The Spatial Organization of the X-chromosome and its Impact on X-inactivation

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

Citation

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:41121361

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, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
The Spatial Organization of the X-chromosome and its Impact on X-inactivation

A dissertation presented
by
John Edward Froberg

to
The Division of Medical Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Genetics and Genomics

Harvard University
Cambridge, Massachusetts
September 2018
The Spatial Organization of the X-chromosome and its Impact on X-inactivation

Abstract

Female mammals have two X-chromosomes, males have one X-chromosome. X-chromosome inactivation (XCI) is a process where most genes on one female X-chromosome is transcriptionally silenced in order to balance gene dosage on the X-chromosome between males and females. It has been known since the discovery of the Barr body in the 1940’s that the active (Xa) and inactive (Xi) X-chromosomes fold differently. However, the precise structures adopted by the two Xs, the mechanisms that cause the Xa and Xi to fold differently, and the role of the unique structure of the Xi in silencing were unknown. In this thesis, I describe our work defining the structures of the Xa and the Xi, the role of Xist RNA and cohesin proteins in shaping the organization of the Xi, and testing whether the “super-structure” of the Xi is necessary for gene silencing. I performed allele-specific Hi-C to map the structures of the Xa and Xi at high resolution and found that while the Xa is organized into megabase-sized topologically associated domains (TADs) like all the other chromosomes, the Xi is largely devoid of TADs. Allele-specific ChIP-seq showed that cohesin proteins, which are necessary for folding chromosomes into TADs, are depleted from the Xi. Deletion of Xist allows partial restoration of both cohesin binding and TADs across large regions of the Xi, suggesting that Xist RNA removes cohesin from chromatin and breaks down TADs on the Xi.

Although the Xi lacks TADs, it folds into two very large “megadomains”, each tens of megabases in size. There is an unusual tandem repeat locus called Dxz4 at the border between the two megadomains. Dxz4 also forms a ~25 Mb looping interaction with another tandem repeat called Firre specifically on the Xi. I generated single and double deletions of Dxz4 and
Firre in mouse embryonic stem cells. Deletion of Dxz4 prevented megadomain formation and deletion of Firre partially disrupted interactions within megadomains. However, deletion of Dxz4, Firre or both loci had no impact on the ability of XCI to initiate, suggesting that the large-scale organization of the Xi is not needed for gene silencing.
# Table of Contents

Dedication .................................................................................................................. vii
Acknowledgements ................................................................................................. viii
List of Figures and Tables ......................................................................................... xi
Chapter 1: Introduction .............................................................................................. 1
  Addendum ................................................................................................................... 2
  Overview ..................................................................................................................... 3
X-chromosome inactivation: balancing gene dosage .................................................. 3
  Mammalian dosage compensation .............................................................................. 4
  Xist: the master regulator of XCI ................................................................................ 6
  Tsix: the negative regulator of Xist ............................................................................ 11
  Ensuring upregulation of Xist on one X: positive regulators of XCI ................. 15
  Xist RNA localization through adaptor proteins ..................................................... 17
  Large-scale genomic features guide Xist spreading ................................................. 20
  Multiple silencing systems establish and maintain XCI ........................................ 21
  Defining the “Xist proteome” is a major goal for understanding mechanisms of silencing during XCI ........................................................... 24
Chromatin conformation: form and function ............................................................. 25
  The structure of the Barr Body: an unsolved feature of the Xi ......................... 26
  Methods for examining chromatin conformation .................................................. 28
  Patterns of higher-order structure: loops, domains and compartments .............. 30
  Loops, domains and compartments are all distinct on the Xi ......................... 32
  Unique large-scale structures on the Xi involve tandem repeats .................... 34
  The role of higher-order structure in gene regulation is unclear ..................... 36
Chapter 2: A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation ............................................................. 40

Addendum .................................................................................................. 41

Abstract ..................................................................................................... 43

Introduction ............................................................................................. 44

Results ....................................................................................................... 45

Discussion ................................................................................................. 67

Materials and methods ........................................................................... 69

Chapter 3: Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape ........................................... 82

Addendum .................................................................................................. 83

Abstract ..................................................................................................... 84

Introduction ............................................................................................. 85

Results ....................................................................................................... 87

Discussion ................................................................................................. 120

Materials and methods ........................................................................... 123

Chapter 4: Conclusions, Perspective and Future Directions ......................... 138

Appendix: Supplemental Figures and Tables ............................................. 149

References .................................................................................................. 207
To Jonathan Widom, who started me on this journey and left us all too soon
Acknowledgements

The work in this dissertation would not be possible without the help and assistance of many others. Science is not a solitary endeavor and I have been incredibly fortunate to have benefited from working under generous mentors and collaborators. My scientific journey began all the way back in high school, where I had the opportunity to do an Inquiry Project on Wednesdays with Dr. Don Dosch. This instilled in me a passion for lab work and familiarity with the scientific process, despite the many disasters I created. I worked in a Cognitive Psychology lab during my freshman year at Northwestern with Dr. Steven Franconeri, who taught me the basics of how to code and how to design an experiment. I decided that I was very interested in biochemistry and that studying chromatin structure would be working on the most exciting problem. Dr. Jonathan Widom took me into his lab as a precocious 19-year-old, and for this I will be forever grateful. He was an astounding mentor; as long as his door was halfway open you could go in and talk to him for two hours about anything. He taught me how to think about biology quantitatively, how to navigate academia and how pick a good scientific problem to work on. He passed away far too soon, and I will always be thankful for the opportunities he provided for me. I also must thank Yvonne Fondufe-Mittendorf, my postdoc in the Widom lab, now Associate Professor at the University of Kentucky. She taught me how to work in a molecular biology lab and answered all 10,000 of my questions every day.

I am deeply grateful for Jeannie Lee for taking me into her lab, first as a precocious 22-year-old 3 weeks out of college and then as a permanent research assistant in the middle of my G1-year. She gave me the freedom to work on something completely outside the lab’s area of expertise at the start of my PhD, and the resources to take risks. When I started, no one
thought chromatin conformation would work, everyone said it is too complicated and requires
too much bioinformatics. Jeannie saw my enthusiasm for the problem and allowed me to
pursue it although I was just at the start of graduate school and had no relevant experience.
She encouraged me to take several bioinformatics courses to give me the skills I needed to
develop chromatin conformation projects, even though it meant missing lab meeting
sometimes. She then facilitated my collaboration with Anand Minajigi, which allowed us to tie
two distinct projects into one complete story. She taught me how to write a scientific paper and
forged our raw data into a finished project. She edited rough drafts of grants and was incredibly
supportive during my time as a PhD student and during my postdoc search. I will be forever
indebted to her unwavering support for my science during my PhD.

I am thankful for 3 years of support from the National Science Foundation from years 2-
4 of the PhD. I also would like to thank my thesis committee, Bob Kingston, Stirling Churchman
and George Church for their guidance during the PhD and support for many applications.

I must thank many members of the Lee lab for their mentorship and collaborations. First
and foremost is Anand Minajigi. We collaborated on his Xist proteomics project, and when his
project began to dovetail nicely with my chromatin conformation project, he had the
graciousness and wisdom to combine our two projects into one story. Stefan Pinter guided me
through the early PhD years and shaped the structure of my chromatin conformation projects
and for this I will always be thankful. He taught me almost everything I know about genomics,
and we have worked together on many projects. Many thanks to Andrea Kriz, who I mentored
at the start of her PhD and has turned from an eager rotation student into one of the most
effective collaborators I have ever seen. Hongjae Sunwoo taught me how to do microscopy
experiments and the nuts and bolts of molecular biology, and Yesu Jeon was my first mentor in the Lee lab and has always helped me see the big picture. Yesu also deserves particular credit for reading and commenting on my thesis introduction. I am very thankful for the other grad students in the Lee Lab, David Colognori, Chen-Yu Wang, Lin Yang, Johnny Kung, Katie Lee & Tom Dial for their moral support and for always being there to talk science. I am thankful for all members of the Lee Lab past and present for their mentorship and support. I am very thankful for my friends in the BBS program outside the Lee lab. I am especially thankful for my former roommates, Murtaza Akhtar and Hao Liu, for many late-night scientific discussions, and to the Snapchat group “BBSome” (formerly “Thesis Therapy Hotline”) for its advice on how to prepare a thesis that fits the requirements. I am very appreciative of the incredible infrastructure the BBS program has provided over the years.

I would finally like to thank my parents, James & Julie Froberg, for their love and support over these past few years. I know it is not easy for them seeing their only child move to a new city and work in a field unfamiliar to them. They have helped me at every step along the and I will always be grateful for what they’ve done for me.
List of Figures and Tables

Figure 1.1: RNA-based regulation in X-chromosome inactivation ........................................ 7
Table 1.1: Summary of IncRNAs and proposed interacting protein partners
for X-inactivation .................................................................................................................. 8
Figure 1.2: Tandem repeats within Xist and their putative functions ................................. 9
Figure 1.3: Model for Xic pairing before XCI onset .......................................................... 14
Figure 1.4: Interaction patterns visible in Hi-C data ............................................................ 30
Figure 2.1: iDRiP-MS reveals a large Xist interactome ....................................................... 48
Figure 2.2: Impact of depleting Xist interactors on H3K27me3 ........................................... 51
Figure 2.3: De-repression of Xi genes by targeting Xist interactors ................................. 55
Figure 2.4: Ablating Xist in cis restores cohesin binding on the Xi .................................... 59
Figure 2.5: Ablating Xist results in Xi reversion to an Xa-like chromosome conformation ................................................................. 64
Figure 2.6: The Xi is suppressed by multiple synergistic mechanisms ............................... 68
Figure 3.1: Dynamics of megadomain formation during XCI ............................................ 90
Figure 3.2: Dynamics of TADs and sub-TADs during XCI in the MeCP2 and DXZ4 regions ................................................................. 95
Figure 3.3: DXZ4 is necessary but not sufficient for megadomain formation ................. 100
Figure 3.4: DXZ4, Firre, and Xi-specific superloops ......................................................... 104
Figure 3.5: Firre is predominantly expressed from the Xa in differentiating female ES cells .................................................................................................................. 108
Figure 3.6: Firre is not required for megadomain formation but deleting Firre weakens megadomains ............................................................................................................. 111
Figure 3.7: DXZ4 and Firre do not affect Xi localization and accessibility ....................... 116
Figure 3.8: Deletion of DXZ4, Firre, or both does not affect accessibility or gene silencing on the Xi ........................................................................................................... 118
Supplementary Figure S2.1: The microscopic images of knock down day 7 ESCs

Supplementary Figure S2.2: Confirmation that the GFP transgene of Xi-TgGFP cells is on the inactive X

Supplementary Figure S2.3: Xi reactivation by inhibiting single versus multiple Xist interactors

Supplementary Figure S2.4: Correlations between biological replicates for allelic-specific RNA-seq analysis

Supplementary Figure S2.5: Correlations between biological replicates for allelic-specific RNA-seq analysis

Supplementary Figure S2.6: Correlations between biological replicates for allelic-specific RNA-seq analysis

Supplementary Figure S2.7: Allelic expression of autosomal genes, including imprinted genes, is not affected by the triple-drug treatments

Supplementary Figure S2.8: Analysis of CTCF and cohesin ChIP-seq replicates demonstrates similar allelic trends on ChrX

Supplementary Figure S2.9: Analysis of biological replicates for cohesin ChIP-seq confirms cohesin restoration in cis when Xist is ablated

Supplementary Figure S2.10: Restored SMC1a peaks are reproducible in biological replicates and occur throughout Xi\textsuperscript{AXist} (Example set 1)

Supplementary Figure S2.11: Restored SMC1a peaks are reproducible in biological replicates and occur throughout Xi\textsuperscript{AXist} (Example set 2)

Supplementary Figure S2.12: Restored RAD21 peaks are reproducible in biological replicates and occur throughout Xi\textsuperscript{AXist}

Supplementary Figure S2.13: Cohesin restored in Xi\textsuperscript{AXist}/Xa\textsuperscript{WT} fibroblasts was Xi-specific and did not occur on autosomes

Supplementary Figure S2.14: Biological replicates of HiC-seq analysis
yield similar findings........................................................................................................................................ 173

Supplementary Figure S2.15: Restored TADs identified in Xi^Xist using Xa TADs of
Dixon et al. as reference................................................................................................................................... 175

Supplementary Table S2.1: iDRip proteomics results...................................................................................... 178

Supplementary Table S2.2: Lentiviral shRNA constructs used for stable knockdowns of
candidate Xist interactors.................................................................................................................................. 184

Supplementary Table S2.3: Table of reagents .................................................................................................. 186

Supplementary Figure S3.1: Superloop anchors are tandem repeats that bind CTCF and
SMC1A on the mouse Xi.................................................................................................................................... 189

Supplementary Figure S3.2: Megadomains are not present on the Xa and
allelic RNA-seq shows Xist upregulation is highly skewed to Xi......................................................... 191

Supplementary Figure S3.3: Biological replicates of the time course Hi-C and
Hi-C² experiments.............................................................................................................................................. 193

Supplementary Figure S3.4: Validation of Dxz4 deletion and replicate Hi-C................................. 195

Supplementary Figure S3.5: Generation of a Xist+Dxz4 transgene........................................................ 197

Supplementary Figure S3.6: A conserved network of superloop interactions
involving Dxz4 on the mouse Xi.......................................................................................................................... 199

Supplementary Figure S3.7: Validation of Firre deletions ................................................................. 201

Supplementary Figure S3.8: Replicate Hi-Cs in Firre deletions .............................................................. 203

Supplementary Figure S3.9: Further analysis of Xist localization and
H3K27me3 foci in Dxz4 and Firre deletions;
metaphase distribution of H3K27me3 in Dxz4^Δ/Δ ........................................................................... 204

Supplementary Figure S3.10: Replicate ATAC-seq analysis in
FirreXi^Δ/+ and Dxz4^Δ/Δ:Firre^Xi/+............................................................................................................. 206
Chapter 1: Introduction
Addendum:

Parts of this chapter were adapted from:


John E. Froberg, Lin Yang and Jeannie T. Lee co-authored and edited this review article.

The adapted portions are presented here with permission from *Journal of Molecular Biology.* *Journal of Molecular Biology* considers theses/dissertations “personal use” and explicitly gives permission for authors of articles to include them in theses (see [https://www.elsevier.com/about/policies/copyright](https://www.elsevier.com/about/policies/copyright)).
OVERVIEW

X-chromosome inactivation is the biological process that ensures one X-chromosome is active and the other is silenced. X-inactivation underlies many human diseases and has proven to be an excellent model system for studying many questions in gene regulation. In particular, since the active and inactive Xs fold into different structures in the nucleus, X-inactivation is a good system to test the role of chromatin structure in gene regulation.

My thesis examines the spatial organization of the X-chromosomes and its impact on X-inactivation. Chapter 1 contains a review X-inactivation itself, then a review of chromatin structure. Chapter 2 describes my work defining the structure of the inactive X-chromosome and how structure is regulated on the X. Chapter 3 describes my test of whether certain types of folding on the inactive X are required for X-inactivation to initiate. Chapter 4 is analysis of how the preceding work contributes to the problem of relating chromatin structure of gene regulation, and suggestions for future directions.

X-CHROMOSOME INACTIVATION: BALANCING GENE DOSAGE

Males have one X-chromosome and females have two Xs in mammals. X-chromosome inactivation is a process where nearly all genes on one X-chromosome are transcriptionally silenced, so that both males and females express genes from only one active X-chromosome. X-inactivation is a dramatic epigenetic process, as hundreds of genes are silenced in cis by the long non-coding RNA (IncRNA) Xist. Xist itself is regulated by several additional IncRNAs, and Xist uses several distinct mechanisms to silence the Xi. In this section, I review dosage compensation, RNA-based gene regulation on the Xi, mechanisms of silencing and outstanding questions in X-inactivation. It should also be noted that this discussion focuses on X-inactivation in rodents unless specifically noted.
Mammalian dosage compensation

Chromosome-based sex determination systems create an imbalance in the dosage of X-linked genes between the two sexes, and this imbalance in gene expression is corrected through X-chromosome inactivation (XCI) [1]. At the time of the evolution of the current X-Y sex determination system in mammals, the X and Y chromosomes were nearly identical except for the male-determining factors on the Y chromosome [2, 3]. However, due to strong selective pressure to link together genes required for maleness on the Y chromosome, a series of large inversions occurred on the Y chromosome and prevented it from recombining with the X-chromosome [2-4]. Failure to recombine quickly led to the loss of most of the Y-chromosome (approximately 1000 genes) [2-4], and created a situation where males carried only one copy of nearly all X-chromosome genes while females carried two copies. This results in two imbalances 1.) an imbalance in the X:Autosome ratio in males and 2.) a 2-fold imbalance in X-linked gene expression between males as females [5]. The second imbalance is corrected through XCI, where the double dosage of X-linked gene expression in females is eliminated by transcriptionally silencing the majority of genes on one X-chromosome [1]. The gene expression imbalance between the X and the autosomes theoretically becomes a problem in both males and females (since XCI produces one active X). It has been proposed there is a process called X-hyperactivation that upregulates X-linked genes 2-fold to equalize dosage between the X and autosomes [5], similarly to the dosage compensation in flies where the male X is upregulated approximately 2-fold [6, 7]. However, some studies have observed X-hyperactivation [8-10] and others have not [11], and often partial upregulation (less than the expected 2X increase) upregulation is observed [10]. These observations suggest that X-hyperactivation may not be universal and may depend upon the biological context. However, X-chromosome inactivation is a robust and widespread phenomenon and has become a model system for many epigenetic processes.
Gene silencing on the X-chromosome is actually cyclical throughout development. In the female germline, the inactive X is reactivated as oocyte precursors differentiate into oocytes. However, in the male germline, both the X and the Y become silenced during spermatogenesis [12-14]. The paternal X is transiently reactivated in the early blastocyst, but then remains inactivated in the extraembryonic tissues of the pre-implantation embryo and remains inactive throughout development [15-17]. Since the paternal X is always inactivated in the extraembryonic tissues, this type of X-inactivation is considered “imprinted X-inactivation”. In contrast, both reactivated Xs remain active in the inner cell mass until approximately the time of implantation when X-inactivation occurs [15, 16]. Unlike the extraembryonic tissues, all cells in the epiblast randomly select one X to be inactive [15], leading to an approximately even ratio of cells that inactivate the paternal X to cells that inactive the maternal X (though there are certain alleles that can cause skew choice [18-20]). Random X-inactivation depends on the X-inactivation center (Xic) [21, 22], which is a region of the X-chromosome harboring several non-coding elements that are responsible for counting the number of X-chromosomes, choosing one to be silent and then silencing the chosen chromosome [23, 24]. Since mouse embryonic stem cells are derived from the inner cell mass, they carry two active Xs [25] and randomly inactivate one X upon differentiating [26, 27], thus mESCs are an excellent model system for studying choice, initiation and establishment of X-inactivation.

X-inactivation is faithfully maintained on the same X across mitotic divisions [28, 29], even in the absence of a functional Xic [30-32], which harbors all the elements necessary to initiate X-inactivation. Since X-inactivation persists across cell divisions even without its initial stimulus, X-inactivation is truly an epigenetic phenomenon [33] and this has important consequences. X-inactivation status can be used as a marker of clonality [34], and can be used to study patterns of clonal expansion and cell migration [35] and explains mosaic animal coat color patterns [1]. In addition, the fact that X-inactivation status can be propagated across many cell divisions is a useful research tool, since clonal cell lines from hybrid mice [36, 37] (or human
cell lines [38]) will show preferential silencing of one allele for most X-linked genes. Although X-inactivation is maintained in somatic cells, the X-inactivation cycle completes in the germ cell progenitors, as the inactive X is reactivated once again.

Understanding XCI is important both because of the utility of XCI as a model system for RNA-based regulation, but also because XCI itself is an important factor in human disease. X-chromosome inactivation plays a role in X-linked dominant diseases, perhaps most notably Rett Syndrome [39]. In Rett Syndrome, females are heterozygous for a diseased MECP2 allele, but due to X-inactivation the wild-type allele is silent in enough cells to lead to a inhibit normal neurological function and cause severe disease [39]. It has been shown in mouse models of Rett Syndrome that re-activating expression of the wild-type allele of the MECP2 gene restores the neurological function, suggesting that X-linked diseases could be cured by re-expression of genes silenced by X-inactivation [40]. There is precedence for reactivation of the silenced wild-type allele to potentially rescue a developmental disorder. Angelman Syndrome is a congenital disorder caused by maternal deletion or mutation of the imprinted Ube3a allele [41-43]. Ube3a is known to be controlled by a long antisense transcript from the Snrpn locus [44-46]. Topoisomerase inhibitors cause loss of imprinting of the silent paternal Ube3a allele by downregulating the antisense transcript in neurons in vitro and in mice [47], which may provide a strategy for rescuing the genetic defect that causes Angelman Syndrome. Similar approaches might be useful for reactivating silent genes on the Xi, such as MECP2 in Rett Syndrome.

**Xist: the master regulator of XCI**

The master regulator of XCI is a long non-coding RNA (IncRNA) called Xist (“X-Inactive Specific Transcript” [21, 48]). Xist was one of the first IncRNAs identified [21, 48-50]. Xist is a 17-kb transcript expressed from the X-inactivation center solely from the inactive X (Xi) [21, 48]. Xist deletions prevent X-inactivation in cis [22, 51] and forced Xist expression is sufficient to induce chromosome-wide silencing [28, 52, 53]. Xist RNA coats the X it is expressed from [54],
and the spreading of Xist RNA along one X-chromosome *in cis* initiates chromosome-wide silencing by recruiting silencing complexes (Fig. 1.1). Whereas Xist expression designates the Xi, *Tsix* expression demarcates the active X (Xa), also *in cis* [52, 55-58]. *Tsix* is antisense to *Xist* and serves as potent antagonist of *Xist* expression. It is therefore an excellent example of a natural antisense transcript that represses its sense partner. Noncoding genes have also been implicated in activating *Xist*, including *Jpx, RepA*, and *Ftx* [59-61].

![Figure 1.1: RNA-based regulation in X-chromosome inactivation](image)

**Figure 1.1: RNA-based regulation in X-chromosome inactivation**

On the inactive X, *Jpx* RNA helps upregulate Xist RNA. Xist then spreads across the chromosome and recruits silencing complexes such as PRC2 to inactive X-linked genes. On the active X, *Tsix* is expressed anti-sense to Xist and prevents Xist upregulation, allowing X-linked genes to continue to be expressed.

Xist RNA is quickly becoming one of the most popular model systems for studying IncRNA function, and the other IncRNAs at the Xic (Table 1.1) also serve as important examples of RNA-based gene regulators. Because of the involvement of many IncRNAs and distinct mechanisms of action, XCI is an excellent model system for studying IncRNA function. There
are several crucial steps in the XCI process, including: blocking Xist expression except on one allele, upregulation of Xist itself, Xist spreading along the X and Xist-mediated silencing and maintenance of silencing. All of these involve distinct aspects of lncRNA biology (antisense regulation, activating RNAs, RNA-chromatin associations, recruitment of silencing complexes by RNA), and all are worth exploring as examples of mechanisms of RNA-based gene regulation.

Table 1.1: Summary of lncRNAs and proposed interacting protein partners for X-inactivation

<table>
<thead>
<tr>
<th>RNA</th>
<th>Function</th>
<th>cis- or trans-acting</th>
<th>Known protein Interactors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xist</td>
<td>Required for initiation of X-inactivation [22].</td>
<td>cis, can in some cases act in trans at autosomal Xist transgenes [62]</td>
<td>PRC2 [60], YY1 [62], hnRNP-U [63], ASF [64], Atrx [65] many others (see Chapters 2 &amp; 4)</td>
</tr>
<tr>
<td>Tsix</td>
<td>Represses Xist expression by silencing the Xist promoter [66-69], also required for X-chromosome pairing, counting the number of X-inactivation centers, and mutually exclusive allelic choice [57, 70, 71]</td>
<td>cis [52]</td>
<td>Dnmt3a [67, 69], CTCF [72]</td>
</tr>
<tr>
<td>RepA</td>
<td>Independent transcript from Xist 5’ end, helps activate Xist [60]</td>
<td>cis [60]</td>
<td>PRC2 [60]</td>
</tr>
<tr>
<td>Jpx</td>
<td>Activator of Xist transcription; counting of X-chromosomes [59, 73]</td>
<td>trans (mild cis preference)  [59]</td>
<td>CTCF[73]</td>
</tr>
<tr>
<td>Ftx</td>
<td>Potential activator of Xist expression [61, 74]</td>
<td>Unknown; possibly cis acting through transcription [61, 74]</td>
<td>unknown</td>
</tr>
</tbody>
</table>
Xist RNA itself is multi-functional and modular. Xist has a unique structure, with two very long exons (exon 1: 9.5 kb, exon 7: 7.7 kb in mouse), and several distinct, conserved tandem repeats are found within Xist [75-77]. These tandem repeats are associated with particular aspects of Xist function (Fig. 1.2).

![Tandem repeats within Xist and their putative functions](image)

**Figure 1.2: Tandem repeats within Xist and their putative functions.** A scale diagram of Xist is shown with exonic sequence labeled black, introns labeled tan and tandem repeats labeled green. Below the diagram are the positions of all tandem repeats, and the function assigned to them (blue: localization, red: silencing, purple: Xist upregulation).

Repeat A is required for silencing [77, 78], and has been implicated in recruiting silencing complexes such as the Polycomb Repressive Complex 2 (PRC2) to the Xi [60]. Repeat A also plays a role in transcriptional upregulation of Xist [78, 79]. Repeat B and its surrounding regions have also been shown to be required for silencing, through recruitment of the Polycomb Repressive Complex 1 (PRC1) complex to the Xi [80, 81]. Repeat C is implicated in Xist RNA localization to chromatin, antisense oligonucleotides against Repeat C prevent Xist from forming a tight “cloud” in the nucleus [82, 83]. Repeat F is also involved in localization, as it contains binding sites for the transcription factor YY1, which tethers nascent Xist transcripts to chromatin [62]. Repeat F is also important for Xist expression, and partially overlaps with one of
Xist’s promoters [79, 84]. Repeat E is another region of Xist required for localization [85-87], as it binds localization factor CIZ1 [85, 86]. Repeat D is a large component of exon 1, but its function in XCI has not been fully determined, though it may also play a role in silencing [88]. The tandem repeats within Xist may each act as a “module”, recruiting particular protein complexes to Xist to carry out defined aspects of XCI such as ensuring its proper localization to the X-chromosome or recruiting one of the many silencing complexes that localize to the Xi.

Due to the tractability of genetic systems and RNA, an understanding of how the different parts of Xist RNA is emerging, however (as will be discussed), less is known about the protein partners of these particular regions.

Much work has centered on the role of Xist in initiating X-inactivation in the early developing embryo. However, it is now clear that Xist has important roles in later development and in adults as well, long after the establishment of XCI. Since XCI silences several hundred genes, some of which are oncogenes, improper XIST regulation could potentially be a mechanism underlying tumorigenesis [89, 90]. Early experiments suggested a tantalizing connection between X-linked gene dosage and cancer. Loss of XCI and downregulation of XIST expression are commonly observed in basal-like cancer, BRCA1-null triple negative breast cancer[91-97], and ovarian cancer lines [98, 99]. Loss of XIST is most commonly caused by X isodisomy, where Xi is lost and Xa is amplified. Reactivation of Xi may be an alternative mechanism leading to loss of XIST and over expression of the X [93, 95, 98, 99]. These observations suggest a correlation between X-chromosome dysfunction and cancer. The link between XCI and cancer was recently determined to be causal, with the finding that deleting Xist in the blood compartment leads to a highly aggressive myeloproliferative neoplasm and myelodysplastic syndrome (mixed MPN-MDS) with 100% penetrance and lethality in mice [100]. This result clearly demonstrates that loss of Xist RNA and overexpression of the X in adult tissues can lead to cancer. As loss of Xist RNA promotes the initiation or progression of cancer, it may be reasonable to propose reactivation of XIST as a therapeutic strategy in cancer.
Somewhat ironically, therapeutic strategies to reactivate XIST on amplified or reactivated X-chromosomes in cancer may resemble the strategies used to reactivate the wild-type copy of X-linked genes in diseases like Rett Syndrome.

In order to correct X-linked diseases through reactivation of the wild-type allele, or to explore the role of XIST as a tumor suppressor, it is necessary to define all the mechanisms used to keep genes on the Xi silent. Much progress has been made in identifying and testing mechanisms of Xist-mediated silencing and epigenetic stability of the Xi, though our understanding is far from complete. Therefore, studying mechanisms of X-inactivation is beneficial both because it is an excellent model for RNA-based gene regulation and multiple epigenetic processes, and because better knowledge of silencing may help guide the next generation of therapies for cancer and X-linked developmental disorders.

**Tsix: the negative regulator of Xist**

The earliest steps in XCI involve counting the number of X-chromosomes and if there is more than one X, ensuring only one of them expresses Xist. In mouse, Tsix is a crucial antagonist of Xist expression and plays critical roles in preventing Xist expression and XCI from occurring on multiple chromosomes. Tsix is a long, non-coding transcript that runs antisense to Xist and prevents Xist transcription in cis. In the early blastocyst, Tsix is expressed only from the maternal allele [57, 58], and maternal inheritance of a Tsix deletion or transcription block cassette causes embryonic lethality due to upregulation of Xist on both alleles in the extraembryonic tissue [57, 58]. However, paternal inheritance of a Tsix loss-of function allele does not cause embryonic lethality [57, 58], and in fact paternal inheritance of an Xist deletion rescues disruption of maternal Tsix [58]. These results demonstrate that Tsix is critical for imprinted XCI, as it is needed to repress Xist on the maternal allele and ensure the maternal X remains active in the extraembryonic tissues. In random XCI, deletion of Tsix skews choice of which chromosome to inactivate towards the deleted allele [57, 101], again demonstrating that
Tsix represses Xist. The fact that XCI can be almost completely skewed towards Tsix loss-of-function alleles enables them to be powerful tools for studying XCI initiation in hybrid mESCs [36].

A considerable amount of genetic analysis has been carried out on Tsix towards understanding the antisense mechanism of action. Tsix controls Xist expression in cis by modulating the chromatin structure and DNA methylation status of the Xist promoter [67, 69]. Anti-sense transcription extending through the Xist promoter is required to silence Xist in cis [66, 68, 102]. It is possible that Tsix acts as a functional RNA and recruits repressors such as the de novo methyltransferase, DNMT3A [67, 69], or titrates activators away from the Xist promoter. The act of antisense transcription through the Xist promoter could also induce a chromatin state that is refractory for sense transcription; alternatively, it could disrupt RNA polymerase function in the sense direction. Definitive experiments to test these hypotheses must separate transcription of the antisense RNA from the action of the antisense transcript.

Tsix’s mechanism of action may extend to other antisense transcripts. There are at least several hundred sense-antisense pairs within mammalian genomes [103-106], many of which are arranged in a structurally similar manner as the Xist-Tsix pair [105]. In these cases, the antisense transcript might similarly regulate expression of the sense transcript with which it overlaps [107-109]. In several well-studied examples within imprinted loci, allele-specific expression patterns have been proposed to be controlled by expression of an antisense transcript. For example, Air silences the paternal Igf2r cluster [110-112] and Kcnq1ot1 silences the paternal Kcnq1 cluster [113, 114]. Both antisense transcripts are implicated in binding of repressive chromatin factors, such as G9a and Polycomb Repressive Complex 2 (PRC2) [111, 114]. There are likely many other examples of sense-antisense transcription modules that may operate like the Xist-Tsix pair. Thus, uncovering the molecular mechanisms that underpin Tsix-mediated regulation of Xist may have broad applicability for understanding the role of antisense transcription.
In addition to its role as an inhibitor of $Xist$ expression on the active $X$, $Tsix$ plays a role in interchromosomal contacts hypothesized to be crucial for X-chromosome choice and for properly demarcating only one active $X$. Before $Xist$ is upregulated, the two $Xic$’s of the female cell transiently come into close contact with each other in the nucleus (Fig. 1.3) [115-118]. This transient “pairing” of the two $Xic$’s may allow competition for transcriptional activators between the two alleles [117, 119-121]. Cooperative interactions between the transcriptional activators such as CTCF [122], YY1 [118] and OCT4 [118, 123] results in asymmetric binding of activators to one X-chromosome. Binding of activators in a cooperative complex on one allele enables expression of Tsix RNA and inhibition of $Xist$ upregulation only on one chromosome. Bringing the X-chromosomes together and enabling redistribution of Tsix activators until only one X-chromosome expresses Tsix enables the cell to efficiently count the number of X-chromosomes that need to be inactivated and choose one as the inactive $X$ [119, 120]. Tsix RNA itself appears to be important for establishing pairing. $Tsix$ and its enhancer, $Xite$, are necessary for pairing [117] and are also each sufficient to induce pairing when integrated onto an autosome [122]. Interestingly, inhibition of transcription with actinomycin D prevents the formation of new pairing complexes but has little effect on the half-life of paired complexes already formed, suggesting that RNA may be required to attract two X-chromosome to each other but not to keep them paired [122]. Tsix RNA can also interact with CTCF, a putative transcriptional activator of Tsix [122] and crucial regulator of chromatin architecture, and knockdown of Tsix RNA disrupts pairing [72]. These observations support the hypotheses that Tsix transcripts mediate pairing by forming a complex with activating factors. In addition to Tsix RNA, another long non-coding RNA, PAR-TERRA may be involved in pairing. PAR-TERRA RNA is expressed from the telomere and is also required for pairing [124]. The $Xic$ interacts with the site of PAR-TERRA transcription, and induces pairing between the telomeres on the $X$. This pairing between the telomeres induced by PAR-TERRA RNA may be necessary to initiate pairing at the $Xic$ [124]. These results provide yet more evidence for RNA-mediated contacts involved in pairing.
Figure 1.3: Model for Xic pairing before XCI onset.

The two X-chromosomes are epigenetically identical and euchromatic in the pre-XCI stage. The two Xs are brought together by Tsix and Xite (pairing) during cell differentiation to enable cross-talk and mutually exclusive choice of Xa and Xi. Because it is thermodynamically favorable to do so, hypothetical transcription factors, potentially OCT4 and CTCF (blue circles), that were previously randomly distributed between the two Tsix/Xite alleles stochastically shift to one X, which would then become future Xa. This shift results in monoallelic Tsix expression and differential chromatin modifications within the Xist region, which lead to repression of Xist on Xa and upregulation of Xist on Xi.
Ensuring upregulation of Xist on one X: positive regulators of XCI

Some IncRNAs of the Xic appear to be transcriptional activators. They include RepA, Jpx, and Ftx, all proposed to be activators of Xist (Table 1.1). The molecular mechanisms that cause these RNAs to induce Xist are currently unknown, though recent studies have suggested several intriguing possibilities.

RepA RNA is transcribed from an independent transcription unit within exon 1 of Xist and consists of a repeated motif (Repeat A) that directly binds and targets PRC2 to the Xist promoter [60]. RepA is believed to induce Xist expression by increasing H3K27 trimethylation of the Xist promoter via its recruitment of PRC2 [60]. As PRC2 is a repressive complex, this may seem to be a counterintuitive way to activate Xist transcription. However, Xist may prefer a heterochromatic environment. Indeed, Xist remains active in the repressive context of the inactive X, it may actually be induced by repressive signals. This intriguing possibility has to be tested more directly to determine whether PRC2 recruitment to the Xist promoter actually induces Xist expression. Consistent with the idea, a mouse deletion of RepA results in loss of Xist induction [78]. It is also possible that the Xist promoter may be bivalent, decorated with both repressive H3K27me3 and active H3K4me3 in the undifferentiated state, poised for transcription like many other developmentally regulated genes [125]. Also, the Repeat A region of the Xist gene itself an important component for full upregulation of Xist [79].

Jpx is an RNA produced from a locus several kilobases upstream from Xist and is another positive regulator of Xist and XCI. Jpx’s activating influence is supported by genetic analysis. When female cells are deleted for Jpx – even on just one allele – the cells can no longer induce Xist expression [59]. Jpx does not apparently function as an enhancer, as standard enhancer assays failed to reveal any activating influence on the Xist promoter. Furthermore, its ability to activate Xist in trans (when Jpx is placed in an autosomal context) also argues against an enhancer mechanism [59]. Thus, Jpx’s action is distinct from enhancer-associated RNAs [126-128]. Because a post-transcriptional knockdown of Jpx phenocopies a
*Jpx* knockout, the transcript itself (not just the *Jpx* gene or transcription) must be the activating force. The latest work indicates that Jpx RNA is part of the X-chromosome counting mechanism and activates *Xist* by titrating away a repressive autosomal factor, CTCF, that normally binds and blocks the *Xist* promoter [73].

*Jpx* could act in other ways as well. *Jpx* is trans-acting, but has a mild cis-preference [59]. Consistent with a cis-preference, chromosome conformation analysis suggests that the *Jpx* locus makes contacts with the *Xist* promoter following the onset of differentiation and XCI [129]. It is possible that Jpx RNA mediates the formation of these chromatin contacts, as has been shown for several other activating RNAs [130]. Another possibility is that the chromatin contacts form independently of Jpx RNA and juxtaposing nascent Jpx transcripts to the *Xist* promoter allows Jpx to recruit activators or titrate repressors from the *Xist* promoter. Studying these activating RNAs may provide several important model systems for evaluating the functions of mechanisms of recently discovered a-RNAs [126] and enhancer RNAs [127, 128].

*Ftx* is another non-coding positive regulator of *Xist*. *Ftx* lies within the XCI, upstream from *Jpx* and transcribed in the same orientation as *Xist*. Like *Jpx*, deletion of *Ftx* decreases *Xist* expression and interferes with initiation of XCI in mESCs [61, 74]. However, it appears that the act of transcription is what is required for *Ftx* to fully activate *Xist*, since targeted degradation of *Ftx* transcripts with LNA gapmers does not affect *Xist* [74].

In addition to non-coding activators of *Xist*, a model for ensuring proper inactivation of only one X-chromosome through feedback loops of dosage-sensitive X-linked genes has also been proposed. The E3 ubiquitin ligase RNF12 (RLIM) is one of the closest coding genes to *Xist* and has been implicated as a dosage-sensitive regulator of *Xist* and one of the links between pluripotency and X-inactivation. *Xist* is upregulated if additional copies of RNF12 are present, and *Rnf12*+/- mESCs are partially defective in the initiation of XCI [131, 132], suggesting that the dose of RNF12 helps set a threshold for *Xist* expression. RNF12 is upregulated upon differentiation and targets REX1 for degradation; REX1 binds to the 5' region of *Xist* and is
implicated in preventing Xist transcription [133]. Thus, RNF12-mediated degradation of REX1 may release the “breaks” on Xist transcription [133, 134], allowing Xist induction. And since RNF12 is X-linked, once one allele is silenced, the RNF12 concentration falls below the threshold to enable Xist expression from the remaining active X [135]. However, much of the work defining this model for RNF12-mediated Xist regulation was performed in mESCs. While further examination of imprinted XCI in extraembryonic tissues shows that maternal RNF12 is required for imprinted XCI (and viability) [136], RNF12 appears to be dispensable for random XCI in the epiblast [137]. In addition the role of RNF12 in mESCs may be dependent upon culture conditions [138], and the studies showing Rnf12 as required for XCI in mESCs generated truncated proteins with possible neomorphic functions rather than clean knockouts [137]. These results suggest that there may be RNF12-dependent and independent mechanisms of Xist upregulation [138], perhaps due to redundancy between mechanisms of Xist upregulation.

Several non-coding loci (Jpx, Ftx, RepA) as well as a nearby protein-coding gene (Rnf12) have been implicated in regulating Xist upregulation. Xist upregulation is a crucial step in XCI but must be tightly regulated to ensure only one X becomes inactive. Further work is needed to characterize these positive regulators of XCI and the molecular mechanisms underlying control of Xist regulation.

**Xist RNA localization through adaptor proteins**

While it is clear that Xist RNA spreads across the entire X-chromosome, mechanisms that localize Xist RNA itself are just beginning to emerge. Several localization factors, including YY1, hnRNP-U (SAF-A) and CIZ1 have been identified. In addition, advances in RNA capture technology have enabled the coverage of Xist RNA across the Xi to be mapped with kilobase resolution and have indicated a role for higher-order structure in the spread of Xist.
Localization begins with loading of Xist RNA at a “nucleation center” located within exon 1 of the Xist locus [62]. The transcription factor, YY1, is required for Xist RNA loading onto the nucleation center. Knocking down YY1 or mutating its three binding sites within the nucleation center eliminates Xist loading and furthermore blocks formation of the prominent cloud of Xist RNA seen in RNA fluorescence in situ hybridization (FISH) experiments. Xist RNA directly binds YY1 in vivo and in vitro, and YY1 in turn directly contacts three YY1-binding sites near “Repeat F” within Xist exon 1 [62]. Thus, YY1 is a “bivalent” protein (capable of binding both RNA and DNA) and acts as “bridge” between Xist DNA and RNA. In this way, YY1 tethers the Xist-PRC2 complex to the nucleation center and positions the complex to spread throughout the rest of the chromosome.

Expressing an Xist transgene carrying mutations in the YY1 binding sites led to the very surprising discovery that Xist RNA can act in trans [62]. In such situations, transgenic Xist RNA can not bind the nucleation site in cis (on the transgene) but was observed to diffuse to the nucleation site of the Xi and then spread along the Xi. Furthermore, when the transgene is not mutated, Xist RNA produced from the Xi could migrate to the transgene nucleation site and spread along the autosome. This observation led to questions about how the Xa does not engage in binding of Xist. Further analysis indicated that YY1 binds only the nucleation site on Xi (not on Xa) [62]. Thus, the cis-limited action of Xist RNA occurs in the normal developmental context likely because of developmental programming which blocks YY1 nucleation sites at all but the Xi. Xist is a cis-acting RNA only to the extent that sites in trans are prevented from binding.

Two other proteins have been shown to be required for proper Xist localization to Xi chromatin. hnRNP-U (also called SAF-A) is a component of the nuclear matrix and is enriched on the Xi [139, 140]. Knockdown of hnRNP-U prevents the formation of an Xist cloud and initiation of XCI. Instead for forming a tight cloud, Xist diffuses throughout the nucleus and appears as dozens of single puncta after knockdown of hnRNP-U [63]. hnRNP-U interacts
directly with Xist RNA, and interestingly; hnRNP-U contains both an RGG RNA binding domain and a SAF-box that binds AT-rich DNA [63]. Both the DNA-binding SAF-box and the RNA-binding RGG domain are required for proper Xist cloud formation. Unlike many nucleic acid binding proteins, hnRNP-U binding to DNA is not disrupted by adding RNA (or vice versa) [141, 142], and hnRNP-U can form large multimeric complexes with nucleic acid [141, 142], suggesting that hnRNP-U may be able to bind DNA and RNA at the same time. These unique biochemical properties of hnRNP-U, combined with the requirement for both RNA and DNA-binding domains in Xist localization implies a model where hnRNP-U constrains Xist to the Xi by acting as a bridge between Xist RNA and the nuclear matrix [63].

CIZ1 is another protein required for proper Xist localization [85, 86]. Knockdown or knockout of CIZ1 causes Xist to exhibit a disperse, punctate distribution in both cultured fibroblasts and somatic cells from Ciz1-/- mice [85, 86]. Ciz1-/- mice also show aberrant upregulation from the X-chromosome, though the level of upregulation is modest suggesting that dosage compensation is partially intact [86]. CIZ1 binds to the Repeat E region of Xist directly [85, 86], and CIZ1 super-resolution microscopy shows CIZ1 protein to be more closely associated with Xist RNA than any other Xist-interacting protein tested [85]. Additionally, CIZ1 remains associated with Xist RNA following hnRNP-U knockout, demonstrating that CIZ1 interacts with Xist independently of hnRNP-U [85]. CIZ1 knockout mice are viable, but exhibit a lymphoproliferative disorder [86] somewhat reminiscent of the blood cancer observed following deletion of Xist in the hematopoietic lineage[100]. CIZ1 is also a nuclear matrix protein, like hnRNP-U [85].

Since CIZ1 and hnRNP-U both bind to Xist RNA independently, it is possible that they act as partially redundant mechanisms to localize Xist to the matrix and to the Xi. In fact, CIZ1 is recruited to the Xi late in differentiation, and may serve to stabilize Xist after XCI initiates [86]. This model would also explain why Ciz1-/- embryos are viable and are partially dosage compensated [86]. Given the possibility that there are at least three pathways (YY1-, hnRNP-U
and CIZ1-dependent pathways), and possible functional redundancy between these pathways, more work is needed to identify and characterize other localization factors.

**Large-scale genomic features guide Xist spreading**

Identifying and studying candidate Xist localization factors such as YY1, hnRNP-U and CIZ1 has provided important insights regarding the mechanisms of Xist localization to the Xi. Sequencing technologies that produce binding profiles of Xist across the Xi at high resolution provide insight into Xist spreading and localization from a different perspective. RAP and CHART are two methodologies analogous to ChIP-seq, but for RNAs, where RNA is first crosslinked to chromatin, chromatin is sonicated into small fragments, and DNA crosslinked to an RNA of interest is capture by hybridizing anti-sense probes to the RNA. Two recent studies have used either Capture Hybridization of RNA Targets (CHART) [143] or RNA Affinity Purification (RAP) [144] to enrich for Xist-binding DNAs and search for possible Xist targeting elements [144, 145]. Both studies failed to find a particular sequence motif or specific chromatin state enriched for Xist, but intriguingly showed a strong correlation between early Xist localization and contact frequency between early sites and the Xic as determined using published Hi-C data from male ESCs [144, 145]. Both RAP and CHART also showed that Xist preferentially localizes to more gene-rich regions of the Xi, avoids gene-poor lamin associated domains (LADs), and that is Xist coverage is highly correlated with H3K27me3 coverage on the Xi. Also, in contrast to the long-standing hypothesis that the enrichment of LINE elements on the X may serve as "booster stations" for spreading, Xist coverage is anti-correlated with LINE density [146]. Both studies highlight a two-step spreading mechanism that is intimately linked with the 3D conformation of the X. Xist first has a strong tendency to bind sites in close proximity to the X, but as Xist expression continues, the RNA invades regions further from the Xic. It is possible that this second spreading step involves re-organization of the structure of the X to bring Xist early sites into closer proximity with distal sites [144], or involves the transfer of
PRC2 complexes and Xist from strong, constitutive PRC2 sites to weaker, Xi-specific sites [36]. Together, RAP and CHART studies of Xist localization show that Xist RNA does not appear to bind to highly specific sites like a transcription factor. Instead the distribution of Xist across the Xi is broad but appears to correlate with other large-scale features of genome organization such as gene density, lamin association and 3D proximity to the Xic. Much of the evidence concerning large-scale genomic features and Xist spreading is correlative; future work disrupting particular elements of nuclear structure will provide tests of the hypothesis that the large-scale genomic organization directs Xist spreading. In addition, experiments interrogating the relationship between known localization factors and Xist spreading by RAP/CHART may further elucidate mechanisms of spreading.

**Multiple silencing systems establish and maintain XCI**

The Xi is a repressive environment, as XCI is established multiple silencing complexes are recruited to the Xi, often directly or indirectly by Xist RNA. Histone modifications associated with active chromatin (histone H4 acetylation [147], H3K4me3 [148]), as well as RNA polymerase II [149] are rapidly lost from the future Xi following Xist upregulation. Repressive histone modification complexes are recruited to the Xi slightly later. Some of the most notable examples are include the PRC2 complex which decorates the Xi with H3K27me3 [150-153], both the canonical [154-156] and non-canonical PRC1 complexes [80, 81, 157] which deposit H2AK119ub, several H3K9me2/me3 complexes [158-160], H4K20me1 [161] and the variant histone macroH2A [162]. In addition, extensive DNA hypermethylation of X-linked promoters [163-165] and hypomethylation [165] of gene bodies also occurs later in X-inactivation. Additional protein complexes such as SMCHD1 [166-168] that play distinct roles in the X-inactivation process are also recruited to the Xi several days after Xist induction in cell culture. Although much progress has been made identifying repressive complexes enriched on the Xi, the exact mechanisms of both their recruitment and silencing itself are still being elucidated.
The silencing complex on the Xi that has received perhaps the most attention is the PRC2 complex and its relationship with Xist RNA. Xist RNA is one of the first examples of an RNA that recruits a chromatin-modifying complex to specific sites (Fig. 1.1). PRC2 is attracted to the X-chromosome by Xist RNA through a repeated motif at the 5’ end of the RNA, known as “Repeat A” [60]. The Repeat A motif has been reported to directly interact with EZH2, the catalytic subunit of PRC2, both in vivo and in vitro. PRC2 in turn decorates the X-chromosome and silences it as it trimethylates histone H3 at lysine 27 (H3K27me3) [150-153]. Along the X, PRC2 first binds ~150 “strong sites”, which have canonical features of known PRC2 binding sites, including a CpG-rich content and presence of bivalent domains [36]. From the strong sites, PRC2 migrates laterally and locally, giving rise to thousands of non-canonical domains which may represent sites of dynamic spreading along the X chromatin [36]. H3K27me3 density also spreads out from the strong sites, and H3K27me3 occupancy is anti-correlated with LINE density [36, 37], an intriguing finding given a long-standing hypothesis that LINE elements serve as “booster elements” that help X-inactivation spread across the whole chromosome [146]. When expressed ectopically from autosomal transgenes, Xist RNA also recruits PRC2 and silences genes located in cis [29, 62, 150], demonstrating that Xist RNA is both necessary and sufficient to recruit PRC2 and inactivates genes on a multi-megabase scale.

The pioneering studies demonstrating an interaction between Xist RNA and the PRC2 complex suggest a simple model for targeting chromatin modifiers to specific genomic locations in cis, and the observations that as a myriad of transcripts associate with PRC2 [169-173] as well as the finding that many other chromatin modifiers have been shown to interact with RNA, such as Dnmt3a, G9a, PRC1, MLL-WDR5 and LSD1-CoREST (reviewed in [174]) led to great enthusiasm for this model as a novel mechanism for controlling gene regulation. However, the mechanisms of PRC2 recruitment and its role in silencing have become less clear upon further examination. Analysis of the full set of transcripts that can be identified through UV-crosslinking and immunoprecipitation in vivo showed that almost all nascent transcripts associate with PRC2
Furthermore, *in vitro* binding experiments suggest that PRC2 binds most RNAs with similar affinities [177-179], though Xist Repeat A RNA has been shown to have a higher affinity than controls [180] and it has been suggested that PRC2 can bind non-specifically to RNA in general but specifically to certain RNAs [178]. RNA binding has also been shown to antagonize both the ability of PRC2 localize to bind chromatin [175] as well as catalyze H3 lysine 27 trimethylation [176, 180]. These results all suggest a relationship contrary to the simple recruitment model suggested by early experiments. It has been proposed that the effect of RNA binding on PRC2 activity is context-dependent: at lowly transcribed regions there isn’t enough nascent RNA to compete it off chromatin, allowing chromatin binding and catalysis of histone trimethylation, but at more highly transcribed regions the nascent RNAs prevent PRC2 from associating with chromatin and blocking transcriptional silencing [175]. Never the less, PRC2 association with the Xi depends upon Xist RNA [29, 150], thus the precise molecular mechanisms that enable Xist to recruit silencing complexes to Xi chromatin remain to be elucidated.

The precise effects of the various silencing complexes on the Xi are also unknown. It has been reported that the non-canonical PRC1 complex associates with Xist RNA first and leads to the recruitment of both canonical PRC1 and PRC2 and gene silencing [80, 81], suggesting a hierarchical relationship between the various PRC1 complexes enriched on the Xi. The temporal dynamics of chromatin modifications on the Xi also suggest a sequential recruitment of complexes to the Xi: initially, active modifications are lost, then “facultative” heterochromatin marks like H3K27me3 and H3K9me2 are acquired, and maintenance factors such as SMCHD1 and DNA methylation become enriched. Perhaps most importantly, many of these silencing pathways are highly redundant. Disruption of neither the PRC2 [181] complex nor the PRC1 [182] complex does not inhibit the initiation of gene silencing. DNA methylation inhibitors alone lead to very little reactivation of gene expression from the Xi [183, 184]. In fact, deletion of *Xist* itself in fibroblasts does not disrupt silencing after many passages [30-32],
suggesting that Xist recruits multiple mutually reinforcing silencing complexes to the Xi that persist through feedback loops without Xist to continually replenish them.

A final complexity in silencing XCI is that fact that some genes escape from X-inactivation and continue to be expressed from both X-chromosomes, at least in a certain number of cells. About 20-30 genes escape from XCI in mouse [185-188], with some tissue-specificity in escape. In contrast, in humans, nearly 10% of X-linked genes escape in all cell types and individuals tested, and an additional 15% of X-linked genes escape in some cell types or in some individuals [189]. Escapee genes are generally clustered towards one end of the human X in the “pseudoautosomal region” [189]; mouse escapees are dispersed throughout the chromosome [185]. Mechanisms of escape are poorly understood; Xist spreading seems to be specifically depleted over escapees [144, 145] but no explanation for this depletion or for continued expression itself is known.

_Defining the “Xist proteome” is a major goal for understanding mechanisms of silencing during XCI_

Although much attention has focused on interactions between Xist and candidate silencing complexes such as PRC2, X-inactivation is clearly a complex process that involves multiple, redundant and mutually reinforcing layers of silencing. The apparent lack of necessity for any one particular silencing complex, as well as the diverse variety of silencing complexes found on the Xi indicates that broader, more unbiased approaches are needed to fully elucidate the mechanisms of X-inactivation. In addition, the apparent heterogeneity of Xist localization mechanisms (tethering the nascent transcript by YY1, association with the nuclear matrix through hnRNP, spread from active regions to inactive regions) also suggests many facets of Xist biology are not well explored. At the start of my PhD, a major problem in X-inactivation was identifying the full suite of proteins that bind to Xist RNA (“the Xist proteome”). Identifying all Xist interacting proteins would also uncover many biological pathways that are important for proper
Xist function and silencing during X-inactivation. Although many initial observations from candidate approaches, such as the enrichment of PRC2 on the Xi and the importance of hnRNPs for localizing Xist have been fruitful, the thought was that finding all Xist interactors in an unbiased manner may uncover a suite of new pathways important for silencing. My colleague Anand Minajigi made a major contribution to understanding the Xist proteome by developing a method (called “iDRiP”, “Identification of Direct RNA-Interacting Proteins”) to isolate and identify proteins that can be UV-crosslinked to Xist RNA. I worked on this problem as well, as I followed up the unexpected observation that Xist RNA interacts with cohesin proteins, which led to a hypothesis that Xist RNA may regulate higher-order chromatin structure on the Xi. Our work identifying Xist interacting proteins and exploring their roles in higher-order chromatin structure is described in Chapter 2 of this dissertation. In addition, the past few years have led to multiple novel approaches to identify Xist interacting proteins or regulators of Xist-mediated silencing from many different groups [190-194]. Our knowledge of Xist RNA-protein interactions has expanded dramatically over the past 6 years and lead to many new insights. This introduction is largely written from the perspective of 2012, before the sea-change in Xist proteomics, and an overview of the “Xist proteomics” experiments and their implications for X-inactivation and are discussed in Chapter 4.

CHROMATIN CONFORMATION: FORM AND FUNCTION

A long-standing hypothesis is that the structure DNA folds into in the nucleus plays a role in regulation of gene expression. X-inactivation is an excellent system for testing this possibility. It has been known for decades that the Xi appears to have a very different structure than the Xa. Thus, examining the structures of the Xa and Xi in detail and working to understand how differences in chromatin organization contribute to silencing on the Xi will provide insight into the role of chromatin structure in gene regulation. In this section, I provide an overview of what was known about the structure of the Xi circa 2012 before genome-wide chromatin
conformation techniques were developed, methods for mapping genome folding and general patterns of genome organization in mammals, the unique structure of the Xi and the emerging picture of the relationship between higher-order structure and gene regulation.

**The structure of the Barr Body: an unsolved feature of the Xi**

There are many differences between the Xi, Xa and autosomes, including transcriptional silencing, histone modification states, replication timing and DNA methylation. In addition to these biochemical differences between the Xi and the autosomes, the Xi folds into a unique structure called the Barr Body. The Barr Body is a dense, compact region of the female mammalian nuclei frequently associated with the nucleolus and visible through simple DNA staining. It was identified in the 1940’s [195] and shortly thereafter recognized as one of the X-chromosomes [196]. The fact that the Barr body appears as a dense structure implies that the Xi is more compact than autosomes or the Xa, however recent imaging studies instead demonstrate that the Xi is only slightly (~20%) more compact than the Xa [197-199]. However, the internal structures of the Xa and Xi appear very different from each other. Although both the Xa and Xi exhibit a “30 nm fiber” chromatin structure, the Xi is more compact at larger length scales, similar to the compaction observed in bulk chromatin upon transcriptional inhibition [200]. The Xi is more spherical and a smoother surface area than the Xa [197-199]. The Xa and autosomes contain alternating areas of dense DNA packaging surrounding “channels” that are relatively DNA free; by contrast, the Xi largely lacks these channels and has a more homogeneous structure [201, 202] (the Xa appears like an open sponge, the Xi like a closed sponge) [202]. Additionally, two chromatin conformation studies [203, 204] suggested that Xi assumes a more “randomized” organization relative to Xa and autosomes, thereby indirectly implicating Xist RNA in generating the less ordered configuration. Indeed, an Xi-specific deletion of Xist partially restored long-range contacts on Xi suggesting that Xist acts to disrupt long-range chromatin contacts [203]. Finally, in addition to these large-scale differences in the
structure of the Xi visible through microscopic analysis, the Xi very frequently is found adjacent
to either the nucleolus [32, 195] or the nuclear envelope [201] (or both), in contrast to the Xa
and autosomal chromosome territories which generally distribute themselves throughout the
nucleus more randomly.

A large body of evidence demonstrates that the Xi adopts a unique structure. However,
at the start of my thesis work in 2012 no one quite knew what this structure looked like at high
resolution. I was intrigued by fact the Xa and Xi fold so differently, and I wanted to map the
structures of the Xa and Xi at high resolution. Preliminary chromatin conformation studies
indicated a role for Xist RNA in regulating chromosome architecture, and I was interested in
rigorously testing the hypothesis that Xist RNA regulates Xi structure. There are several pieces
of evidence implicating RNA as a structural component that determine higher order structures.
RNA has long been known to co-fractionate with chromatin in eukaryotic nuclear extracts [205-
208]. The nuclear matrix consists of a network of ribonucleoprotein particles (reviewed in [209]),
and it has been suggested that Xist RNA is a component of the nuclear matrix, interacting with
factors such as hnRNP-U/SAF-A and ASF [63, 64, 210, 211]. In addition, the apparent
relationship between large-scale chromatin features and Xist spreading, and the proposal that
Xist might remodel the structure of the X to facilitate its spreading motivated further analysis of
Xi structure. I was beginning my PhD just as genomic technologies for mapping chromatin
architecture were reaching maturity, and I decided to use them to map the spatial organization
of the Xi.

I made it my goal to use these technologies to map how the Xi folds and test whether
Xist might regulate the folding of the Xi, the results of which are discussed at length in Chapter
2. Finally, as chromatin conformation was being mapped for the first time, there was a great
deal of interest in understanding how genome folding impacts gene regulation. Once I had
gained insight into how the Xa and Xi fold, I disrupted the organization of the Xi and tested
whether gene expression from the Xi was also disrupted. Those results are described in Chapter 3.

**Methods for examining chromatin conformation**

Advances in “chromatin conformation capture” (hereafter referred to generally as “3C-type”) genomics have provided unrivaled opportunities to map chromatin structure. Chromatin conformation capture is based on proximity ligation techniques [212], where DNA ends in close proximity are ligated together into one molecule. The core idea behind chromatin conformation capture is to use this fact to “capture” sequences close to each other in the nucleus through ligation and identify the sequences in proximity through either PCR, high-throughput sequencing or a combination of the two. In chromatin conformation capture experiments, proteins and nucleic acids are crosslinked together, usually with formaldehyde, and the DNA is sheared with restriction enzymes [213]. DNAs physically close to each other in the nucleus will be crosslinked together, and can be joined together through ligation [213]. The original method, 3C [213], uses inverse PCR to test if two candidate regions interact with each other. 4C is a more high-throughput variant [203, 214, 215], where targeted amplification of a candidate locus is used to enrich for interactions between one “viewpoint” and the rest of the genome. 4C produces a “coverage profile”, a 1-dimensional plot of the interaction frequency between the viewpoint and each position along the chromosome. 5C [216], Capture-C [217] and Hi-C\(^2\) [218] methodologies extend chromatin conformation even further, but selectively enriching for all contacts within a candidate region, usually 1-3 Mb in size. Hi-C [219, 220] is the most general method, as it sequences all interactions in the entire genome, producing large matrices, where the intensity of the heat map at a given position depicts the interaction frequencies between all pairs of loci in any given region (or even the whole genome).

3C-type techniques are quite useful for studying nuclear architecture because of the simplicity and flexibility of high-throughput sequencing. However, it is important to discuss the
validity and limitation of 3C-type techniques to yield meaningful data concerning the organization of the genome in the nucleus. The other major technology for exploring spatial relationships in the nucleus is 3D-DNA FISH, and comparison to FISH serves as the best validator for 3C-type techniques. The advantage of 3C-type techniques over FISH is the ability to use sequencing to specifically identify the interacting loci, rather than looking at a small number of discrete loci or the general shape of a larger region by microscopy, which is all that is possible with the limited number of fluorophores available on most microscopes.

Most 3C-type techniques measure interaction frequencies between pairs of loci across populations, though single-cell methods are becoming available [221]. 3D-DNA FISH, the established microscopy technique for inferring 3d distances between pairs of loci measures distances in single cells. Thus, to truly correctly compare 3d distances from FISH to interaction frequencies from 3C-type data, one must in fact both measure the distribution of distances between pairs of loci across a population of cells and model the expected distribution of interaction frequencies given the 3C-type data [222]. Furthermore, 3C-type techniques and DNA FISH are distinct experiments that each come with their own technical limitations and artifacts; these must be minimized and/or corrected for both types of data before making comparisons [222]. Despite these difficulties in comparing interaction frequencies from 3C-type experiments and DNA FISH, measurements of nuclear structure from 3C-type experiments are generally in good agreement with those from FISH [129, 203, 204, 220, 223, 224], regardless of the 3C-type method employed or the organism of study. One of the most notable exceptions, an apparent discrepancy between the organization of the HoxD locus as measured by 5C and super-resolution microscopy in a Ring1b-/- background, may be due to a major change in the biochemical composition of the chromatin at the locus in Ring1b-/- backgrounds [225]. Furthermore, comparison between results obtained from Hi-C and two new sequencing-based methods [226, 227] for detecting proximity without ligation shows good agreement between Hi-C
and the new methods. Together, a large body of independent evidence supports 3C-type techniques as valid ways to detect and quantify interactions between loci in the nucleus.

**Patterns of higher-order structure: loops, domains and compartments**

![Heatmap showing interaction frequency between all pairs of loci in a region of human chr4, 10kb resolution (data from [220]). “Domains” are visible as squares of elevated interaction frequency at the main diagonal (yellow boxes). “Compartments” are pattern of alternating strong and weak interactions across the matrix (two regions with strong compartment structure are highlighted in grey). “Loops” are strong interactions between two discrete loci (two examples identified with blue arrows).](image)

Figure 1.4: Interaction patterns visible in Hi-C data

Several patterns of interactions have emerged from Hi-C mapping of mammalian genomes (Fig. 1.4). “Domains”, including topologically associating domains (TADs) [204, 228] are continuous regions of the linear genome that contact each other more frequently than and
are insulated from other nearby regions [220]. Domains are visible as “squares” in Hi-C contact maps. Domains are usually observed at length scales ranging from $10^4$-$10^6$ bp. Domains are generally associated with pairs of CTCF sites at the two borders of the domains. CTCF is a DNA binding protein that recognizes a large motif, and has been implicated in higher-order chromatin structure from the advent of 3C-type technologies [229]. “Loops” are pairs of loci that interact more frequently than other nearby pairs of loci; they are visible as strong “dots” in Hi-C contact maps or “peaks” in 4C profiles [220]. Loops frequently occur between CTCF sites, often between the left and right borders of domains [218, 220]. Loops usually form between CTCF-sites where the two CTCF motifs point towards each other (convergent orientation) [218, 220, 230-232]. Loops also generally have cohesin peaks at their anchors [228, 233]. Cohesins are trimeric complexes composed of the ATPases Structural Maintenance of Chromosomes (SMC) SMC1 and SMC3, joined into a putative ring like structure by the RAD21 klesin subunit [234]. Cohesins are protein complexes that classically are involved in holding sister chromatids together during mitosis [234], but have also been shown to be critical for loop formation and stability [235, 236].

In fact, loops and domains are believed to arise from a common process, and this process provides a mechanistic explanation for why loops nearly always form between convergent CTCF sites. The prevailing model is that a molecular motor termed a “loop extrusion complex” [237, 238], possibly cohesin in interphase cells, slides along DNA and extrudes progressively longer loops until it encounters a properly oriented CTCF site, where extrusion stops [218, 233, 239]. Thus, domains emerge in Hi-C by averaging positions of loops that have not been extended to reach the borders yet and loops represent the positions of stalled extrusion complexes at the domain borders [239-242]. There are many computational methods used to detect domains. Two of the most common are finding places where there is an abrupt shift in “directionality index” (the ratio of the number of reads interacting with locus i from upstream to the number of reads interacting with locus i from downstream) [228], and the local
minima of insulation score (the ratio of number of reads that cross over from one side of locus i to the other to the reads immediately adjacent to locus i) [223]. The positions of domain borders, as well as the most rigorously defined loops (those that appear in high resolution Hi-C experiments and obey the convergent rule) are generally the same between cell types and even across species [220, 228], however interactions within domains can vary greatly between different cell types [228, 243].

In addition to domains and loops, a third pattern has also become apparent and arises through a distinct mechanism [233, 244-246]. “Compartments” represent heightened interaction frequencies regions of the genome with similar epigenetic statuses active regions interact with other active regions, inactive regions interact with other inactive regions, and active regions have decreased interactions with inactive regions [219, 220, 233, 244-246]. Compartments can be visualized in Hi-C data sets by an alternating “plaid” pattern of strong and weak interactions. Compartments are usually detected as the 1st principle component of the Hi-C contact map.

**Loops, domains and compartments are all distinct on the Xi**

Given the ability to detect all these folding patterns using Hi-C, the divergent epigenetic statuses of the Xa and the Xi, and the unique structure of the Barr body visible by microscopy, the structure of the X-chromosomes has attracted much attention. At the beginning of my PhD, very little was known about the structure of the X-chromosomes and the chromatin conformation capture tools to map genomic organization at high resolution had just been developed. Initial results from 4C studies demonstrated that the Xi has a more random contact pattern than the Xa [203], and 5C studies suggested that TADs might be absent from Xi [204]. A major focus of my PhD was developing allele-specific Hi-C, to be able to map the structures of the Xa and Xi in the same population across the entire chromosome and to test whether loops, domains and compartments are distinct on the Xi as suggested by the initial 4C and 5C studies. The results of these allele-specific Hi-C experiments are described in depth in Chapter 2 but can be
summarized as follows: My allele-specific Hi-C results [192] suggested that there is a global weakening in TAD strength across the Xi, and that some TADs can be restored following Xist deletion. In addition, I observed a general depletion of architectural proteins CTCF and cohesin across the Xi by allele-specific ChIP-seq, and the regions of the Xi that show restoration of TADs after Xist deletion also show restoration of cohesin binding [192]. These results suggest that Xist RNA antagonizes cohesin binding on the Xi, breaking down TAD and loop structures [192]. The depletion of CTCF from the Xi has been also been observed in other XCI model systems [37, 185], and several additional allele-specific Hi-C analyses of the Xa and Xi have shown strong reduction in TAD strength on the Xi [199, 247, 248].

Hi-C experiments also uncovered additional novel aspects of Xi structure. The Xi is divided into two megadomains with tandem repeat locus Dxz4 as the border [192, 199, 247]; the megadomains have been observed directly using super-resolution imaging [199, 249]. Hi-C also uncovered a network of extremely long-range interactions (superloops) between CTCF-bound tandem repeat loci on the Xi [220]. The interactions between these tandem repeat loci span tens of megabases on both the human and mouse X and are at least an order of magnitude longer than any other loops observed in the rest of the genome, at least under wild-type conditions [220, 233].

Finally, independent work from my colleague, Chen-Yu Wang, showed that compartments on the Xa and Xi are also markedly different. The Xi is devoid of compartments [248], unlike the Xa and the autosomes. However, deletion of SMCHD1, a non-canonical SMC protein that both interacts with Xist RNA and is enriched on the Xi causes the Xi to fold into “S1/S2” compartments that are distinct from the A/B compartments [248]. The S1/S2 compartments form on the Xi early in X-inactivation and disappear later, once SMCHD1 is recruited. Notably, SMCHD1 deletion leads to a major failure of both silencing and Xist spread during X-inactivation [166, 168, 248, 250, 251], suggesting that the disruption of compartments
on the Xi is needed for X-inactivation to occur properly. These results imply a model where Xist RNA first spreads to A compartments, then Xist remodels the compartment organization into S1/S2 compartments and finally SMCHD1 is recruited to fuse together the S1/S2 compartments and enable Xist to fully spread across the entire X [248].

Thus, every feature of chromatin organization is different on the Xi. Local loops and domains are disrupted, compartments are gone, and there are extremely long loops between certain loci and the whole chromosome folds into “megadomains”.

**Unique large-scale structures on the Xi involve tandem repeats**

Tandem repeats appear to be closely associated with the unique structural features of the Xi. *Dxz4* forms the border between megadomains on the Xi in mouse [192, 199, 247], human [220] and chimpanzee [252]. In the human genome, *DXZ4* is a very large, multi-kilobase tandem repeat with dozens of repeat units. Human *DXZ4* assumes very different epigenetic statuses on the Xa and the Xi; on the Xa it is heterochromatinized and methylated but on the Xi, in contrast to the rest of the X-chromosome it is euchromatic and binds CTCF [253, 254]. *Dxz4* is conserved in mouse but exhibits a highly unusual pattern of conservation. Mouse *Dxz4* is not immediately recognizable based on sequence conservation, instead; it was identified based on the presence of a tandem repeat in a syntenic region of the mouse X [255]. Murine *Dxz4* contains fewer repeat units than human *DXZ4* and there is no sequence conservation outside of a short, repeated motif containing a CTCF site. Despite these sequence differences between mouse and human *Dxz4* and dramatic re-arrangement of the mouse and human X-chromosomes, murine *Dxz4* is also the border of the megadomains on the Xi and also binds architectural proteins on the Xi unlike the rest of the chromosome.

In addition to *Dxz4*, several other tandem repeats are involved in unique interaction networks on the Xi. *Dxz4* forms extremely long looping interactions (>10Mb) between other
tandem repeat loci on the human Xi [220, 252, 254], including XIST and another CTCF-bound tandem repeat called FIRRE [256, 257]. These superloops are far longer than almost all other contacts in mammalian genomes; the only comparable interactions across such distance that have been reported are networks of contacts between super-enhancers in cohesin-depleted cells [233]. The interactions between Firre, Dxz4 and Xist have also been observed to occur on the mouse Xi in high-resolution Hi-C experiments [252]. Little is known about whether superloops are important for X-inactivation. Previous studies have suggested that one of the superloop anchors, Firre, regulates gene expression by forming interchromosomal contacts with several autosomal loci and that these interactions require an RNA transcribed from the Firre locus [256]. Another study found this Firre RNA to be required for proper Xist localization to the perinucleolar space and for HK27me3 to be deposited on the Xi, suggesting a possible role for Firre in X-inactivation [258]. Despite these observations regarding Firre RNA, it is unclear whether the superloop between Firre and Dxz4 or any other superloops are required for X-inactivation.

Given the intriguing position at the border of the megadomains and its ability to bind architectural proteins on the Xi, several groups have deleted Dxz4 [199, 252, 259] and all find that that Dxz4 is required for megadomains on the Xi. Despite the clear disruption of the megadomain structure following Dxz4 deletions, different effects on Xi chromatin and gene expression have been reported in different systems, with some reports of silencing of escapee genes [199] or changes in chromatin accessibility in mouse cell lines [199, 259] but others reporting minimal effects in human cells [252]. Thus, Dxz4 appears needed for megadomains, but the role of the megadomains in XCI has not been definitively examined.
**The role of higher-order structure in gene regulation is unclear**

More broadly, the role of higher-order structure has attracted a great deal of attention but remains unclear. Mammalian genomes are mostly non-coding and genes and their regulatory elements are often separated by tens to hundreds of kilobases of non-coding sequence [260]. Since enhancers are frequently very far from their promoters [261], and frequently do not contact the nearest gene [262, 263], it is unclear how the exquisite regulatory specificity needed for mammalian development can be conferred by enhancers. One possibility is that enhancers and promoters form specific looping interactions with each other [264, 265]. The advent of 3C technology enabled the detection of enhancer-promoter loops. One of the best model systems has been the β-globin LCR. The LCR is a large and powerful enhancer needed for upregulation of β-globin genes and in maturing erythrocytes, the LCR forms a specific loop with β-globin promoters [266, 267]. At the β-globin locus, the LCR-promoter loop forms only in maturing erythroid cells [268, 269] and depends upon the transcription factors that bind the LCR [268, 270] to activate the β-globin genes. These observations fit a model where specific factors drive the formation of enhancer-promoter loops, possibly as part of a cooperative complex that forms to fully activate the target gene. However, enhancer-promoter loops can also show a different regulatory logic. TNF-α responsive enhancers form contacts with their promoters prior to TNF-α induction, suggesting that some loops may serve to facilitate rapid gene activation whenever a particular transcription factor is expressed [261]. In addition to possible regulatory roles for loops, the fact that mammalian genomes fold into discrete domains could have many regulatory consequences, such as facilitating co-regulation of co-expressed genes, containing enhancer-promoter contacts to prevent inappropriate gene activation, and providing evolutionarily stable patterns of interaction for new enhancer-promoter partners to evolve [271]. The ability to map chromatin interactions at high resolution led to great excitement that 3C interactions may allow us to “read” the complex regulatory logic of mammalian genomes.
There are multiple lines of evidence to support these models of loops and domains as important components of gene regulatory logic. Gene expression within many TADs can be correlated [204, 272]. More mechanistically, if loops and domains are critical for proper gene regulation, then disrupting loops and domains will lead to improper gene expression and possible phenotypic consequences. This possibility has been tested for several loops and domains, most notably the HoxD cluster. The Hoxd genes are crucial developmental regulators. They are arranged into clusters, where the hox genes are expressed in an anterior to posterior gradient as one moves telomeric to centromeric along the HoxD cluster in the developing embryo, or in a proximal to distal gradient in the limb bud [273]. The HoxD gene cluster lies at the boundary of two TADs [274-276]. This boundary appears to move based on which Hoxd genes are active at a given position [274-276], suggesting that the TAD boundary separates the HoxD cluster into active and inactive domains dynamically, depending on which genes in the cluster must be active at a given time and place during development. Disruption of the HoxD TAD border leads to ectopic expression of the Hoxd genes and inappropriate contacts between the Hox genes and their distal enhancers [277, 278]. Beyond the Hox clusters, genomic rearrangements that cause human limb malformations have been shown to disrupt TAD borders and cause ectopic enhancer promoter interactions [279]. Loss of insulation due to hypermethylation of CTCF sites in glioma has been shown to dysregulate gene expression and promote tumor proliferation [280], suggesting critical roles for domains in establishing proper gene expression that are disrupted in disease. This body of work suggests that TAD boundaries and enhancer promoter loops are necessary for proper gene regulation. Even more striking evidence of the role of loops in gene regulation comes from the β-globin locus. β-globin is upregulated by a large distal enhancer (the LCR) upon binding of the GATA1 transcription factor, which induces LCR-β-globin loops [268]. However, engineering an artificial loop between the LCR and β-globin is also sufficient to induce β-globin upregulation [281], suggesting that
mere proximity between an enhancer and promoter can be sufficient to induce upregulation. A wide body of work characterizing the roles of particular loops and particular TADs has suggested that 3D organization plays a major role in establishing the context for proper gene regulation.

However, the function of 3D genome organization in genome regulation becomes less clear when one steps back and examines it globally rather than at particular loci. It is possible to eliminate loops and domains completely by targeted degradation of CTCF or cohesin. Elimination of domains and loops through targeted degradation of either protein leads to only minimal changes in gene expression [233, 244, 245, 282, 283], suggesting the main function of domains is not necessarily gene regulation. Large-scale assays of reporter gene activity show no correlations between reporter activity and TAD positions [284]. In addition, most looping interactions detected in Hi-C experiments are not enhancer-promoter loops [220]. In fact, enhancer-promoter loops may be more similar to compartments (non-specific associations between regions with similar chromatin status) [243] rather than stable loops present in a large fraction of cells in a population. Finally, a high-resolution analysis of relationships between TNF-α responsive enhancers and their target promoters showed that 84% of responsive enhancer-promoter pairs were simply the closest pairs of enhancers and promoters in the linear genome [261]. Results from these studies suggest that at a global level, chromatin conformation may neither be necessary for proper gene regulation nor provide any unique insight into gene regulatory logic.

Thus, there are two pictures emerging of the role of chromatin conformation in gene regulation. Studies of particular enhancer-promoter pairs and particular domain borders suggest that loops and domains are absolutely critical for gene regulation. Analysis of the global roles of loops, domains and the architectural proteins CTCF and cohesins suggest that loops and
domains are largely dispensable. Reconciling these two views is critical for our understanding of how 3D genome organization does (or does not) affect gene expression.

In Chapter 3, I wade into the debate on the functional role of higher-order chromatin structure, as I test the role of the unique architecture of the Xi in gene silencing during X-inactivation. I explore the role of the megadomains, superloops and tandem repeats in X-inactivation. I first examine the timecourse formation of the megadomains and the loss of the TADs during X-inactivation, then test whether deleting Dxz4, Firre or both tandem repeats prevents the initiation of X-inactivation in mESCs. Since deletion of Dxz4 disrupts both the megadomain and the superloops, and deletion of Firre disrupts the other side of the superloop, these experiments test whether the unique large-scale structures on the Xi are needed for X-inactivation to occur properly. I focus on the megadomains and superloops because they are genuinely unique to the Xi, and because they can be perturbed by deleting the tandem repeats.
Chapter 2: A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation
Addendum:

This chapter and all data contained within it were originally published as is in the following article:


*Co-first authors

*Science* automatically grants permission for authors to use their articles in their thesis (see the “AAAS Author License to Publish Policy”:

https://www.sciencemag.org/site/feature/contribinfo/prep/lic_info.pdf) therefore it is presented here with permission from *Science*.

Attributions:

Xist proteomics experiments were performed by Anand Minajigi in collaboration with Myriam Boukhali and Willi Haas.

UV-RIP and shRNA knockdown experiments were performed by Anand Minajigi.

RNA-seq experiments were performed by Chunyao Wei, and RNA-seq analysis by Chunyao Wei and Barry Kesner.

qRT-PCR experiments were performed by Anand Minajigi and David Colognori.
Microscopy experiments were performed by Hongjae Sunwoo.

Derek Lessing and Bernhard Payer contributed reagents.

The model figure (Figure 2.6) was produced by David Colognori

All other experiments (ChIP-seq, Hi-C) and bioinformatic analysis were performed by John E Froberg.

Anand Minajigi, John Froberg and Jeannie Lee wrote the manuscript.
ABSTRACT

The inactive X chromosome (Xi) serves as a model to understand gene silencing on a global scale. Here, we perform “identification of direct RNA interacting proteins” (iDRiP) to isolate a comprehensive protein interactome for Xist, an RNA required for Xi silencing. We discover multiple classes of interactors, including cohesins, condensins, topoisomerases, RNA helicases, chromatin remodelers and modifiers, which synergistically repress Xi transcription. Inhibiting two or three interactors destabilizes silencing. While Xist attracts some interactors, it repels architectural factors. Xist evicts cohesins from the Xi and directs an Xi-specific chromosome conformation. Upon deleting Xist, the Xi acquires the cohesin-binding and chromosomal architecture of the active X. Our study unveils many layers of Xi repression and demonstrates a central role for RNA in the topological organization of mammalian chromosomes.
INTRODUCTION

The mammalian X chromosome is unique in its ability to undergo whole-chromosome silencing. In the early female embryo, X-chromosome inactivation (XCI) enables mammals to achieve gene dosage equivalence between the XX female and the XY male [285-287]. XCI depends on Xist RNA, a 17-kb long noncoding RNA (lncRNA) expressed only from the inactive X-chromosome (Xi) [288] and that implements silencing by recruiting repressive complexes [29, 153, 170, 289]. While XCI initiates only once during development, the female mammal stably maintains the Xi through her lifetime. In mice, a germline deletion of Xist results in peri-implantation lethality due to a failure of Xi establishment [51], whereas a lineage-specific deletion of Xist causes a lethal blood cancer due to a failure of Xi maintenance [100]. Thus, both the de novo establishment and proper maintenance of the Xi are crucial for viability and homeostasis. There are therefore two critical phases of XCI: (i) A one-time initiation phase in peri-implantation embryonic development that is recapitulated by differentiating embryonic stem [ES] cells in culture, and (ii) a life-long maintenance phase that persists in all somatic lineages.

Once established, the Xi is extremely stable and difficult to disrupt genetically and pharmacologically [30, 183, 290]. In mice, X-reactivation is programmed to occur only twice — once in the blastocyst to erase the imprinted XCI pattern and a second time in the germline prior to meiosis [15, 291]. Although the Xi’s epigenetic stability is a homeostatic asset, an ability to unlock this epigenetic state is of great current interest. The X-chromosome is home to nearly 1000 genes, at least 50 of which have been implicated in X-linked diseases, such as Rett syndrome and Fragile X syndrome. The Xi is therefore a reservoir of functional genes that could be tapped to replace expression of a disease allele on the active X (Xa). A better understanding of Xi repression would inform both basic biological mechanisms and treatment of X-linked diseases.
It is believed that Xist RNA silences the Xi through conjugate protein partners. A major gap in current understanding is the lack of a comprehensive Xist interactome. In spite of multiple attempts to define the complete interactome, only four directly interacting partners have been identified over the past two decades, including PRC2, ATRX, YY1, and HNRPU: Polycomb repressive complex 2 (PRC2) is targeted by Xist RNA to the Xi; the ATRX RNA helicase is required for the specific association between Xist and PRC2 [60, 65]; YY1 tethers the Xist-PRC2 complex to the Xi nucleation center [62]; and the nuclear matrix factor, HNRPU/SAF-A, enables stable association of Xist with the chromosomal territory [63]. Many additional interacting partners are expected, given the large size of Xist RNA and its numerous conserved modular domains. Here, we develop a new RNA-based proteomic method and implement an unbiased screen for Xist’s comprehensive interactome. We identify a large number of high-confidence candidates, demonstrate that it is possible to destabilize Xi repression by inhibiting multiple interacting components, and then delve into a focused set of interactors with the cohesins.

RESULTS

iDRiP identifies multiple classes of Xist-interacting proteins

A systematic identification of interacting factors has been challenging because of Xist's large size, the expected complexity of the interactome, and the persistent problem of high background with existing biochemical approaches [292]. A high background could be particularly problematic for chemical crosslinkers that create extensive covalent networks of proteins, which could in turn mask specific and direct interactions. We therefore developed iDRiP (identification of direct RNA interacting proteins) using the zero-length crosslinker, UV light, to implement an unbiased screen of directly interacting proteins in female mouse
fibroblasts expressing physiological levels of Xist RNA (Fig. 2.1A). We performed in vivo UV crosslinking, prepared nuclei, and solubilized chromatin by DNase I digestion. Xist-specific complexes were captured using 9 complementary oligonucleotide probes spaced across the 17-kb RNA, with a 25-nt probe length designed to maximize RNA capture while reducing non-specific hybridization. The complexes were washed under denaturing conditions to eliminate factors not covalently linked by UV to Xist RNA. To minimize background due to DNA-bound proteins, a key step was inclusion of DNase I treatment before elution of complexes (See Supplemental Discussion). We observed significant enrichment of Xist RNA over highly abundant cytoplasmic and nuclear RNAs (U6, Jpx, 18S rRNA) in eluates of female fibroblasts (Fig. 2.1B). Enrichment was not observed in male eluates or with luciferase capture probes. Eluted proteins were subjected to quantitative mass spectrometry (MS), with spectral counting [293] and multiplexed quantitative proteomics [294] yielding similar enrichment sets (Supplementary Table 2.1).

From three independent replicates, iDRiP-MS revealed a large Xist protein interactome (Fig. 2.1C; Supplementary Table 2.1). Recovery of known Xist interactors, PRC2, ATRX, and HNRPU, provided a first validation of the iDRiP technique. Also recovered were macrohistone H2A (mH2A), RING1 (PRC1), and the condensin component, SmcHD1 — proteins known to be enriched on the Xi [63, 157, 166], but not previously shown to interact directly with Xist. More than 80 proteins were found to be ≥3-fold enriched over background; >200 proteins were ≥2-fold enriched (Supplementary Table 2.1). In many cases, multiple subunits of the epigenetic complex were identified, boosting our confidence in them as interactors. We verified select interactions by performing a test of reciprocity: By baiting with candidate proteins in an antibody capture, RIP-qPCR of UV-crosslinked cells reciprocally identified Xist RNA in the pulldowns (Fig. 2.1D). Called on the basis of high enrichment values, presence of multiple subunits within a candidate epigenetic complex, and tests of reciprocity, novel high-confidence interactors fell into
several functional categories: (i) Cohesin complex proteins, SMC1a, SMC3, RAD21, WAPL, PDS5a/b, as well as CTCF [72], which are collectively implicated in chromosome looping [228, 235, 295]; (ii) histone modifiers such as aurora kinase B (AURKB), a serine/threonine kinase that phosphorylates histone H3 [296]; RING1, the catalytic subunit of Polycomb repressive complex 1 (PRC1) for H2A-K119 ubiquitylation [157]; and SPEN and RBM15, which associate with HDACs; (iii) SWI/SNF chromatin remodeling factors; (iv) topoisomerases, TOP2a, TOP2b, and TOP1, that relieve torsional stress during transcription and DNA replication; (v) miscellaneous transcriptional regulators, MYEF2 and ELAV1; (vi) nucleoskeletal proteins that anchor chromosomes to the nuclear envelope, SUN2, Lamin-B receptor (LBR), and LAP2; (vii) nuclear matrix proteins, hnRPU/SAF-A, hnRPK, and MATRIN3; and (viii) the DNA methyltransferase, DNMT1, known as a maintenance methylase for CpG dinucleotides [297].

To study their function, we first performed RNA immunoFISH of female cells and observed several patterns of Xi coverage relative to the surrounding nucleoplasm (Fig. 2.1E). Like PRC2, RING1 (PRC1) has been shown to be enriched on the Xi [157] and is therefore not pursued further. TOP1 and TOP2a/b appeared neither enriched nor depleted on the Xi (100%, n>50 nuclei). AURKB showed two patterns of localization — peri-centric enrichment (20%, n>50) and a more diffuse localization pattern (80%, data now shown), consistent with its cell-cycle dependent chromosomal localization [296]. On the other hand, while SUN2 was depleted on the Xi (100%, n=52), it often appeared as pinpoints around the Xi in both day 7 differentiating female ES cells (establishment phase; 44%, n=307) and in fibroblasts (maintenance phase; 38.5%, n=52), consistent with SUN2’s function in tethering telomeres to the nuclear envelope. Finally, the cohesins and SWI/SNF remodelers unexpectedly showed a depletion relative to the surrounding nucleoplasm (100%, n=50-100). These patterns suggest that the Xist interactors operate in different XCI pathways.
Figure 2.1: iDRiP-MS reveals a large Xist interactome.

(A) iDRiP schematic.

(B) RT-qPCR demonstrated the specificity of Xist pulldown by iDRiP. Xist and control luciferase probes were used for pulldown from UV-crosslinked female and control male fibroblasts. Efficiency of Xist pulldown was calculated by comparing to a standard curve generated using 10-fold dilutions of input. Mean ± standard error (SE) of three independent experiments shown. P values determined by the Student t-test.

(C) Select high-confidence candidates from three biological replicates are grouped into functional classes. Additional candidates shown in Table S1.

(D) UV-RIP-qPCR validation of candidate interactors. Enrichment calculated as % input, as in (B). Mean ± SE of three independent experiments shown. P values determined by the Student t-test.

(E) RNA immunoFISH to examine localization of candidate interactors (green) in relation to Xist RNA (red). Immortalized MEF cells are tetraploid and harbor two Xi.
Figure 2.1 (Continued): iDRiP-MS reveals a large Xist interactome.
To ask if the factors intersect the PRC2 pathway, we stably knocked down (KD) top candidates using shRNAs (Supplementary Table 2.2) and performed RNA immunoFISH to examine trimethylation of histone H3-lysine 27 (H3K27me3; Fig. 2.2A,B). No major changes to Xist localization or H3K27me3 were evident in d7 ES cells (Supplementary Fig. S2.1). There were, however, long-term effects in fibroblasts: The decreased in H3K27me3 enrichment in shSMARCC1 and shSMARCA5 cells (Fig. 2.2A,B) indicated that SWI/SNF interaction with Xist is required for proper maintenance of PRC2 function on the Xi. Steady state Xist levels did not change by more than 2-fold (Fig. 2.2C) and were therefore unlikely to be the cause of the Polycomb defect. Knockdowns of other factors (cohesins, topoisomerases, SUN2, AURKB) had no obvious effects on Xist localization and H3K27me3. Thus, whereas the SWI/SNF factors intersect the PRC2 pathway, other interactors do not overtly impact PRC2.
Figure 2.2: Impact of depleting Xist interactors on H3K27me3.

(A) RNA immunoFISH of Xist (red) and H3K27me3 (green) after shRNA KD of interactors in fibroblasts (tetraploid; 2 Xist clouds). KD efficiencies (fraction remaining): SMC1a-0.48, SMC3-0.39, RAD21-0.15, AURKB-0.27, TOP2b-0.20, TOP2a-0.42, TOP1-0.34, CTCF-0.62, SMARCA4-0.52, SMARCA5-0.18, SMARCC1-0.25, SMARCC2-0.32, SMARCB1-0.52 and SUN2-0.72. Some factors are essential; therefore, high percentage KD may be inviable. All images presented at the same photographic exposure and contrast.

(B) Quantitation of RNA immunoFISH results. n, sample size. % aberrant Xist/H3K27me3 associations shown.

(C) RT-qPCR of Xist levels in KD fibroblasts, normalized to shControls. Mean± SD of two independent experiments shown.
Figure 2.2 (Continued): Impact of depleting Xist interactors on H3K27me3.
Xi-reactivation via targeted inhibition of synergistic interactors

Given the large number of interactors, we created a screen to analyze effects on Xi gene expression. We derived clonal fibroblast lines harboring a transgenic GFP reporter on the Xi (Supplementary Fig. S2.2) and shRNAs against Xist interactors. Knockdown of any one interactor did not reactivate GFP by more than 4-fold (Fig. 2.3A, shControl+none; Supplementary Fig. S2.3A). Suspecting synergistic repression, we targeted multiple pathways using a combination of drugs. To target DNMT1, we employed the small molecule, 5’-azacytidine (aza) [297] at a nontoxic concentration of 0.3 µM (≤IC₅₀) which minimally reactivated GFP (Fig. 2.3A, shControl + aza). To target TOP2a/b [298], we employed etoposide (eto) at 0.3 µM (≤IC₅₀), which also minimally reactivated GFP (Fig. 2.3A, shControl + eto). Combining 0.3 µM aza + eto led to an 80- to 90-fold reactivation — a level that was almost half of GFP levels on the Xa (Xa-GFP, Fig. 2.3A), suggesting strong synergy between DNMT1 and TOP2 inhibitors. Using aza + eto as priming agents, we designed triple-drug combinations inclusive of shRNAs for proteins that have no specific small molecule inhibitors. In various shRNA + aza + eto combinations, we achieved up to 230-fold GFP reactivation — levels that equaled or exceeded Xa-GFP levels (Fig. 2.3A). Greatest effects were observed for combinations using shSMARCC2 (227x), shSMARCA4 (180x), and shRAD21 (211x). shTOP1 and shCTCF were also effective (175x, 154x). Combinations involving remaining interactors yielded 63x to 94x reactivation.

We then performed allele-specific RNA-seq to investigate native Xi genes. In an F1 hybrid fibroblast line in which the Xi is of Mus musculus (mus) origin and the Xa of Mus castaneus (cas) origin, >600,000 X-linked sequence polymorphisms enabled allele-specific calls [36]. Two biological replicates of each of the most promising triple-drug treatments showed good correlation (Supplementary Fig. S2.4-S2.6). RNA-seq analysis showed reactivation of 75-100 Xi-specific genes in one replicate (Fig. 2.3B) and up to 200 in a second replicate
(Supplementary Fig. S2.3B), representing a large fraction of expressed X-linked genes, considering that only ~210 X-linked genes have an FPKM≥1.0 in this hybrid fibroblast line. Heatmap analysis demonstrated that, for individual Xi genes, reactivation levels ranged from 2x-80x for various combinatorial treatments (Fig. 2.3C). There was a net increase in expression level (ΔFPKM) from the Xi in the triple-drug treated samples relative to the shControl+aza+eto, whereas the Xa and autosomes showed no obvious net increase, thereby suggesting preferential effects on the Xi due to targeting synergistic components of the Xist interactome. Reactivation was not specific to any one Xi region (Fig. 2.3D). Most effective were shRAD21, shSMC3, shSMC1a, shSMARCA4, shTOP2a, and shAURKB drug combinations. Genic examination confirmed increased representation of mus-specific tags (red) relative to the shControl (Fig. 2.3E). Such allelic effects were not observed at imprinted loci and other autosomal genes (Supplementary Fig. S2.7), further suggesting Xi-specific allelic effects. The set of reactivated genes varied among drug treatments, though some genes (RbBP7, G6PDX, FMR1, etc.) appeared more prone to reactivation. Thus, the Xi is maintained by multiple synergistic pathways and Xi genes can be reactivated preferentially by targeting two or more synergistic Xist interactors.
**Figure 2.3: De-repression of Xi genes by targeting Xist interactors.**

(A) Relative GFP levels by RT-qPCR analysis in female fibroblasts stably knocked down for indicated interactors ± 0.3 µM 5’-azacytidine (aza) ± 0.3 µM etoposide (eto). Xa-GFP, control male fibroblasts with X-linked GFP. Mean ± SE of two independent experiments shown. *P* determined by Student *t*-test.

(B) Allele-specific RNA-seq analysis: Number of upregulated Xi genes for each indicated triple-drug treatment (aza+eto+shRNA). Blue, genes specifically reactivated on Xi (fold-change, FC>2); red, genes also unregulated on Xa (FC>1.3).

(C) RNA-seq heat map indicating that a large number of genes on the Xi were reactivated. X-linked genes reactivated in at least one of the triple-drug treatments (aza+eto+shRNA) were shown in the heat map. Color key, Log2 fold-change (FC). Cluster analysis performed based on similarity of KD profiles (across) and on the sensitivity and selectivity of various genes to reactivation (down).

(D) Chromosomal locations of Xi reactivated genes (colored ticks) for various aza+eto+shRNA combinations.

(E) Read coverage of 4 reactivated Xi genes after triple-drug treatment. Xi, mus reads (scale: 0-2). Comp, total reads (scale: 0-6). Red tags appear only in exons with SNPs.
Figure 2.3 (Continued): De-repression of Xi genes by targeting Xist interactors.
Xist interaction leads to cohesin repulsion

To investigate mechanism, we focused on one group of interactors — the cohesins — because they were among the highest-confidence hits and their knockdowns consistently destabilized Xi repression. To obtain Xa and Xi binding patterns, we performed allele-specific ChIP-seq for two cohesin subunits, SMC1a and RAD21, and for CTCF, which works together with cohesins [235, 299-301]. In wildtype cells, CTCF binding was enriched on Xa (cas), but also showed a number of Xi (mus)-specific sites (Fig. 2.4A) [37, 72]. Allelic ratios ranged from equal to nearly complete Xa or Xi skewing (Fig. 2.4A). For the cohesins, 1490 SMC1a and 871 RAD21 binding sites were mapped onto ChrX in total, of which allelic calls could be made on ~50% of sites (Fig. 2.4B,C). While the Xa and Xi each showed significant cohesin binding, Xa-specific greatly outnumbered Xi-specific sites. For SMC1a, 717 sites were called on Xa, of which 589 were Xa-specific; 203 sites were called on Xi, of which 20 were Xi-specific. For RAD21, 476 sites were called on Xa, of which 336 were Xa-specific; 162 sites were called on Xi, of which 18 were Xi-specific. Biological replicates showed similar trends (Supplementary Fig. S2.8A,B).

Cohesin’s Xa preference was unexpected in light of Xist’s physical interaction with cohesins – an interaction suggesting that Xist might recruit cohesins to the Xi. We therefore conditionally ablated Xist from the Xi (Xi^{ΔXist}) and repeated ChIP-seq analysis in the Xi^{ΔXist}/Xa^WT fibroblasts [302]. Surprisingly, Xi^{ΔXist} acquired 106 SMC1a and 48 RAD21 sites in cis, at positions that were previously Xa-specific (Fig. 2.4C,D). Biological replicates trended similarly (Supplementary Fig. S2.8 & S2.9). In nearly all cases, acquired sites represented a restoration of Xa sites, rather than binding to random positions. By contrast, sites that were previously Xi-specific remained intact (Fig. 2.4C,E, Supplementary Fig. S2.8B), suggesting that they do not require Xist for their maintenance. The changes in cohesin peak densities were Xi-specific and significant (Fig. 2.4F). Cohesin restoration occurred throughout Xi^{ΔXist}, resulting in domains of
biallelic binding (Fig. 2.4G, Supplementary Fig. S2.10-S1.12), and often favored regions that harbor genes that escape XCI (e.g., \textit{Bgn})[189, 303]. There were also shifts in CTCF binding, more noticeable at a locus-specific level than at a chromosomal level (Fig. 2.4A,G), suggesting that CTCF and cohesins do not necessarily track together on the Xi. The observed dynamics were ChrX-specific and were not observed on autosomes (Supplementary Fig. S2.13). To determine whether there were restoration hotspots, we plotted restored SMC1a and RAD21 sites (Fig. 2.4H; purple) on Xi^{\Delta Xist} and observed clustering within gene-rich regions. We conclude that Xist does not recruit cohesins to the Xi-specific sites. Instead, Xist actively repels cohesins \textit{in cis} to prevent establishment of the Xa pattern.
**Figure 2.4: Ablating Xist in cis restores cohesin binding on the Xi.**

(A) Allele-specific ChIP-seq results: Violin plots of allelic skew for CTCF, RAD21, SMC1a in wild-type (WT) and Xi^{ΔXist}/Xa^{WT} (ΔXist) fibroblasts. Fraction of mus reads [mus/(mus+cas)] is plotted for every peak with ≥10 allelic reads. *P* values determined by the Kolmogorov-Smirnov (KS) test.

(B) Differences between SMC1a or RAD21 peaks on the Xi^{WT} versus Xa^{WT}. Black diagonal, 1:1 ratio. Plotted are read counts for all SMC1a or RAD21 peaks. Allele-specific skewing is defined as ≥3-fold skew towards either Xa (cas, blue dots) or Xi (mus, red dots). Biallelic peaks, grey dots.

(C) Table of total, Xa-specific, and Xi-specific cohesin binding sites in WT versus ΔXist (Xi^{ΔXist}/Xa^{WT}) cells. Significant SMC1a and RAD21 allelic peaks with ≥5 reads were analyzed. Allele-specific skewing is defined as ≥3-fold skew towards Xa or Xi. Sites were considered “restored” if Xi^{ΔXist}’s read counts were ≥50% of Xa’s. X-total, all X-linked binding sites. Allelic peaks, sites with allelic information. Xa-total, all Xa sites. Xi-total, all sites. Xa-spec, Xa-specific. Xi-spec, Xi-specific. Xi-invariant, Xi-specific in both WT and Xi^{ΔXist}/Xa^{WT} cells. Note: There is a net gain of 96 sites on the Xi in the mutant, a number different from the number of restored sites (106). This difference is due to defining restored peaks separately from calling ChIP peaks (macs2). Allele-specific skewing is defined as ≥3-fold skew towards either Xa or Xi.

(D) Partial restoration of SMC1a or RAD21 peaks on the Xi^{ΔXist} to an Xa pattern. Plotted are peaks with read counts with ≥3-fold skew to Xa^{WT} (“Xa-specific”). x-axis: normalized Xa^{WT} read counts. y-axis: normalized Xi^{ΔXist} read counts. Black diagonal, 1:1 Xi^{ΔXist}/Xa^{WT} ratio; red diagonal, 1:2 ratio.

(E) Xi-specific SMC1a or RAD21 peaks remained on Xi^{ΔXist}. Plotted are read counts for SMC1a or RAD21 peaks with ≥3-fold skew to Xi^{WT} (“Xi-specific”).
Figure 2.4 (Continued): Ablating *Xist in cis* restores cohesin binding on the Xi.

(F) Comparison of fold-changes for CTCF, RAD21, and SMC1 binding in Xi^\Delta Xist^ cells relative to WT cells. Shown are fold-changes for Xi versus Xa. The Xi showed significant gains in RAD21 and SMC1a binding, but not in CTCF binding. Method: Xi^WT^ and Xi^\Delta Xist^ ChIP samples were normalized by scaling to equal read counts. Fold-changes for Xi were computed by dividing the normalized mus read count in Xi^\Delta Xist^ by the mus read count Xi^WT^; fold-changes for Xa were computed by dividing the normalized cas read count in Xi^\Delta Xist^ by the cas read count Xi^WT^.

To eliminate noise, peaks with <10 allelic reads were eliminated from analysis. *P* values determined by a paired Wilcoxon signed rank test.

(G) The representative examples of cohesion restoration on Xi^\Delta Xist^.

(H) Allelic-specific cohesin binding profiles of Xa, Xi^WT^, and Xi^\Delta Xist^.

Shown below restored sites are regions of Xi-reactivation following shSMC1a and shRAD21 triple-drug treatments, as defined in Figure 2.3.
Figure 2.4 (Continued): Ablating *Xist in cis* restores cohesin binding on the Xi.

A: Ablating *Xist in cis* restores cohesin binding on the Xi. (Continued)
**Xist RNA directs an Xi-specific chromosome conformation**

Cohesins and CTCF have been shown to facilitate formation of large chromosomal domains called TADs (topologically associated domains) [228, 229, 232, 235, 299, 300, 304]. The function of TADs is currently not understood, as TADs are largely invariant across development. However, X-linked domains are exceptions to this rule and are therefore compelling models to study function of topological structures [203, 204, 220, 221]. By carrying out allele-specific Hi-C, we asked whether cohesin restoration altered the chromosomal architecture of Xi\(^{\Delta Xist}\). First, we observed that, in wildtype cells, our TADs called on autosomal contact maps at 40-kb resolution resembled published composite (non-allelic) maps [228](Fig. 2.5A, bottom). Our ChrX contact maps were also consistent, with TADs being less distinct due to a summation of Xa and Xi reads in the composite profiles (Fig. 2.5A, top). Using the 44% of reads with allelic information, our allelic analysis yielded high-quality contact maps at 100-kb resolution by combining replicates (Fig. 2.5B, Supplementary Fig. S2.14A) or at 200-kb resolution with a single replicate. In wildtype cells, we deduced 112 TADs at 40-kb resolution on ChrX using the method of Dixon et al. [228]. We attempted TAD calling for the Xi on the 100 kb contact map, but were unable to obtain obvious TADs, suggesting the 112 TADs are present only on the Xa. The Xi instead appeared to be partitioned into two megadomains at the DXZ4 region (Supplementary Fig. S2.14A) [220]. Thus, while the Xa is topologically organized into structured domains, the Xi is devoid of TADs across its full length.

When Xist was ablated, however, TADs were restored in cis and the Xi reverted to an Xa-like conformation (Fig. 2.5B, Supplementary Fig. S2.14B). In mutant cells, ~30 TADs were gained on Xi\(^{\Delta Xist}\) in each biological replicate. Where TADs were restored, Xi\(^{\Delta Xist}\) patterns (red) became nearly identical to those of the Xa (blue), with similar interaction frequencies. These Xi\(^{\Delta Xist}\) regions now bore little resemblance to the Xi of wildtype cells (Xi\(^{WT}\), orange). Overall, the difference in the average interaction scores between Xi\(^{WT}\) and Xi\(^{\Delta Xist}\) was highly significant (Fig.
Intersecting TADs with SMC1a sites on Xi\(^{\Delta Xist}\) revealed that 61 restored cohesin sites overlapped restored TADs (61 did not overlap). In general, restored cohesin sites occurred both within TADs and at TAD borders. TADs overlapping restored peaks had larger increases in interaction scores relative to all other TADs (Fig. 2.5D, Supplementary Fig. S2.15B) and we observed an excellent correlation between the restored cohesin sites and the restored TADs (Fig. 2.5E, Supplementary Fig. S2.15C), consistent with a role of cohesins in re-establishing TADs following Xist deletion. Taken together, these data uncover a role for RNA in establishing topological domains of mammalian chromosomes and demonstrate that Xist must actively and continually repulse cohesins from the Xi, even during the maintenance phase, to prevent formation of an Xa chromosomal architecture.
Figure 2.5: Ablating Xist results in Xi reversion to an Xa-like chromosome conformation.

(A) Chr13 and ChrX contact maps showing triangular domains representative of TADs. Purple shades correspond to varying interaction frequencies (dark, greater interactions). TADs called from our composite (non-allelic) HiC data at 40-kb resolution (blue bars) are highly similar to those (gray bars) called previously [228].

(B) Allele-specific HiC-seq analysis: Contact maps for three different ChrX regions at 100-kb resolution comparing Xi^ΔXist (red) to Xi^WT (orange), and Xi^ΔXist (red) versus Xa (blue) of the mutant cell line. Our Xa TAD calls are shown with RefSeq genes.

(C) Fraction of interaction frequency per TAD on the Xi (mus) chromosome. The positions of our Xa TAD borders were rounded to the nearest 100 kb and submatrices were generated from all pixels between the two endpoints of the TAD border for each TAD. We calculated the average interaction score for each TAD by summing the interaction scores for all pixels in the submatrix defined by a TAD and dividing by the total number of pixels in the TAD. We then averaged the normalized interaction scores across all bins in a TAD in the Xi (mus) and Xa (cas) contact maps and computed the fraction of averaged interaction scores from mus chromosomes. ChrX and a representative autosome, Chr5, are shown for the WT cell line and the Xist^ΔXist/+ cell line. P value determined by paired Wilcoxon signed rank test.

(D) Violin plots showing that TADs overlapping restored peaks have larger increases in interaction scores relative to all other TADs. We calculated the fold-change in average interaction scores on the Xi for all X-linked TADs and intersected the TADs with SMC1a sites (Xi^ΔXist/ Xi^WT). 32 TADs occurred at restored cohesin sites; 80 TADs did not overlap restored cohesin sites. Violin plot shows distributions of fold-change average interaction scores between Xi^WT and Xi^ΔXist. P-value determined by Wilcoxon ranked sum test.
Figure 2.5 (Continued): Ablating Xist results in Xi reversion to an Xa-like chromosome conformation.

(E) Restored TADs overlap regions with restored cohesins on across XiΔXist. Several datasets were used to call restored TADs, each producing similar results. Restored TADs were called in two separate replicates (Rep1, Rep2) where the average interaction score was significantly higher on XiΔXist than on XiWT. We also called restored TADs based on merged Rep1+Rep2 datasets. Finally, a consensus between Rep1 and Rep2 was derived. Method: We calculated the fold-change in mus or cas for all TADs on ChrX and on a control, Chr5; then defined a threshold for significant changes based on either the autosomes or the Xa. We treated Chr5 as a null distribution (few changes expected on autosomes) and found the fraction of TADs that crossed the threshold for several thresholds. These fractions corresponded to a false discovery rate (FDR) for each given threshold. An FDR of 0.05 was used.
Figure 2.5 (Continued): Ablating Xist results in Xi reversion to an Xa-like chromosome conformation.

A

B

X\textsuperscript{\text{XI}^{Rt}}

X\textsuperscript{\text{XI}^{Wt}}

X\textsuperscript{\text{XI}^{SMC1a}}

X\textsuperscript{\text{Xa}}

C

D

E

X-chromosome (186 mB)

Restored TADs

Consensus

Rep1

Rep2

Restored SMC1a

RefSeq genes

Nhp

Six

Mef2

Zic3

Fmr1

Dmd

Arx

Ogl

Hmx

Grn31f

Nik

Otf

Phe6

Thf8

$P<0.01$

$P<0.001$

All other TADs (n=80)

TADs at restored cohesion peaks (n=32)
DISCUSSION

Using iDRiP, we have identified a comprehensive Xist interactome and revealed multiple synergistic pathways to Xi repression (Fig. 2.6). With Xist physically contacting 80-250 proteins at any given time, the Xist ribonucleoprotein particle may be as large as the ribosome. Our study supports a model in which Xist RNA simultaneously acts as (i) scaffold for the recruitment of repressive complexes (such as PRC1, PRC2, ATRX, mH2A, and SmcHD1) to establish and maintain the inactive state; and as (ii) a repulsion mechanism to extrude architectural factors such as cohesins in order to avoid acquisition of a transcription-favorable chromatin conformation. Without Xist, cohesins return to their default Xa binding state. Repulsion could be based on eviction, with Xist releasing cohesins as it extrudes them, or on sequestration, with Xist sheltering cohesins to prevent Xi binding. Our study shows that the Xi harbors three types of cohesin sites: (i) Xi-specific sites that do not depend on Xist; (ii) biallelic sites that are also Xist-independent; and (iii) Xa-specific sites, many of which cannot be established on the Xi because of active repulsion by Xist. The type i and type iii sites likely explain the paradoxical observations that, on the one hand, depleting cohesins leads to Xi reactivation but, on the other, loss of Xist-mediated cohesin recruitment leads to an Xa-like chromosome conformation that is permissive for transcription. In essence, modulating the Type i and Type iii sites both have the effect of destabilizing the Xi, rendering the Xi more accessible to transcription. Disrupting Type i sites by cohesin knockdown would change the repressive Xi structure, while ablating Xist would restore the Type iii sites that promote an Xa-like conformation. Our study has focused on cohesins, but RNA-mediated repulsion may be an outcome for other Xist interactors and may be as prevalent an epigenetic mechanism as RNA-mediated recruitment [73].

The robustness of Xi silencing is demonstrated by the observation that we destabilized the Xi only after pharmacologically targeting two or three distinct pathways. The fact that the triple-drug treatments varied with respect to reactivated loci and depth of de-repression creates
the possibility of treating X-linked disease in a locus-specific manner by administering unique drug combinations. Given the existence of many other disease-associated lncRNAs, the iDRiP technique could be applied systematically towards identifying new drug targets for other diseases and generally for elucidating mechanisms of epigenetic regulation by lncRNA.

Figure 2.6: The Xi is suppressed by multiple synergistic mechanisms.

Xist RNA (red) suppresses the Xi by either recruiting repressive factors (e.g., PRC1, PRC2) or expelling architectural factors (e.g., cohesins).
MATERIALS AND METHODS

Identification of Direct RNA interacting Proteins (iDRiP)

Mouse Embryonic Fibroblasts (MEFs) were irradiated with UV light at 200 mJ energy (Stratagene 2400) after rinsing with PBS. The pellets were resuspended in CSKT-0.5% (10 mM PIPES, pH 6.8, 100 mM NaCl, 3 mM MgCl₂, 0.3 M sucrose, 0.5% Triton X-100, 1 mM PMSF) for 10 min at 4 ºC followed by a spin. The pellets were again resuspended in Nuclear Isolation Buffer (10 mM Tris pH 7.5, 10 mM KCl, 0.5% Nonidet-P 40, 1x protease inhibitors, 1 mM PMSF), and rotated at 4ºC for 10 min. The pellets were collected after a spin, weighed, flash frozen in liquid nitrogen, and stored at -80 ºC until use.

Approximately, equal amounts of female and male UV cross linked pellets were thawed and resuspended for treatment with Turbo DNase I in the DNase I digestion buffer (50 mM Tris pH 7.5, 0.5% Nonidet-P 40, 0.1% sodium lauroyl sarcosine, 1x protease inhibitors, SuperaseIn). The tubes were rotated at 37 ºC for 45 min. The nuclear lysates were further solubilized by adding 1% sodium lauroyl sarcosine, 0.3 M lithium chloride, 25 mM EDTA and 25 mM EGTA to final concentrations and continued incubation at 37 ºC for 15 min. The lysates were mixed with biotinylated DNA probes (Table S3) prebound to the streptavidin magnetic beads (MyOne streptavidin C1 Dyna beads, Invitrogen) and incubated at 55 ºC for 1 hr before overnight incubation at 37 ºC in the hybridization chamber. The beads were washed three times in Wash Buffer (10 mM Tris, pH 7.5, 0.3 M LiCl, 1% LDS, 0.5% Nonidet-P 40, 1x protease inhibitor) at room temperature followed by treatment with Turbo DNase I in DNase I digestion buffer with the addition of 0.3 M LiCl, protease inhibitors, and SuperaseIn at 37 ºC for 20 min. Then, beads were washed two more times in the Wash Buffer. For MS analysis, elution was done in Elution Buffer (10 mM Tris, pH 7.5, 1 mM EDTA) at 70 ºC for 4 min followed by
brief sonication in Covaris. For the quantification of pulldown efficiency, MEFs, without crosslinking, were used and elution was done at 95 °C. The elute was used for RNA isolation and RT-qPCR. When crosslinked MEFs were used, elute was subjected for proteinase-K treatment (50 mM Tris pH 7.5, 100 mM NaCl, 0.5% SDS, 10 µg proteinase K) for 1 hr at 55 °C. RNA was isolated by Trizol and quantified with SYBR green qPCR. Input samples were used to make standard curve by 10 fold dilutions, to which the RNA pulldown efficiencies were compared and calculated. The efficiency of Xist pulldown was relatively lower after UV crosslinking, similar to [305, 306].

Quantitative proteomics

Proteins co-enriched with Xist from female or male cells were quantitatively analyzed either using a label-free approach based on spectral-counting [293] or by multiplexed quantitative proteomics using tandem-mass tag (TMT) reagents [307, 308] on an Orbitrap Fusion mass spectrometer (Thermo Scientific). Disulfide bonds were reduced with dithiothreitol (DTT) and free thiols alkylated with iodoacetamide as described previously [294]. Proteins were then precipitated with trichloroacetic acid, resuspended in 50 mM HEPES (pH 8.5) and 1 M urea and digested first with endoproteinase Lys-C (Wako) for 17 hours at room temperature and then with sequencing-grade trypsin (Promega) for 6 hours at 37 °C. Peptides were desalted over Sep-Pak C18 solid-phase extraction (SPE) cartridges (Waters), the peptide concentration was determined using a BCA assay (Thermo Scientific). For the label-free analysis peptides were then dried and re-suspended in 5 % formic acid (FA) and 5 % acetonitrile (ACN) and 5 µg of peptides were analyzed by mass spectrometry as described below. For the multiplexed quantitative analysis a maximum of 50 µg of peptides were labeled with one out of the available TMT-10plex reagents (Thermo Scientific) [307]. To achieve this, peptides were
dried and resuspended in 50 µl of 200 mM HEPES (pH 8.5) and 30 % (ACN) and 10 µg of the TMT in reagent in 5 µl of anhydrous ACN was added to the solution, which was incubated at room temperature (RT) for one hour. The reaction was then quenched by adding 6 µl of 5 % (w/v) hydroxylamine in 200 mM HEPES (pH 8.5) and incubation for 15 min at RT. The labeled peptide mixture was then subjected to a fractionation using basic pH reversed phase liquid chromatography (bRPLC) on an Agilent 1260 Infinity HPLC system equipped with an Agilent Extend-C18 column (4.6x250 mm; particle size, 5 µm) basically as described previously [309]. Peptides were fractionated using a gradient from 22-35 % ACN in 10 mM ammonium bicarbonate over 58 min at a flowrate of 0.5 ml/min. Fractions of 0.3 ml were collected into a 96-well plate to then be pooled into a total twelve fractions (A1-A12, B1-B12, etc.) that were dried and re-suspended in 8 µl of 5 % FA and 5 % ACN, 3 of which were analyzed by microcapillary liquid chromatography tandem mass spectrometry on an Orbitrap Fusion mass spectrometer and using a recently introduced multistage (MS3) method to provide highly accurate quantification [310].

The mass spectrometer was equipped with an EASY-nLC 1000 integrated autosampler and HPLC pump system. Peptides were separated over a 100 µm inner diameter microcapillary column in-house packed with first 0.5 cm of Magic C4 resin (5 µm, 100 Å, Michrom Bioresources), then with 0.5 cm of Maccel C18 resin (3 µm, 200 Å, Nest Group) and 29 cm of GP-C18 resin (1.8 µm, 120 Å, Sepax Technologies). Peptides were eluted applying a gradient of 8-27 % ACN in 0.125 % formic acid over 60 min (label-free) and 165 min (TMT) at a flow rate of 300 nl/min. For label-free analyses we applied a tandem-MS method where a full-MS spectrum (MS1; m/z 375-1500; resolution 6x10^4; AGC target, 5x10^5; maximum injection time, 100 ms) was acquired using the Orbitrap after which the most abundant peptide ions where selected for linear...
ion trap CID-MS2 in an automated fashion. MS2 scans were done in the linear ion trap using the following settings: quadrupole isolation at an isolation width of 0.5 Th; fragmentation method, CID; AGC target, 1x10^4; maximum injection time, 35 ms; normalized collision energy, 30 %). The number of acquired MS2 spectra was defined by setting the maximum time of one experimental cycle of MS1 and MS2 spectra to 3 sec (Top Speed). To identify and quantify the TMT-labeled peptides we applied a synchronous precursor selection MS3 method [294, 310, 311] in a data dependent mode. The scan sequence was started with the acquisition of a full MS or MS1 one spectrum acquired in the Orbitrap (m/z range, 500-1200; other parameters were set as described above), and the most intense peptide ions from detected in the full MS spectrum were then subjected to MS2 and MS3 analysis, while the acquisition time was optimized in an automated fashion (Top Speed, 5 sec). MS2 scans were performed as described above. Using synchronous precursor selection, the 10 most abundant fragment ions were selected for the MS3 experiment following each MS2 scan. The fragment ions were further fragmented using the HCD fragmentation (normalized collision energy, 50 %) and the MS3 spectrum was acquired in the Orbitrap (resolution, 60,000; AGC target, 5x10^4; maximum injection time, 250 ms).

Data analysis was performed on an on an in-house generated SEQUEST-based [312] software platform. RAW files were converted into the mzXML format using a modified version of ReAdW.exe. MS2 spectra were searched against a protein sequence database containing all protein sequences in the mouse UniProt database (downloaded 02/04/2014) as well as that of known contaminants such as porcine trypsin. This target component of the database was followed by a decoy component containing the same protein sequences but in flipped (or reversed) order [313]. MS2 spectra were matched against peptide sequences with both termini consistent with trypsin specificity
and allowing two missed trypsin cleavages. The precursor ion m/z tolerance was set to
50 ppm, TMT tags on the N-terminus and on lysine residues (229.162932 Da, only for
TMT analyses) as well as carbamidomethylation (57.021464 Da) on cysteine residues
were set as static modification, and oxidation (15.994915 Da) of methionines as variable
modification. Using the target-decoy database search strategy [313] a spectra
assignment false discovery rate of less than 1 % was achieved through using linear
discriminant analysis with a single discriminant score calculated from the following
SEQUEST search score and peptide sequence properties: mass deviation, XCorr, dCn,
number of missed trypsin cleavages, and peptide length [314]. The probability of a
peptide assignment to be correct was calculated using a posterior error histogram and
the probabilities for all peptides assigned to a protein were combined to filter the data set
for a protein FDR of less than 1 %. Peptides with sequences that were contained in
more than one protein sequence from the UniProt database were assigned to the protein
with most matching peptides [314].

For a quantitative estimation of protein concentration using spectral-counts we
simply counted the number of MS2 spectra assigned to a given protein (Table S1). TMT
reporter ion intensities were extracted as that of the most intense ion within a 0.03 Th
window around the predicted reporter ion intensities in the collected MS3 spectra. Only
MS3 with an average signal-to-noise value of larger than 28 per reporter ion as well as
with an isolation specificity [294] of larger than 0.75 were considered for quantification.
Reporter ions from all peptides assigned to a protein were summed to define the protein
intensity. A two-step normalization of the protein TMT-intensities was performed by first
normalizing the protein intensities over all acquired TMT channels for each protein
based to the median average protein intensity calculated for all proteins. To correct for
slight mixing errors of the peptide mixture from each sample a median of the normalized
intensities was calculated from all protein intensities in each TMT channel and the protein intensities were normalized to the median value of these median intensities.

**UV RIP**

The protocol followed is similar to the one described in [62]. Briefly, MEFs were crosslinked with UV light at 200 mJ and collected by scraping in PBS. Cell pellets were resuspended in CSKT-0.5% for 10 min at 4 °C followed by a spin. The nuclei were resuspended in the UV RIP buffer (PBS buffer containing 300 mM NaCl (total), 0.5% Nonidet-P 40, 0.5% sodium deoxycholate, and 1x protease inhibitors) with Turbo DNase I 30 U/IP for 30 min at 37 °C. Supernatants were collected after a spin and incubated with 5 µg specific antibodies prebound to 40 µl protein-G magnetic beads (Invitrogen) at 4 °C overnight. Beads were washed three times with cold UV RIP buffer. The beads were resuspended in 200 µl Turbo DNase I buffer with 20 U Turbo DNase, Superasein, 1x protease inhibitors) for 30 min at 37 °C. The beads were resuspended and washed three more times in the UV RIP washing buffer containing 10 mM EDTA. The final 3 washes were given after three fold dilution of UV RIP washing buffer. The beads were resuspended in 200 µl proteinase-K buffer with 10 µg proteinase-K and incubated at 55 °C for 1 hr. RNA was isolated by Trizol and pulldown efficiencies were calculated by SYBR qPCR using input for the standard curve.

*Generation of Xi-TgGFP clonal fibroblasts*

Xi-TgGFP (68-5-11) tail-tip fibroblasts (TTF) were initially derived from a single female pup, a daughter of a cross between a *M. castaneus* male and a *M. musculus* female, homozygous for an X-linked GFP transgene driven by a strong, ubiquitous promoter
The fibroblasts were immortalized by SV40 transformation, and clonal lines were derived from individual GFP-negative cells selected by fluorescence-activated cell sorting. In our experience, occasional clones with undetectable GFP expression nevertheless have the transgene located on the active X chromosome. Thus, we confirmed the GFP transgene location on the inactive X for the particular clone used here, 68-5-11 (see Supplementary Fig. S2.2).

Generation of stable KD of Xi-TgGFP TTF and 16.7 ES cells

The protocol is as described in:

http://www.broadinstitute.org/rna/public/resources/protocols

A cocktail of 3 shRNA viruses were used for infections (Supplementary Table 2.2) followed with puromycin selection. In all the experiments, non-clonal knockdown cells were used.

Assay for the reactivation of Xi-TgGFP

Approximately, 125,000-150,000 Xi-TgGFP (68-5-11) cells were plated along with control (shNegative control, i.e., shNC) cells treated with DMSO or stable KD cells treated with 0.3 μM azacytidine and 0.3 μM Etoposide for 3 days in 6 well plates. RNA was isolated by Trizol twice, with an intermittent Turbo DNase treatment after the first isolation for 30 min at 37 °C. One μg RNA was used for each of the RT+ and RT-reactions (Superscript III, Invitrogen) followed by the SYBR green qPCR using the primers listed in supplementary Table S3, with annealing temperature of 60 °C for 45
cycles. The relative efficiency of Xi-TgGFP reactivations was calculated by comparing to U1 snRNA as the internal control.

**ImmunoFISH**

Cells were grown on coverslips, rinsed in PBS, pre-extracted in 0.5% CSKT on ice, washed once in CSK, followed by fixation with 4% paraformaldehyde in PBS at room temperature. After blocking in 1% BSA in PBS for 20 min supplemented with 10 mM VRC (New England Biolabs) and RNase inhibitor (Roche), incubation was carried out with primary antibodies (Table S3) at room temperature for 1 hr. Cells were washed three times in PBST-0.02% Tween-20. After incubating with secondary antibody at room temperature for 30 min, cells were washed three times by PBS/0.02% Tween-20. Cells were fixed again in 4% paraformaldehyde and dehydrated in ethanol series. RNA FISH was performed using a pool of Cy3B or Alexa 568 labeled Xist oligonucleotides for 4-6 hours at 42°C in a humid chamber. Cells were washed three times in 2X SSC and nuclei were counter-stained by Hoechst 33342. Cells were observed under Nikon 90i microscope equipped with 60X/1.4 N.A. objective lens, Orca ER CCD camera (Hamamatsu), and Volocity software (Perkin Elmer). Xist RNA FISH probes, a set of total 37 oligonucleotides with 5’ amine modification (IDT), were labeled with NHS-Cy3B (GE Healthcare) overnight at room temperature followed by ethanol precipitation. In the case of confirmation of Xi-TgGFP cells, probes were made by nick-translation of a GFP PCR product with Cy3-dUTP and of a plasmid containing the first exon of the mouse Xist gene, with FITC-dUTP.
Allelic ChIP-seq

Allele-specific ChIP-seq was performed according to the method of Kung et al [72], in two biological replicates. To increase available read depth, we pooled together two technical replicates for $X_i^{\Delta xist}/X_a^{WT}$ Rad21 replicate 1 sequenced on a 2x50bp HiSeq2500 rapid run and we also pooled two technical replicates of wild-type Rad21 replicate 1, one sequenced on a HiSeq 2x50 bp run and one on a MiSeq 2x50 bp run. All other libraries were sequenced on using 2x50 bp HiSeq2500 rapid runs. To visualize ChIP binding signal, we generated fpm-normalized bigWig files from the raw ChIP read counts for all reads (comp), mus-specific (mus) and cas-specific reads separately. For Smc1a, CTCF and Rad21, peaks were called using macs2 with default settings. To generate consensus peak sets for all three epitopes, peaks for the two wild-type and $X_i^{\Delta xist}/X_a^{WT}$ replicates were pooled and peaks present in at least two experiments were used as the common peak set. To make comparisons between allelic read counts between different experiments, we defined a scaling factor as the ratio of the total read numbers for the two experiments and multiplied the allelic reads for each peak in the larger sample by the scaling factor. We plotted the number of reads on Xi vs Xa in wild-type for all peaks on the X-chromosome to determine if there is a general bias towards binding to the Xa or the Xi. To evaluate allelic skew on an autosome, we generated plots of mus read counts vs cas read counts for all peaks on chromosome 5 from 1-140,000,000. We used this particular region of chromosome 5 because $X_i^{\Delta xist}/X_a^{WT}$ is not fully hybrid, and this is a large region of an autosome that is fully hybrid based on even numbers of read counts from input and from our Hi-Cs over this region in $X_i^{\Delta xist}/X_a^{WT}$ (data not shown). To identify peaks that are highly Xa-skewed in wild-type but bind substantially to the Xi in $X_i^{\Delta xist}/X_a^{WT}$ (restored peaks), for Xa-skewed peaks in wild-type, we plotted normalized read counts on Xi in $X_i^{\Delta xist}/X_a^{WT}$ versus read counts on Xa in wild-
type. We defined restored peaks as peaks that are 1.) more than 3X Xa-skewed in wild-type 2.) have at least 5 allelic reads in wild-type 3.) exhibit normalized read counts on Xi in $\frac{Xi^{\Delta xist}}{Xa^{WT}}$ that are at least half the level of Xa in wild-type. This threshold ensures that all restored peaks have at least a 2X increase in binding to the Xi in $\frac{Xi^{\Delta xist}}{Xa^{WT}}$ relative to wild-type. We identified restored peaks using these criteria in both replicates of Smc1a and Rad21 ChIP separately, and to merge these calls into a consensus set for each epitope, we took all peaks that met criteria for restoration in at least one replicate and had at least 50% wild-type Xa read counts on Xi in $\frac{Xi^{\Delta xist}}{Xa^{WT}}$ in both replicates.

**Allele specific RNA-seq**

Xi-TgGFP TTFs (68-5-11) with the stable knock down of candidates were treated with 5'-azacytidine and etoposide at 0.3 µM each for 3 days. Strand-specific RNA-seq, the library preparation, deep sequencing, and data analysis was followed as described in [72]. Two biological replicates of each drug treatment were produced. All libraries were sequenced with Illumina Hiseq 2000 or 2500 using 50 cycles to obtain paired end reads. To determine the allelic origin of each sequencing read from the hybrid cells, reads were first depleted of adaptors dimers and PCR duplicates, followed by the alignment to custom mus/129 and cas genomes to separate mus and cas reads. After removal of PCR duplicates, ~90% of reads were mappable. Discordant pairs and multi-mapped reads were discarded. Reads were then mapped back to reference mm9 genome using Tophat v2.0.10 (-g 1 --no-coverage-search --read-edit-dist 3 --read-mismatches 3 --read-gap-length 3 --b2-very-sensitive --mate-inner-dist 50 --mate-std-dev 50 --library-type fr-firststrand), as previously described [36, 72, 316]. Following alignment, gene expression levels within each library were quantified using Homer v4.7 (rna mm9 -count genes -
strand + -noadj -condenseGenes) [316] and the normalized differential expression analyses across samples were performed by using EdgeR [317].

**Hi-C library preparation and analysis**

Hi-C libraries were generated according to the protocol in Lieberman-Aiden et al., 2009 [318]. Two biological replicate libraries were prepared for wild-type and $Xi^{\Delta xist}/Xa^{WT}$ fibroblasts each. We obtained 150-220 million 2x50bp paired-end reads per library. The individual ends of the read-pairs were aligned to the mus and cas reference genomes separately using novoalign with default parameters for single-end alignments, and the quality score of the alignment was used to determine whether each end could be assigned to either the mus or the cas haplotype [36]. The single-end alignments were merged into a Hi-C summary file using custom scripts. Reads were filtered for self-ligation events and short fragments (less than 1.5X the estimated insert length) likely to be random shears using Homer [316, 319]. Hi-C contact maps were generated using Homer. “Comp” maps were made from all reads. “Xi” and “Xa” reads were from reads where at least one read-end could be assigned to either the mus or cas haplotype, respectively. A small fraction of reads (~5% of all allelic reads) aligned such that one end aligned to mus, the other to cas. These “discordant” reads were excluded from further analysis, as they are likely to be noise arising due to random ligation events and/or improper SNP annotation [220, 320]. All contact maps were normalized using the matrix balancing algorithm of Knight and Ruiz [321], similar to iterative correction [220, 322], using the MATLAB script provided at the end of their paper. We were able to generate robust contact maps using the comp reads in one replicate at 40kb resolution, but due to the fact that only ~44% of reads align allele-specifically, we were only able to generate contact maps for the cas and mus haplotypes at 200kb. To increase our resolution, we
pooled together both biological replicates and analyzed the comp contact map at 40kb resolution and the mus and cas contact maps at 100kb. We called TADs at 40kb on chrX, chr5 and chr13 using the method of Dixon et al. [228]. Specifically, we processed the normalized comp 40kb contact maps separately into a vector of directionality indices using DI_from_matrix.pl with a bin size of 40000 and a window size of 200000. We used this vector of directionality indices as input for the HMM calls.m script and following HMM_generation, we processed the HMM and generated TAD calls by passing the HMM output to file_ends_cleaner.pl, converter_7col.pl, hmm_probability_correcter.pl, hmm-state_caller.pl and finally hmm-state_domains.pl. We used parameters of min=2, prob=0.99, binsize=40000 as input to the HMM probability correction script.

To create a general metric describing interaction frequencies within TADs at resolution available in the allele-specific interaction maps, for each TAD, on chrX and chr5 we averaged the normalized interaction scores for all bins within each TAD, excluding the main diagonal. To make comparisons between interaction frequency over TADs between the cas (Xa) and mus (Xi) haplotypes at the resolution available with our current sequencing depth, we defend the “fraction mus” as the average interaction score for a TAD in the mus contact map divided by the sum of the average interaction scores in the mus and cas contact maps.

To discover TADs that show significantly increased interaction frequency in XiΔxist/XaWT, we generated a null distribution of changes in average normalized interaction scores for all TADs on chromosome 5, 1-140Mb using the cas and mus contact maps. We reasoned that there would be few changes in interaction frequency on an autosome between the mus or cas contact maps for wild-type and XiΔxist/XaWT, thus the distribution of fold changes in interaction score on an autosome constitutes a null distribution. Using this distribution of fold changes allowed us to calculate a threshold fold change for an empirical FDR of 0.05, and all TADs that had a greater increase in
average normalized interaction score on Xi between wild-type and Xi^{Axist}/Xa^{WT} were considered restored TADs. We performed this analysis of restored TADs separately in each biological replicate using the 200kb contact maps to generate interaction scores over TADs and using the combined data at 100kb resolution.

*RNA-seq, ChIP-seq, and HiC-seq data are deposited in GEO under series GSE6751*
Chapter 3: Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape
**Addendum:**

This chapter and all data contained within it have been submitted for publication and are currently under review at *Nature Communications*. In addition, we have made this chapter publicly available as a pre-print manuscript on bioRxiv.org:


**Attributions:**

Stefan F Pinter performed all 4C experiments and 4C bioinformatic analysis.

Andrea J Kriz performed all Hi-C\(^2\) experiments and contributed to Hi-C\(^2\) bioinformatic analysis.

Teddy Jegu generated ATAC-seq libraries.

The remaining experiments and computational work (Hi-C, Hi-C\(^2\) analysis, RNA-seq, ATAC-seq analysis, microscopy, CRISPR deletions and genotyping) were performed by John E Froberg.
ABSTRACT

The mammalian inactive X-chromosome (Xi) is structurally distinct from all other chromosomes and serves as a model for how the 3D genome is organized. The Xi shows weakened topologically associated domains and is instead organized into megadomains and superloops directed by the noncoding loci, *Dxz4* and *Firre*. Their functional significance is presently unclear, though one study suggests that they permit Xi genes to escape silencing. Here, we find that megadomains do not precede Xist expression or Xi gene silencing. Deleting *Dxz4* disrupts megadomain formation, whereas deleting *Firre* weakens intra-megadomain interactions. Surprisingly, however, deleting *Dxz4* and *Firre* has no impact on Xi silencing and gene escape. Nor does it affect Xi nuclear localization, stability, or H3K27 methylation. Additionally, ectopic integration of *Dxz4* and Xist is not sufficient to form megadomains on autosomes, further uncoupling megadomain formation from chromosomal silencing. We conclude that *Dxz4* and megadomains are dispensable for Xi silencing and escape from X-inactivation.
INTRODUCTION

A longstanding principle in gene regulation invokes significance of higher order chromosome structures and specificity of 3D interactions between distant genetic elements. Advances in genomics have provided new opportunities to probe chromosome architecture and resulted in discovery of three types of long-range intra-chromosomal interactions. First, “topologically associating domains” (TADs) define continuous regions with extensive cis-contacts [220]. TADs are usually observed at length scales from 10^4-10^6 bp [228], depend on cohesins [233, 244, 282], and are generally flanked by convergent CTCF sites at TAD borders [218, 220, 230, 323]. TADs are visible as squares along the diagonal of Hi-C contact heat maps. Second, “loops” define enhanced contacts between pairs of distant loci that interact via CTCF. Loops can exist within or between TADs, and are visible as strong “dots” within a TAD square or between separate TADs in Hi-C contact maps [220]. Third, “compartments” transcend TADs and loops and exist as orthogonal structures formed by interactions between chromatin of similar epigenetic states. A-compartments harbor discontinuous chromosomal regions enriched for active genes, whereas B-compartments harbor discontinuous regions enriched for repressed genes [219, 220, 233, 244-246]. A/B compartments are visualized in Hi-C correlation maps by alternating “plaid” patterns of strong and weak interactions. Rapid depletion of CTCF [245] or cohesin [233] leads to genome-wide loss of TADs and loops, more pronounced A/B compartments (in the case of cohesin depletion) and only modestly affects transcription in the short term [233, 244, 282, 283]. Thus, while loops are thought to be important for long-range gene regulation (such as enhancer-promoter interactions), the functional organization into TADs and compartments is presently less well understood.

Recent conformational studies of the inactive X (Xi) has provided new insight into 3D chromosomal structure-function relationships [324, 325]. X-chromosome inactivation (XCI) occurs in female cells as part of a dosage compensation mechanism that equalizes dosage of
X-linked genes between males and females [285, 292, 326]. Chromosome conformation capture studies have demonstrated that, whereas the active X (Xa) resembles autosomes in having defined TADs, loops, and compartments, the Xi adopts a distinct structure seen on no other mammalian chromosome [192, 199, 203, 204, 220, 247]. ChIP-seq studies have shown that binding of architectural proteins including CTCF[37, 192] and cohesins[192] are relatively depleted on the Xi, providing a mechanistic explanation for the attenuation of TADs. The Xi also lacks the characteristic separation between transcriptionally active A compartments and silent B compartments. Instead, during XCI, the Xi is partitioned into transitional Xist-rich S1 and Xist-poor S2 compartments, which are later merged into a single compartment by the non-canonical SMC protein, SMCHD1 [248]. The merging of S1/S2 structure has physiological consequence, as ablating SMCHD1 precludes this fusion and leads to failure of silencing of >40% of genes on the Xi [248]. Thus, on the Xi, compartmentalization appears to play an important role in gene silencing.

On the other hand, the significance of domains on the Xi is under debate. Studies have shown that the Xi folds into two “megadomains” separated by a non-coding locus bearing tandem repeats known as “Dxz4” [192, 199, 247, 252]. In humans, DXZ4 is heterochromatinized and methylated on the Xa but is euchromatic and unmethylated on the Xi, where it binds CTCF [253, 254]. Murine Dxz4 is not well-conserved at the sequence level, but the syntenic region harbors a unique tandem repeat harboring strong CTCF binding sites [255]. In both mouse and human, Dxz4/DXZ4 resides at the strong border between the two megadomains of the Xi and binds CTCF and cohesin in an allele-specific manner (Supplementary Fig. S3.1). Deleting the Dxz4/DXZ4 region in both species results in loss of megadomains and increased frequency of interaction across the border [199, 252, 327]. Despite clear disruption of the Xi super-structure, there is presently no agreement regarding functional consequences. One group reported loss of ability of “escapees” to avoid silencing on the Xi [199]. Changes in repressive chromatin marks
and accessibility have also been reported in the mouse [199, 327]. Still others reported minimal effects, or even opposite effects, such as a partial loss of Xi heterochromatin in human cells [252]. Thus, there exists major disagreement as to whether the Dxz4 region and megadomains enable or oppose silencing.

Additionally, the Xi is characterized by a network of extremely long-range loops termed “superloops” [220] and the importance of these structures is also unknown. Superlooping occurs between Xist, DXZ4, and another tandem repeat element on the Xi called FIRRE [256, 257], another CTCF-bound noncoding locus (Supplementary Fig. S3.1) [220, 252, 254]. Far longer than almost all other contacts in mammalian genomes, the loops between Dxz4 and Firre extrude up to 25 Mb of DNA, a scale typically seen only in perturbed states, such as between super-enhancers of cohesin-depleted cells [233]. One study suggests that Firre RNA may direct Xist to the perinucleolar space and influence H3K27me3 deposition on the X [258]. However, despite the fact that the Firre locus falls at the border between two TADs and contains many CTCF binding sites, a recent study found that Firre is neither necessary nor sufficient to form borders between TADs, though it is required for superlooping with Dxz4 [328]. Here we combine genetic, epigenomic, and cell biological methods and study the impact of large-scale 3D structures on Xi biology.

RESULTS

Megadomains appear after Xist expression but not before Xi gene silencing

It is presently unknown how the formation of Xi megadomains relates to the timeline of XCI. To assess whether megadomains precede or follow XCI, we performed allele-specific Hi-C in female mouse embryonic stem cells (ES), which model different steps of XCI when they are induced to differentiate in culture. We examined timepoints day 0 (before XCI), day 3 (early
XCI, day 7 (mid-XCI), and day 10 (late-XCI) in the mESC line, $T_{six}^{TST/+}$, and compared the megadomain timeline to the time course of Xist upregulation and Xi silencing across the differentiating population in two biological replicates. Allelic analysis was made possible in $T_{six}^{TST/+}$ in two ways. First, it carries one X-chromosome of *M. castaneus* (cas) origin and one of *M. musculus* 129 origin (mus) [36, 101], the combination of which enabled employment of >600,000 polymorphisms to distinguish alleles. Second, the cell line carries a stop-mutation in the mus $T_{six}$ allele [101] that ensures that the mus X-chromosome is chosen as the Xi in >95% of cells [36, 72, 145].

For Hi-C analysis, we sequenced to a depth of 25-50 million reads, as megadomains are large (>70 Mb), prominent structures and can be sensitively detected at a resolution of 2.5 megabases (Mb). Allele-specific Hi-C contact maps and corresponding Pearson correlation heatmaps showed that, as expected, megadomains did not appear on the Xa during any stage of differentiation (Supplementary Fig. S3.2A,.). Focusing in on the Xi, we observed that, in pre-XCI cells (day 0) and in cells undergoing XCI (day 3), $X_{\text{mus}}$ resembled $X_{\text{cas}}$ (the Xa) in also lacking detectable megadomains (Fig. 3.1A,B). RNA FISH analysis of these timepoints showed that 30-60% of cells showed robust Xist RNA clouds by day 3 (Fig. 3.1C,D). Allele-specific RNA-seq analysis also showed robust upregulation of Xist starting on day 3 and continuing throughout differentiation (Fig. 3.1E, Supplementary Fig. S3.2C,D). Importantly, Xist was upregulated almost exclusively from $X_{\text{mus}}$ as expected, consistent with the $T_{six}^{TST}$ allele carried in cis[101]. [Note: The nonrandom pattern at day 3 agrees with $T_{six}$ being a primary determinant of allelic choice [56] rather than being a secondary selection mechanism following a stochastic choice process [329, 330]. A small fraction of reads coming from $X_{\text{cas}}$ is likely to be artifactual, as virtually all the $X_{\text{cas}}$ reads fell into one peak near the 5’ end of Xist, rather than being distributed across the entire gene body (Fig. 3.1E). This peak fell within a repetitive region of Xist (repeat A) and contained only one SNP ($\text{rs225651233}$) — a 129 G -> Cast/EiJ T variant
falling within a low complexity 24 bp poly-T tract. Thus, the X^{ass} reads are likely to be from an improperly-defined SNP.] Despite highly skewed Xist upregulation, allele-specific RNA-seq analysis showed that X-linked gene expression remained relatively unskewed (Fig. 3.1F), implying that de novo silencing or turnover of preexisting mRNA lagged behind Xist upregulation. A Hi-C “mixing” experiment indicated that our allelic Hi-C assay could detect megadomains when present in 25% of cells (Fig. 3.1H,I). Thus, the fact that megadomains were not readily visible on day 3 suggests that <25% of cells harbored them. Given that Xist spreading had taken place in 30-60% of cells and little silencing had taken place at this time point, megadomains were unlikely to have preceded Xist spreading and gene silencing.

On the other hand, analysis of day 7 cells revealed Xist expression in >80% of cells (Fig. 3.1C) and robust Xi silencing (Fig. 3.1F). It was at this timepoint that strong megadomains were first observed (Fig. 3.1A,B). Analysis of day 10 cells showed similarly strong Xist expression, Xi silencing, and megadomain formation. To quantify megadomain signals, we computed the Pearson correlation for the Xi contact maps and performed principal component analysis (PCA). On day 7 and day 10, there was a sharp transition in the 1st principal component score (PC1) at Dxz4, indicating changed interaction patterns on each side of Dxz4 at later timepoints but not in days 0 or 3 (Fig. 3.1G), consistent with appearance of megadomains. By contrast, the PC1 score distribution for the Xa was nearly identical for all timepoints without a sharp transition at Dxz4 (Fig. 3.1G). In addition, the sharp transition in the PC1 curve is a valid measure of megadomain strength, as our mixing experiment showed that the slope of the curve at Dxz4 was directly proportional to the fraction of day 10 cells in the mixing experiment (Supplementary Fig. S3.2E). The dynamics of megadomain formation were highly reproducible between two biological replicates (Supplementary Fig. S3.3A). Taken together, these data suggest that megadomains do not precede XCI and appear either concurrently with or (more likely) only after Xist has spread and silenced the Xi.
Figure 3.1: Dynamics of megadomain formation during XCI.

(A) KR-normalized Hi-C matrices on future Xi (mus) in female ES cells on days 0, 3, 7 and 10 of differentiation (2.5 Mb resolution) (B) Pearson correlation of Hi-C matrices on days 0, 3, 7, 10 of differentiation (1 Mb resolution). (C) Xist RNA FISH on day 0, 3, 7, 10 of differentiation in Tsix^{TST/+} ES cell line. (D) Fraction of cells showing Xist clouds on day 3 of differentiation in two biological replicates. Error bars represent standard deviation of counts over 6 fields in one imaging session. (E) Allele-specific expression from the Xi (red) or Xa (blue) over Xist in Tsix^{TST/+} during differentiation. (F) Density plots of the number of X-linked genes with a given level of allelic expression from the Xi on day 0, 3, 7, 10 of differentiation in Tsix^{TST/+}. Wilcoxon p-values comparing the mean allelic expression levels from the Xi between d0 and all other timepoints are indicated. (G) 1st principal component of the 1Mb correlation matrix plotted for all bins on the future Xi (mus, left) and future Xa (cas, right) for Days 0, 3, 7, 10 of differentiation. Dotted lines correspond to the bin containing Dxz4. All Hi-C data in this figure are generated from merging together reads from two biological replicates. (H) In silico mixing experiment to determine the sensitivity of the Hi-C assay for detecting megadomains. KR-normalized Hi-C matrices at 2.5 Mb resolution (top) and Pearson correlation of Hi-C matrices at 1 Mb (bottom) for the Xi from data sets generated by mixing an indicated ratio of day 0 (d0, megadomain-negative) and day 10 (d10, megadomain-positive) datasets. (i) 1st principal component of the 1Mb correlation matrix plotted for all bins on the Xi for datasets within varying ratios of d10:d0 reads.
Figure 3.1 (Continued)
Time course of TAD attenuation on the Xi

Recent analyses indicate that TADs are not abolished on the Xi but are instead attenuated [248, 327]. Here we investigate the time course of TAD attenuation during XCI. To enrich for interactions and obtain higher resolution allele-specific contact maps, we performed Hi-C², a variation of the Hi-C protocol that focuses analysis on defined regions through hybrid capture using high density probe sets [331]. We investigated ~1.5 Mb regions around (i) Dxz4 to assess the behavior of the strong megadomain border and (ii) the TAD harboring the disease locus and inactivated gene, Mecp2, to examine how topological domains are weakened during XCI.

Interestingly, despite megadomains appearing only late during the XCI time course, the Dxz4 region showed a strong boundary during all time points and on both alleles (Fig. 3.2A,B). Therefore, Dxz4 acts as a border irrespective of XCI status and presence/absence of megadomains. At days 7 and 10, the proximal TAD flanking Dxz4 strengthened and expanded on the Xi but not the Xa (Fig. 3.2A,B), leaving only two “boxes” on either side of Dxz4 by day 10, rather than the patchwork of smaller sub-TADs present on the Xa and Xi at earlier timepoints. This finding indicated that the emergence of a megadomain correlates with increased insulation by Dxz4, indicating that Dxz4 insulates interactions at increasingly larger distances when megadomains form. The temporal dynamics of chromatin conformation surrounding Dxz4 were quite similar in two independent replicates of the differentiation timecourse and Hi-C² enrichment (Supplementary Fig. S3.3B).

Within the region containing Mecp2, Hi-C² contact maps showed that X^{mus} and X^{cas} behaved similarly on days 0 and 3, in that both were organized into several sub-TADs (Fig. 2C,D). However, once Xist spread over X^{mus} and the Xi formed as a consequence on days 7 and 10, both TAD and sub-TAD organization become obscured compared to the Xa, where these domains stayed similar to earlier timepoints. In contrast to a previous analysis performed
in neural progenitor cells (NPCs) [199], we did not observe the persistence of a small domain around Mecp2 in differentiating female ES cells. The loss of domain organization in the Mecp2 region was observed in two distinct biological replicates (Supplementary Fig. S3.3C).

To quantify these changes, we computed insulation scores using standard methods [199, 223] (Fig. 3.2E,F). In brief, insulation scores quantify how strongly a given locus acts as a border[223, 332, 333], and are calculated by running sliding windows across a chromosomal region and measuring the log ratio of reads crossing over a locus to reads neighboring a locus (Supplementary Fig. S3.3D). Loci in the interiors of a TAD would be expected to have similar numbers of cross-over interactions and local interactions on each side, leading to insulation scores near zero, whereas loci at borders would register as a local minimum of crossover interactions than local interactions, leading to strong negative insulation scores at domain boundaries. We observed several interesting facets of the insulation score curves on the Xa and Xi at Dxz4 and Mecp2 during the differentiation timecourse. There was a strong decrease in insulation scores near Dxz4 on both alleles at all timepoints, consistent with Dxz4 acting as a boundary throughout differentiation (Fig. 3.2E,G). The variance of the insulation scores is a measure of the global strength of insulation, with smaller variance corresponding to weaker insulation [334]. Near Dxz4, the variance was slightly smaller on the Xi than the Xa for all timepoints. There was no statistically significant difference in variance of insulation scores on the Xi on day 0 compared with the variance insulation scores on the Xi for the later timepoints (pairwise F-test p-values >0.05 for all comparisons between day 0 Xi and later Xi timepoints) and this observation held across two biological replicates (Fig. 3.2G, Supplementary Fig. S3.3E). Thus, the Dxz4 region is a strong boundary regardless of XCI status, but that Dxz4 insulates interactions from increasingly larger distances to form megadomains.

The Mecp2 region showed a different pattern of insulation score changes across differentiation. The distribution of insulation scores across the Mecp2 region significantly
narrowed on days 7 and 10 on the Xi but not Xa (Fig. 3.2H). Indeed, across the Mecp2 region, the variance on the day 7 or day 10 Xi was significantly lower than on day 0 (day 7 vs day 0 p-value=0.008443; day 10 vs day 0 p-value=0.001819). There was no significant difference in the variance on the Xi between day 0 and day 3 (p=0.7369). This clear decrease in the variance on the Xi at day 7 and day 10 relative to d0 was observed in two biological replicates (Supplementary Fig. S3.3F). These results indicate that TAD and sub-TAD structures of the Mecp2-containing TAD region are reduced in strength in the same timeframe that megdomains are gained.
Figure 3.2: Dynamics of TADs and sub-TADs during XCI in the *Mecp2* and *Dxz4* regions.

(A,B) Hi-C² contact maps around *Dxz4* (mm9 coordinates chrX:71,832,976-73,511,687) and on the future Xi (A) or Xa (B) on days 0, 3, 7, 10 of ES differentiation (50 kb resolution). (C,D) Hi-C² contact maps around *Mecp2* (mm9 coordinates chrX:70,370,161-71,832,975) and on the future Xi (C) or Xa (D) on Days 0, 3, 7, 10 of differentiation (50 kb resolution). Green bars indicate positions of sub-TAD borders determined from 25 kb d0 comp Hi-C² matrices; dark blue track shows Dixon et al [228]. TAD calls in mESCs, light blue track shows Marks et al [186]. TAD calls in mESCs, red bars indicate positions of either *Dxz4* or *Mecp2*. In addition, the regions of the contacts corresponding to sub-TADs (green), Dixon et al. TADs (dark blue), Marks et al. TADs (light blue) have been indicated with boxes on the day 0 Xi contact maps for reference. (E,F) Insulation scores across the *Dxz4* region (E) or Insulation scores across the *Mecp2* region (F). Insulation scores on the Xa are in blue and insulation scores on the Xi are red. For reference, CTCF ChIP-seq in d0 F1-2.1 mESCs (black) and TAD calls from Dixon et al. (grey) are shown [228]. (G,H) Violin plots showing the distributions of insulation scores across the *Dxz4* region (G) and *Mecp2* region (H). All data in this figure are generated from merging together reads from two biological replicates. Note: to generate violin plots and evaluate the significance of differences in variance between timepoints we excluded the 6 bins on each edge of the Hi-C² region because the regions needed to calculate insulation score fall partly outside the Hi-C² region and have far lower read counts than sequences targeted by the capture probes.
Figure 3.2 (Continued)
**Dxz4 is necessary but not sufficient for megadomain formation**

There presently exist three deletions containing *Dxz4/DXZ4* — two in mouse [199, 252, 327], one in human [252]. One of the mouse deletions [199] is a large deletion that contains more than just the noncoding element, *Dxz4/DXZ4* (Fig. 3.3A). We generated a new deletion of *Dxz4* and its flanking sequences that left untouched a small cluster of CTCF motifs with very high CTCF coverage and an unusual satellite repeat (Fig. 3.3A and Supplementary Fig. S3.4A)[255], both of which were deleted in a previous 200-kb *Dxz4* deletion. We generated a smaller 100-kb deletion spanning *Dxz4* (*Dxz4<sup>Δ100</sup>*) and validated our deletion by Sanger sequencing, by DNA fluorescence in situ hybridization (FISH) with a probe internal to the deleted region, and by genomic DNA sequencing to examine read distributions over the deleted region (Supplementary Fig. S3.4, Methods). Importantly, to distinguish Xi from Xa, we performed the deletion analysis in *Tsix<sup>TST</sup>+* mESCs.

To test the impact of removing *Dxz4*, we differentiated wild-type and homozygously deleted (*Dxz4<sup>ΔΔ</sup>*) cells for 10 days and performed Hi-C. Whereas wild-type cells showed strong megadomains, *Dxz4<sup>ΔΔ</sup>* cells showed disrupted megadomain structures in Hi-C contact maps (Fig. 3.3B) and corresponding Pearson correlation maps. Most prominently, the sharp border around *Dxz4* was eliminated, though some intra-megadomain interactions remained on either side of the deletion. A disrupted megadomain border was confirmed by loss of the sharp transition in PC1 score around the *Dxz4* locus (Fig. 3.3C). This effect was observed in two biological replicates (Supplementary Fig. S3.4D). These results agree with prior reports [199, 252, 327] that the 200-300 kb region around *Dxz4* is required for megadomain organization. Additionally, our work delineates the required region to a 100-kb domain containing the *Dxz4* tandem repeat itself (as opposed to the CTCF motif cluster and the proximal satellite repeats).

Given its necessity for megadomain organization, we asked whether *Dxz4* is also sufficient to form a megadomain on an autosome when Xist RNA is expressed in *cis*. We co-
transfected a dox-inducible full-length Xist construct along with a BAC containing Dxz4 into male fibroblasts and used RNA and DNA FISH to identify Xist-inducible clones (Supplementary Fig. S3.5A) where both Xist and Dxz4 had co-inserted (XPDxz4.4; Supplementary Fig. S3.5B). To localize the transgene, we adapted the 4C technique [214, 215] that is ordinarily used to view 3D interactions from a single locus. We reasoned that, by placing the 4C viewpoint anchor at the transgene backbone, we could map the transgene through the pattern of cis-interactions on the same chromosome (Fig. 3.3D), as interaction frequencies are typically highest near the viewpoint position. This fact has previously been used to aid genome assembly [320, 335, 336]. Indeed, in addition to interaction peaks at Xist and Dxz4 as expected (Supplementary Fig. S3.5C), the only other strong 4C peak in the genome appeared on chr14 near Stc1 (Fig. 3.3E). This finding contrasts with both a control Xist-only transgene line which showed a peak only on chr10 (Fig. 3.3F) and the parental rtTA fibroblasts which showed no peaks anywhere in the genome. We confirmed insertion of both Dxz4 and the Xist construct into Stc1 by observing co-localization between Stc1, Xist and Dxz4 DNA FISH probes at one spot in the transgenic cell line (Supplementary Fig. S3.5D).

To test whether induction of Xist expression could induce megadomain formation at Dxz4 ectopically, we induced Xist and performed Hi-C to determine whether co-insertion of Xist and Dxz4 induced formation of megadomains on transgenic chr14 in fibroblasts. We induced Xist for 2 days because a previous report suggested that induction of Xist from the male X for 2 days was sufficient to at least initiate megadomain formation [199]. No megadomains formed and the overall chr14 contact maps looked highly similar to the non-transgenic and Xist-only controls (Fig. 3.3G). We then extended the time frame and induced for 9 days, given that our ES cell time course suggested that several days of Xist upregulation were needed to form megadomains on the Xi. Still, no megadomains formed in these post-XCI cells (Fig. 3.3G). To assess whether Xist and Dxz4 could do so in cells undergoing de novo XCI, we attempted three
times to create the $Xist$-$Dxz4$ transgene line in a female ES cell background, but such a line could not be generated, due to potential lethal consequences of the $Xist$ transgene. These results indicate that $Xist$ and $Dxz4$ together are not sufficient for megadomain formation in a cell line that had already undergone XCI (fibroblasts). We conclude that $Dxz4$ and $Xist$ expression are necessary but not sufficient for megadomain formation in post-XCI cells.
Figure 3.3: *Dxz4* is necessary but not sufficient for megadomain formation. 

(A) Schematic of prominent features in the *Dxz4* region and deletions generated in this study and previous studies. (B) Top: Contact maps for the Xi at 2.5Mb resolution for wild-type (left) and *Dxz4* Δ/Δ (right) cells on d10 of differentiation. Bottom: Pearson correlation matrices at 1Mb resolution for the wild-type (left) and *Dxz4* Δ/Δ Xi (right). Heatmaps generated by merging reads together from two biological replicates. (C) 1st principal component of the 1Mb correlation matrix plotted for wild-type (black) and *Dxz4* Δ/Δ (red) across all bins on the Xi. The dotted line corresponds to the bin containing *Dxz4*. (D) Strategy for using 4C to localize the transgene insertion site. (E,F) 4C interaction profiles using a viewpoint in the backbone of an Xist transgene in either wild-type (top), Xist only Tg (middle) or Xist+Dxz4 Tg (bottom) lines across chr14 (E) or chr10 (F). (G) Hi-C contact maps for chr14 at 1Mb resolution in either wild-type (left), Xist only Tg (middle left), Xist+Dxz4 Tg Dox induced for 2 days (middle right) or 9 days (right). The position of the transgene insertion site at ~69.8Mb (near Stc1) is indicated by a green bar in the Xist+Dxz4 Tg contact maps.
Figure 3.3 (Continued)

A

![Diagram of chromosome position and unique repeats]

B

WT Xa Hi-C contacts

Dox4 bin

Dox44444 Hi-C contacts

Dox4 bin

WT Xa Hi-C correlation

Dox44444 Hi-C correlation

C

![Graph of PC1 score over chromosome position]

D

4C method for identifying Tg insertion site

E

Transgene backbone 4C, chr14

XY rTA

0-20000

XsΔ+ Tg

0-20000

XsΔ+ Dox4

0-20000

F

Transgene backbone 4C, chr10

XY rTA

0-20000

XsΔ+ Tg chr14 Hi-C, 2 days +Dox

XsΔ+ Dox4 Tg chr14 Hi-C, 9 days +Dox

G

WT XY rTA chr14 Hi-C

XsΔ+ Tg chr14 Hi-C

XsΔ+ Dox4 Tg chr14 Hi-C, 2 days +Dox

XsΔ+ Dox4 Tg chr14 Hi-C, 9 days +Dox
**Dxz4, Firre, and Xi-specific superloops**

In addition to serving as the border between the megadomains, Dxz4 has been shown to form extremely long (>10 Mb) looping interactions with other loci on the human Xi [220, 252, 254]. To further dissect the role of Dxz4 in establishing the large-scale structure of the Xi, we performed 4C using a viewpoint within the core of the Dxz4 tandem repeats in post-XCI fibroblasts to identify interacting loci that may be important for helping to establish the unique structure of the mouse Xi. Dxz4 generally interacted with the chromosome telomeric to Dxz4 and formed few long-range interactions towards the centromeric side of the chromosome (Fig. 3.4A, Supplementary Fig. S3.6A). However, Dxz4 interacted strongly with another non-coding tandem repeat, Firre (Fig. 3.4A). The two loci formed an extremely strong loop despite the fact that Firre is 25 Mb centromeric to Dxz4. The strength of their interaction was equivalent to that of two loci separated by < 200 kb (data not shown). To verify the Dxz4:Firre interaction, we performed a reciprocal 4C using a viewpoint within the core of the Firre tandem repeats and confirmed a strong reciprocal interaction (Fig. 3.4A). On the Xa, Firre also formed a broad domain of interactions with nearly all sequences within several Mb of itself, as reported for the Firre RNA contact map previously [256]. In contrast to this prior study, however, we did not observe any evident interchromosomal contacts from either Xa or Xi allele in fibroblasts.

Our allele-specific analysis revealed that the Dxz4:Firre interaction is primarily detected on the Xi (mus) allele. Indeed, when we repeated this reciprocal 4C experiment in another hybrid fibroblast line that chose to inactive X^{Cas}, the Dxz4:Firre interaction was detected on X^{Cas}, rather than X^{mus}. For these experiments, we used both a unique 4C anchor in the 3’ flanking region of Firre that provides allelic information, as well as an allele-agnostic anchor in the core of Firre repeat. Because Dxz4 and Firre are both highly repetitive, we also examined multiply-aligning reads. The Firre-Dxz4 interaction was only observed on the Xi regardless of whether
the Xi was the mus or cas chromosome (Fig. 3.4B-D). Thus, \textit{Firre} and \textit{Dxz4} formed an Xi-specific superloop conserved between mouse and primate [220, 252].

Other superloops have been identified on the human Xi using high-resolution Hi-C. \textit{FIRRE, DXZ4, XIST, ICCE} and \textit{X75} are all repetitive loci that bind CTCF on the Xi in human cells, and all form long-range interactions with each other [220, 252]. We examined whether these superloops also occur in mouse cells. Indeed, in addition to \textit{Firre}, we observed elevated interaction frequencies between \textit{Dxz4} and a region spanning \textit{Xist} to \textit{Ftx} (Supplementary Fig. S3.6A,B) and a region syntenic with human \textit{X75} (Supplementary Fig. S3.6A,C). However, the \textit{Xist-Dxz4} and \textit{x75-Dxz4} contacts were less prominent than the \textit{Firre-Dxz4} contact.
Figure 3.4: *Dxz4*, *Firre*, and Xi-specific superloops.

(A) Reciprocal interaction between *Dxz4* and *Firre* in post-XCI fibroblasts. Top: 4C interaction profiles with the core of the *Firre* tandem repeat as the viewpoint. Bottom: 4C interaction profiles with the core of the *Dxz4* tandem repeat as the viewpoint. Black: all unique reads (comp). Grey: repetitive (non-unique) reads (comp). Blue: Xa (cas) specific reads. Red: Xi (mus) specific reads.

(B) 4C from a unique, allele-specific viewpoint at the 3’ end of *Firre* in cells with either an inactive mus (cas Xa/mus Xi) or an inactive cas (cas Xi/mus Xa). Red: interaction profile for unique (comp) reads on the mus allele. Blue: interaction profile for unique (comp) reads on the cas allele. Pink: interaction profile for non-unique (reps) reads from the mus allele. Light Blue: interaction profile for non-unique (reps) reads from the cas allele. Positions of *Firre*, *Dxz4* and *Xist* are shaded light grey and indicated with red bars.

(C) 4C interaction profile over *Dxz4* from a unique, allele-specific viewpoint at the 3’ end of *Firre* in cells with either a mus Xi (top tracks) or a cas Xi (bottom tracks). Light blue: interaction profile from the cas *Firre* viewpoint. Pink: interaction profile from the mus *Firre* viewpoint.

(D) 4C interaction profile over *Firre* using the *Dxz4* core viewpoint in cells either with a mus Xi (top tracks) or a cas Xi (bottom tracks). Black: all unique (comp) reads. Grey: non-unique (repetitive) reads. Blue: cas reads. Red: mus reads.

Note: (c) and (d) zoom into *Dxz4* and *Firre* respectively to better highlight the allele-specific superloop interactions.
Figure 3.4 (Continued)
**Firre is predominantly expressed from the Xa**

Previous reports have suggested that *Firre* escapes from X-inactivation [185, 256, 258], and that Firre RNA is necessary for *Xist* localization and deposition of H3K27me3 on the Xi [258]. By allele-specific RNA-seq, we observed Firre reads from both Xa and Xi during differentiation, and expression appeared to be predominantly though not exclusively exonic (Fig. 3.5A,B). Because *Firre* is highly repetitive, SNP calls may be not be fully reliable. To confirm allele-specific expression, we used genetic means to examine expression from the Xa and Xi. With allele-specific guide RNAs, we generated an Xi-specific *Firre* deletion in $T_{S}x^{TST/+}$ ("*Firre*$_{Xi/+}$ clone D1"), an Xa-specific Firre deletion in $T_{S}x^{TST/+}$ ("*Firre*$_{Xa/+}$ clone H6"), and a homozygous deletion ("*Firre*$_{∆/∆}$") in female ES cells (Supplementary Fig. S3.7). We then measured *Firre* expression on day 10 of differentiation using 4 published sets of primers [256, 258] and one new intronic primer set and deduced the expressed allele(s) by examining differences in expression pattern between the reciprocal heterozygous clones. First, by quantitative RT-PCR of wild-type female ES cells, we inferred that Firre expression was expressed at <10% of Xist RNA overall. The variability between amplicons suggested that there could be multiple isoforms of Firre (Fig. 3.5D). Second, deleting *Firre* on the Xi abolished expression of the two lowest expressed amplicons (Fig. 3.5E,F). By contrast, deleting *Firre* on the Xa abolished expression of the two most highly expressed exonic amplicons and the intronic amplicon. Finally, a homozygous deletion abolished all expression measured from these primer pairs.

To determine if *Dxz4* influences *Firre* expression, we also performed RT-PCR in the $D_{x}z_{4}^{Δ/Δ}$ cell line. No changes were evident, indicating that the *Dxz4-Firre* superloop does not impact transcriptional regulation of *Firre* (Fig. 3.5E-I). Finally, to examine whether either repeat locus regulates *Xist* expression, we performed RT-PCR in cell lines carrying deletions of either *Firre* or *Dxz4*, and also generated and tested a double ES cell knockout of *Dxz4* and *Firre* on
the Xi carrying the TsixTST allele (Dxz4Δ/Δ:FirreXΔ+) (Supplementary Fig. S3.7, Methods). None of these deletions affected Xist expression (Fig. 3.5J). Together, our results suggest that different isoforms of Firre RNA are expressed from the Xa and Xi, but expression may predominate on the Xa. The results also indicate that the superloops are irrelevant for expression of Dxz4, Firre, and Xist.
Figure 3.5: *Firre* is predominantly expressed from the Xa in differentiating female ES cells.

(A) Expression over *Firre* during differentiation. (B) Allelic expression over *Firre* during differentiation, red=Xi reads, blue=Xa reads. (C) Positions of Firre primers within the *Firre* locus. (D) Expression of Xist or Firre amplicons in d10 wild-type cells normalized to Gapdh (logarithmic scale). (E-J) Expression normalized to Gapdh measured with JR1 (E), CD (F), intronic (G), JR2 (H), JR4 (I) Firre primers or Xist primers (j) in WT, *Dxz4Δ/Δ FirreΔ/+*, *Dxz4Δ/Δ:FirreΔ/+*, *FirreΔΔ/+* and *FirreΔΔ* cells. Blue bars are +RT, orange bars are -RT. Asterisks indicate a statistically significant (*p* < 0.05, t-test) difference between normalized expression levels in WT and normalized expression levels in the deletion. Error bars show standard error of the mean, 3 biological replicates.
Firre is not strictly required for megadomain formation, but works together with Dxz4 to strengthen megadomains

Prior work had not examined whether Xi superloops contribute to the formation of the megadomain boundary at Dxz4. Our Firre deletion lines allowed us to use Hi-C to test whether loss of the other anchor in the Dxz4-Firre superloop perturbed megadomains. To exclude possible trans-effects relating to Firre RNA [337] expressed from the Xa and to focus on the role of Firre locus as a superloop anchor, we conducted all analysis in our Xi-specific Firre deletion. We performed Hi-C on day 10 of differentiation in wild-type, Firre\(^{Xi/+}\) and Dxz4\(^{\Delta/\Delta}\):Firre\(^{Xi/+}\) cells. Importantly, despite disruption to the other side of the Dxz4-Firre superloop, the Firre\(^{Xi/+}\) cells retained the sharp megadomain boundary on the Xi. However, in two biological replicates, the intra-megadomain interactions appeared attenuated compared to wild-type (Fig. 3.6A,B; Supplementary Fig. S3.8). Plotting PC1 scores for all bins on the Xi showed a sharp transition in PC1 score at Dxz4 in wild-type and Firre\(^{Xi/+}\) (Fig. 3.6C), in agreement with the continued presence of a bipartite structure. Thus, Firre may influence the strength of interactions within each megadomain but is not strictly required for formation of the Dxz4 border and the bipartite mega-structures. Given the phenotypes of the Dxz4 and Firre single deletions, we predicted that the double deletion would obliterate all trace of megadomains. Indeed, the Dxz4\(^{\Delta/\Delta}\):Firre\(^{Xi/+}\) cells showed a loss of the sharp border at Dxz4 and a depletion of intra-megadomain interactions (Fig. 3.6A-C). Based on the Pearson correlation heatmap and PC1 analysis, the double deletion may potentially reduce the intra-megadomain interactions on either side of Dxz4 more than in the Dxz4\(^{\Delta/\Lambda}\) single deletion, though the sharp border at Dxz4 was lost in both cases (Fig. 3.3, Supplementary Fig. S3.4 versus Fig. 3.6, Supplementary Fig. S3.8). We conclude that Firre is not absolutely required for megadomain formation. However, Firre and the Firre-Dxz4 superloop may work together with Dxz4 to strengthen the megadomain structure.
Figure 3.6: *Firre* is not required for megadomain formation but deleting *Firre* weakens megadomains.

(A) Contact maps for the Xi at 2.5Mb resolution for wild-type (left), *Firre*<sup>XlΔ/+</sup> (center) and *Dxz4ΔΔ*:Firre<sup>XlΔ/+</sup> cells on d10 of differentiation. (B) Pearson correlation matrices at 1Mb resolution for the wild-type (left) *Firre*<sup>XlΔ/+</sup> (center) and *Dxz4ΔΔ*:Firre<sup>XlΔ/+</sup>. (C) 1<sup>st</sup> principal component of the 1Mb correlation matrix plotted for wild-type (black) and *Firre*<sup>XlΔ/+</sup> (blue) and *Dxz4ΔΔ*:Firre<sup>XlΔ/+</sup> (purple) across all bins on the Xi. The dotted line corresponds to the bin containing *Dxz4*. All data in this figure are generated from merging together reads from two biological replicates.
Figure 3.6 (Continued)

A

WT X1 H-C contacts  
Dox4

Finn\(^{lo}\) clone D1  
Dox4\(^{lo}\) Finn\(^{lo}\) X1

B

WT X1 H-C correlation  
Dox4

Finn\(^{lo}\) clone D1  
Dox4\(^{lo}\) Finn\(^{lo}\)

C

Dox4

PC1 score

chrX position (Mb)
Megadomains and superloops can be uncoupled from XCI and escape

Recent studies have diverged on the effect of deleting the \textit{Dxz4} region on XCI, as one study suggested a loss of escape from XCI for many escapees in mouse NPCs [199] and another suggested a partial loss of H3K27me3 on the human fibroblast Xi [252]. A prior report also suggested that knockdown of \textit{Firre} RNA disrupts localization of the Xi to the nucleolus and maintenance of H3K27me3 on the Xi [258]. Here we assessed the effect of deleting \textit{Dxz4}, \textit{Firre}, or both on various aspects of XCI. First, we examined effects on the Xist RNA cloud that normally forms over the Xi but observed no obvious changes in Xist cloud morphology or number of cells exhibiting \textit{Xist} upregulation on day 10 of differentiation in \textit{Dxz4}\textsuperscript{-/-}, \textit{Firre}\textsuperscript{+/-}, \textit{Firre}\textsuperscript{+/-}, \textit{Dxz4}\textsuperscript{-/-}:\textit{Firre}\textsuperscript{+/-} or \textit{Firre}\textsuperscript{-/-} (Fig. 3.7A,B, Supplementary Fig. S3.9A,B) versus wildtype female cells. There was also no effect on the characteristic localization of Xist RNA/Xi to the perinucleolar region [302] (Fig. 3.7A,B, Supplementary Fig. S3.9A,B). We also observed no difference in the enrichment of the H3K27me3 repressive mark on the Xi after 7 days of differentiation in any of the \textit{Dxz4} or \textit{Firre} deletions (Fig. 3.7A,C) or after 10 days in the \textit{Dxz4} and Xi-specific \textit{Firre} deletions (Supplementary Fig. S3.9A,C). This suggests that neither \textit{Firre} nor \textit{Dxz4} is required for Xist to be expressed, localized, and deposit H3K27me3 on the Xi. To test whether there is a partial loss of H3K27me3 across a macroscopic region of the Xi, as observed in a human DXZ4 deletion [252], we produced metaphase spreads in WT and \textit{Dxz4}\textsuperscript{-/-} cells and performed H3K27ac and H3K27me3 immunofluorescence to visualize the Xi. The Xi stood out as the chromosome with almost no H3K27ac signal and very strong H3K27me3 signal [338] (Supplementary Fig. S3.9D). However, we observed no obvious difference between the H3K27me3 banding pattern on the WT or \textit{Dxz4}\textsuperscript{-/-} Xi in metaphase spreads, suggesting no loss of H3K27me3 across a large region of the mouse Xi following \textit{Dxz4} deletion (Supplementary Fig. S3.9E).
We next used ATAC-seq [339] to assay chromatin accessibility on the Xi. In wild-type cells, ATAC signal was indeed heavily skewed towards the Xa, with >85% of all peaks on the X binding specifically to the Xa (Fig. 3.7D). There were ~20 biallelic sites (e.g., promoters of escapee genes) and only one Xi-specific peak (at Firre) (Fig. 3.7E). If Dxz4 impairs X-chromosome accessibility as previously proposed for escapee genes [199], we would expect nearby biallelic peaks near in wild-type to become Xa-specific in the deletion. On the other hand, if Dxz4 or Firre were required to inhibit chromatin accessibility, we would expect many ATAC-peaks to appear on the Xi near genes subject to XCI. Significantly, the overall ATAC-seq patterns were highly similar between wild-type and all mutant genotypes — Dxz4\(\Delta/\Delta\), Firre\(^{\Delta/\Delta}\), and Dxz4\(\Delta/\Delta\):Firre\(^{\Delta/\Delta}\) (Fig. 3.8A). We did not observe any “restored” sites on the mutant Xi, when plotting mutant Xi read counts vs. wild-type Xa read counts for peaks that reached at least one half of the wild-type Xa read count (Fig. 3.8B-D, Supplementary Fig. S3.10A). We also compared the Xi read counts for biallelic peaks and observed no changes in Dxz4\(\Delta/\Delta\) cells (Fig. 3.8E), Firre\(^{\Delta/\Delta}\) cells, and Dxz4\(\Delta/\Delta\):Firre\(^{\Delta/\Delta}\) cells (Fig. 3.8F, Supplementary Fig. S3.10B). Thus, we found no decrease in chromatin accessibility at escapee genes, in contrast to a previous study [199]. Altogether, our results demonstrate that deletion of either Dxz4, Firre, or both tandem repeats has no impact on chromatin accessibility on the Xi, at either inactive genes or escapees.

Finally, we asked whether deleting both Dxz4 and Firre disrupts the pattern of gene silencing or escape on the Xi. We performed allele-specific RNA-seq in wild-type Tsix\(^{TST/+}\) and Dxz4\(^{\Delta/\Delta}\):Firre\(^{\Delta/\Delta}\) and looked for changes on either a genome-wide or Xi-scale in the mutant relative to wildtype. Surprisingly, no significant differences were detected on a global or Xi-wide scale (Fig. 3.8G-I), in contrast to previous deletions of Dxz4 [199]. Cumulative frequency plots showed balanced X-to-autosomal gene dosages when comparing mutant to wildtype cells (Fig. 3.8G). The overall number of genes escaping XCI was similar in wild-type and Dxz4\(^{\Delta/\Delta}\):Firre\(^{\Delta/\Delta}\),
with about half of the escapees being shared between them (Fig. 3.8H). Examination of allelic contributions to overall X-chromosomal expression showed a predominance of Xa expression in both wild-type and \(Dxz4^{\Delta/\Delta}:Firre^{X\Delta/+}\), with the pattern being highly similar in two biological replicates (Fig. 3.8I; \(p > 0.3\) for all pairwise comparisons). We conclude that neither \(Firre\) nor \(Dxz4\) significantly perturbs Xi silencing and escape from silencing. Thus, the unique superloops of the Xi can be uncoupled from XCI biology.
Figure 3.7: *Dxz4* and *Firre* do not affect Xi localization and accessibility.

(A) Xist RNA FISH (red) combined with nucleophosmin immunofluorescence (green) (top) and H3K27me3 IF (bottom) in WT, *Dxz4*Δ/Δ *Firre*Δ+/Δ, *Dxz4*Δ/Δ:*Firre*Δ+/+, *Firre*Δ+/Δ and *Firre*Δ/Δ cells. (B) Fraction of cells with Xist clouds (red) and fraction of Xist clouds in the perinucleolar space (cyan). (C) Fraction of cells with an H3K27me3 focus. (D) ATAC-seq coverage across the X (all unique reads, black), Xa (blue) and Xi (red) in WT cells, 3 biological replicates. (E) Allelic status of ATAC peaks in WT cells.
Figure 3.7 (Continued)

A

WT  \(\text{D}x\text{z}4^\Delta\)  \(\text{Firre}^{\text{X}\Delta}\)  \(\text{D}x\text{z}4^\Delta:\text{Firre}^{\text{X}\Delta}\)  \(\text{Firre}^{\text{X}\Delta}\)

B

\[
\begin{array}{cc}
\text{Xist clouds} & \text{perinucleolar} \\
\end{array}
\]

\[
\begin{array}{c}
\% \text{cells with Xist expression} \\
\end{array}
\]

C

\[
\begin{array}{c}
\% \text{cells with H3K27me3 tag} \\
\end{array}
\]

D

AVAC-seq of WT cells

E

<table>
<thead>
<tr>
<th></th>
<th>Xa-specific</th>
<th>Biallelic</th>
<th>Xi-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT rep1</td>
<td>214</td>
<td>24</td>
<td>1 (Firre)</td>
</tr>
<tr>
<td>WT rep2</td>
<td>206</td>
<td>21</td>
<td>1 (Firre)</td>
</tr>
<tr>
<td>WT rep3</td>
<td>158</td>
<td>25</td>
<td>1 (Firre)</td>
</tr>
</tbody>
</table>
Figure 3.8: Deletion of Dxz4, Firre, or both does not affect accessibility or gene silencing on the Xi.

(A) Comparison of Xi ATAC-seq coverage, WT (black), Dxz4\(^{\Delta/\Delta}\), Firre\(^{Xi/+}\) (blue) and Dxz4\(^{\Delta/\Delta}\).Firre\(^{Xi/+}\) (purple). (B) Comparison between WT Xa and Dxz4\(^{\Delta/\Delta}\) Xi ATAC coverage for peaks that are Xa-specific in wild-type. The black line corresponds to a Dxz4\(^{\Delta/\Delta}\) Xi:WT Xa ratio of 1:1, the red line corresponds to a ratio of 1:2. (C) Comparison between WT Xa and Firre\(^{Xi/+}\) (left) or Dxz4\(^{\Delta/\Delta}\).Firre\(^{Xi/+}\) (right) Xi ATAC coverage for peaks that are Xa-specific in wild-type. The black line corresponds to a Dxz4\(^{\Delta/\Delta}\) Xi:WT Xa ratio of 1:1, the red lines correspond to a ratio of 1:2. (D) Number of restored ATAC peaks (Xa-skewed in wild-type but biallelic in deletion) for all deletions. Concordant peaks are peaks reproducibly restored across replicates. (E) Comparison between WT Xi and Dxz4\(^{\Delta/\Delta}\) Xi ATAC coverage for peaks that are Xi-specific in wild-type. The black line corresponds to a Dxz4\(^{\Delta/\Delta}\) Xi:WT Xi ratio of 1:1, the blue line corresponds to a ratio of 1:2. (F) Comparison between WT Xi and Firre\(^{Xi/+}\) (left) or Dxz4\(^{\Delta/\Delta}\).Firre\(^{Xi/+}\) (right) Xi ATAC coverage for peaks that are Xi-specific in wild-type. The black lines correspond to a Dxz4\(^{\Delta/\Delta}\) Xi:WT Xi ratio of 1:1, the blue lines correspond to a ratio of 1:2. (G) CDF of the fold change in gene expression between WT and Dxz4\(^{\Delta/\Delta}\).FirreXi\(^{Xi/+}\) for autosomal genes (teal) and X-linked genes with fpm >1 in all experiments. The Kolmogorov–Smirnov p-value is indicated. (H) Overlap between escapees in WT and Dxz4\(^{\Delta/\Delta}\).FirreXi\(^{Xi/+}\). (i) Density plots of the number of genes with a given level of expression from the Xi in two wild-type and two Dxz4\(^{\Delta/\Delta}\).FirreXi\(^{Xi/+}\) replicate RNA-seq experiments. Grey and black: wild-type, purple: Dxz4\(^{\Delta/\Delta}\).FirreXi\(^{Xi/+}\).
Figure 3.8 (Continued)
DISCUSSION

An outstanding question in genome and nuclear organization is how higher-order chromatin structure regulates gene expression. Here, using the Xi as a model, we have tested the relevance of two higher order structures — superloops and megadomains — for the biology of XCI. While indeed *Dxz4* is required to form megadomains and *Firre* is required for superloops and for full strength of megadomains, we unexpectedly observed that abolishing these structures had no impact whatsoever when assaying a range of XCI phenotypes, including (i) chromosome-wide silencing as determined by RNA-seq, (ii) ability of escapees to avoid inactivation, (iii) subnuclear localization of the Xi, (iv) enrichment of H3K27me3 mark along the Xi, and (v) general chromatin accessibility as measured by ATAC-seq. By analyzing the time course of megadomain and superloop formation, we determined that these structures do not precede Xist spreading or XCI. Instead, they either occur concurrently with Xist spreading and XCI, or they may be a consequence thereof. We also observe that *Dxz4* and *Firre* may work together to strengthen megadomains through superloop formation: Deleting *Firre* weakens intra-megadomain interactions without affecting the strong *Dxz4* border, deleting *Dxz4* abolishes the sharp megadomain border, and deleting both loci has a more severe effect on overall megadomain organization than deleting either singly. The significance of attenuated megadomains is unclear, however, given that there was no perturbation to XCI when either *Dxz4* or *Firre* or both were deleted. Taken together, our data argue that the superstructures are not necessary for XCI biology, at least in the ex vivo cellular context.

Our findings therefore beg a number of interesting questions. First, what is the purpose of *Dxz4*, *Firre*, megadomains, and superloops, and why are they conserved across 80 million years of mammalian radiation? While our observation that *Dxz4* is required for megadomains on the Xi is in agreement with three other studies [199, 252, 327], our results are at odds with the previous proposal that *Dxz4* enables genes to escape XCI [199]. Our study also finds no loss of
accessibility on ~35 escapee genes when $Dxz4$ is deleted on the Xi. The different conclusions may result from either use of different cell types (mESC versus NPCs) or clonal variation. Notably, the previous study observed the effects only in one of four NPC clones and the clone showed an unusually high number of escapees — ~100 escapees, a number that is 2-4 times greater than reported by any other study [185, 186, 188]. If this NPC clone were an oddity, comparing its expression state to a $Dxz4$-deleted cell line of a different NPC background would lead to the impression that the unusual escapees (which ordinarily do not escape XCI) had become silenced.

Additionally, our data suggest that loss of insulation and formation of the megadomains does not occur before Xist spreading and XCI are completed, and thereby providing further evidence for a decoupling of megadomain structures from epigenetic silencing on the Xi. Indeed, our results are in line with other deletions of $Dxz4$, all of which failed to find an effect on gene silencing after deletion of the megadomains [199, 327]. Whatever function $Dxz4$ might serve, our study indicates that it is necessary but not sufficient for megadomain formation in post-XCI cells, even when $Xist$ is present ectopically together with $Dxz4$. Because we could not derive the transgenic line in a female ES cell background, we do not formally know whether $Dxz4$ and $Xist$ together might be sufficient during de novo XCI in an ectopic context. However, given that $Dxz4$ has no impact on XCI and escape in any measurable way, the sufficiency during the XCI establishment phase seems moot.

$Firre$ was also of interest. What is its function and why do superloops form on the Xi via this CTCF-enriched repeat? Although we failed to find an XCI-related function for $Firre$ after ablating it on the Xi, we found that $Firre$, $Dxz4$, $Xist$, and $X75$ form a conserved network of superloops on the mammalian Xi, in agreement with a previous study [252]. Yet, superloops are dispensable for establishing XCI, since both superloop anchors $Dxz4$ and $Firre$ can be deleted on the future Xi with no impact on XCI establishment. Perhaps Firre RNA has a role in X-
inactivation, as suggested from Firre knockdown experiments implicating Firre RNA in maintaining perinucleolar localization of the of the Xi and enrichment of H3K27me3 [188]. We leveraged new Xa- and Xi-specific Firre deletions to test which allele expresses Firre RNA and find distinct isoforms associated with the Xa versus Xi. Quantitative RT-PCR suggests that transcription from the Xa may predominate, though we cannot be certain without knowing all potential isoforms. Deleting Firre on either Xi or Xa does not perturb Xi localization, Xist expression or H3K27me3 deposition on the Xi. Thus, neither the Firre superloop anchor nor any transcript produced by Firre is needed for XCI.

Could megadomains and superloops be default consequences of Xist-mediated attenuation of TADs and compartments? It is important to note that our study follows XCI only in the ex vivo cellular context. It is possible that Dxz4, Firre, megadomains, and superloops play an important role in long-term maintenance of the Xi and that this role would only be revealed by following mice over their lifespan. It is also possible that these Xi megadomains and superloops are incidental organizational structures with no primary impact on gene regulation. While these structures do not disrupt XCI, other macrostructures do. In particular, the recently identified S1/S2 compartments that are revealed by loss of SMCHD1 function play an essential role during de novo Xi silencing [248]. Why the Xi would be folded in these ways with or without function is unclear. For Dxz4 and Firre superloops, a role in a non-XCI pathway — critical in a whole-organism context and not measurable by our present assays — must also be entertained. Irrespective of function, the megadomains and superloops represent the largest architectural structures identified by Hi-C to date in mammals, and both are clearly unique to the Xi. Their evolutionary conservation across 80 million years suggests that the superstructures likely persist for reasons that will only become clear with further study.
MATERIALS AND METHODS

Cell lines and growth conditions

ES cells were grown in regular ES+LIF medium (500 ml DMEM with the addition of 1 ml of β-mercaptoethanol, 6 ml of MEM NEAA, 25 ml of 7.5% NaHCO₃, 6 ml of GlutaMAX-1, 15 ml of 1M HEPES, 90 ml of FBS, 300 µl of LIF, 6 ml of PEN/STREP) on irradiated feeders. To differentiate mESCs and allow them to undergo X-inactivation, mESCs were harvested by trypsinization and quenched in ES medium without LIF. Feeders were removed by adding the cell suspension to tissue culture plates for 45 minutes at 37°C. Differentiating embryoid bodies were cultured for 4 days on low-adherence plates in ES medium without LIF. On the 4th day, the embryoid bodies were plated onto gelatinized tissue culture plates and allowed to attach. ES cells for experiments were harvested by extensive trypsinization to detach them from the plates. Unless otherwise noted, all experiments were performed after 10 days of differentiation.

Fibroblasts were grown on un-gelatinized tissue culture plates in MEF media (500 ml DMEM with the addition of 1 ml of β-mercaptoethanol, 6 ml of MEM NEAA, 25 ml of 7.5% NaHCO₃, 6 ml of GlutaMAX-1, 15 ml of 1M HEPES, 60 ml of FBS, 300 µl of LIF, 6 ml of PEN/STREP).

Generation of Dxz4 and Firre deletion cell lines

Oligos encoding gRNAs flanking either Dxz4 or Firre were cloned into wild-type Cas9+GFP plasmid PX458 [340] by first linearizing the plasmid with BbsI, purifying the plasmid using the Qiagen PCR Purification Kit, then ligating annealed and phosphorylated oligos into the plasmid using T4 DNA ligase for 1 hour at room temperature, then transforming into OneShot Top10 chemically competent E. coli. gRNA plasmid DNA was prepared using the Qiagen Miniprep Kit and gRNA sequences were verified by Sanger Sequencing.
To delete *Dxz4* or *Firre*, pairs of gRNAs flanking either *Dxz4* or *Firre* were transfected into mESCs using Lipofectamine LTX. Briefly, for each transfection, 50 uL of OptiMEM media was added to 2.5 uL LTX reagent and 50 uL OptiMEM was added to 0.5 uL PLUS reagent. The OptiMEM+PLUS mix was added to a mixture of 250 ng of each gRNA, then the OptiMEM+LTX mix was added and incubated at room temperature for 5 minutes to generate the transfection mixture. Meanwhile, 2x10^5 mESCs were harvested by trypsinization and brought to a volume of 900 uL ES+LIF media. Once the transfection mixture was ready, the mESCs were layered dropwise on top of it and allowed to incubate for 20 minutes at room temperature. Following incubation, the entire transfection mixture was added to one well of a 12-well dish containing feeders and 1 mL ES+LIF media. The transfected cells were allowed to grow for 16-48 hours.

To screen for Cas9-transfected cells, the transfected cells were harvested with trypsin, washed 2X in PBS and resuspended in 300 uL FACS media (1X Leibowitz’s+5% FBS) and passed through a cell strainer. The GFP-positive cells were isolated by FACS selection and plated on 10 cm feeder plates (~2000-10,000 GFP+ cells/plate). The FACS-sorted GFP positive cells were allowed to grow into large colonies, typically after about 6-8 days of growth. 192 colonies were manually picked and transferred to 96 well plates covered in feeders. Once the 96 well plates were nearly confluent, they were passaged onto 3 new gelatinized plates (no feeders). Freezing media (MEF media+ final concentration 10% DMSO) was added to two plates and they were left at -80°C for storage. The third plate was grown until most wells were fully confluent.

We used a PCR screen to identify *Dxz4* or *Firre* deletion clones. Genomic DNA was prepared from the colonies by incubating them overnight in Laird buffer+proteinase K (50 uL buffer per well) at 55°C. The genomic DNA was transferred to a new 96 well plate and diluted it 1:10 in H2O, then heated at 95°C for 10 minutes to denature it and inactivate the proteinase K. Next, PCR reactions using primers flanking *Dxz4* or *Firre* were prepared in 96 well plates using 20 uL PCR mix+2 uL denature genomic DNA. 40 cycles of amplification were used, and the
PCR reactions were run on 2% agarose gels and visualized by ethidium bromide staining. Deletion clones were identified by PCR reactions that produced a band at the expected size. Deletion clones were thawed onto 12-well plates with feeders, and deletions were verified by Sanger sequencing the PCR product, performing DNA FISH using a fosmid probe entirely within the deleted region, and examining reads over the deleted regions from our genomics experiments.

To generate Xa-specific and homozygous Firre deletions, we employed a restriction assay to determine whether clones carried a deletion on the Xa or Xi (or both). We took advantage of a cas- (Xa-) specific polymorphism that creates a new TaqI restriction site within the Firre deletion PCR product to screen clones for deletions on particular alleles. We performed PCR amplification as before, but then added 30 uL of 1X Cutsmart buffer+10U TaqI (NEB) to each PCR reaction, then incubated the reactions at 65°C for 45 minutes before running the reactions on a 2% agarose gel.

Preparation of high molecular weight DNA

Briefly, 500 mL cultures of E. coli containing either Xist+P or RP23-161K4 were grown and spun at 4000 rpm for 15 minutes. Alkaline lysis was performed by re-suspending in 20 mL Buffer 1, aliquoting the cell suspension into two Oak Ridge polypropylene centrifuge tubes, adding 10 mL of Buffer 2 to each tube and inverting 20 times to mix, then adding 12 mL buffer 3 and inverting 20 times and incubating on ice for 5-10 minutes. Protein and genomic contaminants were removed by centrifugation at 10,000 rpm in a JA-20 rotor at 4°C. DNA was precipitated by adding 35 mL isopropanol to 15 mL centrifuged lysate in a 50 mL Falcon tube, incubating 20 minutes at room temperature, then spinning at 3500 rcf for 20 minutes at 4°C. Pellets were resuspended in 500 uL TE+1%SDS and 15 uL 20 mg/mL Proteinase K was added and the DNA mixture was incubated at 55°C for 1.5 hours to remove protein contaminants. The DNA was phenol:chloroform extracted by adding phenol:chloroform:isoamyl alcohol and shaking by hand.
for 20 seconds, then the DNA was precipitated with 40 uL 3M NaOAc and 1 mL isopropanol per 500 uL DNA mixture for 10 minutes at -20°C. At this time, the precipitated DNA formed a stringy white mass, the excess liquid was removed from this mass and 1 mL 70% ethanol was added to the DNA. The DNA was centrifuged for 5 minutes at 16300 g, the supernatant removed, and 1 mL 70% ethanol was added to the pellet and the pellet was spun again for 5 minutes at 16300 g. Supernatant was removed and excess ethanol was allowed to evaporate for 5 minutes, then the DNA pellet was resuspended in 200 uL 10 mM Tris by gentle pipetting with a cut tip.

Generation of a Xist+Dxz4 transgene

To generate an autosomal Xist+Dxz4 transgene, we co-transfected a doxycycline-inducible Xist construct and a BAC containing mouse Dxz4 into male fibroblasts containing rtTA [62]. First, we prepared DNA from our “Xist+P” construct and the Dxz4-containing BAC RP23-161K4 using a custom high-molecular weight purification protocol. DNA for transfection was only used if it gave the expected digest pattern with either XhoI+BamHI for Xist+P of XhoI for RP23-161K4 and was not excessively smeared. To co-transfect Xist+P and RP23-161K4 into fibroblasts, 2x10⁶ fibroblasts were harvested, washed twice in 1X PBS and resuspended in 700 uL 1X PBS. We then added 5 ug Xist+P and 20 ug RP23-161K4 to the cell suspension and electroporated in a 1 mm cuvette at 200V, 1050 uF using a Bio-RAD Xcell GenePulser electroporation system. Electroporated cells were plated onto 3 10 cm dishes in MEF media made with tet-free FBS and grown for one day. The Xist+P construct contains a hygromycin selectable marker, and to select for Xist transgenes, starting one day after transfection, we added 200 ug/mL hygromycin to the media and changed the media every day for 10 days. Once colonies were grown, we manually picked them and transferred them to 96-well plate. We only obtained about 10 hygromycin resistant colonies. Once confluent, we split colonies onto 3 wells of a 24-well plate. One well was kept for maintenance, the other two were used for screening.
To screen for transgenic lines with both inducible \textit{Xist} and \textit{Dxz4} inserted at the same ectopic site, we used the following strategy. We induced each clone with 1000 \text{ug/mL} doxycycline overnight and performed \textit{Xist} RNA FISH to test if \textit{Xist} could be induced. We also performed \textit{Xist} RNA FISH in the same clones without dox induction to ensure \textit{Xist} expression is inducible. We kept clones that could induce robust \textit{Xist} RNA FISH clouds. We then used DNA FISH to check whether the \textit{Xist}+P construct inserted at the same site as the \textit{Dxz4}-containing BAC. We simultaneously performed DNA FISH using an \textit{Xist} probe, a probe within the \textit{Xist} construct backbone, and a fosmid probe against \textit{Dxz4}. We obtained one clone where all 3 probes co-localize at one spot, indicating co-insertion of \textit{Xist} and \textit{Dxz4} into an autosome. We then used 4C to localize the candidate insertion site into \textit{Stc1} on chr14. We then performed DNA FISH using a fosmid probe overlapping \textit{Stc1} combined with a \textit{Dxz4} fosmid and a probe overlapping the backbone of the \textit{Xist} transgenic construct to confirm co-localization of \textit{Xist}, \textit{Dxz4} and \textit{Stc1} at one spot.

\textit{DNA FISH}

BAC or fosmid DNA was prepared using the high molecular weight DNA preparation procedure. Probes were labeled using the Roche Nick Translation kit. 75,000-150,000 cells were cytospun onto slides for 5 minutes at 1000 rpm. Cells were pre-extracted and fixed by passing the slides through CSK-T for 3 minutes at 4°C, CSK for 3 minutes at 4°C, 1X PBS+4% formaldehyde for 10 minutes at room temperature. RNA was removed by digestion with 0.1 mg/mL RnaseA in 1X PBS for 1 hr at 37 degrees. Slides were dehydrated by passage through 70%, 90%, 100% ethanol for two minutes at each concentration, then allowed to dry. Probe was added to hybridization mix (50% formamide, 2X SSC, 10% dextran sulfate, 0.1 mg/mL mouse Cot-1 DNA) and added directly to the slides. Slides were denatured at 92°C for 10 minutes on a PCR block, then incubated in a humid chamber at 37°C overnight. Slides were washed once in 2X
SSC, once in 2X SSC+Hoechst 33342 and once in 2X SSC. Mounting media was added and the slides were imaged.

**RNA FISH**

Slides were prepared for RNA FISH using the same protocol as for DNA FISH but with the RnaseA treatment omitted. Xist RNA FISH was performed using a mixture of Cy3-labeled DNA oligos covering Repeats A, B and C within Xist. The RNA FISH protocol was the same as the DNA FISH protocol, except that the denaturing step was omitted and the hybridization buffer+probe mixture was heated at 92°C for 5 minutes then 37°C for 5 minutes and then added directly to the slides. Slides were incubated at 42°C for 4-8 hours and then were washed once in 2X SSC, once in 2X SSC+Hoechst 33342 and once in 2X SSC. Mounting media was added and the slides were imaged.

**Immunofluorescence**

75,000-150,000 cells were cytospun onto slides for 5 minutes at 1000 rpm. Slides were washed once with 1X PBS, then 1X PBS+4% formaldehyde was added for 10 minutes at room temperature, then 1X PBS+0.5% Triton-X 100 for 10 minutes at room temperature to remove un-crosslinked proteins. Slides were washed once in 1X PBS, excess buffer was removed from cell spots and 1% BSA in 1X PBS was added for 45 minutes. Block solution was removed and a 1:200 dilution of H3K27me3 antibody (Active Motif 39155) in 1X PBS+1% BSA was added for 1 hour. Slides were washed 3X in 1X PBS+0.02% Tween-20. Excess liquid was removed and a 1:2000 dilution of goat-Anti-Rabbit Alexa 555 conjugated antibody (Thermo Fisher) was added for 1 hour in the dark. Slides were washed once in 1X PBS+0.02% Tween-20, then twice in 1X PBS and then imaged.

**ImmuoFISH**
Slides were prepared the same way as for immunofluorescence; with 0.5 U/uL Protector RNase Inhibitor (Sigma) added to the blocking buffer. To visualize the nucleolus, we used a 1:200 dilution of Nucleophosmin antibody (abcam 10530) in blocking solution as the primary antibody. After immunofluorescence, we post-fixed the slides for 10 minutes in 4% formaldehyde+PBS, and then Xist RNA FISH was performed starting at the dehydration step.

**Metaphase immunofluorescence**

We added 50 ng/mL Karyomax to the media of day 10 differentiating embryoid bodies for 4 hours to arrest cells in metaphase. We harvested the cells via trypsinization, and trypsin was quenched by addition of media. We spun the cells at 1000 rpm for 5 minutes, aspirated the media, then washed twice in 1X PBS. Cells were then resuspended to a concentration of 5x10^5 cell/mL in 75 mM KCl and placed at 37°C for 10 minutes for swelling. 1x10^5 cells were then cytospun onto a microscope slide at 1000 rpm for 5 minutes. The cells were fixed in PFA and immunofluorescence was performed as described for interphase cells. We stained H3K27me3 with a 1:200 dilution of Active Motif 39535 and H3K27ac with a 1:200 dilution of Cell Signaling D5E4.

**Hi-C library preparation**

We used the *in situ* Hi-C method of Rao et al. [220] to prepare all libraries, using 5-10 million cells. Importantly, we sequenced 20-40 million reads per library. This is a lower sequencing depth than many published Hi-Cs, however since the megadomains are large and prominent feature of the organization of the Xi, this depth is appropriate for detecting the megadomains efficiently and economically. We performed a timecourse of Hi-C experiments at 4 timepoints during differentiation (days 0, 3, 7 & 10). To test whether megadomains form in the absence of *Dxz4* or *Firre*, we differentiated cells for 10 days and performed Hi-C in wild-type, *Dxz4^+/+*, *Firre^+/+*, or *Dxz4^+/+;Firre^+/+*. Finally, to test whether megadomains can form on an autosome
with Xist and Dxz4 ectopically inserted, we performed Hi-C in the Xist+Dxz4 transgene line after 2 days of induction with 1000 ng/mL dox, as well as an Xist+P only transgene line after 2 days of induction with 1000 ng/mL dox and the parental male XY rtTA line (no induction).

**Hi-C analysis**

Hi-C alignment to mm9 was performed according to the method of Minajigi & Froberg et al. [192] The allele-specific Hi-C reads were filtered for quality and uniqueness with HOMER. Custom scripts were used to convert HOMER tag directories into the format accepted by Juicebox; contact maps were generated using the Juicer tools ‘pre’ command. All Hi-C contact maps visualized in this study are KR-normalized contact maps generated by Juicebox.

The first principal component of the Hi-C correlation matrix has been used as a quantitative measure of the presence or absence of megadomains[259]. We used R to generate the Pearson correlation of 1Mb KR-normalized allele-specific chrX Hi-C matrices, and we plot the first principal component as a function of position along the X-chromosome. Hi-C matrices with a megadomain exhibit a sharp transition in the first principal component score at the bin containing Dxz4.

**Hi-C mixing experiment**

We mixed together aligned reads from the day 0 and day 10 Hi-C libraries such that 0%, 10%, 25%, 50%, 75% and 100% of reads were from the day 10 Hi-C. We the generated HOMER tag directories and normalized contact maps in Juicebox as described for the Hi-C experiments. We plotted PC1 scores across the Xi at 1Mb resolution and defined the PC1 slope at Dxz4 as the PC1 score @ bin 74 – PC1 score @ bin 72.

**HYbrid Capture Hi-C (Hi-C²)**
HYbrid Capture Hi-C (Hi-C$^2$) probes were designed and hybridization to in-situ Hi-C libraries carried out as described previously [218]. Probe sets were designed to enrich interactions in two regions of interest: chrX:70,370,161-71,832,975 and chrX:71,832,976-73,511,687 (mm9). Briefly, 120 bp probes were designed around the MboI restriction sites of the regions of interest as previously described [218] 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 [218]. 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 Hi-C summary files and filtered using HOMER as previously described [192]. For the chrX:70,370,161-71,832,975 captures, 3-4% of mapped and paired reads fell within the target region (0.05% expected based on size of capture region versus genome) and for the chrX:71,832,976-73,511,687 captures, 1-2% of mapped and paired reads fell within the target region (0.06% expected based on size of capture region versus genome). 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, a custom script 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 [341]. The resulting Hi-C contact maps in .hic format were visualized and normalized with the ‘Coverage (Sqrt)’ option in Juicebox [342].

*Explanation of insulation score analysis*

Insulation scoring has become a standard in the Hi-C field for identifying borders and quantifying boundary strength [223, 332, 333]. In brief, insulation scores quantify how strongly a given locus acts as a border [223, 332, 333], and are calculated by running sliding windows across a chromosomal region and measuring changes in interaction frequencies between successive windows. We show an explanatory diagram in Supplementary Fig. S3.3D. The main idea is that the raw insulation score at a particular locus is the ratio of the number of reads that “cross over” that locus (where one end falls to the left and one end falls to the right of the locus) to the number of reads nearby that do not cross over the locus (reads where both ends fall to the right or where both fall to the left of the locus). Once raw insulation scores are calculated for all loci in a region, the insulation score for a given locus is the log-ratio of the raw insulation score at the given locus divided by the mean insulation score for all loci across the region. At loci in the middle of domains, there are a similar number of reads that cross over the locus to the number of reads right beside the locus but that don’t cross over, thus insulation scores in the middle of domains tend to be near zero or perhaps slightly positive. However, at the borders (“insulators”) of domains, there are few reads that cross over the border relative to the reads to the left or to the right in the two domains separated by the border. This means insulation scores are strongly negative at domain borders, and in fact domain borders are defined as the local minima of insulation scores across a region.

*Insulation score analysis with Hi-C² data*

We computed insulation score across the *Mecp2* and *Dxz4* regions to quantitatively measure changes in domain organization during the timecourse of X-inactivation. To do this, we output
the ‘Coverage (Sqrt)’ normalized Hi-C contact maps at 25 kb resolution across either the Mecp2 or Dxz4 regions using Juicer tools ‘dump’ command. We used custom shell and R scripts to convert the densematrix format output from Juicer into the full matrix format accepted by the cworld suite of Hi-C tools (https://github.com/dekkerlab/cworld-dekker). We computed insulation scores across the captured regions using the cworld perl script ‘matrix2insulation.pl’ using the parameters ‘-v --is 125000 --ids 75000 –im sum’. This set of options uses a smaller number of bins to calculate insulation scores, which we found to be optimal for analyzing insulation over small regions with just a few dozen bins. We plotted the distribution of insulation scores across each region and each timepoint. We evaluated changes in insulation across regions by testing whether there was a difference in the variance of insulation scores between timepoints or between the Xa and the Xi using the F-test. This is appropriate as a loss of insulation by definition is a decrease in the variance of insulation across a region [334], which can be visualized as a “flatter” insulation score curve. To generate violin plots and calculate F-test p-values, we excluded the 6 bins on the left and right edges of each Hi-C^2 region because the windows used to calculate insulation score at these loci fall partially outside the region covered by Hi-C^2 probes and have far less read coverage than the regions covered by the probes.

**4C library preparation and analysis**

We previously developed a modified 4C protocol[124] to examine chromatin conformation from repetitive viewpoints. Our protocol has several advantages over existing 4C profiles: 1.) It sequences the genomic region amplified by the 4C primers, ensuring that on-target priming events can be identified and filtered from numerous off-target priming events 2.) Sequencing the viewpoint allows every read to be assigned to a particular allele if the viewpoint is near a variant, 3.) We use a random barcode to identify PCR duplicates, which previously has not been possible in 4C experiments. We performed our modified 4C using the protocol and analysis pipeline previously described for viewpoints within PAR-TERRA repeats[124]. We used it for
several viewpoints within *Firre* and *Dxz4*. Some viewpoints were in the core tandem repeats. For these viewpoints, we use the read outside the viewpoint for allelic determination. Others were in unique regions near the tandem repeats; for these we could use known variants to assign every read to the Xa or the Xi. We performed our analysis in two fibroblast lines, one where the mus X is inactive (mus Xi cas Xa), the other where the cas X is inactive (mus Xa cas Xi).

**Assay for Transposase-Accessible Chromatin with high-throughput sequencing**

50,000 cells were washed in cold PBS and lysed in cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630) containing proteinase inhibitor cocktail (Roche). Nuclei were resuspended in 1X TD Buffer (Illumina FC-121-1030) and 2.5uL of Tn5 Transposase (Illumina FC-121-1030) were added. Transposition reaction was performed at 37°C for 30min, and DNA was purified using a Qiagen MinElute Kit. DNA libraries were amplified for a total of 8 cycles. Libraries were assessed for quality control on the BioAnalyzer 2100 (Aglient) to ensure nucleosomal phasing and complexity. Sequencing was performed on the HiSeq 2500 (Illumina), using 50 bp paired-end reads.

**ATAC-seq analysis**

Attack seq alignment to mm9 was performed exactly as ChIP-seq alignment was performed in Minajigi & Froberg et al. [192]. Peaks were called using macs2 with default parameters. Biallelic peaks were identified as peaks with at least 10 allelic reads in a sample and an Xi:Xa ratio greater than 1/3. Xi-specific peaks were defined as peaks with at least 10 allelic reads and an Xi:Xa ratio less than 1/3. To test whether Xi-specific peaks in wild-type are “restored” (that is: acquire appreciable accessibility on the Xi) in either the *Dxz4* or *Firre* deletion, we plot the wild-
type Xa reads on the x-axis and the deletion Xi reads on the y-axis and identify peaks where the deletion Xi/wild-type Xa ratio is greater than ½ (these are peaks where the deletion accessibility level reaches at least half the wild-type accessibility ratio). We also examine the biallelic peaks and plot the wild-type Xi reads on the x-axis and the deletion Xi reads on the y-axis to determine whether the accessibility on the Xi changes for the peaks that are bi-allelic in wild-type.

**RNA-seq library preparation**

Total RNA was isolated from 2-5 million trypsinized cells using Trizol extraction. polyA+ mRNA was isolated using the NEBNext® Poly(A) mRNA Magnetic Isolation Module using 5 ug of total RNA as input. Isolated mRNA was reverse-transcribed using Superscript III and actinomycin D to inhibit template switching. Second-strand synthesis was performed using the NEBNext Ultra Directional RNA Second Strand Synthesis Module. Library preparation and NEBNext® ChIP-Seq Library Prep Master Mix Set for Illumina. A USER enzyme treatment was performed following adaptor ligation to specifically degrade the second strand and allow a stranded analysis. Libraries were amplified for 10-15 cycles of PCR using Q5 polymerase and NEBNext multiplex oligos.

**RNA-seq analysis**

RNA-seq reads were aligned to the cas (Xa) and mus (Xi) genomes allele-specifically using a previously published pipeline [10, 36, 72, 192]. Following alignment, gene expression levels for each gene were defined using HOMER. Differential expression and fold changes between conditions were calculated using DESeq2. We plotted the cumulative distributions of fold changes for autosomal and X-linked genes and evaluated the significance of any differences between the distributions of the fold changes using the Kolmogorov-Smirnov (KS) test. To examine allele-specific expression from the Xa and the Xi, we summed together allelic reads across both biological replicates and filtered for genes with at least 12 allelic reads in both wild-
type and $Dxz^4_{1\alpha}:Firre^{X_{1\alpha}}$ and fpm > 0 in all replicates. We also used RNA-seq done in pure hybrid mus or cas fibroblasts to identify and eliminate genes that have incorrect SNP information. We defined escapee genes in a particular condition as genes where at least 10% of allelic reads came from the Xi in either replicate of that condition. We plotted the distribution of expression levels from the Xi $(X_i/(X_i+X_a)$ read counts) for all genes passing our filtered for each replicate. We evaluated the significance in differences of the mean expression level from the Xi using the Wilcoxon Signed Rank Test with Bonferroni correction for multiple hypothesis testing.

**qRT-PCR**

Total RNA was isolated from cells using Trizol extraction. 500 ng RNA was heated at 70 degrees C for 10 minutes then cooled to 4 degrees in the presence of 50 ng random primers in 5 uL total volume. The RNA was reverse-transcribed in a 10 uL reaction containing 1X First Strand buffer, 10 mM DTT, 500 uM dNTPs, 6U Protector RNase inhibitor and 100U Superscript III. The reaction was incubated for 5 minutes at 25 degrees, then 1 hr at 50 degrees and 15 minutes at 85 degrees. Reverse transcription reactions were diluted to 100 uL with water before qPCR. 500 ng RNA was added to 100 uL water as a -RT control. 1 uL template was used per 15 uL qPCR reaction prepared with 1X Taq UniverSYBR Green (Bio-Rad) master mix and 200 nM primers, and reactions were performed in triplicate. All qPCR primers were run using an annealing temperature of 55 degrees. The primers used for qRT-PCR are listed below:

<table>
<thead>
<tr>
<th>Name</th>
<th>Amplicon</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>JR_Firre1_F</td>
<td>JR1</td>
<td>GGAGAAAGGCAGAAATGCAG</td>
</tr>
<tr>
<td>JR_Firre1_R</td>
<td>JR1</td>
<td>CAGTGTTCCAGCTCCAGTGA</td>
</tr>
<tr>
<td>JR_Firre2_F</td>
<td>JR2</td>
<td>AGGTATGCTTACCTCTCCT</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Gene</td>
<td>Sequence</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>JR_Firre2_R</td>
<td>JR2</td>
<td>CAAATTCAAGCAGGCAAGGG</td>
</tr>
<tr>
<td>JR_Firre4_F</td>
<td>JR4</td>
<td>TTTTTTCATGCAGGGTGATTG</td>
</tr>
<tr>
<td>JR_Firre4_R</td>
<td>JR4</td>
<td>AACAGTGCCCATTTGAGCTCC</td>
</tr>
<tr>
<td>Disteche_Firre_F</td>
<td>CD</td>
<td>ACCAGGTACCGTGAGCAATC</td>
</tr>
<tr>
<td>Disteche_Firre_R</td>
<td>CD</td>
<td>TTCCTCATCCCCTTCTCCT</td>
</tr>
<tr>
<td>Firre_int_set1_F</td>
<td>Firre_intronic</td>
<td>CCTGCCTACACATGCTACAA</td>
</tr>
<tr>
<td>Firre_int_set1_R</td>
<td>Firre_intronic</td>
<td>CAGGTCTTTGGGTCTTCTATC</td>
</tr>
<tr>
<td>Xist_ex1-3_F</td>
<td>Xist_exon1-3</td>
<td>GCTGGTTCGTCTATCTTGAGGG</td>
</tr>
<tr>
<td>Xist_ex1-3_R</td>
<td>Xist_exon1-3</td>
<td>CAGAGTAGCGAGGACTTGAGAG</td>
</tr>
<tr>
<td>Gapdh_F</td>
<td>Gapdh</td>
<td>ATGAATACGGCTACGCAACAGG</td>
</tr>
<tr>
<td>Gapdh_R</td>
<td>Gapdh</td>
<td>GAGATGCTCAGGTGTGGGGGG</td>
</tr>
</tbody>
</table>

**Code availability**

The custom analysis pipelines for all genomic analyses are available upon request with no restrictions.

**Data availability**: All sequencing data that support the findings of this study have been deposited in the National Center for Biotechnology Information GEO repository under accession GSE116649.
Chapter 4: Conclusions, Perspective and Future Directions
In the preceding chapters, I have provided an overview of X-inactivation including what was known regarding Xist-interacting proteins circa 2012-13, an analysis of the conformation of the two X-chromosomes and the role of Xist in establishing higher-order chromatin structures and tested whether the unique “megadomain and superloop” organization of the Xi is required for XCI initiation. In Chapter 2, I found that unlike all other mammalian chromosomes, the Xi largely lacks both TADs and the architectural proteins CTCF and cohesin believed to fold chromosomes into TADs. Deletion of Xist partially restores both cohesin binding to the Xi and TADs across several multi-megabase regions of the Xi, and unbiased identification of Xist-interacting proteins through the novel iDRiP method showed that Xist RNA binds to both CTCF and cohesin proteins. These observations suggest a model where Xist RNA binds to architectural proteins and prevents them from binding to chromatin, thereby breaking down TAD structures on the Xi [192].

In addition to the diminution of TADs on the Xi, allele-specific Hi-C showed that the Xi folds into two giant megadomains (chapter 2 and 3, [192, 199, 220, 247, 252]) . An unusual tandem repeat that binds CTCF and cohesins on the Xi lies at the border of the megadomains and forms a 25 Mb loop with another tandem repeat called Firre. To test the role of the megadomains and superloops in XCI, I deleted both Dxz4 and Firre in mESCs and tested whether the megadomains can form and XCI can initiate in the absence of these tandem repeats (Chapter 3). Deletion of Dxz4 completely eliminates the megadomains, and Firre deletion appears to weaken interactions within the megadomains. However, neither Dxz4 nor Firre are required for XCI initiation, suggesting that the unique “megadomain and superloop” superstructure of the Xi is not needed for XCI.

Together, my results provide new insight into the higher order structure of the Xi and its impact on XCI. In this chapter, I review these results, provide perspective on how they fit into our understanding of the interplay between higher-order chromatin structure and gene regulation, and emphasize suggestions for future research directions.
Disruption of TADs by Xist: direct or indirect effect?

The Xi is largely devoid of TADs, unlike all other mammalian chromosomes. The Xi is also strongly depleted in CTCF and cohesin, the architectural proteins necessary for TAD formation. The absence of CTCF and cohesin binding on the Xi provides a mechanistic explanation for why TADs are dramatically weakened on the Xi relative to the Xa. It would be useful to test this explanation. One possible avenue would be to disrupt WAPL, the protein that removes cohesin from the chromatin arms during prophase [343, 344] and acts to restrict cohesin binding to chromatin during interphase [344, 345]. WAPL depletion strengthens cohesin binding to interphase chromatin [345] and causes longer loops to form [240, 346], thus if the lack of cohesin is what underlies the lack of TADs on the Xi, the TADs should partially re-emerge in WAPL depleted cells. Another possibility would be to overexpress cohesin loader NIPBL, as this should also increase cohesin binding on the Xi [346].

It would also be very interesting to more directly test whether the interaction between Xist RNA and cohesin proteins uncovered by iDRiP and confirmed by UV-RIP is directly responsible for the loss of cohesin binding to the Xi and loss of TADs. The model proposed in Chapter 2 suggests that Xist RNA directly binds to any cohesins that diffuse into the Xi territory and prevents them from binding to chromatin [192], explaining the observed depletion of cohesins on the Xi chromatin. Identifying the domain of Xist RNA that binds cohesin (and perhaps the domain of the cohesin proteins that binds Xist RNA) and specifically perturbing the interaction between Xist RNA and cohesin would more directly test the model presented in Chapter 2. Detailed characterization of the Xist RNA-cohesin interaction would be quite valuable. Of the three Xist proteomics methods, only iDRiP identified cohesin proteins as Xist interactors [190-192]. Little is known about cohesin-RNA interactions and their biological significance. Interestingly, CTCF, cohesin’s partner in forming loops and domains, has been shown to bind many RNAs [72] and Jpx RNA has been shown to titrate CTCF away from the Xist promoter to favor Xist upregulation [73]. Exploring the RNAs that bind architectural proteins
and their biological functions may provide new mechanisms and insight into control of chromatin organization.

There are several other aspects of regulation of Xi chromatin structure by Xist that are worth further study. The Xist deletion experiments in Chapter 2 show that removal of Xist allows partial restoration of TADs and cohesin binding, demonstrating that Xist RNA antagonizes cohesin binding to chromatin. However, the deletion experiments do not shed light on whether Xist directly binds to cohesin to remove it from chromatin, or whether some other aspect of the Xi that depends upon Xist indirectly restricts cohesin binding. There are several other testable indirect models for how Xist could disrupt cohesin binding on the Xi. iDRiP showed that cohesin remover WAPL is also an Xist interactor \[192\]; perhaps Xist recruits WAPL to remove cohesin.

The Xi is also depleted of accessible chromatin (Chapter 3, \[199, 327\]), and cohesin loader NIPBL localizes to accessible chromatin especially promoters of active genes \[240, 295, 347\]. The general lack of accessible sites on the Xi may simply prevent cohesin from being loaded in the first place, and removal of Xist may restore accessibility at certain sites on the Xi. In fact, the clusters of restored TADs and cohesin sites in the absence of Xist generally contain at least one escapee gene, though they can extend for several megabases away from any escapee. Since the escapee genes are accessible on the Xi (Chapter 3, \[199, 327\]), the absence of Xist may permit cohesin loading at the escapees, which then extrude loops out from the escapees and bring together other potentially accessible sites where more cohesin is loaded, creating a positive feedback of accessibility, cohesin loading and domain formation extending for several hundred kilobases. This model can be tested by deletion of the escapee prior to Xist deletion; this should block cohesin loading and may prevent restoration of cohesin binding within several hundred kb. Cohesin-independent models may also be at work. CTCF is depleted from the Xi \[37, 185, 192\], though to a lesser extent then cohesin (Chapter 2, \[192\]). It is possible that the depletion of CTCF underlies loss of TADs, perhaps due to the fact that the Xi is hypermethylated and methylation is known to block CTCF binding \[348, 349\]. This model can be
tested by seeing whether CTCF, cohesin and TADs are restored after inhibiting DNA methylation. Finally, the non-canonical SMC protein SMCHD1 is an Xist interactor, is enriched on the Xi and deletion of SMCHD1 leads to partial TAD restoration on the Xi [248]. It is possible SMCHD1 plays a role in antagonizing TADs on the Xi. None of these potential mechanisms of TAD disruption are mutually exclusive, and further defining them will provide stronger insight into both how Xist regulates the X-chromosome, and what pathways control higher-order structure more generally.

The observations that the Xi lacks TADs and that Xist RNA is responsible for the lack of TADs has important implications for chromosome biology. Further elucidating the mechanisms of TAD loss on the Xi may provide additional insight into whether TADs are necessary for proper gene regulation. If TADs and or architectural protein binding could be fully restored on the Xi, this would provide a powerful test of whether the lack of TADs and cohesins is necessary for silencing on the Xi. This in turn would allow for a test of the role of TADs and architectural proteins in mammalian gene regulation in XCI, an important model system and biological context.

**Additional regulators of chromosome conformation and silencing within the Xist interactome**

During the course of my PhD, advances in RNA capture technology enabled experiments to define all Xist interacting proteins (the “Xist proteome”). Apart from hnRNP-U, PRC2, PRC1, YY1, and ATRX, many new interaction partners have been identified, and many of these proteins may be crucial regulators of both XCI and the unique higher order structure of the Xi. The next few years of X-inactivation research will certainly entail detailed examination of the mechanisms of novel Xist interactors in XCI. Because of the importance of novel Xist interactors for the field, I will review the major new interacting proteins identified by the various
Xist proteomics approaches (iDRiP [192], RNA Affinity Purification (RAP) [190], and ChIRP-MS [191]).

All three studies identified SPEN (also called SHARP or MINT) as an Xist interactor [190-192], and two genetic screens for silencing factors also found SPEN [193, 194]. SPEN is a very large protein and contains several RNA recognition motifs [350]. SPEN associates with multiple silencing complexes, especially HDAC3 and promotes histone deacetylation [350-352]. One study showed that the Repeat A region of Xist is required for SPEN recruitment [190], and crosslinking studies localize SPEN binding to Repeat A [353]. Several studies showed that SPEN is required for Xist to initiate silencing [190, 191, 193, 194]. SPEN appears to act as a recruiter of HDAC activity to Xist RNA and the Xi. This provides a mechanistic link between Xist and histone deacetylation, which is one of the earliest events upon Xist induction but an event that has been poorly understood. Since histone deacetylation may alter accessibility on the Xi, it will be interesting to examine chromatin conformation changes on the Xi in the absence of SPEN or HDAC3. Moreover, biochemical studies to examine the relationship between Xist and SPEN would be desirable. It is known that the region of Xist that binds SPEN, Repeat A, is absolutely required for Xist-mediated silencing [77], and this repeat may be important for recruiting a number of different factors, including PRC2 and SPEN. However, the precise mechanism of action is not clear at present. Indeed, it is also known from genetic studies that Repeat A is required for the upregulation of Xist expression [77, 78]. It will also be interesting to test whether TADs, CTCF and cohesin are lost and if megadomains form in \( \Delta \)Repeat A Xist cell lines, as these experiments will provide a further test of the relationship between silencing and structure.

Two other proteins (WTAP and RBM15) that bind to Repeat A have attracted attention for their possible role in novel Xist processing and silencing pathways. WTAP binds to RBM15 [354] (which was also identified as an Xist interactor by all three studies), and both WTAP [355] and RBM15 [356] are involved in recruiting the METTL3 N\(^6\)-methyladenosine (m\(^6\)A) complex to
RNA. Indeed, another study reported that XIST is heavily m\textsuperscript{6}A-methylated, that m\textsuperscript{6}A recruiter and Xist interactor RBM15 is required for silencing, and that the m\textsuperscript{6}A reader YTHCD1 is required for Xist-mediated silencing [356]. These results suggest a novel role for RNA methylation in XCI, however the study exploring the role of RNA methylation performed some experiments in human cell lines and others in a male mESC inducible Xist system, making their generality to more natural XCI systems unclear. The possibility that RNA methylation may control the Xist proteome deserves attention and may help explain how particular proteins are recruited to Xist under particular conditions.

On this note, many of the proteins identified in Xist proteomics experiments are general RNA processing factors, such as PTBP1 or various hnRNPs. This is expected, since Xist is a large, abundant spliced transcript it will certainly need to associate with general RNA processing factors. Perhaps its association with recruiters and readers of RNA methylation is the same; general binding of an RNA processing factor to an RNA that needs to be processed in a particular manner. However, general RNA processing factors should definitely not be overlooked as potential regulators of critical aspects of XCI. For instance, HNRNPU is a common RNA binding protein, and all studies identified it as an Xist interactor [190-192]. HNRNPU is required for Xist localization [63], and the fact that all groups identified HNRNPU as an Xist interactor and as required for Xist localization is an important validation of the various Xist proteomics techniques. Furthermore, all studies identified hnRNP-K as a novel Xist interactor [190-192]. HNRNPK has since been shown to be required for Xist-mediated silencing [80, 191] and for directly recruiting the PRC1 complex and indirectly recruiting the PRC2 complex to the Xi [80]. HNRNPK, like HNRNPU is another general, common RNA-binding protein that binds thousands of transcripts, but yet has a crucial role in XCI. Thus, general RNA binding proteins like the hnRNPs deserve study as novel regulators of XCI.

Another Xist interactor that falls in the category of general RNA binders is Lamin B Receptor (LBR). LBR interacts with Xist in iDRIP and RAP proteomics studies but not ChIRP-
MS [190-192]. LBR binds the nuclear lamina [357], and has proposed to serve as a bridge between the nuclear lamina and Xist RNA [358], explaining why the Xi preferentially localizes to the nuclear lamina. The LBR-Xist interaction was proposed to be required for silencing and Xist spreading across the X. However, further analysis (which I co-authored [359]) of many of the cell lines used in these experiments showed that they did not have the correct genotypes, raising doubts about the validity of the results suggesting Xist must bind LBR to be able to spread and silence. The possibility that Xist interactors such as LBR drive the nuclear localization of the Xi, and that the nuclear localization of the Xi is important for silencing deserves further testing. In addition, studies of the Xi conformation in situations where its localization is perturbed (perhaps in LBR knockdown/knockout, for instance), will also be very interesting.

Finally, the results from all the Xist proteomics studies taken together suggest that Xist interacts with a wide variety of proteins involved in many biological processes. The most heavily studied novel interactors tend to fall into two categories: either the strongest hits (SPEN, HNRNPU, HNRNPK, LBR are among the most highly enriched proteins across all techniques) or proteins involved in biological processes attracting a great deal of attention in the field (SPEN for histone deacetylation, RBM15/WTAP/YTHCD1 for RNA methylation, cohesins for chromatin conformation). Systematic evaluation of all interactors would be very helpful and will likely uncover novel biology. One previously identified interactor, YY1 [62], did not turn up as a hit in the proteomic screens. This could be due to its sub-stoichiometric interaction, as indeed it was shown to interact perhaps only at the Xic nucleation site and may not interact with Xist throughout the X chromosome. Some new candidate interactors (topoisomerases, chromatin remodelers, nuclear matrix components) identified by the recent proteomic screens may play novel roles in helping the Xi adopt its unique higher-order structure. The standard workflow for evaluating an Xist interactor should be 1.) validate the interaction through RNA-IP, 2.) test whether it is direct or indirect through biochemical tests, 3.) identify the region of Xist where the
protein binds (through RNA-IP in various deletions and/or crosslinking IP), 4.) knockdown or knockout the interactor of interest and assay XCI phenotypes (Xist localization, chromatin modifications on the Xi, silencing) 5.) compare knockdown/knockout of the interactor to knockout of its binding region of Xist. The lists generated from Xist proteomics experiments likely suffer from false positives due to the technical limitations of the underlying techniques and are likely incomplete. They will improve with time, and the data we have now will prove invaluable for XCI researchers to test out new putative interactors and new pathways in X-chromosome biology.

**The elusive function of megadomains on the Xi highlights limitations to studying the roles of domain organization in gene regulation**

In addition to lacking TADs and compartments, the Xi is organized into a “megadomain and superloop” structure by the tandem repeats *Dxz4* and *Firre*. *Dxz4* is the locus at the border between the two megadomains (Chapter 3, [192, 220, 224, 247, 252]). It binds both CTCF and cohesin on the Xi, unlike the rest of the chromosome. It is a tandem repeat harboring multiple CTCF sites all in the same orientation; in fact, nothing is conserved between mouse and human *DXZ4* aside from the fact that both species have a tandem repeat containing multiple CTCF motifs at syntenic regions [255]. *Dxz4* forms an extremely long “superloop” with *Firre* specifically on the Xi (Chapter 3, [220, 252, 254]. *Firre* is in some ways similar to *Dxz4*. It is also a tandem repeat on the X-chromosome, and it also harbors a large cluster of CTCF sites (Chapter 3, [256, 328]). The megadomains and the superloops are two striking features of the structure of the Xi and no similar megabase-scale structures are seen on autosomes.

However, this unique “super structure” on the Xi appears dispensable for gene silencing during XCI. I deleted *Dxz4* and *Firre* individually and together (Chapter 3). Deletion of *Dxz4* breaks down the megadomain structure and allows interactions between the two normally separate megadomains. *Firre* deletion appears to slightly weaken interactions within each
megadomain. Deleting both has the most severe effect on the organization of the Xi; loss of megadomains and altered contact patterns within the remnants of the megadomains. Deletion of Firre disrupts the superloop with Dxz4 [328]. These results show that Dxz4 and Firre are necessary for both megadomains and superloops, thus perturbing them disrupts the super structure of the Xi. However, deleting Dxz4, Firre or both has no effect on the ability of Xist RNA to spread, H3K27me3 recruitment to the Xi, chromatin accessibility on the Xi or gene silencing during XCI initiation (Chapter 3). Despite dramatic effects on the structure of the Xi, deleting Dxz4 and/or Firre does not inhibit XCI initiation in the slightest. My results are generally in line with previous Dxz4 deletions in mouse [199, 327] and human [252]. These previous studies indicated some subtle alterations to the chromatin context on the Xi, such as increased CTCF binding and accessibility [327], partial loss of H3K27me3 [252] or loss of escape [199].

However, I was unable to observe any of these deviations from the wild-type ground state of the Xi in my experiments, and indeed none of these small deviations have been replicated between two studies, suggesting that they are unique to each experimental system. Never the less, both my work and published work all agree that disrupting the megadomains does not cause any defect in the maintenance or initiation of gene silencing due to XCI.

The observation that the megadomains are not required for XCI fits into a larger debate about the role of the domain organization in mammalian gene regulation. On the one hand, studies of particular domains, such as the HoxD cluster [277, 278, 360] or TAD borders disrupted by structural variants in human polydactyly [279], indicate that domain borders are essential for preventing ectopic enhancer-promoter interactions and improper gene expression patterns. On the other hand, removal of all domains by depletion of cohesin or CTCF has only very minimal effects on gene expression genome-wide [233, 244, 245, 282, 283]. The dispensability of Dxz4 for silencing on the Xi fits with the more global picture, that domains generally are not needed for proper gene regulation. However, both studies of particular
domains and of the global role of domains suffer from limitations, which are relevant for interpreting the functions of \textit{Dxz4} and the megadomains.

The current studies concerning the role of domains in gene regulation all suffer from a lack of biological context. Studies showing dramatic phenotypes and effects on gene expression upon genetic disruption of TADs at polydactyly associated loci or the \textit{HoxD} cluster may be specific only to those particular developmental contexts. Perhaps precise inhibition of ectopic enhancer-promoter contexts through precisely organized domains only matters in contexts like the developing limb, where careful spatiotemporal control of gene expression is needed to properly pattern the limb. Disruption of intricate enhancer-promoter networks may not have any obvious phenotype in the cell line system generally used for experiments that test the global role of domains through depletion of architectural proteins. Perhaps only a few dozen genes are dysregulated in each cell type if domains are disrupted, but if domains were disrupted in all cell types this may add up to a significant fraction of the genome and lead to multiple detrimental phenotypes. In order to truly assess the function of domains in gene regulation, more specific deletions of more borders need to be made in more biological contexts, and global domain disruption experiments need to be performed in many more cell types, ideally in whole organisms.

These limitations in the breadth of biological context in analysis of both particular domains and the global role of domains is also a limitation to interpreting the function of \textit{Dxz4} and the megadomains. My work in Chapter 3 only examines the tandem repeats and super structures in the biological context of XCI initiation and establishment. While the finding that \textit{Dxz4} and \textit{Firre} are not needed for XCI initiation and establishment is important for researchers considering what loci on the Xi are important for XCI and for the chromatin conformation field in its evaluation of the function of megabase-scale structures, my results speak only to the context of XCI initiation and establishment. To fully assess possible functions of \textit{Dxz4} and \textit{Firre}, it is necessary to look more broadly at more biological contexts, perhaps by removing both loci in
animals and carefully examining them for any phenotypes. Furthermore, it has been proposed that domains may not simply regulate which enhancer-promoter interactions occur, but instead provide landscapes for new regulatory interactions to evolve [271, 361]. It may be necessary to look at *Dxz4* and *Firre* and particular domains more generally in an evolutionary context to infer their functions, rather than examine them only in particular systems in particular organisms. Although the function of *Dxz4*, *Firre* and the Xi super structure, along with domain organization in general, still remains elusive, studying these phenomena in more cell types, model systems and biological contexts will ultimately provide insight into their functions.
Appendix: Supplemental Figures and Tables
SUPPLEMENTAL DISCUSSION

Here we discuss similarities and differences between our findings and those of two recent studies that were published while our manuscript was under review. One study performed Xist proteomic analysis on formaldehyde cross-linked chromatin of mouse cells and identified 30 Xist-specific proteins, via the ChIRP-MS methodology [362]. A second study carried out Xist proteomic analysis on UV-crosslinked chromatin and identified 10 proteins, via the RAP-MS methodology [363]. The proteins identified by the two studies are similar, with each highlighting SPEN. Our iDRiP method identified a larger set of interactors (80-250 proteins), inclusive of the proteins identified by the two groups. However, in addition, our study uncovered numerous other factors, such as multiple cohesin subunits, CTCF, multiple SWI/SNF subunits, three topoisomerases, various chromatin modifiers, condensins, many RNA helicases, SUN2, AURKB, and macroH2A. Notably, our proteomics also isolated the known interactors, PRC2 and ATRX, whereas neither factor was identified using ChIRP-MS and RAP-MS [362, 363]. The identification of these positive controls suggests that iDRiP is highly sensitive and further confirms a direct and specific interaction between Xist RNA, PRC2, and ATRX.

The iDRiP-MS, ChIRP-MS, and RAP-MS methodologies have significant technical differences. Firstly, to minimize background due to DNA-bound proteins, a key step employed by iDRiP is inclusion of DNase I treatment before elution of complexes. We believe that this enhanced the specificity and sensitivity of iDRiP. Secondly, ChIRP-MS employed 3% formaldehyde crosslinking for 30 minutes, which could have resulted in large crosslinked networks of chromatin and favored identification of abundant factors (such as nuclear matrix factors and other general RNA-binding proteins) that do not necessarily directly interact with Xist. On the other hand, RAP-MS employed UV crosslinking, but a major difference is the use of a male cell line in which Xist is
overexpressed from a doxycycline-inducible promoter. Our study was performed on a female cell line expressing endogenous, physiological levels of Xist. In our hands, doxycycline induction leads to >30-fold overexpression of Xist, which in turn could similarly favor identification of the highly abundant nuclear matrix factors and other general RNA-binding proteins over specific chromatin factors, such as cohesins, SWI/SNF, PRC2, ATRX, and topoisomerases. Furthermore, because the male cells die after hours of Xist induction (due to loss of X-linked expression), the RAP-MS approach could have also selected against repressive complexes pertinent to XCI. Here, we report all candidates obtained from the mass spectrometry, without applying any filters, and provide enrichment scores. The roles of numerous other interactors can be investigated in future.
Figure S2.1. The microscopic images of knock down day 7 ESCs.

The stable knock down embryonic stem cells (ESCs) were differentiated after the withdrawal of LIF for seven days. On day 4, the cells were plated on the gelatin coated coverslips until day 7 of differentiation. The coverslips were prepared for immunoFISH, as described in methods, followed by imaging for Xi markers, Xist (Red) and H3K27me3 (Green).
**Figure S2.2. Confirmation that the GFP transgene of Xi-TgGFP cells is on the inactive X.**

**(A)** Fluorescent in Situ Hybridization (FISH) indicates the location of the GFP transgene (DNA FISH, red) relative to the inactive X (characterized by a cloud of Xist RNA, identified by RNA FISH in green). In primary fibroblasts selected for high GFP expression (top panels), the transgene is on the active X and does not colocalize with the inactive X (examples indicated by white arrowheads). However, in Xi-TgGFP cells the GFP transgene does colocalize with the inactive X (bottom panels, arrowheads indicate one cell as an example. Xi-TgGFP cells are tetraploid; thus, two inactive X chromosomes are seen per cell).

**(B)** Allele-specific expression of the X-linked gene *Mecp2* shown by RT-PCR. Hybrid Xi-TgGFP cells have one *M. musculus* (mus) X chromosome with the GFP transgene, and one *M. castaneus* (cas) X. A mus-cas single nucleotide polymorphism is detected by Ddel digest, yielding a 179-bp band for expression from the cas allele, or a 140-bp band for expression from the mus allele. A 200-bp band is common to both alleles. Only the expected cas allele of *Mecp2* is expressed in Xi-TgGFP cells (lanes 1, 2, 5), as for purely cas cells (lanes 3, 4, 6), and in contrast to cells of a pure mus background (lane 8), or from a non-clonal hybrid cell population with expression from both alleles (lane 7).
Figure S2.2 (Continued)

A

GFP transgene

GFP transgene + Xist RNA

GFP-positive primary cells

Xi-TgGFP cells

B

| 1. +RT | Xi-TgGFP cells | undigested |
| 2. -RT | cas cells | |
| 3. +RT | Xi-TgGFP cells | |
| 4. -RT | cas cells | |
| 5. hybrid, nonclonal cells | |
| 6. mus cells | Dde I digest |
**Figure S2.3.** Xi reactivation by inhibiting single versus multiple Xist interactors.

(A) Quantitative RT-PCR demonstrated that shRNA knockdown of single Xist interactors resulted in a maximum of 4-fold GFP upregulation.

(B) Biological replicates for allele-specific RNA-seq analysis: Number of upregulated Xi genes for triple-drug treated cells (aza+eto+shRNA). Blue, genes specifically reactivated on Xi; red, genes also upregulated on Xa. There was a net increase in expression level (ΔFPKM) from the Xi in the triple-drug treated samples relative to the shControl+aza+eto, whereas the Xa and autosomes showed no obvious net increase, thereby suggesting direct effects on the Xi as a result of disrupting the Xist interactome.
Figure S2.4. Correlations between biological replicates for allelic-specific RNA-seq analysis.

Shown are allelic (mus) FPKM values for replicate 1 (Rep1) and replicate 2 (Rep2) for indicated triple-drug treatment (orange text) for all genes, Xi genes, and Chr13 genes.
Figure S2.5. Correlations between biological replicates for allelic-specific RNA-seq analysis.

Shown are allelic (mus) FPKM values for replicate 1 (Rep1) and replicate 2 (Rep2) for indicated triple-drug treatment (orange text) for all genes, Xi genes, and Chr13 genes.
Figure S2.6. Correlations between biological replicates for allelic-specific RNA-seq analysis.

Shown are allelic (mus) FPKM values for replicate 1 (Rep1) and replicate 2 (Rep2) for indicated triple-drug treatment (orange text) for all genes, Xi genes, and Chr13 genes.
Figure S2.7. Allelic expression of autosomal genes, including imprinted genes, is not affected by the triple-drug treatments.

Read coverages of three representative autosomal genes (A) and four representative imprinted genes (B) after triple-drug treatment. Mus, *Mus musculus* allele. Comp, total reads. Tracks are shown at the same scale within each grouping. Red tags appear only in exons with SNPs.
Figure S2.7 (Continued)
Figure S2.8. Analysis of CTCF and cohesin ChIP-seq replicates demonstrates similar allelic trends on ChrX.

(A) Allele-specific ChIP-seq results of biological replicates: Violin plots of allelic skew for CTCF, RAD21, SMC1a in wild-type (WT) and X\(^{\Delta Xist}/Xa^{WT}\) (\(\Delta Xist\)) fibroblasts. Fraction of mus reads \([\text{mus}/(\text{mus+cas})]\) is plotted for every peak with \(\geq10\) allelic reads. \(P\) values determined by the Kolmogorov-Smirnov (KS) test.

(B) Table of total, Xa-specific, and Xi-specific cohesin binding sites in WT versus \(\Delta Xist\) (Xi\(^{\Delta Xist}/Xa^{WT}\)) cells. Significant SMC1a and RAD21 allelic peaks with \(\geq5\) reads were analyzed. Allele-specific skewing is defined as \(\geq3\)-fold skew towards Xa or Xi. Sites were considered “restored” if Xi\(^{\Delta Xist}\)’s read counts were \(\geq50\%\) of Xa’s. X-total, all X-linked binding sites. Allelic peaks, sites with allelic information. Xa-total, all Xa sites. Xi-total, all sites. Xa-spec, Xa-specific. Xi-spec, Xi-specific. Xi-invariant, Xi-specific in both WT and Xi\(^{\Delta Xist}/Xa^{WT}\) cells. Note: The net gain of sites on the Xi in the mutant does not equal the number of restored sites. This difference is due to defining restored peaks separately from calling ChIP peaks (macs2). Allele-specific skewing is defined as \(\geq3\)-fold skew towards either Xa or Xi.

(C) Correlation analysis showing Log2 Xi\(^{\Delta Xist}\) to Xa\(^{WT}\) ratios of SMC1a coverage in replicates 1 and 2 (Rep1, Rep2). Rep1, blue dots. Rep2, red dots. Both, purple dots. Consensus, upper right quadrant.

(D) Correlation analysis showing Log2 Xi\(^{\Delta Xist}\) to Xa\(^{WT}\) ratios of RAD21 coverage in replicates 1 and 2 (Rep1, Rep2). Rep1, blue dots. Rep2, red dots. Both, purple dots. Consensus, upper right quadrant.
Figure S2.8 (Continued)

A

<table>
<thead>
<tr>
<th>Fraction X\textit{i}(\textit{mus})</th>
<th>Replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT \Boxist</td>
</tr>
<tr>
<td>Cell line:</td>
<td></td>
</tr>
<tr>
<td>CTCF</td>
<td></td>
</tr>
<tr>
<td>RAD21</td>
<td></td>
</tr>
<tr>
<td>SMC1a</td>
<td></td>
</tr>
</tbody>
</table>

\[ P = 0.0001 \quad P = 0.0999 \quad P = 1.2 \times 10^{-6} \]

B

<table>
<thead>
<tr>
<th>cell line</th>
<th>SMC1a</th>
<th>RAD21</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT \Boxist</td>
<td>X-total</td>
<td>1490</td>
</tr>
<tr>
<td>\Boxist</td>
<td>Allelic peaks</td>
<td>589</td>
</tr>
<tr>
<td>WT \Boxist</td>
<td>Xa-total</td>
<td>597</td>
</tr>
<tr>
<td>\Boxist</td>
<td>Xa-spec</td>
<td>384</td>
</tr>
<tr>
<td>WT \Boxist</td>
<td>Xi-total</td>
<td>267</td>
</tr>
<tr>
<td>\Boxist</td>
<td>Xi-spec</td>
<td>22</td>
</tr>
<tr>
<td>Restored on \textit{Xi}^\textit{Xist}</td>
<td>-</td>
<td>100</td>
</tr>
</tbody>
</table>

C

D
Figure S2.9. Analysis of biological replicates for cohesin ChIP-seq confirms cohesin restoration in cis when Xist is ablated.

Allele-specific ChIP-seq analysis of SMC1a and RAD21 biological replicates. Top panels: Differences between SMC1a or RAD21 peaks on the $X_i^{WT}$ versus $X_a^{WT}$. Black diagonal, 1:1 ratio. Plotted are read counts for all SMC1a or RAD21 peaks. Allele-specific skewing is defined as $\geq$3-fold skew towards either $X_a$ (cas, blue dots) or $X_i$ (mus, red dots). Biallelic peaks, grey dots. Middle panels: Partial restoration of SMC1a or RAD21 peaks on the $X_i^{\Delta Xist}$ to an $X_a$ pattern. Plotted are peaks with read counts with $\geq$3-fold skew to $X_a^{WT}$ (“$X_a$-specific”). x-axis: normalized $X_a^{WT}$ read counts. y-axis: normalized $X_i^{\Delta Xist}$ read counts. Black diagonal, 1:1 $X_i^{\Delta Xist}/X_a^{WT}$ ratio; red diagonal, 1:2 ratio. Bottom panels: $X_i$-specific SMC1a or RAD21 peaks remained on $X_i^{\Delta Xist}$. Plotted are read counts for SMC1a or RAD21 peaks with $\geq$3-fold skew to $X_i^{WT}$ (“$X_i$-specific”).
Figure S2.9 (Continued)
Figure S2.10. Restored SMC1a peaks are reproducible in biological replicates and occur throughout Xi\textsuperscript{\text{\textregistered}Xist} (Example set 1).

The representative examples of SMC1a restoration on Xi\textsuperscript{\textregisteredXist}. “Restored” peaks shown as ticks under each biological replicate (Rep1, Rep2). The “consensus” restored peaks are shown in the last track of each grouping.
Figure S2.10 (Continued)
Figure S2.11. Restored SMC1a peaks are reproducible in biological replicates and occur throughout $X_i^\Delta X_{\text{list}}$ (Example set 2).

The representative examples of SMC1a restoration on $X_i^\Delta X_{\text{list}}$. “Restored” peaks shown as ticks under each biological replicate (Rep1, Rep2). The “consensus” restored peaks are shown in the last track of each grouping.
Figure S2.11 (Continued)
Figure S2.12. Restored RAD21 peaks are reproducible in biological replicates and occur throughout $X_i^{\Delta Xist}$.

The representative examples of RAD21 restoration on $X_i^{\Delta Xist}$. “Restored” peaks shown as ticks under each biological replicate (Rep1, Rep2). The “consensus” restored peaks are shown in the last track of each grouping.
Figure S2.12 (Continued)
**Figure S2.13.** Cohesin restored in \( \Xi^{\text{Xist}}/\Xi^{\text{WT}} \) fibroblasts was \( \Xi \)-specific and did not occur on autosomes.

Correlation plots comparing SMC1a or RAD21 coverages on the mus versus cas alleles in wildtype fibroblasts (WT) versus \( \Xi^{\text{Xist}}/\Xi^{\text{WT}} \) fibroblasts (\( \Delta \text{Xist} \)). Representative autosome, Chr5, is shown. Equation shows the slope and y-intercepts for the black diagonals as a measure of correlation. Pearson’s \( r \) also shown.
Figure S2.14. Biological replicates of HiC-seq analysis yield similar findings.

(A) Allele-specific contact map for the X-chromosome in wild-type fibroblasts at 100 kb resolution. Orange, Xi. Blue, Xa. DXZ4 location is indicated. Xa TADs are not apparent at the scale shown. The Xi appears to be partitioned into megadomains at DXZ4.

(B) Contact maps for various ChrX regions at 40-kb resolution comparing Xi$^{\Delta Xist}$ (red) to Xi$^{WT}$ (orange), and Xi$^{\Delta Xist}$ (red) versus Xa (blue) of the mutant cell line. Our Xa TAD calls are shown with RefSeq genes. Rep1 and Rep2 contact maps are shown side by side.
Figure S2.14 (Continued)
Figure S2.15. Restored TADs identified in $X_i^{\Delta Xist}$ using Xa TADs of Dixon et al. as reference.

(A) Using TADs called by Dixon et al. [364] (rather than our own called TADs, as shown in Figure 5C) as a basis for identifying restored TADs, we calculated the fraction of interaction frequency per TAD on the Xi (mus) chromosome. Highly similar results were obtained. The positions of our Xa TAD borders were rounded to the nearest 100 kb and submatrices were generated from all pixels between the two endpoints of the TAD border for each TAD. We calculated the average interaction score for each TAD by summing the interaction scores for all pixels in the submatrix defined by a TAD and dividing by the total number of pixels in the TAD. We then averaged the normalized interaction scores across all bins in a TAD in the Xi (mus) and Xa (cas) contact maps and computed the fraction of averaged interaction scores from mus chromosomes. ChrX and a representative autosome, Chr5, are shown for the WT cell line and the Xist$^{\Delta Xist}$/+ cell line. P value determined by KS test. P-value determined by paired Wilcoxon signed rank test.

(B) Using TADs called by Dixon et al. [364] (rather than our own called TADs, as shown in Figure 5C) as a basis for identifying restored TADs, violin plots also showed that TADs overlapping restored peaks have larger increases in interaction scores relative to all other TADs. We calculated the fold-change in average interaction scores on the Xi for all X-linked TADs and intersected the TADs with SMC1a sites ($Xi^{\Delta Xist}$/ $Xi^{WT}$). 32 TADs occurred at restored cohesin sites; 80 TADs did not overlap restored cohesin sites. Violin plot shows distributions of fold-change average interaction scores between $Xi^{WT}$ and $Xi^{\Delta Xist}$. P-value determined by Wilcoxon ranked sum test.

(C) Using TADs called by Dixon et al. [364] (rather than our own called TADs, as shown in Figure 5C) as a basis for identifying restored TADs, we also found that restored TADs
overlapped regions with restored cohesins on across XΔXist. Note highly similar results obtained here relative to Figure

**Figure S2.15 (Continued)**

5E. Several datasets were used to identify restored TADs, each producing similar results. Restored TADs were called in two separate replicates (Rep1, Rep2) where the average interaction score was significantly higher on XΔXist than on XWT. We also called restored TADs based on merged Rep1+Rep2 datasets. Finally, a consensus between Rep1 and Rep2 was derived. Method: We calculated the fold-change in mus or cas for all TADs on ChrX and on a control, Chr5; then defined a threshold for significant changes based on either the autosomes or the Xa. We treated Chr5 as a null distribution (few changes expected on autosomes) and found the fraction of TADs that crossed the threshold for several thresholds. These fractions corresponded to a false discovery rate (FDR) for each given threshold. An FDR of 0.05 was used.
Figure S2.15 (Continued)
Supplementary Table S2.1: iDRiP proteomics results.

All proteins with >2X enrichment over male using multiplex quantitation

<table>
<thead>
<tr>
<th>Reference</th>
<th>log2_avgFX/avgMX</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>Q62504</td>
</tr>
<tr>
<td>sp</td>
<td>Q8K0E8</td>
</tr>
<tr>
<td>sp</td>
<td>Q01149</td>
</tr>
<tr>
<td>sp</td>
<td>Q9DBZ1</td>
</tr>
<tr>
<td>sp</td>
<td>Q9WVM1</td>
</tr>
<tr>
<td>sp</td>
<td>P35601</td>
</tr>
<tr>
<td>sp</td>
<td>Q60847</td>
</tr>
<tr>
<td>sp</td>
<td>Q61391</td>
</tr>
<tr>
<td>sp</td>
<td>Q8CEC0</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VDF2</td>
</tr>
<tr>
<td>sp</td>
<td>Q65Z40</td>
</tr>
<tr>
<td>sp</td>
<td>O88532</td>
</tr>
<tr>
<td>sp</td>
<td>O08734</td>
</tr>
<tr>
<td>sp</td>
<td>Q8R0G9</td>
</tr>
<tr>
<td>tr</td>
<td>Q8BVY0</td>
</tr>
<tr>
<td>sp</td>
<td>P11087</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CRB2</td>
</tr>
<tr>
<td>sp</td>
<td>Q60848</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VEK3</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BUI3</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VE37</td>
</tr>
<tr>
<td>sp</td>
<td>Q7TPV4</td>
</tr>
<tr>
<td>sp</td>
<td>Q8C854</td>
</tr>
<tr>
<td>sp</td>
<td>Q91ZX7</td>
</tr>
<tr>
<td>sp</td>
<td>Q99JX7</td>
</tr>
<tr>
<td>sp</td>
<td>Q9D8M4</td>
</tr>
<tr>
<td>sp</td>
<td>P09021</td>
</tr>
<tr>
<td>sp</td>
<td>Q6P5D8</td>
</tr>
<tr>
<td>sp</td>
<td>P70255</td>
</tr>
<tr>
<td>sp</td>
<td>P02340</td>
</tr>
<tr>
<td>sp</td>
<td>Q9Z0H4</td>
</tr>
<tr>
<td>sp</td>
<td>Q924C1</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CYL5</td>
</tr>
<tr>
<td>sp</td>
<td>P43247</td>
</tr>
</tbody>
</table>
Supplementary Table S2.1: (continued)

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Log2 FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>Q9CPS7</td>
</tr>
<tr>
<td>sp</td>
<td>P35441</td>
</tr>
<tr>
<td>sp</td>
<td>Q3U9G9</td>
</tr>
<tr>
<td>sp</td>
<td>P28653</td>
</tr>
<tr>
<td>sp</td>
<td>Q61398</td>
</tr>
<tr>
<td>sp</td>
<td>O35730</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CX86</td>
</tr>
<tr>
<td>sp</td>
<td>Q6PHZ5</td>
</tr>
<tr>
<td>sp</td>
<td>Q9WVJ9</td>
</tr>
<tr>
<td>sp</td>
<td>Q00P9</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CXK8</td>
</tr>
<tr>
<td>tr</td>
<td>J3QQ16</td>
</tr>
<tr>
<td>sp</td>
<td>Q91YK2</td>
</tr>
<tr>
<td>sp</td>
<td>Q9JLM8</td>
</tr>
<tr>
<td>sp</td>
<td>Q07417</td>
</tr>
<tr>
<td>sp</td>
<td>Q9WTX8</td>
</tr>
<tr>
<td>sp</td>
<td>Q9DBR1</td>
</tr>
<tr>
<td>sp</td>
<td>Q02788</td>
</tr>
<tr>
<td>sp</td>
<td>Q921F2</td>
</tr>
<tr>
<td>sp</td>
<td>Q69ZN7</td>
</tr>
<tr>
<td>sp</td>
<td>O88322</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CQL1</td>
</tr>
<tr>
<td>sp</td>
<td>Q61235</td>
</tr>
<tr>
<td>tr</td>
<td>H3BJG4</td>
</tr>
<tr>
<td>sp</td>
<td>P59997</td>
</tr>
<tr>
<td>sp</td>
<td>Q9DC23</td>
</tr>
<tr>
<td>sp</td>
<td>Q99KE1</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BJS4</td>
</tr>
<tr>
<td>tr</td>
<td>Q921K2</td>
</tr>
<tr>
<td>sp</td>
<td>P11352</td>
</tr>
<tr>
<td>sp</td>
<td>P00375</td>
</tr>
<tr>
<td>tr</td>
<td>G5E924</td>
</tr>
<tr>
<td>sp</td>
<td>Q9JL15</td>
</tr>
<tr>
<td>sp</td>
<td>P28301</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BJW6</td>
</tr>
<tr>
<td>sp</td>
<td>Q91Z31</td>
</tr>
<tr>
<td>sp</td>
<td>Q3TDQ1</td>
</tr>
<tr>
<td>sp</td>
<td>Q9DOE1</td>
</tr>
<tr>
<td>sp</td>
<td>F8VPU2</td>
</tr>
</tbody>
</table>
### Supplementary Table S2.1: (continued)

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>P84089</td>
<td>ERH_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P62317</td>
<td>SMD2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>B0V2N1</td>
<td>PTPRS_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q5SYD0</td>
<td>MYO1D_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9DCN2</td>
<td>NB5R3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9EQI8</td>
<td>RM46_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P46935</td>
<td>NE_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P35550</td>
<td>FBRL_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P70202</td>
<td>LXN_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9R0M6</td>
<td>RAB9A_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P38060</td>
<td>HMGCL_MOUSE</td>
</tr>
<tr>
<td>tr</td>
<td>Q8VHM5</td>
<td>Q8VHM5_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P70227</td>
<td>ITPR3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>O70503</td>
<td>DHB12_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VDD9</td>
<td>PHIP_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BHD7</td>
<td>PTBP3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P59235</td>
<td>NUP43_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q99020</td>
<td>ROAA_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9WTP7</td>
<td>KAD3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8C2Q3</td>
<td>RBM14_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q5SX40</td>
<td>MYH1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P97868</td>
<td>RBBP6_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9WUK4</td>
<td>RFRC2_MOUSE</td>
</tr>
<tr>
<td>tr</td>
<td>Q0VBL3</td>
<td>Q0VBL3_MOUSE</td>
</tr>
<tr>
<td>tr</td>
<td>E9Q5G3</td>
<td>E9Q5G3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q64012</td>
<td>RALY_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P42230</td>
<td>STA5A_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P83870</td>
<td>PHF5A_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q61578</td>
<td>ADRO_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9DCA2</td>
<td>RT11_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9Z777</td>
<td>BAZ1B_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CW46</td>
<td>RAVR1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>O70318</td>
<td>E41L2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q6PAC3</td>
<td>DCA13_MOUSE</td>
</tr>
<tr>
<td>tr</td>
<td>Q3TIX6</td>
<td>Q3TIX6_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>O35492</td>
<td>CLK3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q80TH2</td>
<td>LAP2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8K4P0</td>
<td>WDR33_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CW03</td>
<td>SMC3_MOUSE</td>
</tr>
</tbody>
</table>
Supplementary Table S2.1: (continued)

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene Symbol</th>
<th>Description</th>
<th>Log2 Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>Q8K2A1</td>
<td>GULP1_MOUSE</td>
<td>1.32</td>
</tr>
<tr>
<td>sp</td>
<td>Q8K2F8</td>
<td>LS14A_MOUSE</td>
<td>1.31</td>
</tr>
<tr>
<td>sp</td>
<td>Q3UMR5</td>
<td>MCU_MOUSE</td>
<td>1.31</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BX02</td>
<td>KANK2_MOUSE</td>
<td>1.31</td>
</tr>
<tr>
<td>sp</td>
<td>P47738</td>
<td>ALDH2_MOUSE</td>
<td>1.29</td>
</tr>
<tr>
<td>sp</td>
<td>P08074</td>
<td>CBR2_MOUSE</td>
<td>1.29</td>
</tr>
<tr>
<td>sp</td>
<td>Q9WVL0</td>
<td>MAAI_MOUSE</td>
<td>1.28</td>
</tr>
<tr>
<td>sp</td>
<td>Q6PFR5</td>
<td>TRA2A_MOUSE</td>
<td>1.28</td>
</tr>
<tr>
<td>sp</td>
<td>Q8CGB6</td>
<td>TENC1_MOUSE</td>
<td>1.28</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VCW8</td>
<td>ACSF2_MOUSE</td>
<td>1.28</td>
</tr>
<tr>
<td>sp</td>
<td>Q99KP6</td>
<td>PRP19_MOUSE</td>
<td>1.27</td>
</tr>
<tr>
<td>sp</td>
<td>P10404</td>
<td>ENV1_MOUSE</td>
<td>1.25</td>
</tr>
<tr>
<td>sp</td>
<td>Q4FK66</td>
<td>PR38A_MOUSE</td>
<td>1.25</td>
</tr>
<tr>
<td>sp</td>
<td>Q6NS46</td>
<td>RRP5_MOUSE</td>
<td>1.25</td>
</tr>
<tr>
<td>sp</td>
<td>Q9R112</td>
<td>SQRD_MOUSE</td>
<td>1.24</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VE80</td>
<td>THOC3_MOUSE</td>
<td>1.24</td>
</tr>
<tr>
<td>sp</td>
<td>Q921H8</td>
<td>THIKA_MOUSE</td>
<td>1.24</td>
</tr>
<tr>
<td>sp</td>
<td>Q922Q4</td>
<td>P5CR2_MOUSE</td>
<td>1.23</td>
</tr>
<tr>
<td>sp</td>
<td>Q922H2</td>
<td>PDK3_MOUSE</td>
<td>1.23</td>
</tr>
<tr>
<td>tr</td>
<td>Q8BGJ5</td>
<td>Q8BGJ5_MOUSE</td>
<td>1.23</td>
</tr>
<tr>
<td>sp</td>
<td>P55012</td>
<td>S12A2_MOUSE</td>
<td>1.22</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VC70</td>
<td>RBMS2_MOUSE</td>
<td>1.22</td>
</tr>
<tr>
<td>sp</td>
<td>Q922V4</td>
<td>PLRG1_MOUSE</td>
<td>1.21</td>
</tr>
<tr>
<td>sp</td>
<td>Q91VI7</td>
<td>RINI_MOUSE</td>
<td>1.21</td>
</tr>
<tr>
<td>sp</td>
<td>P30285</td>
<td>CDK4_MOUSE</td>
<td>1.21</td>
</tr>
<tr>
<td>sp</td>
<td>P45952</td>
<td>ACADM_MOUSE</td>
<td>1.21</td>
</tr>
<tr>
<td>sp</td>
<td>P61979-3</td>
<td>HNRPK_MOUSE</td>
<td>1.20</td>
</tr>
<tr>
<td>sp</td>
<td>O70325</td>
<td>GPX41_MOUSE</td>
<td>1.20</td>
</tr>
<tr>
<td>sp</td>
<td>O89086</td>
<td>RBM3_MOUSE</td>
<td>1.20</td>
</tr>
<tr>
<td>sp</td>
<td>Q6PE01</td>
<td>SNR40_MOUSE</td>
<td>1.20</td>
</tr>
<tr>
<td>sp</td>
<td>Q60749</td>
<td>KHDR1_MOUSE</td>
<td>1.20</td>
</tr>
<tr>
<td>sp</td>
<td>O55222</td>
<td>ILK_MOUSE</td>
<td>1.19</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CY66</td>
<td>GAR1_MOUSE</td>
<td>1.19</td>
</tr>
<tr>
<td>sp</td>
<td>Q99LC2</td>
<td>CSTF1_MOUSE</td>
<td>1.18</td>
</tr>
<tr>
<td>sp</td>
<td>Q6P5E4</td>
<td>UGGG1_MOUSE</td>
<td>1.18</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BQZ5</td>
<td>CPSF4_MOUSE</td>
<td>1.18</td>
</tr>
<tr>
<td>sp</td>
<td>Q91VC3</td>
<td>IF4A3_MOUSE</td>
<td>1.18</td>
</tr>
<tr>
<td>sp</td>
<td>Q61990</td>
<td>PCBP2_MOUSE</td>
<td>1.18</td>
</tr>
<tr>
<td>sp</td>
<td>Q9QYS9</td>
<td>QKI_MOUSE</td>
<td>1.18</td>
</tr>
</tbody>
</table>
### Supplementary Table S2.1: (continued)

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>P50544</td>
<td>ACADV_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P70372</td>
<td>ELAV1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P11276</td>
<td>FINC_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BHB4</td>
<td>WDR3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9D0B0</td>
<td>SRSF9_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q61937</td>
<td>NPM_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q3U0V1</td>
<td>FUBP2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q60668</td>
<td>HNRPD_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8C7V3</td>
<td>UTP15_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8CAQ8</td>
<td>IMMT_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P51480</td>
<td>CD2A1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P27048</td>
<td>RSMB_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P52293</td>
<td>IMA1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8QZT1</td>
<td>THIL_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q80X85</td>
<td>RT07_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>O88559</td>
<td>MEN1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9Z2X1</td>
<td>HNRPF_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BG05</td>
<td>ROA3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q91W39</td>
<td>NCOA5_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P33174</td>
<td>KIF4_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q08879</td>
<td>FBLN1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9CYK1</td>
<td>SYWM_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P13020</td>
<td>GELS_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q5XG71</td>
<td>UTP20_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q80YX1</td>
<td>TENA_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9EP97</td>
<td>SENP3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P52825</td>
<td>CPT2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q60973</td>
<td>RBBP7_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q64133</td>
<td>AOFA_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q99JY0</td>
<td>ECHB_MOUSE</td>
</tr>
<tr>
<td>tr</td>
<td>E9QNN1</td>
<td>E9QNN1_MOUSE</td>
</tr>
<tr>
<td>tr</td>
<td>Q91VA7</td>
<td>Q91VA7_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q05920</td>
<td>PYC_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P13864</td>
<td>DNMT1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>O88569</td>
<td>ROA2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q05CL8</td>
<td>LARP7_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8K411</td>
<td>PREP_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q3UJB9</td>
<td>EDC4_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8BP71</td>
<td>RFOX2_MOUSE</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
<td>Score</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>-------</td>
</tr>
<tr>
<td>sp</td>
<td>P62320</td>
<td>SMD3_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P50136</td>
<td>ODBA_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8VE22</td>
<td>RT23_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9ERU9</td>
<td>RBP2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q80YV2</td>
<td>NIPA_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9R0Y5</td>
<td>KAD1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q9Z2I8</td>
<td>SUCB2_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q99PV0</td>
<td>PRP8_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P37040</td>
<td>NCPR_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>P14733</td>
<td>LMNB1_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8QZY9</td>
<td>SF3B4_MOUSE</td>
</tr>
<tr>
<td>tr</td>
<td>A2ANY6</td>
<td>A2ANY6_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q61029</td>
<td>LAP2B_MOUSE</td>
</tr>
<tr>
<td>sp</td>
<td>Q8CI11</td>
<td>GNL3_MOUSE</td>
</tr>
</tbody>
</table>
Supplementary Table S2.2: Lentiviral shRNA constructs used for stable knockdowns of candidate Xist interactors.

<table>
<thead>
<tr>
<th>RefSeq_shRNA viruses</th>
<th>Xist interacting candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRCN0000011883</td>
<td>Top1</td>
</tr>
<tr>
<td>TRCN0000321370</td>
<td>Ctcf</td>
</tr>
<tr>
<td>TRCN0000071385</td>
<td>Smarca4</td>
</tr>
<tr>
<td>TRCN0000295773</td>
<td>Smarca5</td>
</tr>
<tr>
<td>TRCN0000321371</td>
<td>Ctcf</td>
</tr>
<tr>
<td>TRCN0000109008</td>
<td>Smc3</td>
</tr>
<tr>
<td>TRCN0000276847</td>
<td>Rad21</td>
</tr>
<tr>
<td>TRCN0000174832</td>
<td>Rad21</td>
</tr>
<tr>
<td>TRCN0000321718</td>
<td>Aurkb</td>
</tr>
<tr>
<td>TRCN0000317702</td>
<td>Smarcb1</td>
</tr>
<tr>
<td>TRCN0000071383</td>
<td>Smarca4</td>
</tr>
<tr>
<td>TRCN0000325493</td>
<td>Top2a</td>
</tr>
<tr>
<td>TRCN0000295713</td>
<td>Smarca5</td>
</tr>
<tr>
<td>TRCN0000309135</td>
<td>Kif4</td>
</tr>
<tr>
<td>TRCN0000321651</td>
<td>Aurkb</td>
</tr>
<tr>
<td>TRCN0000109007</td>
<td>Smc3</td>
</tr>
<tr>
<td>TRCN0000090909</td>
<td>Kif4</td>
</tr>
<tr>
<td>TRCN0000321444</td>
<td>Ctcf</td>
</tr>
<tr>
<td>TRCN0000071388</td>
<td>Smarcc1</td>
</tr>
<tr>
<td>TRCN0000288446</td>
<td>Smarca5</td>
</tr>
<tr>
<td>TRCN0000072181</td>
<td>GFP</td>
</tr>
<tr>
<td>TRCN0000071389</td>
<td>Smarcc1</td>
</tr>
<tr>
<td>TRCN0000070988</td>
<td>Top2b</td>
</tr>
<tr>
<td>TRCN0000011884</td>
<td>Top1</td>
</tr>
<tr>
<td>TRCN0000070990</td>
<td>Top2b</td>
</tr>
<tr>
<td>TRCN0000229486</td>
<td>Pds5a</td>
</tr>
<tr>
<td>TRCN0000011886</td>
<td>Top1</td>
</tr>
<tr>
<td>TRCN0000085541</td>
<td>Smarcc2</td>
</tr>
<tr>
<td>TRCN0000317622</td>
<td>Smarcb1</td>
</tr>
<tr>
<td>TRCN0000324673</td>
<td>Smc1a</td>
</tr>
<tr>
<td>TRCN0000229484</td>
<td>Pds5a</td>
</tr>
<tr>
<td>TRCN0000085540</td>
<td>Smarcc2</td>
</tr>
<tr>
<td>TRCN0000070987</td>
<td>Top2a</td>
</tr>
<tr>
<td>TRCN0000071386</td>
<td>Smarca4</td>
</tr>
</tbody>
</table>
**Supplementary Table S2.2: (continued).**

<table>
<thead>
<tr>
<th>TRCN0000109009</th>
<th>Smc3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRCN0000246806</td>
<td>Sun2</td>
</tr>
<tr>
<td>TRCN0000276903</td>
<td>Rad21</td>
</tr>
<tr>
<td>TRCN0000071391</td>
<td>Smarcc1</td>
</tr>
<tr>
<td>TRCN0000070992</td>
<td>Top2b</td>
</tr>
<tr>
<td>TRCN0000317701</td>
<td>Smarcb1</td>
</tr>
<tr>
<td>TRCN0000085542</td>
<td>Smarcc2</td>
</tr>
<tr>
<td>TRCN0000321719</td>
<td>Aurkb</td>
</tr>
<tr>
<td>TRCN0000246805</td>
<td>Sun2</td>
</tr>
<tr>
<td>TRCN0000246804</td>
<td>Sun2</td>
</tr>
<tr>
<td>TRCN0000217996</td>
<td>Pds5a</td>
</tr>
<tr>
<td>TRCN0000090908</td>
<td>Kif4</td>
</tr>
<tr>
<td>TRCN0000324674</td>
<td>Smc1a</td>
</tr>
<tr>
<td>TRCN0000324672</td>
<td>Smc1a</td>
</tr>
<tr>
<td>TRCN0000353984</td>
<td>Top2a</td>
</tr>
<tr>
<td>TRCN0000231782_pLKO_TRC021</td>
<td>Control</td>
</tr>
<tr>
<td>TRCN0000231782_pLKO_TRC021</td>
<td>Control</td>
</tr>
</tbody>
</table>
**Supplementary Table S2.3: Table of reagents**

iDRiP oligonucleotides, primer pairs used for RT-qPCR of candidate proteins following stable knockdown and various antibodies and small molecules used for immunofluorescence and X-reactivation studies.

<table>
<thead>
<tr>
<th>5'-Biotinylated DNA Oligos used in Xist interactome capture</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>CATTTAAGAGCAAAGTCGTTTTC</td>
</tr>
<tr>
<td>X2</td>
<td>AATATTTTACATTACAGGGTGGCA</td>
</tr>
<tr>
<td>X3</td>
<td>TAAAGACCAAGCAAAGATACTTGTC</td>
</tr>
<tr>
<td>X4</td>
<td>ATGCTTCATATATTGCAGTGGTTCAC</td>
</tr>
<tr>
<td>X5</td>
<td>TGTATTAAGTAAATTCCATGACCC</td>
</tr>
<tr>
<td>X6</td>
<td>AACTTAGCAATTAATTCTGGGACTC</td>
</tr>
<tr>
<td>X7</td>
<td>ATGCATATCTGTATGCATGCTTATT</td>
</tr>
<tr>
<td>X8</td>
<td>CATATTACGGGGACTAAGGACT</td>
</tr>
<tr>
<td>X9</td>
<td>ATGGGCACTGCATTITTAGCAATA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers used in qPCR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>U1 snRNA-F</td>
<td>CCAGGGCGAGGCTTATCCATT</td>
</tr>
<tr>
<td>U1 snRNA-R</td>
<td>GCAGTCCACCTACACAAAT</td>
</tr>
<tr>
<td>eGFP-F</td>
<td>GAC GTA AAC GGC CAC AAG TT</td>
</tr>
<tr>
<td>eGFP-R</td>
<td>AAG TCG TG CTG CTT CAT GTG</td>
</tr>
<tr>
<td>U6 snRNA-F</td>
<td>CTC GCT TCG GCA GCA CA</td>
</tr>
<tr>
<td>U6 snRNA-R</td>
<td>AAC GCT TCA CGA ATT TGC GT</td>
</tr>
<tr>
<td>Smc1a-F</td>
<td>TCG GAC CAT TTC AGA GGT TTA CC</td>
</tr>
<tr>
<td>Smc1a-R</td>
<td>CAG GTG CTC CAT GTA TCA GGT</td>
</tr>
<tr>
<td>Smc3-F</td>
<td>CGA AGT TAC CGA GAC CAA ACA</td>
</tr>
<tr>
<td>Smc3-R</td>
<td>TCA CTG AGA ACA AAC TGG ATT GC</td>
</tr>
<tr>
<td>Rad21-F</td>
<td>ATG TTC TAC GCA CAT TTT GTC CT</td>
</tr>
<tr>
<td>Rad21-R</td>
<td>TGC ACT CAA ATA CAT GGG CTT T</td>
</tr>
<tr>
<td>Aurkb-F</td>
<td>CAG AAG GAG AAC GCC TAC CC</td>
</tr>
<tr>
<td>Aurkb-R</td>
<td>GAG AGC AAG CGC AGA TGT C</td>
</tr>
<tr>
<td>Top2b-F</td>
<td>CTG ACC TGG GTG AAC AAT GCT</td>
</tr>
<tr>
<td>Top2b-R</td>
<td>TGG CTC CAC TGA TCC AAT GTA T</td>
</tr>
<tr>
<td>Top2a-F</td>
<td>GAG AGG CTA CGA CTC TGA CC</td>
</tr>
<tr>
<td>Top2a-R</td>
<td>CTC CAG GTA GGG GGA TGT TG</td>
</tr>
<tr>
<td>Top1-F</td>
<td>AAG ATC GAG AAC ACC GGC ATA</td>
</tr>
<tr>
<td>Top1-R</td>
<td>CTT TTC CTC CTT CGG TCT TTC C</td>
</tr>
</tbody>
</table>
### Supplementary Table S2.3: (continued)

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctcf-F</td>
<td>GAT CCT ACC CTT CTC CAG ATG AA</td>
</tr>
<tr>
<td>Ctcf-R</td>
<td>GTA CCG TCA CAG GAA CAG GT</td>
</tr>
<tr>
<td>Smarca4-F</td>
<td>CAA AGA CAA GCA TAT CCT AGC CA</td>
</tr>
<tr>
<td>Smarca4-R</td>
<td>CAC GTA GTG TGT GTT AAG GAC C</td>
</tr>
<tr>
<td>Smarca5-F</td>
<td>GAC ACC GAG ATG GAG GAA GTA</td>
</tr>
<tr>
<td>Smarca5-R</td>
<td>CGA ACA GCT CTG TCT GCT TTA</td>
</tr>
<tr>
<td>Smarc1-F</td>
<td>AGC TAG ATT CGG TGC GAG TCT</td>
</tr>
<tr>
<td>Smarc1-R</td>
<td>CCA CCA GTC CAG CTA GTG TTT T</td>
</tr>
<tr>
<td>Smarc2-F</td>
<td>GCT GCC TAC AAA TTC AAG AGT GA</td>
</tr>
<tr>
<td>Smarc2-R</td>
<td>AGG AAA ATG TTA GGT CGT GAC AG</td>
</tr>
<tr>
<td>Smarcb1-F</td>
<td>TCC GAG GTG GGA AAC TAC CTG</td>
</tr>
<tr>
<td>Smarcb1-R</td>
<td>CAG AGT GAG GGG TAT CTC TTG T</td>
</tr>
<tr>
<td>Sun2-F</td>
<td>ATC CAG ACC TTC TAT TTC CAG GC</td>
</tr>
<tr>
<td>Sun2-R</td>
<td>CCC GGA AGC GGT AGA TAC AC</td>
</tr>
<tr>
<td>Jpx</td>
<td>GCA CCA CCA GCC TTC TGT AAC</td>
</tr>
<tr>
<td></td>
<td>GGG CAT GTT CAT TAA TTG GCC AG</td>
</tr>
<tr>
<td>18S</td>
<td>CGC GGT TCT ATT TTG TTG GT</td>
</tr>
<tr>
<td></td>
<td>AGT CGG CAT CGT TTA TGG TC</td>
</tr>
<tr>
<td>Xist1.1</td>
<td>TCA TCC ACC GAG CTA CTC TT</td>
</tr>
<tr>
<td></td>
<td>AGC AAA GCC TGA ACT GGT ATA G</td>
</tr>
<tr>
<td>Xist1.2</td>
<td>CCT GCA AGG GAT ACC GTT TAT</td>
</tr>
<tr>
<td></td>
<td>ATG AAA GGC GAA GGA GTA TGG</td>
</tr>
<tr>
<td>Xist1-3</td>
<td>GCT GGT TCG TCT ATC TTG TGG G</td>
</tr>
<tr>
<td></td>
<td>CAG AGT AGC GAG GAC TTG AAG AG</td>
</tr>
<tr>
<td>Xist3-4</td>
<td>AGC TTA CAG GCC ACA TGG AGA A</td>
</tr>
<tr>
<td></td>
<td>ATC GTC AAA GGG AAT AGG TCG CCA</td>
</tr>
<tr>
<td>Xist4-6</td>
<td>ACA AAG GCT CAA AGA ATG CCA CCC</td>
</tr>
<tr>
<td></td>
<td>TTG CTT GAT CAC GCT GAA GAC CCC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Brand</th>
<th>Antibodies and Catalog #</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOVUS BIOLOGICALS INC</td>
<td>SMC3 antibody (NB100-207)</td>
</tr>
<tr>
<td>NOVUS BIOLOGICALS INC</td>
<td>SMC1 Antibody (A300-055A)</td>
</tr>
<tr>
<td>BETHYL LABORATORIES INC</td>
<td>TOP1 Antibody (A302-589A)</td>
</tr>
<tr>
<td>SIGMA-ALDRICH INC</td>
<td>ANTI-SUN2 antibody (HPA001209-100UL)</td>
</tr>
<tr>
<td>SANTA CRUZ BIOTECHNOLOGY</td>
<td>Brg-1 (H-88) (sc-10768)</td>
</tr>
<tr>
<td>PROTEINTECH GROUP INC</td>
<td>TOP2A-Specific Antibody (20233-1-AP)</td>
</tr>
<tr>
<td>ABCAM INC</td>
<td>Anti-Aurora B Kinase antibody (ab2254)</td>
</tr>
</tbody>
</table>
**Supplementary Table S2.3: (continued)**

<table>
<thead>
<tr>
<th>Brand</th>
<th>Chemicals and Catalog #</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCAM INC</td>
<td>Anti-Rad21 antibody - ChIP Grade (ab992); Novus NB100-203</td>
</tr>
<tr>
<td>ACTIVE MOTIF</td>
<td>Histone H3K27me3 antibody (pAb) (39155)</td>
</tr>
<tr>
<td>PROTEINTECH GROUP INC</td>
<td>TOP2B Polyclonal Antibody (20549-1-AP)</td>
</tr>
<tr>
<td>CELL SIGNALING TECHNOLOGY</td>
<td>SMARCC2/BAF170 (D8O9V) Rabbit mAb (12760)</td>
</tr>
<tr>
<td>E M D MILLIPORE</td>
<td>Anti-CTCF Antibody (07-729)</td>
</tr>
<tr>
<td>SIGMA-ALDRICH INC</td>
<td>Etoposide (E1383-25MG)</td>
</tr>
<tr>
<td>SIGMA-ALDRICH INC</td>
<td>5-aza-2'-deoxycytidine (A3656)</td>
</tr>
</tbody>
</table>
Figure S3.1: Superloop anchors are tandem repeats that bind CTCF and SMC1A on the mouse Xi.

SMC1A and CTCF ChIP-seq coverage across the superloop anchors in mouse (A) Dxz4, (B) Firre, (C) The X-Inactivation Center, and (D) x75. From top to bottom within a panel: SMC1A ChIP-seq in MEF, black=comp (all unique reads including neutral, cas, and mus), blue=Xa, red=Xi, CTCF ChIP-seq in MEF (same color scheme), CTCF ChIP-seq in mESCs, black=comp, grey=repetitive alignments.
Figure S3.1 (Continued)
Figure S3.2: Megadomains are not present on the Xa and allelic RNA-seq shows Xist upregulation is highly skewed to Xi.

(A) KR-normalized Hi-C matrices on future Xa (cas) on days 0, 3, 7, 10 of differentiation (2.5 Mb resolution). (B) Pearson correlation of Hi-C matrix on future Xa (cas) on days 0, 3, 7, 10 of differentiation (1Mb resolution). (C) Numbers of Xist reads expressed from the mus (red) or cas (blue) during differentiation. (D) Fraction of allelic Xist reads expressed from X^mus (carrying the Tsix^{TST} allele) during differentiation, in two biological replicates. (E) Slope of the PC1 score curve at Dxz4 for Hi-Cs generated with varying ratios of d10:d0 reads. Dashed line represents linear best fit; linear model parameters are included.
Figure S3.2 (Continued)

A

Day 0
Day 3
Day 7
Day 10

B

C

Allelic read counts from Xist locus

D

Fraction Xist reads from mus allele

E

PCR slope = 0.33
R² = 0.051
p-value: 0.0003829

fraction megadomain positive (Day0/Day10)
Figure S3.3: Biological replicates of the time course Hi-C and Hi-C$^2$ experiments.

(A) Pearson correlation matrices at 2.5 Mb resolution for the Xi (right) days 0, 3, 7, 10 of differentiation in two biological replicate Hi-C experiments. (B) Hi-C$^2$ contact maps around $Dxz4$ (mm9 coordinates chrX:71,832,976-73,511,687) and on the Xi on days 0, 3, 7, 10 of differentiation (50 kb resolution) in two biological replicates. (C) Hi-C$^2$ contact maps around $Mecp2$ (mm9 coordinates chrX:70,370,161-71,832,975) and on the future Xi on days 0, 3, 7, 10 of differentiation (50 kb resolution) in two biological replicates. In (B) and (C), Green bars indicate positions of domain borders determined from 25 kb d0 comp Hi-C$^2$ matrices; dark blue track shows Dixon et al. TAD calls in mESCs, light blue track shows Marks et al. TAD calls in mESCs, red bars indicate positions of either $Dxz4$ or $Mecp2$. (D) Conceptual diagram of insulation score. Top: formula for calculating insulation score at region i. $R_i$ refers to the sum of interactions within the region to the right of i, $L_i$ refers to the sum of interactions within the region to the left of i, and $C_r_i$ represents the sum of interactions that “cross over” i. Middle: diagram of the window used to calculate insulation score for an example (non-border) i in the day 0 Xi $Mecp2$ contact map. Bottom: plot of insulation score across the day 0 Xi $Mecp2$ region. The shaded green bars indicate the positions of the borders of the Mecp2 sub-TADs across all diagrams. (E, F) Violin plots showing the distributions of insulation scores across the $Dxz4$ region (E) and $Mecp2$ region (F) in both biological replicates. Note: to generate violin plots and evaluate the significance of differences in variance between timepoints we excluded the 6 bins on each edge of the Hi-C$^2$ region because the regions needed to calculate insulation score fall partly outside the Hi-C$^2$ region and have far lower read counts than sequences targeted by the capture probes.
Figure S3.3 (Continued)
Figure S3.4: Validation of Dxz4 deletion and replicate Hi-C.

(A) Scheme of the Dxz4 deletion. The large red bar shows the region deleted in this study; the smaller red bar shows the coordinates of the Dxz4 tandem repeat array. Interspersed repeats are shown in gold, and CTCF ChIP-seq in mESCs is shown in black for all unique reads and grey for all repetitive reads. The “ΔDxz4 Sanger BLAT” track is the BLAT result for Sanger sequencing result from the PCR product generated with primers flanking the deleted region. (B) Hi-C coverage over the Dxz4 deletion region for Dxz4^{Δ/Δ} clone E5 (top) and WT (bottom). The near-absence of reads in Dxz4^{Δ/Δ} over the deletion implies a biallelic deletion. Note: we observed reads in Dxz4^{Δ/Δ} within the left end of the deletion (grey box), roughly chrX:72,940,913-72,948,483. BLAT analysis of this region indicates that it is repeated several times elsewhere in the genome (data not shown), thus reads within this region are to be expected even in Dxz4^{Δ/Δ}. (C) DNA FISH in WT (left) and Dxz4^{Δ/Δ} (right) using either Dxz4 (red) or Xist (green) probes. (D) Pearson correlation matrices at 2.5 Mb resolution for the wild-type (left) and Dxz4^{Δ/Δ} Xi (right) in two biological replicate Hi-C experiments. Note: the WT control Hi-C in replicate 2 in this figure also serves as the replicate 2 day 10 timepoint in FigS3a, thus this Pearson correlation matrix is present in both FigS3a and this figure.
Figure S3.4 (Continued)

A

Schematic of Dcx4 deletion

B

RefSeq Genes

C

WT  Dcx4

D

Xi pearson correlation, rep1

Xi pearson correlation, rep2
Figure S3.5: Generation of a Xist+Dxz4 transgene.

(A) Xist RNA FISH in Xist+Dxz4 Tg cells (clone XPDxz4.4) +/-dox after either 2 days of induction (top) or 9 days of induction (bottom panels). (B) Co-localization between transgenic Xist and Dxz4. Left, DNA FISH for the transgene’s P1 backbone (cyan)+Xist (red). Right: DNA FISH for Xist (red) and Dxz4 (green). (C) 4C contact profiles in XPDxz4.4 or a separate Xist-only transgenic cell line (XY X+P) on chrX from a viewpoint within the backbone of the Xist construction. The positions of Firre, Dxz4 and Xist are indicated on the X-chromosome. (D) Co-localization between Xist and Dxz4 and the candidate insertion region (Stc1, chr14) obtained from 4C. Left: DNA FISH for Stc1 (red). Middle: overlap between DNA FISH for Stc1 (red) and Dxz4 (green). Right: overlap between DNA FISH for Stc1 (red) and the Xist transgene’s P1 backbone (cyan).
Figure S3.5 (Continued)

A

-Dox (2 days)  +Dox (2 days)

-Dox (9 days)  +Dox (9 days)

B

Transgene backbone

Xist

C

<table>
<thead>
<tr>
<th>chrX position</th>
<th>20 Mb</th>
<th>40 Mb</th>
<th>60 Mb</th>
<th>80 Mb</th>
<th>100 Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superloop anchors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XP Dxd4.4 4C profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xist+P 4C profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D

Stl

Stl  Dxd4

Stl  Transgene backbone
Figure S3.6: A conserved network of superloop interactions involving *Dxz4* on the mouse Xi.

(A) 4C coverage from the *Dxz4* viewpoint over the whole X in mus Xi MEFs. Black=comp, grey=reps, blue=Xa, red=Xi. The positions of putative superloop anchors *Firre*, *Dxz4*, *Xist* and *x75* are shown. (B) 4C coverage from the *Dxz4* viewpoint over the *Xist* region. (C) 4C coverage from the *Dxz4* viewpoint over the region homologous to human *x75* in mouse. (D) 4C coverage from the *Dxz4* viewpoint over a putative singleton artifact near *Dmd*. Many of the other large peaks on the X or elsewhere are “singletons”, amplified sequences likely due to mispriming, this is an example of a singleton.
Figure S3.6 (Continued)
Figure S3.7: Validation of Firre deletions.

(A) Allele-specific ATAC-seq coverage over Firre. The coordinates of the Firre deletion are shown by the dark red bar, and the two “Sanger BLAT” tracks represent the coordinates of the BLAT alignment of the PCR products flanking Firre generated from Firre\(^{\text{Xi}\Delta/+}\) clone D1 and \(Dxz4^{\Delta/\Delta}:\text{Firre}^{\text{Xi}\Delta/+}\) clone H12. (B) SNP information from the Mouse Genomes Project (cite Sanger) for the SNP that falls within the Firre deletion amplicon. (C) Scheme of restriction assay to determine which allele of Firre is deleted. (D) Results of restriction assay in Xa- and Xi- Firre targeted pools and Firre\(^{\text{Xa}\Delta/+}\) and Firre\(^{\text{\Delta/\Delta}}\) candidate clones (E) DNA FISH in Firre\(^{\text{Xi}\Delta/+}\) clone D1 for Dxz4 (red), and Firre (green). (F) DNA FISH for Dxz4 (red) and Firre (cyan) in WT, Firre\(^{\text{Xa}\Delta/+}\) and Firre\(^{\text{\Delta/\Delta}}\) clones. In Firre\(^{\text{Xi}\Delta/+}\) and Firre\(^{\text{Xa}\Delta/+}\) most cells exhibit one Firre spot but two Dxz4 spots, consistent with a heterozygous Firre deletion, whereas Firre\(^{\text{\Delta/\Delta}}\) does not exhibit Firre spots, consistent with a homozygous Firre deletion. (G) DNA FISH in Dxz4\(^{\Delta/\Delta}:\text{Firre}^{\text{Xi}\Delta/+}\). Left, Firre (green). Right, X-linked gene Hprt (Cyan). Most cells exhibit one Firre spot but two Hprt spots, consistent with a heterozygous Firre deletion.
Figure S3.7 (Continued)

A

<table>
<thead>
<tr>
<th>chrX position</th>
<th>47,900 kb</th>
<th>47,940 kb</th>
<th>47,980 kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>[0.5]</td>
<td></td>
<td>[0.5]</td>
</tr>
<tr>
<td>Dz444</td>
<td>[0.5]</td>
<td></td>
<td>[0.5]</td>
</tr>
<tr>
<td>Fima&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>[0.5]</td>
<td></td>
<td>[0.5]</td>
</tr>
<tr>
<td>Dz444&lt;sup&gt;−/−&lt;/sup&gt;Fima&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanger BLAT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RefSeq Genes

B

<table>
<thead>
<tr>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>dbSNP&lt;sup&gt;+&lt;/sup&gt;</th>
<th>Ref</th>
<th>129S1_SvImJ</th>
<th>CAST_EiJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>47,988,853</td>
<td>6720401G&lt;sup&gt;+&lt;/sup&gt;39Rik</td>
<td>rs258858030</td>
<td>T</td>
<td>-</td>
<td>G</td>
</tr>
</tbody>
</table>

C

PCR assay for Xa or homozygous Fima deletion

**PCR primer:** ATCTAA

**PCR primer:** ATCGAA

**Taq I restriction site**

Fima deletion PCR expected results:

<table>
<thead>
<tr>
<th>-TaqI</th>
<th>Xa&lt;sup&gt;+/+&lt;/sup&gt;</th>
<th>Xa&lt;sup&gt;ΔΔ&lt;/sup&gt;</th>
<th>Xa&lt;sup&gt;ΔΔ&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>+TaqI</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D

DNA FISH in <i>Fima<sup>−/−</sup></i> D1

E

DNA FISH in WT

F

DNA FISH in <i>Fima<sup>−/−</sup></i><sup>+</sup>

G

DNA FISH in Dz444<sup>−/−</sup>Fima<sup>−/−</sup> H12
Figure S3.8: Replicate Hi-Cs in *Firre* deletions.

(A) Pearson correlation matrices at 2.5 Mb resolution for the wild-type (left) and *Firre*<sup>Xi/+</sup> (middle, clone D1) and *Dxz4<sup>Δ/Δ</sup>:Firre*<sup>Xi/+</sup> Xi (right) in two biological replicate Hi-C experiments. Note: the WT control Hi-C in replicates 1 and 2 in this figure also serve as the replicates 1 and 2 day 10 timepoints in FigS3a, thus these Pearson correlation matrices are present in both FigS2a and the current figure. (B) KR-normalized Hi-C matrix for the independently derived *Firre*<sup>Xi/+</sup> clone (E9) at 2.5 Mb resolution. (C) Pearson correlation matrix for the independently derived *Firre*<sup>Xi/+</sup> clone (E9) at 2.5 Mb resolution.
Figure S3.9: Further analysis of Xist localization and H3K27me3 foci in Dxz4 and Firre deletions; metaphase distribution of H3K27me3 in Dxz4^{Δ/Δ}.

(A) Xist RNA FISH (red) combined with rRNA FISH (green) (top), H3K27me3 IF (red, bottom) in WT and Dxz4^{Δ/Δ} Firre^{X0/+} and Dxz4^{Δ/Δ}.Firre^{X0/+} cells on day 10. (B) Fraction of cells with Xist clouds (red) and fraction of Xist clouds in the perinucleolar space (cyan) on day 10. (C) Fraction of cells with an H3K27me3 focus. (D) Immunofluorescence for H3K27ac (red, left) or H3K27me3 (green, middle) on metaphase spreads from WT cells after 10 days of differentiation. The Xi is readily detectable as the one chromosome in the spread depleted of H3K27ac and enriched in H3K27me3 (merge, right). (E) Coating of H3K27me3 on several metaphase inactive Xs in WT (left) or Dxz4^{Δ/Δ} (right) on day 10.
Figure S3.9 (Continued)

A

WT   Dz4Δ

Xist RNA

H3K27me3

B

\[
\begin{array}{c}
\% \\
\hline
\text{Xist clouds} & \text{H3K27me3} \\
\text{WT} & \text{WT} \\
\text{Dz4Δ} & \text{Dz4Δ} \\
\text{Fm4Δ} & \text{Fm4Δ} \\
\text{Dz4Δ,Fm4Δ} & \text{Dz4Δ,Fm4Δ} \\
\end{array}
\]

C

\[
\begin{array}{c}
\% \text{cells with H3K27me3 H3K27ac} \\
\hline
\text{WT} & \text{WT} \\
\text{Dz4Δ} & \text{Dz4Δ} \\
\text{Fm4Δ} & \text{Fm4Δ} \\
\text{Dz4Δ,Fm4Δ} & \text{Dz4Δ,Fm4Δ} \\
\end{array}
\]

D

H3K27ac IF

H3K27me3 IF

H3K27ac IF + H3K27me3 IF

E

WT metaphase Xi

Dz4Δ metaphase Xi
Figure S3.10: Replicate ATAC-seq analysis in *Firre*<sup>Xi<sub>∆/+</sub></sup> and *Dxz4<sub>∆/∆</sub>:Firre<sup>Xi<sub>∆/+</sub></sup>*.

(A) Comparison between WT Xa and mutant Xi for ATAC coverage for peaks that are Xa-specific in wild-type for replicate ATAC-seqs. Top left: *Dxz4<sub>∆/∆</sub> Xi* vs WT Xa, top right: *Dxz4<sub>∆/∆</sub>:Firre<sup>Xi<sub>∆/+</sub></sup> Xi* vs WT Xa, Bottom left: *Firre<sup>Xi<sub>∆/+</sub></sup> Xi* vs WT Xa, replicate 2, Bottom left: *Firre<sup>Xi<sub>∆/+</sub></sup> Xi* vs WT Xa, replicate 3. Black lines correspond to a *Dxz4<sub>∆/∆</sub> Xi*:WT Xa ratio of 1:1, red lines correspond to a ratio of 1:2. (B) Comparison between WT Xa and mutant Xi ATAC coverage for peaks that are Xi-specific in wild-type for replicate ATAC-seqs. Top left: *Dxz4<sub>∆/∆</sub> Xi* vs WT Xi, top right: *Dxz4<sub>∆/∆</sub>:Firre<sup>Xi<sub>∆/+</sub></sup> Xi* vs WT Xi, Bottom left: *Firre<sup>Xi<sub>∆/+</sub></sup> Xi* vs WT Xi, replicate 2, Bottom left: *Firre<sup>Xi<sub>∆/+</sub></sup> Xi* vs WT Xi, replicate 3. Black lines correspond to a *Dxz4<sub>∆/∆</sub> Xi*:WT Xi ratio of 1:1, blue lines correspond to a ratio of 1:2.
Figure S3.10 (Continued)
References


255. Horakova, A.H., et al., The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. Genome biology, 2011. 13(8).


