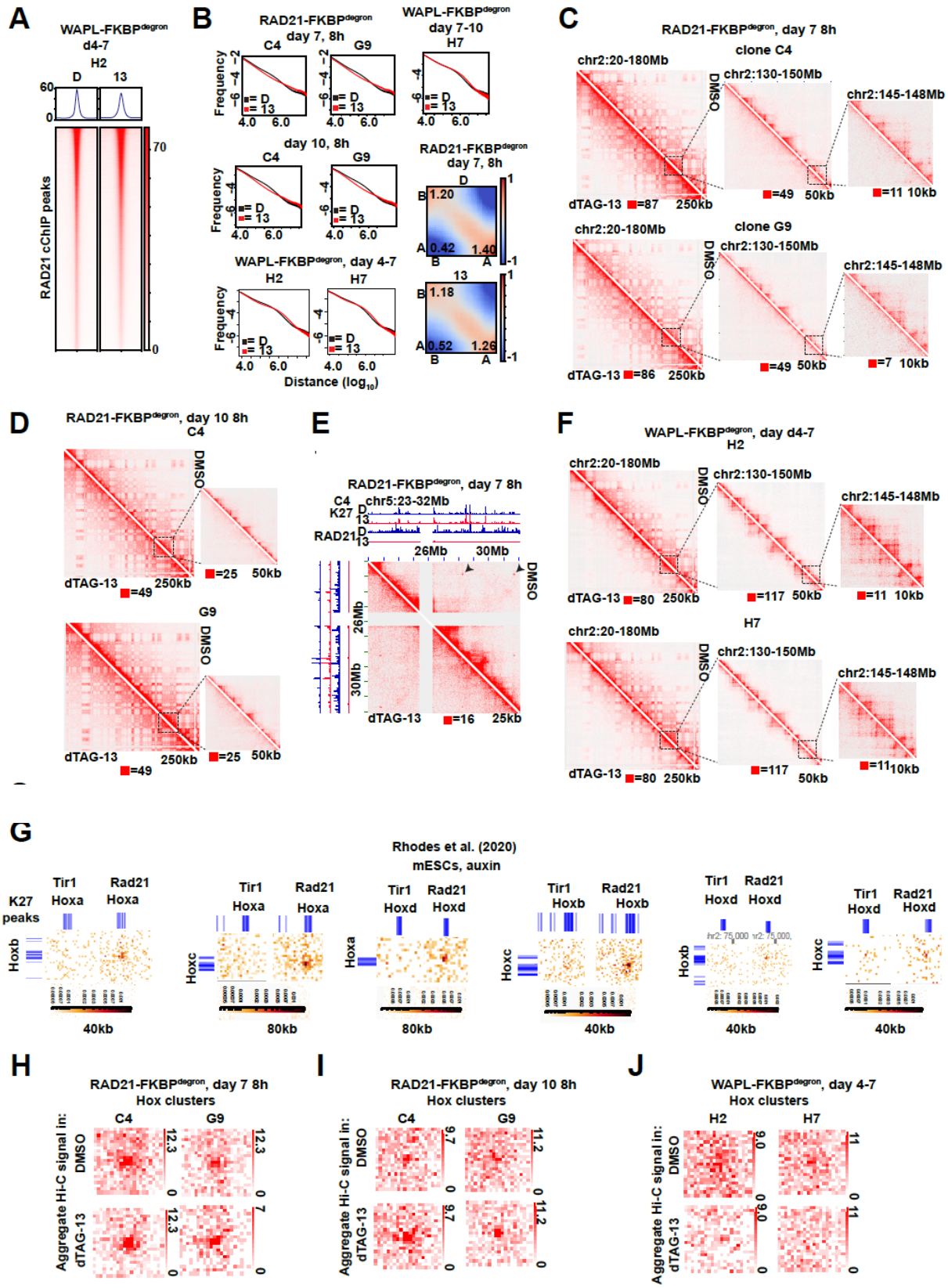

A
WAPL-FKBP^degron
d4-7
H2
D
13
60
0
70
0
RAD21 cChIP peaks
B
RAD21-FKBP^degron
day 7, 8h
C4
G9
WAPL-FKBP^degron
day 7-10
H7
day 10, 8h
C4
G9
RAD21-FKBP^degron
day 7, 8h
WAPL-FKBP^degron, day 4-7
H2
H7
Frequency
D
13
4.0
6.0
1.20
0.42
1.40
1.18
0.52
1.26
Distance (log10)
C
RAD21-FKBP^degron, day 7 8h
clone C4
chr2:20-180Mb
chr2:130-150Mb
chr2:145-148Mb
DMSO
dTAG-13
=87
250kb
=49
50kb
=11
10kb
clone G9
=86
=49
=7
D
RAD21-FKBP^degron, day 10 8h
C4
G9
=25
=16
E
RAD21-FKBP^degron, day 7 8h
C4
chr5:23-32Mb
K27
RAD21
26Mb
30Mb
F
WAPL-FKBP^degron, day d4-7
H2
H7
=80
=117
=11
G
Rhodes et al. (2020)
mESCs, auxin
K27 peaks
Tir1
Rad21
Hoxa
Hoxb
Hoxc
Hoxd
40kb
80kb
H
RAD21-FKBP^degron, day 7 8h
Hox clusters
C4
G9
Aggregate Hi-C signal in:
DMSO
dTAG-13
12.3
I
RAD21-FKBP^degron, day 10 8h
Hox clusters
6.7
9.7
11.2
J
WAPL-FKBP^degron, day 4-7
Hox clusters
H2
H7
9.0
11

Figure S2.3. Replication of architectural changes across FKBP^{degron} clones.

(A) Metaplot and heatmap of RAD21 binding by cChIP-seq in WAPL-FKBP^{degron} clone H2.

(B) Frequency versus distance plots for Hi-C interaction maps across individual FKBP^{degron} clones at all time points used in this study. Samples treated with DMSO vehicle control 500nM dTAG-13 as indicated. Saddle plot analysis of compartment strength in RAD21-FKBP^{degron} Hi-C maps at day 7 (clones merged) also shown. Scores represent B-B, B-A, and A-A compartment interaction strength.Representative sections of day 7 Hi-C maps at different resolutions for RAD21-FKBP^{degron} clones.

(C) Representative sections of day 10 Hi-C maps at different resolutions for RAD21-FKBP^{degron} clones.

(D) Section of day 7 RAD21-FKBP^{degron} Hi-C maps (clones merged) displaying example of cohesin-independent loops not anchored by H3K27me3 (arrows).

(E) Example of non-cohesin and non-H3K27me3 anchored loops (black arrows) persisting after degradation in day 7 RAD21-FKBP^{degron} Hi-C maps (clones merged).

(F) Representative sections of day 7 Hi-C maps at different resolutions for WAPL-FKBP^{degron} clones.

(G) Aggregate Peak Analysis plots of aggregate RAD21-FKBP^{degron} Hi-C signal at day 7 (clones merged) over regions anchored by Hox clusters.

(H) Aggregate Peak Analysis plots of aggregate RAD21-FKBP^{degron} Hi-C signal at day 10 (clones merged) over regions anchored by Hox clusters.

(I) Aggregate Peak Analysis plots of aggregate WAPL-FKBP^{degron} Hi-C signal at day 7 (clones merged) over regions anchored by Hox clusters.

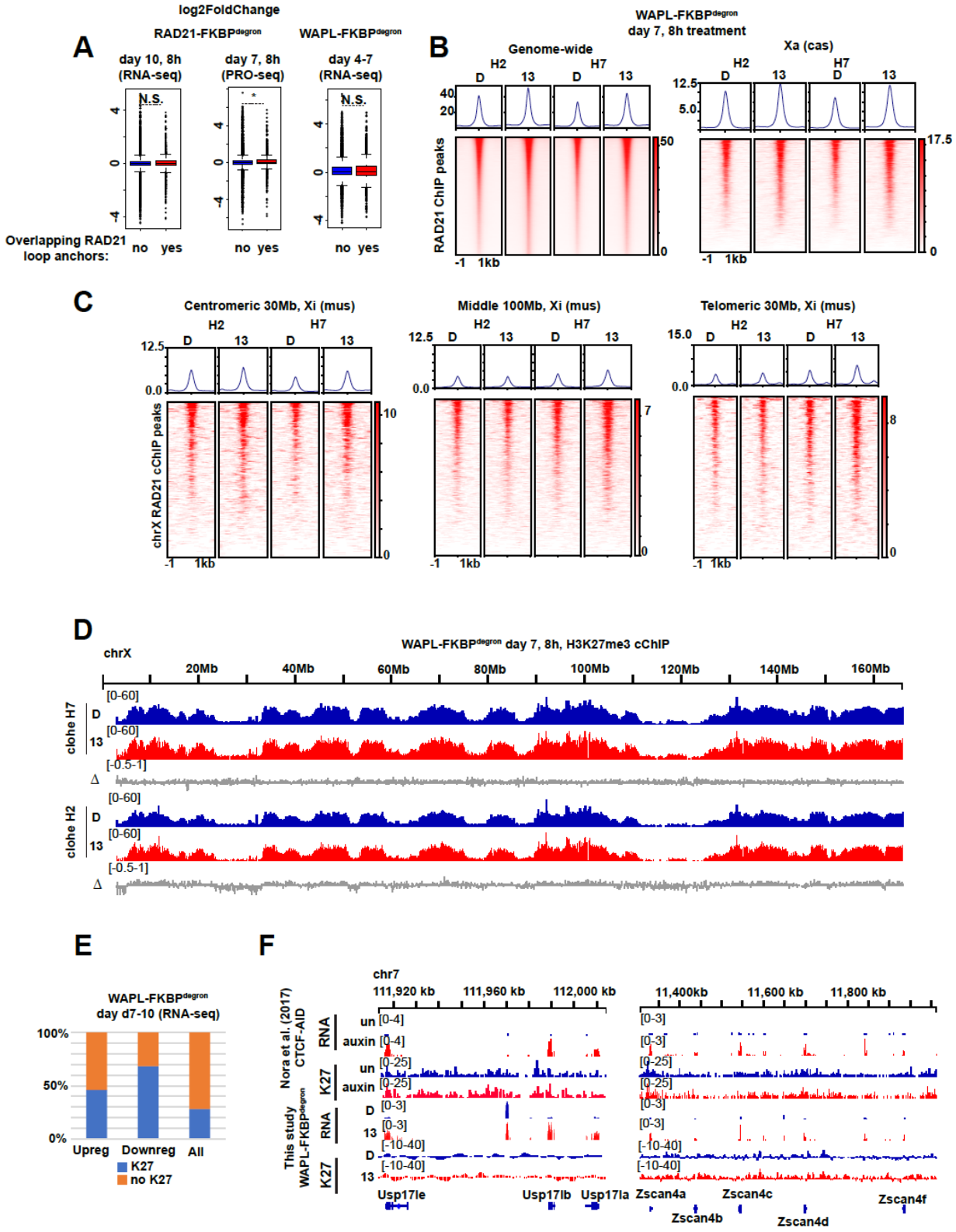

A
log2FoldChange
RAD21-FKBP^degron
WAPL-FKBP^degron
day 10, 8h (RNA-seq)
day 7, 8h (PRO-seq)
day 4-7 (RNA-seq)
N.S.
Overlapping RAD21 loop anchors:
no yes
B
WAPL-FKBP^degron day 7, 8h treatment
Genome-wide
Xa (cas)
H2
H7
D
13
RAD21 ChIP peaks
-1 1kb
C
Centromeric 30Mb, Xi (mus)
Middle 100Mb, Xi (mus)
Telomeric 30Mb, Xi (mus)
chrX RAD21 cChIP peaks
D
chrX
20Mb 40Mb 60Mb 80Mb 100Mb 120Mb 140Mb 160Mb
WAPL-FKBP^degron day 7, 8h, H3K27me3 cChIP
clone H7
clone H2
E
WAPL-FKBP^degron day d7-10 (RNA-seq)
100%
50%
0%
Upreg Downreg All
K27
no K27
F
chr7
111,920 kb 111,960 kb 112,000 kb
11,400kb 11,600 kb 11,800 kb
Nora et al. (2017) CTCF-AID
This study WAPL-FKBP^degron
RNA
K27
un
auxin
Usp17le Usp17lb Usp17la Zscan4a Zscan4c Zscan4f Zscan4b Zscan4d

Figure S2.4: Additional analysis of cChIP after short term WAPL degradation and gene expression changes after 3 day degradation.

(A) Log2-fold (dTAG-13/DMSO) change in FKBP^{degron} lines of genes overlapping RAD21-bound anchors of loops. Loops called in respective DMSO control samples. Significant p-values by Wilcoxon rank test indicated by * (p<2.772e-06). N.S., not significant.

(B) Metaplot and heatmaps of RAD21 binding by H3K27me3 cChIP-seq after 8 hours of WAPL degradation genome-wide (left) and on the active X chromosome (right).

(C) Metaplot and heatmaps of RAD21 binding by H3K27me3 cChIP after 8 hours of WAPL on the centromeric 30Mb, middle 100Mb, and telomeric 30Mb of the inactive X chromosome.

(D) Tracks of H3K27me3 by cChIP-seq on the X chromosome after 8 hours of WAPL degradation or DMSO control treatment. Log fold ratio tracks (dTAG-13/DMSO) are also shown (Δ).

(E) Genes upregulated or downregulated upon WAPL degradation from days 7-10 (relative to DMSO control) overlapping with H3K27me3 cChIP-seq peaks. All genes shown for reference.

(F) RNA-seq and H3K27me3 ChIP-seq over the *Usp17* and *Zscan4* gene clusters in CTCF-AID degron lines after 4 days of auxin treatment or untreated cells (Nora et al., 2017) or after 3 days of DMSO or dTAG-13 treatment (500nM) in WAPL-FKBP^{degron} cells early in differentiation (days 4-7). Note that cohesin was recently implicated in regulation of such 2-cell genes (Zhang et al., 2020).

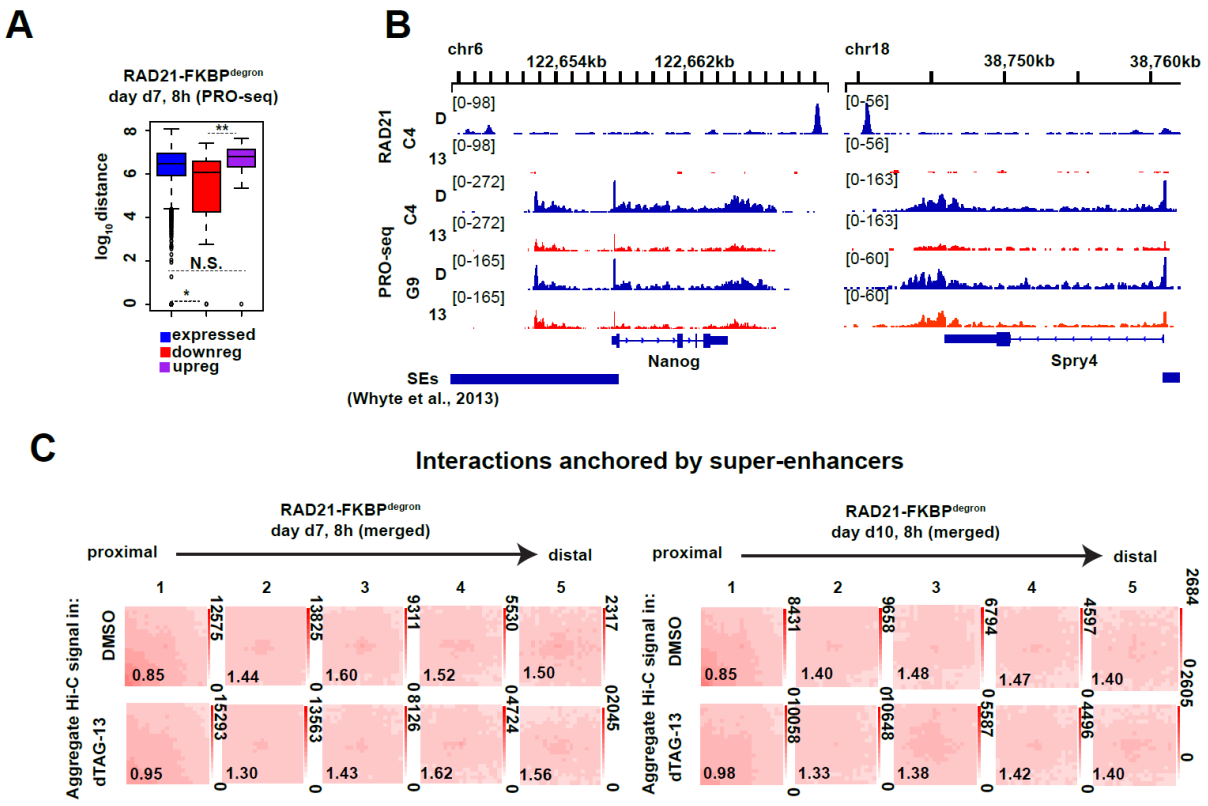


Figure S2.5. Cohesin significantly contributes to proper upregulation of proximal super-enhancer targets without affecting long-range super-enhancer interactions.

(A) Distance to nearest super-enhancer (Whyte et al. 2013) for expressed genes upregulated or downregulated upon RAD21 degradation (relative to DMSO). Expressed genes shown for reference. Significant p-values by Wilcoxon rank test indicated by * ($p<0.006057$), ** ($p<0.01569$). N.S., not significant.

(B) Day 7 RAD21-FKBP^{degron} PRO-seq and RAD21 cChIP-seq signal over representative super-enhancer targets downregulated upon 8 hours RAD21 degradation (relative to DMSO).

(C) Aggregate Peak Analysis plots of aggregate RAD21-FKBP^{degron} Hi-C signal (clones merged) over regions anchored by super-enhancers. Super-enhancers segregated by distance into 5 quantiles.

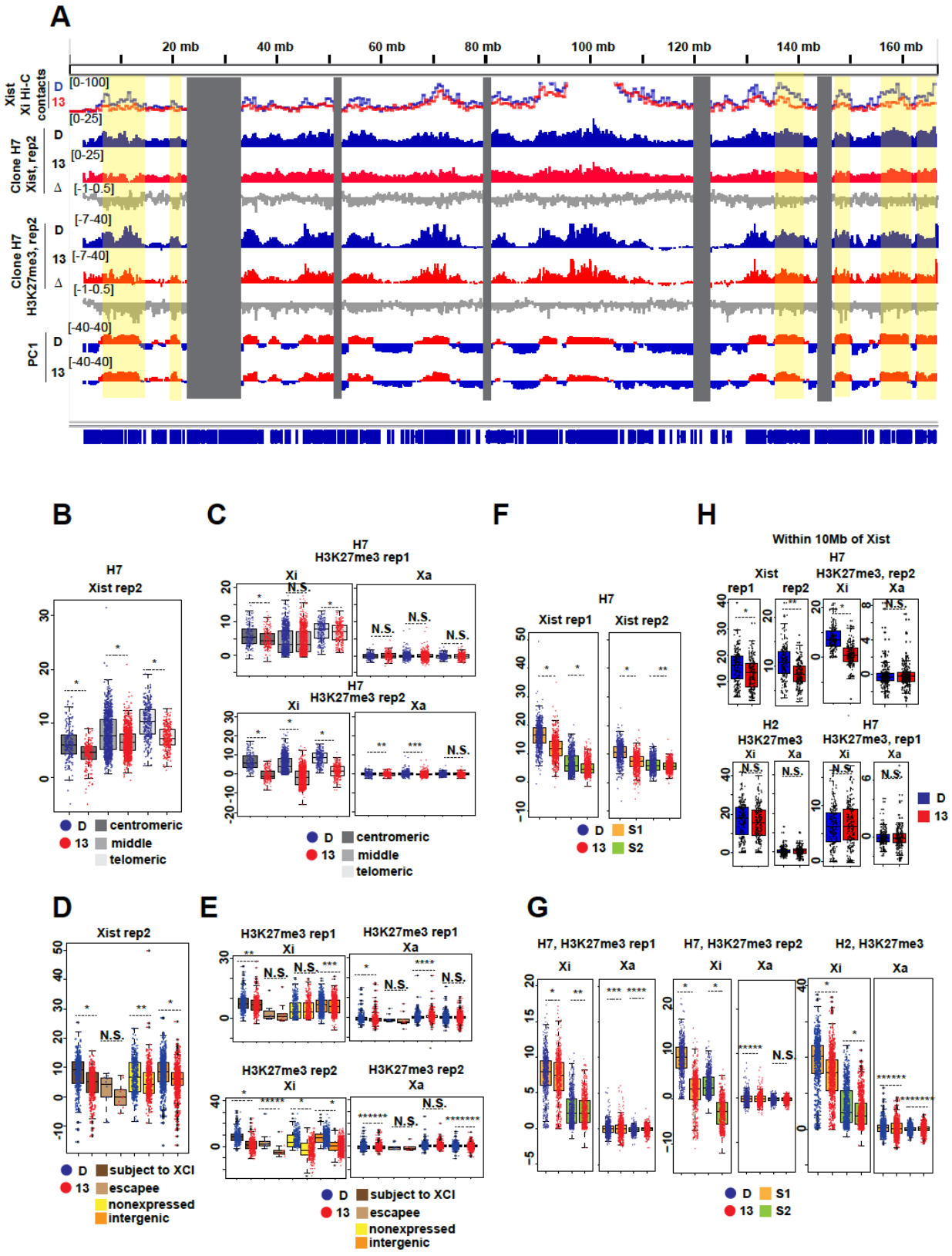

A
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
140 mb
160 mb
Xist Xi Hi-C contacts
D
13
[0-100]
[0-25]
Clone H7 Xist, rep2
[-1-0.5]
Clone H7 H3K27me3, rep2
[-7-40]
[-40-40]
PC1
B
H7
Xist rep2
centromeric
middle
telomeric
C
H7
H3K27me3 rep1
H7
H3K27me3 rep2
Xi
Xa
N.S.
D
13
S1
S2
D
E
H3K27me3 rep1
H3K27me3 rep2
subject to XCI
escapee
nonexpressed
intergenic
F
H7
Xist rep1
Xist rep2
G
H7, H3K27me3 rep1
H7, H3K27me3 rep2
H2, H3K27me3
H
Within 10Mb of Xist
H7
Xist
rep1
rep2
H7
H3K27me3, rep2
H2
H3K27me3
H7
H3K27me3, rep1

Figure S2.6. Additional biological replicate for Xist CHART-seq and H3K27me3 cChIP-seq to examine spreading defects.

(A) Tracks of CHART-seq (Xist) and H3K27me3 cChIP-seq over the X chromosome and allelic RAD21 cChIP over the Xa and Xi in WAPL-FKBP^{degron} lines treated from days 4-7 of differentiation with DMSO (D) or 500nM dTAG-13 (13). Log fold ratio tracks (dTAG-13/DMSO) are also shown (Δ). Principal component corresponding to compartments on the Xi in same samples shown for reference (PC1). Hi-C contacts to the 1Mb bin containing the Xist transcription locus in same conditions also shown.

(B) Boxplot showing Xist coverage over 100kb bins in the centromeric most 30Mb, middle 100Mb, and telomeric most 30Mb of the X chromosome in WAPL-FKBP^{degron} lines treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * (p<2.2e-16).

(C) Boxplots showing allelic H3K27me3 coverage over 100kb bins in the centromeric most 30Mb, middle 100Mb, and telomeric most 30Mb of the X chromosome in WAPL-FKBP^{degron} lines treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * (p<2.2e-16), ** (p<0.03), and *** (p<2.2e-05). N.S., not significant.

(D) Boxplot showing average Xist density over different classes of genes on the X chromosome in WAPL-FKBP^{degron} lines treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * (p<2.2e-16) and ** (p<1.049e-14). N.S., not significant.

(E) Boxplot showing average allelic H3K27me3 density over different classes of genes on the X chromosome in WAPL-FKBP^{degron} lines treated as in (A). Significant p-values by Wilcoxon signed rank test indicated by * (p< 2.2e-16), ** (p< 2.79e-15), *** (p< 9.04e-13), **** (p< 0.001573), ***** (p< 0.001954), ****** (p< 2.239e-05), ******* (p< 1.308e-05). N.S., not significant.

(F) Boxplot showing Xist average density over 100kb bins in S1 and S2 compartments. Significant p-values by Wilcoxon signed rank test indicated by * ($p<2.2e-16$) and ** ($p<4.13e-09$).

(G) Boxplot showing average allelic H3K27me3 density over 100kb bins in S1 and S2 compartments. Significant p-values by Wilcoxon signed rank test indicated by * ($p<2.2e-16$), **($p<0.0082$), *** ($p< 0.005$), ****($p<1.86e-10$), ***** ($p<1.89e-08$), ****** ($p<2.127e-05$), ******* ($p<0.0008859$).

(H) Boxplots showing average Xist and allelic H3K27me3 density over 100kb within 10Mb of the Xist transcription locus. Significant p-values by Wilcoxon signed rank test indicated by * ($p<1.543e-05$), **($p<4.026e-13$).

Sample	no. of read pairs	% of read pairs with chimeric MboI site (GATCGATC)	no. of PE150 read pairs with chimeric MboI site	GATCGATC) read pairs with chimeric MboI site (	alignment rate	no. of aligned reads	unique alignment rate	no. of uniquely aligned reads	of contacts in filtered HOMER	contacts that a	nondistra-d
RAD21-FKBP(degron) clone C4, DMSO d10 8h	120118372	47% (PE150 reads), 18% (PE50 reads)	18998084	14345445	99%	118917188	90%	108106535	119488850	33%	80057530
RAD21-FKBP(degron) clone C4, dTAG-13 d10 8h	124312167	46% (PE150 reads), 18% (PE50 reads)	16052619	16094731	99%	123009101	89%	110951355	126756382	33%	84926722
RAD21-FKBP(degron) clone G9, DMSO d10 8h	216164058	34% (PE150 reads), 13% (PE50 reads)	35434391	14552884	99%	214109341	90%	193676470	183410233	33%	122884856
RAD21-FKBP(degron) clone G9, dTAG-13 d10 8h	107706572	45% (PE150 reads), 18% (PE50 reads)	20968916	10999616	99%	106488513	89%	95769574	119511806	35%	77682674
RAD21-FKBP(degron) clone C4, DMSO d7 8h	281194186	43% (PE150 reads)	120913500	N/A	98%	275570302	88%	248575660	259364409	21%	205197242
RAD21-FKBP(degron) clone C4, dTAG-13 d7 8h	227028344	39% (PE150 reads)	88541054	N/A	98%	221806692	88%	199330886	184868298	21%	146240234
RAD21-FKBP(degron) clone G9, DMSO d7 8h	223560051	44% (PE150 reads)	98366422	N/A	98%	219759530	89%	198521325	210196517	21%	166252329
RAD21-FKBP(degron) clone G9, dTAG-13 d7 8h	141971850	46% (PE150 reads)	65307051	N/A	98%	138706497	88%	124367341	128115247	22%	100543142
WAPL-FKBP(degron) clone H2, DMSO d4-7	359384550	47% (PE150 reads)	168910739	N/A	97%	347165475	87%	312664559	368215403	20%	292933801
WAPL-FKBP(degron) clone H2, dTAG-13 d4-7	253407556	45% (PE150 reads)	114033400	N/A	98%	248339405	87%	219957759	216182229	21%	171024399
WAPL-FKBP(degron) clone H7, DMSO d4-7	246580744	44% (PE150 reads), 15% (PE50 reads)	37255141	24286495	96%	236224353	87%	213785505	271185427	24%	206695905
WAPL-FKBP(degron) clone H7, dTAG-13 d4-7	253955228	44% (PE150 reads), 15% (PE50 reads)	62916492	16644480	96%	243035153	86%	219163362	260295369	24%	198307823
WAPL-FKBP(degron) clone H7, DMSO d7-10	91025468	15% (PE50 reads)	N/A	13653820	98%	89412171	82%	74481884	98157383	25%	74001911
WAPL-FKBP(degron) clone H7, dTAG-13 d7-10	133187015	15% (PE50 reads)	N/A	19978052	98%	130641770	82%	108926950	139191310	22%	108689079

Supplemental Table S2.1 Statistics for Hi-C datasets

References

Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311-316.

Andergassen, D., Smith, Z.D., Lewandowski, J.P., Gerhardinger, C., Meissner, A., and Rinn, J.L. (2019). In vivo Firre and Dxz4 deletion elucidates roles for autosomal gene regulation. bioRxiv, 612440.

Arnould, C., Rocher, V., Finoux, A.L., Clouaire, T., Li, K., Zhou, F., Caron, P., Mangeot, P.E., Ricci, E.P., Mourad, R.*, et al.* (2021). Loop extrusion as a mechanism for formation of DNA damage repair foci. Nature *590*, 660-665.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research *37*, W202-208.

Balaton, B.P., and Brown, C.J. (2016). Escape Artists of the X Chromosome. Trends Genet *32*, 348-359.

Ballas, N., Lioy, D.T., Grunseich, C., and Mandel, G. (2009). Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. Nature neuroscience *12*, 311-317.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.

Bansal, P., Kondaveeti, Y., and Pinter, S.F. (2019). Forged by DXZ4, FIRRE, and ICCE: How Tandem Repeats Shape the Active and Inactive X Chromosome. Front Cell Dev Biol *7*, 328.

Barr, F.A., Sillje, H.H., and Nigg, E.A. (2004). Polo-like kinases and the orchestration of cell division. Nat Rev Mol Cell Biol *5*, 429-440.

Barr, M.L., and Bertram, E.G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature *163*, 676.

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R.*, et al.* (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature *543*, 519-524.

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

Belyaev, N., Keohane, A.M., and Turner, B.M. (1996). Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells. Hum Genet *97*, 573-578.

Berletch, J.B., Ma, W., Yang, F., Shendure, J., Noble, W.S., Disteche, C.M., and Deng, X. (2015). Escape from X inactivation varies in mouse tissues. PLoS genetics *11*, e1005079.

Berletch, J.B., Yang, F., Xu, J., Carrel, L., and Disteche, C.M. (2011). Genes that escape from X inactivation. Hum Genet *130*, 237-245.

Bintu, B., Mateo, L.J., Su, J.H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science *362*.

Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N.*, et al.* (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nature genetics *40*, 663-669.

Boggs, B.A., Connors, B., Sobel, R.E., Chinault, A.C., and Allis, C.D. (1996). Reduced levels of histone H3 acetylation on the inactive X chromosome in human females. Chromosoma *105*, 303-309.

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

Bonora, G., Deng, X., Fang, H., Ramani, V., Qiu, R., Berletch, J.B., Filippova, G.N., Duan, Z., Shendure, J., Noble, W.S.*, et al.* (2018). Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. Nat Commun *9*, 1445.

Bravo, R., and Macdonald-Bravo, H. (1985). Changes in the nuclear distribution of cyclin (PCNA) but not its synthesis depend on DNA replication. EMBO J *4*, 655-661.

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*, 329-331.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38-44.

Buheitel, J., and Stemmann, O. (2013). Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. EMBO J *32*, 666-676.

Burke, L.J., Zhang, R., Bartkuhn, M., Tiwari, V.K., Tavoosidana, G., Kurukuti, S., Weth, C., Leers, J., Galjart, N., Ohlsson, R.*, et al.* (2005). CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. The EMBO journal *24*, 3291-3300.

Calabrese, J.M., Sun, W., Song, L., Mugford, J.W., Williams, L., Yee, D., Starmer, J., Mieczkowski, P., Crawford, G.E., and Magnuson, T. (2012). Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell *151*, 951-963.

Carrette, L.L.G., Wang, C.Y., Wei, C., Press, W., Ma, W., Kelleher, R.J., 3rd, and Lee, J.T. (2018). A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. Proc Natl Acad Sci U S A *115*, E668-E675.

Chan, K.L., Roig, M.B., Hu, B., Beckouet, F., Metson, J., and Nasmyth, K. (2012). Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. Cell *150*, 961-974.

Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S., and Lee, J.T. (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. Science (New York, NY *295*, 345-347.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes & development *20*, 2223-2237.

Colognori, D., Sunwoo, H., Kriz, A.J., Wang, C.Y., and Lee, J.T. (2019). Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. Mol Cell *74*, 101-117 e110.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240-244.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet *2*, 292-301.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. The Journal of cell biology *153*, 773-784.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nature genetics *22*, 323-324.

Cuadrado, A., Gimenez-Llorente, D., Kojic, A., Rodriguez-Corsino, M., Cuartero, Y., Martin-Serrano, G., Gomez-Lopez, G., Marti-Renom, M.A., and Losada, A. (2019). Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and Polycomb Domains in Embryonic Stem Cells. Cell Rep *27*, 3500-3510 e3504.

Curran, T.G., Bryson, B.D., Reigelhaupt, M., Johnson, H., and White, F.M. (2013). Computer aided manual validation of mass spectrometry-based proteomic data. Methods *61*, 219-226.

Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.C., Sanborn, A.L., Machol, I., Shamim, M.*, et al.* (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proceedings of the National Academy of Sciences of the United States of America *113*, E4504-4512.

Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.M. (2019). DNA loop extrusion by human cohesin. Science *366*, 1338-1345.

de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M.*, et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell *7*, 663-676.

de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol Cell *60*, 676-684.

Del Rosario, B.C., Kriz, A.J., Del Rosario, A.M., Anselmo, A., Fry, C.J., White, F.M., Sadreyev, R.I., and Lee, J.T. (2019). Exploration of CTCF post-translation modifications uncovers Serine-224 phosphorylation by PLK1 at pericentric regions during the G2/M transition. Elife *8*.

Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J.B., Blau, C.A., Shendure, J., Duan, Z.*, et al.* (2015). Bipartite structure of the inactive mouse X chromosome. Genome biology *16*, 152.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602-616.

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A *105*, 10762-10767.

Deschamps, J., and Duboule, D. (2017). Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. Genes Dev *31*, 1406-1416.

Despang, A., Schopflin, R., Franke, M., Ali, S., Jerkovic, I., Paliou, C., Chan, W.L., Timmermann, B., Wittler, L., Vingron, M.*, et al.* (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet *51*, 1263-1271.

Disteche, C.M. (2016). Dosage compensation of the sex chromosomes and autosomes. Semin Cell Dev Biol *56*, 9-18.

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

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst *3*, 99-101.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst *3*, 95-98.

El-Kady, A., and Klenova, E. (2005). Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. FEBS Lett *579*, 1424-1434.

Elia, A.E., Cantley, L.C., and Yaffe, M.B. (2003). Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. Science (New York, NY *299*, 1228-1231.

Erb, M.A., Scott, T.G., Li, B.E., Xie, H., Paulk, J., Seo, H.S., Souza, A., Roberts, J.M., Dastjerdi, S., Buckley, D.L.*, et al.* (2017). Transcription control by the ENL YEATS domain in acute leukaemia. Nature *543*, 270-274.

Falco, G., Lee, S.L., Stanghellini, I., Bassey, U.C., Hamatani, T., and Ko, M.S. (2007). Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Dev Biol *307*, 539-550.

Fang, H., Bonora, G., Lewandowski, J.P., Thakur, J., Filippova, G.N., Henikoff, S., Shendure, J., Duan, Z., Rinn, J.L., Deng, X.*, et al.* (2020). Trans- and cis-acting effects of Firre on epigenetic features of the inactive X chromosome. Nat Commun *11*, 6053.

Fang, J., Chen, T., Chadwick, B., Li, E., and Zhang, Y. (2004). Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J Biol Chem *279*, 52812-52815.

Ficarro, S.B., Adelmant, G., Tomar, M.N., Zhang, Y., Cheng, V.J., and Marto, J.A. (2009). Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. Anal Chem *81*, 4566-4575.

Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.P., Hu, Y.J., Nguyen, D.K., Tsuchiya, K.D., and Disteche, C.M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. Developmental cell *8*, 31-42.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110-114.

Fraser, J., Williamson, I., Bickmore, W.A., and Dostie, J. (2015). An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol Mol Biol Rev *79*, 347-372.

Froberg, J.E., Pinter, S.F., Kriz, A.J., Jegu, T., and Lee, J.T. (2018). Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. Nat Commun *9*, 5004.

Gandhi, R., Gillespie, P.J., and Hirano, T. (2006). Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. Curr Biol *16*, 2406-2417.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science *360*, 102-105.

Gdula, M.R., Nesterova, T.B., Pintacuda, G., Godwin, J., Zhan, Y., Ozadam, H., McClellan, M., Moralli, D., Krueger, F., Green, C.M.*, et al.* (2019). The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. Nat Commun *10*, 30.

Gendrel, A.V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S.*, et al.* (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Developmental cell *23*, 265-279.

Gimenez-Abian, J.F., Sumara, I., Hirota, T., Hauf, S., Gerlich, D., de la Torre, C., Ellenberg, J., and Peters, J.M. (2004). Regulation of sister chromatid cohesion between chromosome arms. Curr Biol *14*, 1187-1193.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E.*, et al.* (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575-579.

Golsteyn, R.M., Mundt, K.E., Fry, A.M., and Nigg, E.A. (1995). Cell cycle regulation of the activity and subcellular localization of Plk1, a human protein kinase implicated in mitotic spindle function. The Journal of cell biology *129*, 1617-1628.

Golsteyn, R.M., Schultz, S.J., Bartek, J., Ziemiecki, A., Ried, T., and Nigg, E.A. (1994). Cell cycle analysis and chromosomal localization of human Plk1, a putative homologue of the mitotic kinases Drosophila polo and Saccharomyces cerevisiae Cdc5. J Cell Sci *107 (Pt 6)*, 1509-1517.

Gorlick, R., Kolb, E.A., Keir, S.T., Maris, J.M., Reynolds, C.P., Kang, M.H., Carol, H., Lock, R., Billups, C.A., Kurmasheva, R.T.*, et al.* (2014). Initial testing (stage 1) of the Polo-like kinase inhibitor volasertib (BI 6727), by the Pediatric Preclinical Testing Program. Pediatr Blood Cancer *61*, 158-164.

Goto, Y., and Kimura, H. (2009). Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene. Nucleic Acids Res *37*, 7416-7428.

Greenberg, J.R. (1972). High stability of messenger RNA in growing cultured cells. Nature *240*, 102-104.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y.*, et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900-910.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yanez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B.*, et al.* (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693-707 e614.

Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R.*, et al.* (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol *21*, 198-206.

Hauf, S., Roitinger, E., Koch, B., Dittrich, C.M., Mechtler, K., and Peters, J.M. (2005). Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. PLoS biology *3*, e69.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc Natl Acad Sci U S A *109*, 17507-17512.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Hendzel, M.J., Wei, Y., Mancini, M.A., Van Hooser, A., Ranalli, T., Brinkley, B.R., Bazett-Jones, D.P., and Allis, C.D. (1997). Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. Chromosoma *106*, 348-360.

Hirano, T. (2005). Condensins: organizing and segregating the genome. Curr Biol *15*, R265-275.

Horakova, A.H., Moseley, S.C., McLaughlin, C.R., Tremblay, D.C., and Chadwick, B.P. (2012). The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. Hum Mol Genet *21*, 4367-4377.

Hu, B., Petela, N., Kurze, A., Chan, K.L., Chapard, C., and Nasmyth, K. (2015). Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. Nucleic Acids Res *43*, e132.

Huang, H., Zhu, Q., Jussila, A., Han, Y., Bintu, B., Kern, C., Conte, M., Zhang, Y., Bianco, S., Chiariello, A.M.*, et al.* (2021). CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. Nat Genet.

Huis in 't Veld, P.J., Herzog, F., Ladurner, R., Davidson, I.F., Piric, S., Kreidl, E., Bhaskara, V., Aebersold, R., and Peters, J.M. (2014). Characterization of a DNA exit gate in the human cohesin ring. Science *346*, 968-972.

Jaeger, M.G., Schwalb, B., Mackowiak, S.D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B.*, et al.* (2020). Selective Mediator dependence of cell-type-specifying transcription. Nat Genet *52*, 719-727.

Jaeger, M.G., and Winter, G.E. (2021). Fast-acting chemical tools to delineate causality in transcriptional control. Mol Cell *81*, 1617-1630.

Jansz, N., Keniry, A., Trussart, M., Bildsoe, H., Beck, T., Tonks, I.D., Mould, A.W., Hickey, P., Breslin, K., Iminitoff, M.*, et al.* (2018a). Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. Nat Struct Mol Biol *25*, 766-777.

Jansz, N., Nesterova, T., Keniry, A., Iminitoff, M., Hickey, P.F., Pintacuda, G., Masui, O., Kobelke, S., Geoghegan, N., Breslin, K.A.*, et al.* (2018b). Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway. Cell Rep *25*, 1912-1923 e1919.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science (New York, NY *293*, 1074-1080.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119-133.

Jeppesen, P., and Turner, B.M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell *74*, 281-289.

Joshi, O., Wang, S.Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P.J., Habibi, E., Shaik, J., Saeed, S., Handoko, L.*, et al.* (2015). Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency. Cell Stem Cell *17*, 748-757.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kettenbach, A.N., Schweppe, D.K., Faherty, B.K., Pechenick, D., Pletnev, A.A., and Gerber, S.A. (2011). Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. Sci Signal *4*, rs5.

Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol *26*, 1351-1359.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231-1245.

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. Science *366*, 1345-1349.

Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., Leon, J*., et al.* (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Molecular and cellular biology *21*, 2221-2234.

Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Buhler, M., Liu, P*., et al.* (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell *17*, 471-485.

Kriz, A.J., Colognori, D., Sunwoo, H., Nabet, B., and Lee, J.T. (2021). Balancing cohesin eviction and retention prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-inactivation. Mol Cell.

Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M*., et al.* (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nat Struct Mol Biol *28*, 152-161.

Kueng, S., Hegemann, B., Peters, B.H., Lipp, J.J., Schleiffer, A., Mechtler, K., and Peters, J.M. (2006). Wapl controls the dynamic association of cohesin with chromatin. Cell *127*, 955-967.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol Cell *65*, 432-446 e435.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F*., et al.* (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol Cell *57*, 361-375.

Lake, R.J., and Jelinek, W.R. (1993). Cell cycle- and terminal differentiation-associated regulation of the mouse mRNA encoding a conserved mitotic protein kinase. Molecular and cellular biology *13*, 7793-7801.

Lanctot, C., Kaspar, C., and Cremer, T. (2007). Positioning of the mouse Hox gene clusters in the nuclei of developing embryos and differentiating embryoid bodies. Exp Cell Res *313*, 1449-1459.

Lee, J.T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nature reviews Molecular cell biology *12*, 815-826.

Li, Y., Haarhuis, J.H.I., Sedeno Cacciatore, A., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D*., et al.* (2020). The structural basis for cohesin-CTCF-anchored loops. Nature *578*, 472-476.

Liskay, R.M., and Evans, R.J. (1980). Inactive X chromosome DNA does not function in DNA-mediated cell transformation for the hypoxanthine phosphoribosyltransferase gene. Proc Natl Acad Sci U S A *77*, 4895-4898.

Liu, N.Q., Maresca, M., van den Brand, T., Braccioli, L., Schijns, M.M.G.A., Teunissen, H., Bruneau, B.G., Nora, E.P., and de Wit, E. (2020). WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. Nature Genetics.

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene *5*, 1743-1753.

Lock, L.F., Melton, D.W., Caskey, C.T., and Martin, G.R. (1986). Methylation of the mouse hprt gene differs on the active and inactive X chromosomes. Mol Cell Biol *6*, 914-924.

Lock, L.F., Takagi, N., and Martin, G.R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell *48*, 39-46.

Losada, A., Hirano, M., and Hirano, T. (2002). Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. Genes & development *16*, 3004-3016.

Lowery, D.M., Mohammad, D.H., Elia, A.E., and Yaffe, M.B. (2004). The Polo-box domain: a molecular integrator of mitotic kinase cascades and Polo-like kinase function. Cell cycle (Georgetown, Tex *3*, 128-131.

Lu, Y., Shan, G., Xue, J., Chen, C., and Zhang, C. (2016). Defining the multivalent functions of CTCF from chromatin state and three-dimensional chromatin interactions. Nucleic Acids Res *44*, 6200-6212.

Luikenhuis, S., Wutz, A., and Jaenisch, R. (2001). Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. Molecular and cellular biology *21*, 8512-8520.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R*., et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

Luppino, J.M., Park, D.S., Nguyen, S.C., Lan, Y., Xu, Z., Yunker, R., and Joyce, E.F. (2020). Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. Nat Genet *52*, 840-848.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372-373.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

MacPherson, M.J., Beatty, L.G., Zhou, W., Du, M., and Sadowski, P.D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. Molecular and cellular biology *29*, 714-725.

McLaughlin, K., Flyamer, I.M., Thomson, J.P., Mjoseng, H.K., Shukla, R., Williamson, I., Grimes, G.R., Illingworth, R.S., Adams, I.R., Pennings, S.*, et al.* (2019). DNA Methylation Directs Polycomb-Dependent 3D Genome Re-organization in Naive Pluripotency. Cell Rep *29*, 1974-1985 e1976.

Meggio, F., and Pinna, L.A. (2003). One-thousand-and-one substrates of protein kinase CK2? FASEB J *17*, 349-368.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu Rev Genomics Hum Genet *17*, 17-43.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285-1297.

Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J Cell Biol *147*, 1399-1408.

Meyer, B.J. (2010). Targeting X chromosomes for repression. Curr Opin Genet Dev *20*, 179-189.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res *45*, D183-D189.

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W.*, et al.* (2015a). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science (New York, NY *349*.

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W.*, et al.* (2015b). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393-396.

Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanenkov, V.V.*, et al.* (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS One *7*, e34915.

Morales, C., and Losada, A. (2018). Establishing and dissolving cohesion during the vertebrate cell cycle. Current opinion in cell biology *52*, 51-57.

Nabet, B., Ferguson, F.M., Seong, B.K.A., Kuljanin, M., Leggett, A.L., Mohardt, M.L., Robichaud, A., Conway, A.S., Buckley, D.L., Mancias, J.D.*, et al.* (2020). Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. Nat Commun *11*, 4687.

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

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59-64.

Nagano, T., Lubling, Y., Varnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N., Wingett, S., Fraser, P., and Tanay, A. (2017). Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature *547*, 61-67.

Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A.*, et al.* (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep *3*, 1678-1689.

Nakajima, H., Toyoshima-Morimoto, F., Taniguchi, E., and Nishida, E. (2003). Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. J Biol Chem *278*, 25277-25280.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annu Rev Genet *43*, 525-558.

Nikolsky, Y., and Bryant, J. (2009). Protein networks and pathway analysis. Preface. Methods Mol Biol *563*, v-vii.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods *6*, 917-922.

Nora, E.P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., Uebersohn, A., Hajj, B., Saux, A.L., Coulon, A.*, et al.* (2020). Molecular basis of CTCF binding polarity in genome folding. Nat Commun *11*, 5612.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944 e922.

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

Norris, D.P., Brockdorff, N., and Rastan, S. (1991). Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. Mamm Genome *1*, 78-83.

Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nature structural & molecular biology *20*, 566-573.

Ogawa, Y., Sun, B.K., and Lee, J.T. (2008). Intersection of the RNA interference and X-inactivation pathways. Science *320*, 1336-1341.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A*., et al.* (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal *3*, ra3.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet *15*, 234-246.

Pal, K., Forcato, M., Jost, D., Sexton, T., Vaillant, C., Salviato, E., Mazza, E.M.C., Lugli, E., Cavalli, G., and Ferrari, F. (2019). Global chromatin conformation differences in the Drosophila dosage compensated chromosome X. Nat Commun *10*, 5355.

Paliou, C., Guckelberger, P., Schopflin, R., Heinrich, V., Esposito, A., Chiariello, A.M., Bianco, S., Annunziatella, C., Helmuth, J., Haas, S*., et al.* (2019). Preformed chromatin topology assists transcriptional robustness of Shh during limb development. Proc Natl Acad Sci U S A *116*, 12390-12399.

Paulson, J.R., and Laemmli, U.K. (1977). The structure of histone-depleted metaphase chromosomes. Cell *12*, 817-828.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat Rev Genet *6*, 893-904.

Perea-Resa, C., Bury, L., Cheeseman, I.M., and Blower, M.D. (2020). Cohesin Removal Reprograms Gene Expression upon Mitotic Entry. Mol Cell *78*, 127-140 e127.

Peters, J.M., Tedeschi, A., and Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. Genes Dev *22*, 3089-3114.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Plath, K., Talbot, D., Hamer, K.M., Otte, A.P., Yang, T.P., Jaenisch, R., and Panning, B. (2004). Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. J Cell Biol *167*, 1025-1035.

Plys, A.J., Davis, C.P., Kim, J., Rizki, G., Keenen, M.M., Marr, S.K., and Kingston, R.E. (2018). Phase separation and nucleosome compaction are governed by the same domain of Polycomb Repressive Complex 1. bioRxiv, 467316.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res *20*, 110-121.

Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A.V., Zentner, G.E., Ren, B., and Lobanenkov, V.V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc Natl Acad Sci U S A *117*, 2020-2031.

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

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D.*, et al.* (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320 e324.

Reyes-Turcu, F.E., Ventii, K.H., and Wilkinson, K.D. (2009). Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. Annu Rev Biochem *78*, 363-397.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

Rhodes, J.D.P., Feldmann, A., Hernandez-Rodriguez, B., Diaz, N., Brown, J.M., Fursova, N.A., Blackledge, N.P., Prathapan, P., Dobrinic, P., Huseyin, M.K.*, et al.* (2020). Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. Cell Rep *30*, 820-835 e810.

Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal *4*, rs3.

Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat Rev Genet *19*, 789-800.

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol Cell *67*, 837-852 e837.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proceedings of the National Academy of Sciences of the United States of America *105*, 8309-8314.

Rudolph, D., Impagnatiello, M.A., Blaukopf, C., Sommer, C., Gerlich, D.W., Roth, M., Tontsch-Grunt, U., Wernitznig, A., Savarese, F., Hofmann, M.H.*, et al.* (2015). Efficacy and mechanism of action of volasertib, a potent and selective inhibitor of Polo-like kinases, in preclinical models of acute myeloid leukemia. J Pharmacol Exp Ther *352*, 579-589.

Rudolph, D., Steegmaier, M., Hoffmann, M., Grauert, M., Baum, A., Quant, J., Haslinger, C., Garin-Chesa, P., and Adolf, G.R. (2009). BI 6727, a Polo-like kinase inhibitor with improved pharmacokinetic profile and broad antitumor activity. Clin Cancer Res *15*, 3094-3102.

Sakakibara, Y., Nagao, K., Blewitt, M., Sasaki, H., Obuse, C., and Sado, T. (2018). Role of SmcHD1 in establishment of epigenetic states required for the maintenance of the X-inactivated state in mice. Development *145*.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A *112*, E6456-6465.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335-348.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M.*, et al.* (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat Genet *47*, 1179-1186.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., C, H.H., Mirny, L.*, et al.* (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51-56.

Sekiya, T., Murano, K., Kato, K., Kawaguchi, A., and Nagata, K. (2017). Mitotic phosphorylation of CCCTC-binding factor (CTCF) reduces its DNA binding activity. FEBS Open Bio *7*, 397-404.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605-2606.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S.*, et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell *4*, 481-495.

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465-469.

Spector, D.L., Ochs, R.L., and Busch, H. (1984). Silver staining, immunofluorescence, and immunoelectron microscopic localization of nucleolar phosphoproteins B23 and C23. Chromosoma *90*, 139-148.

Spencer, R.J., del Rosario, B.C., Pinter, S.F., Lessing, D., Sadreyev, R.I., and Lee, J.T. (2011). A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation. Genetics *189*, 441-454.

Splinter, E., de Wit, E., Nora, E.P., Klous, P., van de Werken, H.J., Zhu, Y., Kaaij, L.J., van Ijcken, W., Gribnau, J., Heard, E.*, et al.* (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes & development *25*, 1371-1383.

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

Starmer, J., and Magnuson, T. (2009). A new model for random X chromosome inactivation. Development *136*, 1-10.

Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S., and Lieberman, P.M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. The EMBO journal *27*, 654-666.

Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A.*, et al.* (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature *544*, 59-64.

Sumara, I., Vorlaufer, E., Stukenberg, P.T., Kelm, O., Redemann, N., Nigg, E.A., and Peters, J.M. (2002). The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. Molecular cell *9*, 515-525.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA activates Xist by evicting CTCF. Cell *153*, 1537-1551.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032-2034.

Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H.N., Huynh, T.N., Zhen, C.Y., Ma, B., Wang, H., and Ren, X. (2018). Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. J Biol Chem.

Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I.F., Tang, W., Cisneros, D.A., Bhaskara, V.*, et al.* (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature *501*, 564-568.

Terrenoire, E., McRonald, F., Halsall, J.A., Page, P., Illingworth, R.S., Taylor, A.M., Davison, V., O'Neill, L.P., and Turner, B.M. (2010). Immunostaining of modified histones defines high-level features of the human metaphase epigenome. Genome Biol *11*, R110.

Torrano, V., Navascues, J., Docquier, F., Zhang, R., Burke, L.J., Chernukhin, I., Farrar, D., Leon, J., Berciano, M.T., Renkawitz, R.*, et al.* (2006). Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. J Cell Sci *119*, 1746-1759.

Valsecchi, C.I.K., Basilicata, M.F., Georgiev, P., Gaub, A., Seyfferth, J., Kulkarni, T., Panhale, A., Semplicio, G., Manjunath, V., Holz, H.*, et al.* (2021). RNA nucleation by MSL2 induces selective X chromosome compartmentalization. Nature *589*, 137-142.

Vassilev, L.T., Tovar, C., Chen, S., Knezevic, D., Zhao, X., Sun, H., Heimbrook, D.C., and Chen, L. (2006). Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1. Proceedings of the National Academy of Sciences of the United States of America *103*, 10660-10665.

Vian, L., Pekowska, A., Rao, S.S.P., Kieffer-Kwon, K.R., Jung, S., Baranello, L., Huang, S.C., El Khattabi, L., Dose, M., Pruett, N.*, et al.* (2018a). The Energetics and Physiological Impact of Cohesin Extrusion. Cell.

Vian, L., Pekowska, A., Rao, S.S.P., Kieffer-Kwon, K.R., Jung, S., Baranello, L., Huang, S.C., El Khattabi, L., Dose, M., Pruett, N.*, et al.* (2018b). The Energetics and Physiological Impact of Cohesin Extrusion. Cell *175*, 292-294.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep *10*, 1297-1309.

Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proc Natl Acad Sci U S A *112*, 4672-4677.

Wang, C.Y., Colognori, D., Sunwoo, H., Wang, D., and Lee, J.T. (2019). PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. Nat Commun *10*, 2950.

Wang, C.Y., Jegu, T., Chu, H.P., Oh, H.J., and Lee, J.T. (2018). SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. Cell *174*, 406-421 e425.

Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. Nat Genet *28*, 371-375.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L.*, et al.* (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588 e1528.

Wellik, D.M. (2007). Hox patterning of the vertebrate axial skeleton. Dev Dyn *236*, 2454-2463.

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

Williamson, I., Kane, L., Devenney, P.S., Flyamer, I.M., Anderson, E., Kilanowski, F., Hill, R.E., Bickmore, W.A., and Lettice, L.A. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. Development *146*.

Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet *12*, 542-553.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Wutz, G., Ladurner, R., St Hilaire, B.G., Stocsits, R.R., Nagasaka, K., Pignard, B., Sanborn, A., Tang, W., Varnai, C., Ivanov, M.P.*, et al.* (2020). ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin(STAG1) from WAPL. Elife *9*.

Wutz, G., Varnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J.*, et al.* (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J *36*, 3573-3599.

Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. Molecular and cellular biology *31*, 2174-2183.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A *104*, 7145-7150.

Yesbolatova, A., Natsume, T., Hayashi, K.I., and Kanemaki, M.T. (2019). Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. Methods *164-165*, 73-80.

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. Cell *152*, 727-742.

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C*., et al.* (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. Nat Genet *36*, 1105-1110.

Yusufzai, T.M., and Felsenfeld, G. (2004). The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. Proc Natl Acad Sci U S A *101*, 8620-8624.

Zhang, K., Wu, D.Y., Zheng, H., Wang, Y., Sun, Q.R., Liu, X., Wang, L.Y., Xiong, W.J., Wang, Q., Rhodes, J.D.P*., et al.* (2020). Analysis of Genome Architecture during SCNT Reveals a Role of Cohesin in Impeding Minor ZGA. Mol Cell *79*, 234-250 e239.

Zhang, L.F., Huynh, K.D., and Lee, J.T. (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell *129*, 693-706.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750-756.

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