A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary

Graphical Abstract

The promoter of Linx (LinxP) is a novel regulator of Xist

- The Tsix-TAD regulates not only Tsix but also Xist, in part via LinxP
- LinxP influences choice making during random XCI by regulating Xist expression in cis
- Linx transcription affects local topology but is not necessary for Xist regulation
- LinxP is conserved in sequence and synteny across placental mammals

Highlights

Authors
Rafael Galupa, Elphège Pierre Nora, Rebecca Worsley-Hunt, ..., Uwe Ohler, Luca Giorgetti, Edith Heard

Correspondence
edith.heard@embl.org

In Brief
Galupa et al. uncover elements important for Xist regulation in its neighboring TAD and reveal that these elements can influence gene regulation both within and between topological domains. These findings, in a context where dynamic, developmental expression is necessary, challenge current models for TAD-based gene-regulatory landscapes.

Galupa et al., 2020, Molecular Cell 77, 352–367
January 16, 2020 © 2020 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.molcel.2019.10.030
A Conserved Noncoding Locus Regulates Random Monoallelic $X_{ist}$ Expression across a Topological Boundary

Rafael Galupa,1,13 Ephègue Pierre Nora,1,12,15 Rebecca Worsley-Hunt,2,12 Christel Picard,1,12 Chris Gard,1 Joke Gerard van Bemmel,1,14 Nicolas Servant,5,4 Yinxiu Zhan,5,6 Fatima El Marjou,7 Colin Johanneau,7 Patricia Diabangouaya,1 Agnès Le Saux,1 Sonia Lameiras,8 Juliana Pipoli da Fonseca,8 Friedemann Loos,9 Joost Gribnau,9 Sylvain Baulande,8 Uwe Ohler,2,10 Luca Giorgetti,5 and Edith Heard1,11,13,16,*

1Mammalian Developmental Epigenetics Group, Genetics and Developmental Biology Unit, Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Paris, France
2Berlin Institute for Medical Systems Biology, Max Delbruck Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
3Bioinformatics, Biostatistics, Epidemiology and Computational Systems Unit, Institut Curie, PSL Research University, INSERM U900, Paris, France
4MINES ParisTech, PSL Research University, Centre for Computational Biology (CBIO), Paris, France
5Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
6University of Basel, Basel, Switzerland
7Transgenesis Facility, Institut Curie, Paris, France
8Institut Curie Genomics of Excellence (ICGex) Platform, Institut Curie, Paris, France
9Department of Developmental Biology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands
10Department of Biology, Humboldt University, Berlin, Germany
11Collège de France, Paris, France
12These authors contributed equally
13Present address: European Molecular Biology Laboratory, Heidelberg, Germany
14Present address: Gladstone Institute of Cardiovascular Diseases, San Francisco, CA, USA
15Present address: Cardiovascular Research Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA
16Lead Contact
*Correspondence: edith.heard@embl.org
https://doi.org/10.1016/j.molcel.2019.10.030

SUMMARY

cis-Regulatory communication is crucial in mammalian development and is thought to be restricted by the spatial partitioning of the genome in topologically associating domains (TADs). Here, we discovered that the $X_{ist}$ locus is regulated by sequences in the neighboring TAD. In particular, the promoter of the noncoding RNA $Linx$ ($LinxP$) acts as a long-range silencer and influences the choice of X chromosome to be inactivated. This is independent of $Linx$ transcription and independent of any effect on $Tsix$, the antisense regulator of $X_{ist}$ that shares the same TAD as $Linx$. Unlike $Tsix$, $LinxP$ is well conserved across mammals, suggesting an ancestral mechanism for random monoallelic $X_{ist}$ regulation. When introduced in the same TAD as $X_{ist}$, $LinxP$ switches from a silencer to an enhancer. Our study uncovers an unsuspected regulatory axis for X chromosome inactivation and a class of cis-regulatory effects that may exploit TAD partitioning to modulate developmental decisions.

INTRODUCTION

Expression of most X-linked genes in placental mammals is equalized in XX and XY individuals through X chromosome inactivation (XCI). This involves transcriptional silencing of one of the two X chromosomes during female development (Lyon, 1961). In mice, XCI is triggered by upregulation of the long noncoding RNA (lncRNA) $X_{ist}$, which is conserved across placental mammals and is expressed in female somatic cells from either the paternal or maternal inactive X chromosome (reviewed in Galupa and Heard, 2018). Embryonic XCI can be recapitulated ex vivo in differentiating mouse embryonic stem cells (mESCs). These represent a powerful system to study the regulatory mechanisms of XCI, since $X_{ist}$ transcription is repressed in the pluripotent, undifferentiated state, while upon differentiation, $X_{ist}$ is robustly upregulated from one X chromosome in XX mESCs.

How the initial choice to inactivate one of two X chromosomes is made remains an open question. A minimal regulatory network has recently been proposed (Mutzel et al., 2019), but the underlying molecular actors and mechanisms remain unknown. In mice, several genetic loci influence $X_{ist}$ expression in cis, including the elusive X-controlling element ($Xce$) (Cattanach and Papworth, 1981) as well as several control elements within the X-inactivation center ($Xic$) (for review, see Galupa and Heard, 2018).
These include Tsix, the antisense repressor of Xist, and its enhancer, Xite; deleting either of these loci skews XCI entirely or partially, respectively, in favor of the mutant allele (Lee, 2000; Lee and Lu, 1999; Ogawa and Lee, 2003; Sado et al., 2001). Tsix function seems to be mouse specific (Migeon et al., 2001, 2002), and both Tsix and Xite are poorly conserved across placental mammals (Galuppa and Heard, 2018), suggesting that other cis-regulatory elements are probably implicated in the regulation of choice across mammals.

The set of genomic elements that participate in Xist cis regulation at the onset of random XCI is still unknown. The longest single-copy transgenes tested (~460 kb), including Xist, Tsix, and Xite, failed to induce Xist upregulation in differentiating female mESCs (Heard et al., 1999), suggesting that further cis regulators exist. Chromosome conformation analysis of the murine Xic (Nora et al., 2012) revealed that the Xist/Tsix locus lies at the boundary between two topologically associating domains (TADs), which in total span ~850 kb (Figure 1A). TADs spatially partition mammalian genomes (Dixon et al., 2012; Nora et al., 2012) and represent a structural scale of chromosomes at which functional properties such as transcriptional co-regulation and promoter-enhancer communication are maximized (Zhan et al., 2017). The boundary at the Xist/Tsix locus, which is conserved in mouse and human (Galuppa and Heard, 2018), seems to partition two different cis-regulatory landscapes (van Bemmelen et al., 2019; Nora et al., 2012). Genes within each of the two Xic-TADs show different functions in the regulation of Xist as well as opposite transcriptional behaviors during mESC differentiation (Nora et al., 2012). The “Xist-TAD” (~550 kb) contains the Xist promoter and some of its known positive regulators, such as Ftx (Furlan et al., 2016), which all become upregulated during differentiation; this domain has probably evolved as a hub of positive regulators of Xist. On the other hand, the “Xts-TAD” (~300 kb) includes loci that seem to have evolved as negative cis regulators of Xist to modulate XCI choice, such as the Tsix promoter and Xite; genes within this TAD are downregulated during differentiation (Nora et al., 2012).

Previous transgenic studies in vivo defined an interval within the Tsix-TAD that seems important for Tsix expression in cis (Figure 1B; see figure legend); this region excludes Xite and the Tsix promoter, but harbors a poorly characterized IncRNA locus, Linx, the in vivo expression of which is restricted to cells that will undergo random XCI (Nora et al., 2012). Linx binds pluripotency factors such as Nanog and Oct4, and its expression in mESCs is downregulated during differentiation (Nora et al., 2012). The patterns of expression of Linx in mESCs and during development, together with the fact that it shares the same TAD as Tsix, led to the suggestion that Linx might be a regulator of Tsix (Giorgetti et al., 2014; Nora et al., 2012). However, the role of Linx in the regulation of XCI was not so far addressed.

Here, we genetically dissect the contribution of the Tsix-TAD as well as different elements within it, in particular of the Linx locus, to the regulation of Tsix and Xist during random XCI. Our results reveal that the cis-regulatory landscape of Xist is not restricted to its own TAD but includes elements located in the adjacent TAD. We find that the Tsix-TAD is important for Tsix regulation as expected but that it is also critical for regulating Xist in a Tsix-independent manner. We show that this occurs, at least in part, via the Linx locus, which harbors cis-regulatory elements that modulate Xist expression and XCI choice; this Xist-regulatory action of Linx is not via the non-coding Linx transcript. Instead, we define a cis-regulatory DNA element, which unlike Tsix is conserved across placental mammals.

RESULTS

The Tsix-TAD Regulates Xist Expression and XCI Independently of Tsix

To determine whether the Tsix-TAD harbors essential elements for endogenous Tsix and Xist regulation, we deleted a 245-kb region encompassing all the loci within the Tsix-TAD except Xite and Tsix (Figure 1B). This deletion does not seem to disrupt the TAD boundary or the Xist-TAD (Figure S1A). Transcriptional profiling of both control and Δ245-kb male mESCs during differentiation revealed that Xist expression, which is normally very low in male mESCs, was aberrantly upregulated in the mutants upon differentiation (10-fold after 2 days of differentiation; Figure 1C). This was associated with Xist cloud formation in ~6% of mutant male cells, which is not observed in wild-type male mESCs (Figure 1D). Concomitantly, Xite and Tsix expression...
Figure 2. The Linx Locus Harbors cis-Regulatory Elements that Control XCI Choice

(A) ATAC-seq data for the Tsix-TAD region in differentiating XX mESCs. For each time point, results of peak calling are represented by gray marks below the data. Green marks depict differential peak analysis. Identical results were found for day 0 versus day 2 and day 1 versus day 2 (p < 0.01) within the region of interest, while no differential peaks were found for day 0 versus day 1. Gray box highlights the promoter of Linx, the only differential peak within the 245-kb region. Normalized data are shown for one replicate (second replicate in Figure S2A); peak analysis was performed on both replicates. See STAR Methods for more details.

(B) The Linx locus and its chromatin features (see STAR Methods for sources of datasets represented). The position of introns and exons is based on Nora et al. (2012) and mESC RNA Scripture (Guttman et al., 2010). Targeted region LinxP (∼2 kb) is indicated.

Figure legend continued on next page.
were reduced (Figure 1C). Expression levels of markers for pluripotency, differentiation, and proliferation were not affected (Figures S1B–S1D). Therefore, the Δ245-kb region contains elements that repress Xist and/or activate Xite and Tsix, either directly or indirectly.

To understand whether the 245-kb deletion affects random XCI, we analyzed heterozygous Δ245-kb female ESCs (Figure S1E) and postimplantation embryos derived from polymorphic mouse strains (Figure 1E). Allelic ratio analyses showed that the presence of the Δ245-kb region skews Xist expression in favor of the mutant allele (0.88 versus 0.56, p < 0.001; Figures 1F and S1F) and triggers preferential inactivation in cis, as evaluated by the expression of an X-linked gene, Atpt7a (Figures 1G and S1F). Early differentiating female mESCs also displayed preferential expression of Xist from the Δ245-kb allele (Figure S1E). We conclude that this 245-kb region is critical for controlling Xist upregulation and choice during the initiation of random XCI (see also the notes in the Figure S1 legend).

We next assessed whether the Δ245-kb allele affects Xist expression via dysregulation of its antisense repressor Tsix (Lee and Lu, 1999; Lee et al., 1999; Luikenhuis et al., 2001; Stavropoulos et al., 2001). For this, we used a system that uncouples Xist and Tsix regulation; in the Xist-GFP/Tsix-mCherry (XGTC) female mESC line (Loos et al., 2016), Tsix and Xist are both truncated on the same chromosome and unable to repress each other. Tsix transcription is prematurely truncated, so it does not repress the Xist promoter in cis, and Xist transcription is also prematurely truncated, so there is no Xist RNA to silence Tsix expression in cis. It is still possible, however, to monitor the activity of the Tsix and Xist thanks to fluorescent reporters cloned downstream of each promoter. The other X chromosome in this line remains unmodified. We deleted the 245-kb region on the Xist-GFP/Tsix-mCherry (XGTC) female mESC line (Figure 1H). We found that mCherry (Tsix) levels were markedly reduced in Δ245-kb XGTC cells compared to controls, before and after differentiation (Figure 1I). The Δ245-kb allele thus influences Tsix expression, and this is not a result of aberrant Xist activation and Xist RNA silencing (absent in this system). However, we found that GFP (Xist) levels were also affected, with a significantly higher proportion of cells upregulating GFP from the Δ245-kb allele upon differentiation (66% versus 38%; p < 0.001) (Figure 1J). Given the absence of Tsix/Xist mutual regulation in this cell line, Xist upregulation cannot be a result of Tsix downregulation. These results indicate that the Tsix-TAD contains not only regulators of Tsix but also elements that repress Xist independently of Tsix. This occurs despite the fact that the Xist promoter is located in the adjacent TAD.

**Linx Harbors cis-Regulatory Elements that Modulate XCI Choice Independently of Linx Transcription or RNA**

Next, we set out to define the elements within the Δ245-kb region that could account for the misregulation of Xist on the one hand and Tsix on the other (which would ultimately affect Xist as well; in fact, Xist upregulation in the Δ245-kb allele is most likely a consequence of both downregulation of Tsix and loss of other regulatory elements that act on Xist in a Tsix-independent manner). Within the 245-kb interval, the only sequences previously implicated in the regulation of XCI are Tsx, which stimulates Tsix expression but the deletion of which only mildly affects Xist (Anguera et al., 2011), and Linx, the function of which has not been investigated genetically (Nora et al., 2012). To identify putative candidate cis-regulatory elements in this region that could account for the dramatic skewing of XCI in the Δ245-kb allele, we performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro et al., 2013) in differentiating XX cells (day 0, day 1, and day 2) (Figures 2A and S2A). We found strong open-chromatin sites at all known promoters within the 245-kb interval, as well as at an intergenic, non-annotated region between Chic1 and Tsx. This region displays chromatin marks of active transcription (e.g., H3K27Ac), hereby named as putative enhancer element Orix. Deletion of Orix in mESCs or in mice did not reveal any significant effect on Tsix or Xist expression (Figures S2B–S2D).

None of the identified ATAC-seq peaks within the 245-kb region (including Orix) showed significant changes during differentiation, except the promoter region of Linx, which showed reduced accessibility at day 2 compared to day 0 or day 1 (p < 0.01; Figure 2A). The dynamic behavior of the Linx promoter at the onset of XCI, together with its proposed role in regulating Tsix, prompted us to further investigate the Linx locus in the context of random XCI regulation. We abrogated Linx transcription and RNA by deleting a ~2-kb region centered on Linx TSS (ΔLinxP) in male and female mESCs, as well as in mice (Figures 2B and S3A–S3C; see also the note in the Figure S3 legend). Differentiating (day 4) ΔLinxP-heterozygous polymorphic female mESCs displayed modest but significant skewing in Xist allelic ratios in favor of the mutant allele (1.2-fold, p < 0.01; Figure 2C), similar to the intermediate Xce alleles reported to date (Galupa and Heard, 2015). Our results were consistent in both clones analyzed, regardless of the strain origin of the mutated allele. We also detected preferential Xist cloud formation on the ΔLinxP...
chromosome by RNA-DNA fluorescence in situ hybridization (FISH) (Figure 2D), implying skewed XCI choice. We observed similar results in three independent mutant clones generated in isogenic female mESCs (Figure S3D). Analysis of Xist allelic ratios in postimplantation heterozygous female embryos also revealed a slight but significant preference for Xist expression from the ∆LinxP allele (0.54 versus 0.48, p < 0.01; Figures 2E, 2F, and S3E) and corresponding preferential Atp7a inactivation (0.59 versus 0.64, p < 0.01; Figures 2G and S3E). We conclude that LinxP is a negative cis regulator of Xist that modulates the probability of XCI choice. We found very similar results for another element within Linx, the LinxE element (Figures S2E–S2G and the note in the Figure S2 legend). To distinguish the contribution of the Linx transcript/transcription from the LinxP element itself, we inverted LinxP in mice and mESCs (Figure 2H), which similarly to ∆LinxP abolished Linx IncRNA and transcription across the Linx locus (Figure S3F). Unlike ∆LinxP, heterozygous LinxP-inv female embryos did not show bias of Xist or Atp7a allelic ratios compared to wild type (Figures 2I, 2J, and S3G). Together, these results imply that transcription across the Linx locus or the Linx IncRNA is not mediating the effect of the LinxP deletion in Xist regulation (see also the note in the Figure S3 legend); these effects are therefore most likely a consequence of losing important cis-regulatory genomic elements, which seem to work in an orientation-independent manner. LinxP (and LinxE) thus acts as a cis-regulatory element that negatively modulates Xist expression during differentiation and influences choice at the onset of XCI. Xist expression is affected to a greater extent in ∆245-kb mutants than in ∆LinxP mutants, indicating that other regulators remain to be discovered.

The LinxP Element Represses Xist Independently of Tsix

Given that Linx shares the same TAD as Tsix, we next explored whether LinxP modulates XCI choice by acting as a classic enhancer of Tsix, and therefore negatively affecting Xist expression. However, the LinxP deletion did not downregulate Tsix expression in differentiating male mESCs (Figure 3A; see also the first note in the Figure S4 legend). In fact, in the undifferentiated state (day 0), Tsix is slightly upregulated in ∆LinxP mutants (Figures 3A and S4A), in line with previous observations that Linx and Tsix expression levels from the same allele are anti-correlated (Giorgetti et al., 2014). Together, our results argue against a role for LinxP as an active enhancer of Tsix expression. In female mESCs (day 0), Tsix allelic ratios are also not affected by LinxP heterozygous deletion (Figure 3B). However, we did detect modest but significant differences in Xist allelic ratios prior to differentiation (Figure 3B), implying that the effects on Xist might precede effects on Tsix. This raises the possibility that Linx regulates Xist in a Tsix-independent manner, which could account, at least partially, for the effects observed with the ∆245-kb allele. Differences in Xist allelic ratios between mutant and wild-type alleles became stronger upon differentiation (Figure 3B). Tsix allelic ratios eventually became significantly different as well (Figure 3B), which may be due to silencing in cis by Xist RNA. To uncouple Tsix and Xist regulation, we generated heterozygous ∆LinxP mutants in the XGTC cell line (Figure 3C). Cherry (Tsix) levels were slightly upregulated in the ∆LinxP XGTC cells compared to controls at day 0 and day 2 (Figure 3D), consistent with the results on ∆LinxP male mESCs (Figure 3A) and again arguing against a role for LinxP as an enhancer of Tsix. However, the proportion of cells upregulating GFP from the ∆LinxP allele upon differentiation was slightly but significantly increased (38% versus 30%, p = 0.008) (Figure 3G), supporting that LinxP represses Xist in cis independently of Tsix. We have thus identified a specific element within the Tsix-TAD that regulates Xist, but not via Tsix. Moreover, this controlling element acts as long-range cis repressor, not as an enhancer, to regulate the Xist promoter ~170 kb away in the adjacent TAD.

Topological Changes Associated with Linx Expression Are Not Involved in Xist Regulation

Distal regulatory elements are generally thought to act on their target genes through physical contacts. A major regulator of these contacts is the protein CTCF (Nora et al., 2017). The Linx locus harbors three CTCF-bound sites between the regulatory elements LinxP and LinxE (Figure 4A), which anchor strong loops with other CTCF sites within the Tsix-TAD (Giorgetti et al., 2014; Nora et al., 2012). To explore a possible role for these sites in mediating the regulation of Xist by LinxP/LinxE, we deleted a large intrinsic interval containing the CTCF sites in male ESCs (∆Linx-int1, ~51 kb) and mice (∆Linx-CBS, ~25 kb) (Figure 4A). Chromosome conformation capture carbon copy (5C) analysis of the mutant mESCs revealed disruption of local 3D organization. Increased contacts were found between the Linx 3’ end region and the Chic1 locus, which harbors CTCF sites in convergent orientation to those within the Linx 3’ end region (Figure 4B). Furthermore, the Linx 3’ end region lost contacts with Xite (Figure 4B) and displayed decreased basal contacts throughout the Xist-TAD (Figure 4C, black arrow). The interaction frequencies were reduced between LinxE and the Xist promoter and unaltered between LinxP and the Xist promoter (Figure S6A). However, in heterozygous female embryos, we did not observe any effect on Xist or Atp7a allelic ratios (Figures 4D, S5A, and S5B). This indicates that Xist-mediated regulation of Xist does not require the intronic CTCF sites and can operate in the context of a disrupted chromatin topology of the Tsix- and Xist-TADs.

We then wished to determine whether LinxP itself could directly contact the Xist promoter prior or during XCI initiation. To obtain high-resolution interaction profiles for Linx and Xist promoters, we performed Capture-C (Hughes et al., 2014) in differentiating female ESCs (days 0, 1, 2, and 4). We observed no preferential interaction peaks with Xist when capturing the Linx promoter (Figure 4E) or vice versa (Figure 4F); in fact, their topological landscapes seem rather stable during early differentiation. We also investigated the global organization of the Xic-TADs at the onset of XCI, by performing 5C on the same samples, but we found that the structure of the Xic-TADs remained mostly unaffected upon differentiation (Figure S5C). Together, these data do not reveal any differentiation-specific differences in the topological organization of the Xic that could explain how LinxP regulates the Xist promoter during the initiation of XCI.

Finally, we wondered whether the ∆LinxP allele itself could be affecting the structural landscape of the Xic and thereby influencing Xist expression in cis. We performed 5C on wild-type and mutant ∆LinxP male mESCs as well as LinxP-inv and ∆LinxE
male mESCs for comparison. Differential analysis of 5C maps comparing ΔLinxE to wild-type cells revealed no obvious alterations in the structural organization of the Xic TADs (Figures 4G and S5D), even though ΔLinxE leads to skewing in Xist expression (Figures S2F and S2G). However, ΔLinxP led to marked differences in contact frequencies throughout the Xic-TADs, in particular a gain of contacts between the Tsix- and the Xist-TADs (Figures 4H, 4J, and S5D). Similar results were observed for the LinxP-inv allele (Figures 4I, 4J, and S5D), implying the involvement of Linx transcription and/or Linx lncRNA in the structural changes observed. To further test this hypothesis, without disturbing the LinxP element, we knocked in a poly(A) cassette downstream of LinxP, which abolishes Linx transcription (Figures S6B and S6C). 5C analysis revealed that early truncation of Linx transcription also led to a significant gain of contacts between the Tsix- and the Xist-TADs (Figures S6D and S6E), further supporting that loss Linx transcription or lncRNA is associated with the structural phenotype. We note, however, that this gain is not as high as in the LinxP deletion or inversion, raising the possibility that the LinxP element itself might also contribute to the Xic topological organization. These changes, however, are not correlated with an effect on Xist regulation, as the LinxP-inv allele does not impact Xist expression or XCI choice (Figures 2J and S3G). The interaction frequency

Figure 3. The LinxP Element Is Not an Enhancer of Tsix but Regulates Xist Expression

(A) Gene expression analysis during differentiation. Data are normalized to wild-type day 0 for each gene, and represents the average of two biological replicates for each genotype.

(B) Allelic quantification of Xist (top) and Tsix (bottom) RNA during early differentiation. See legend of Figure 2C for more information on the clones. Data are presented as means and error bars represent SEM (six biological replicates). Statistical analysis was performed using a two-tailed paired t test with Bonferroni’s correction (**p < 0.01).

(C) XGTC female line (129/Cast) as in Figure 1H. We generated ΔLinxP mutant clones on the Cast allele.

(D and E) Median fluorescence intensity (FI) of mCherry (normalized to WT, day 0) or percentage of EGFP positive cells (as in Figure 1J). Wild-type data represent an average of five wild-type clones, with four experimental replicates for each. ΔLinxP data represent an average of five independent clones, with four experimental replicates for each. Statistical analysis was performed using a paired two-tailed t test (**p < 0.01; ***p < 0.001; ****p < 0.0001).
between LinxP and the Xist promoter in the Linx-inv allele does not seem to be significantly altered (Figure S6A); this could be the reason for not seeing an effect on Xist regulation in this mutant, if we are to assume that the interaction frequency between LinxP and Xist is important for how LinxP regulates Xist. Our data do not allow us to conclude whether this is indeed the case, and this assumption remains an open question that merits further investigation. In conclusion, our data show that the Linx locus is independently involved, on the one hand, in helping to shape Xic folding via its transcription or IncRNA (at least partly) and, on the other hand, in modulating Xist expression and XCI choice via its cis-regulatory elements.

The LinxP Element Acts as a cis Activator of Xist When Sharing the Same TAD

To further explore how LinxP might regulate Xist, we performed knockins of LinxP (~2 kb) into the Xist-TAD, in polymorphic female cells, and we determined allelic ratios of Xist expression from the modified or wild-type X chromosomes. We inserted LinxP at two different, independent locations within the Xist-TAD: one was between Jpx and Ftx (Figure 5A), ~60 kb away from the Xist promoter and within the high-frequency contact region upstream of Xist (see Figure 1A), and the other was between Ftx and Xpct (Figure 5B), ~170 kb away from the Xist promoter, which corresponds to the same distance between the endogenous LinxP and the Xist promoter. In both locations, LinxP was inserted in both orientations and included a transcriptional stop cassette to prevent potential LinxP-mediated transcription spreading into the new loci. As controls, we also introduced the transcriptional stop cassette alone in both locations and in the two possible orientations. We differentiated these cell lines and determined Xist allelic ratios at days 0, 2, and 4. Our results consistently showed that the presence of LinxP in the Xist-TAD, regardless of its orientation or position, leads to preferential Xist expression from that chromosome at each differentiation time point (Figures 5A and 5B; see also the second note in the Figure S4 legend). The controls showed no such effects. The action of LinxP on Xist seems therefore to be TAD dependent (or context dependent); LinxP acts as a repressive modulator of Xist expression at its original location in the neighboring TAD and as an enhancer of Xist when lying within the same TAD as the Xist promoter.

The LinxP Element Is Conserved in Sequence and Synteny across Mammals

The Linx locus is poorly conserved overall (Figure 6A), similarly to many IncRNA loci (Chodroff et al., 2010). However, we observed a high degree of sequence conservation for the LinxP element across mammals, from mouse to cetaceans and primates, including humans (Figure 6B). In particular, two conserved modules within LinxP show shared synteny across placental mammals, but not in the marsupial opossum (Figure 6C). One of these modules coincides with binding of Nanog and Oct4 in mESCs (Figure 6B). The pluripotency factors are known repressors of Xist expression, but their repressive mechanisms remain to be determined (Minkovsky et al., 2013; Navarro et al., 2008; Sousa et al., 2018; reviewed in Minkovsky et al., 2012). It is therefore possible that the pluripotency factors are implicated in the cis repression of Xist by LinxP. We note that LinxP is the first regulator of choice described to date that is conserved in sequence and position across placental mammals; the other known regulators of choice, Tsix, Xite, and Xce, seem in fact poorly conserved across mammals (Galupa and Heard, 2018; Peeters et al., 2016). Therefore, LinxP may mediate an ancestral mechanism of Xist negative regulation and choice making during random XCI. Random XCI and the presence of both Xist and LinxP within the Xic are all specific features of placental mammals.

DISCUSSION

In a quest to understand cis regulation at the Xic in the light of its topological organization, we found that the cis-regulatory landscape of Xist actually includes sequences separated from the Xist promoter by a TAD boundary and located almost 200 kb away in the neighboring TAD. This was surprising, as current views posit that TAD boundaries prevent communication between cis-regulatory elements and genes in neighboring TADs, thus working as powerful insulator elements. While this is the case for a subset of loci investigated to date (Flavahan et al., 2016; Franke et al., 2016; Gröschel et al., 2014; Hnizd et al., 2016; Lupiánnez et al., 2015; Northcott et al., 2014; Vicente-García et al., 2017), including the Xic (van Bemmel et al., 2019; Nora et al., 2012), our results suggest that TAD boundaries are not completely impermeable to cis-regulation, a concept that
is supported as well by other studies (Despang et al., 2019; Diao et al., 2017; Groff et al., 2018; Kragesteen et al., 2018; Tsujimura et al., 2015). Depending on the nature of cis-regulatory elements (i.e., the factors they bind), the topological organization of the genome might be more or less important for their activity. Our study reveals that the Tsix-TAD is a Xist-repressive landscape and that this landscape is presumably required to temper the activation of Xist during the onset of XCI, where Xist expression must be rendered monoallelic. Our discovery that a conserved element can act as a Xist repressor in the Tsix-TAD and a Xist activator in the Xist-TAD highlights the importance of Xic topological partitioning (further discussed below).

We have identified that the promoter region of the Linx lncRNA locus (LinxP), which lies within the Tsix-TAD, negatively regulates Xist expression, and it does this independently of any effect on Tsix expression. Furthermore, unlike other regulators of Xist, such as Jpx, Ftx, and Tsix, which have been reported to regulate Xist in cis via their transcripts or transcription (reviewed in Galup and Heard, 2015), LinxP regulates Xist in cis in a manner independent of Linx transcripts or transcription. Thus, even though Linx produces an 80-kb-long lncRNA, the element that regulates Xist appears to act independently of this RNA. We found that the LinxP element acts as a long-range, negative regulator of Xist. However, whether this inter-TAD cis-regulation between neighboring TADs involves physical contacts still remains an open question. Contacts between TADs have been detected ever since their discovery; the difference between interaction frequency within TADs and across TAD boundaries is ~2-fold only. Inter-TAD contacts have also been observed with single-cell Hi-C (Nagano et al., 2013), high-resolution microscopy (Bintu et al., 2018; Giorgetti et al., 2014) and a crosslink-free and ligation-free approach (Redolfi et al., 2019). We were able to detect contacts between LinxP and the Xist promoter, but these do not occur at higher frequency than between neighboring sequences (Figures 4E and 4F). It should also be noted that inter-TAD contacts do not imply inter-TAD regulation, as illustrated by a recent study (Despang et al., 2019), and that inter-TAD regulation does not have to require inter-TAD contacts. Indeed, it has recently been suggested that cis-regulatory elements can employ a variety of mechanisms to control their targets, some independent of 3D proximity with their target (Alexander et al., 2019; Benabdallah et al., 2019). Thus, it is possible that LinxP-mediated regulation of Xist happens without direct physical proximity between the loci (although it is nevertheless influenced by the topological organization of the Xic, as discussed below). The communication between LinxP and Xist might rely on alternative mechanisms, such as nuclear microenvironments and/or phase-transition domains (Furlong and Levine, 2018). Indeed, the pluripotency factor Oct4, which binds LinxP, has been implicated in such phase-separation mechanisms (Boija et al., 2018).
Our finding that the LinxP cis-regulatory element has a different effect on Xist depending on which side of the TAD boundary it is located is very intriguing. In its endogenous location, within the Tsix-TAD, LinxP acts as a silencer. We show that this silencing effect acts independently of Tsix’s repression of Xist. Silencers have been largely underappreciated in the transcriptional regulation field, despite the first examples being reported more than 30 years ago in yeast, flies, birds, and mammals (Baniahmad et al., 1987; Brand et al., 1985; Cao et al., 1989; Doyle et al., 1989; Nakamura et al., 1989; Saffer and Thurston, 1989) and a recent attempt to map silencers across the mouse and human genomes (Jayavelu et al., 2018). Silencers are similar to enhancers in that they normally act in an orientation-independent way and overlap DNA hypersensitive sites, but they repress, rather than activate, their target genes; we did observe these properties for LinxP. Silencers’ mechanisms of action are not fully understood, but they can act either at short or long distances (or both) (Gray and Levine, 1996; Li and Arnosti, 2011; Perry et al., 2011; Studer et al., 1994; Weintraub et al., 1995). LinxP’s repressive action occurs at a distance of ~170 kb and across a TAD boundary. Consistent with this action on Xist, LinxP binds two known repressors of Xist, the pluripotency factors Nanog and Oct4. How these factors repress Xist has remained unclear (reviewed in Minkovsky et al., 2012). Linx expression is actually positively regulated by the pluripotency network, and this may be linked to the way it represses Xist. It will be interesting to understand and dissect how a transcriptionally active promoter can act as a long-range silencer of

Figure 6. The LinxP Element Is Conserved across Placental Mammals and Overlaps the Binding Site for Pluripotency Factors
(A) Sequence conservation analysis. Conservation score across placental mammals shows poor sequence conservation for Linx (compared to Cdx4), except for a few regions. Multiz alignment shows conserved stretches in green.
(B) Zoom-in from (A) of the Linx promoter region, showing two highly conserved modules across placental mammals. Nanog and Oct4 ChIP-seq, as well as DNase-seq (DNase I hypersensitive sites sequencing), are represented below (same as in Figure 2B).
(C) Synteny analysis across placental mammals and opossum of the two conserved modules identified in (B). Note that they are highly syntenic in placental mammals, lying close to Cdx4 and Xist on the X chromosome. In the marsupial opossum, the conserved element (half of one LinxP module) lies on chromosome 2, while Cdx4 and Rsx (the marsupial equivalent to Xist) lie on the X chromosome. Genomes of species marked with an asterisk (*) are shown here in inverse orientation to what is annotated in UCSC for clarity purposes. Each species is designated by the first letter of its genus (in capital) and the first three letters of its specific epithet; the order of the species is the same as in (B), where they are designated by their common names. Evolutionary distance is represented in million years (Ma).
another gene, especially in the light of recent models of gene expression that involve the clustering of cis-regulatory elements and promoters into condensates (Pliss and Kingston, 2018). It is important to note that LinxP is a negative modulator of Xist activity rather than a complete repressor, as its deletion leads not to Xist activation in all cells but simply to a bias in random monoallelic Xist expression.

When we inserted LinxP in the same TAD as the Xist promoter (and also at the same distance of ~170 kb), it actually enhanced Xist expression in cis rather than repressing it. cis-Regulatory elements that can act as both silencers and enhancers have already been reported, and this behavior has been shown to depend on the combination of factors binding to them at different developmental stages (Brand et al., 1987; Jiang et al., 1993; Kirov et al., 1993; Gisselbrecht et al., 2019). In the case of LinxP, this dual activity is present in the same cell type, but it is dependent on the TAD in which the LinxP element is located. We speculate that the different ways the Xist promoter responds to LinxP are associated to topology; the TAD boundary at the Xic might not be merely separating cis repressors and cis activators on each side of the Xist promoter but might actually be determining whether they act as silencers or enhancers. In other words, different environments created by different TADs may define how certain controlling elements mediate their effects. This could have important implications in the context of cell-to-cell variability and fluctuations of the topological structure of chromosomes over time (Fudenberg and Minny, 2012; Giorgetti et al., 2014), implying that a cis-regulatory element could be exploited as either a silencer or an enhancer depending on the topological organization of the locus at a given time point. Further functional studies will allow us to test such hypotheses.

Besides harboring a long-range regulator of Xist, the Linx locus is also involved in (1) regulating Cdx4, located ~10 kb upstream of Linx; and (2) shaping the topological organization of the Xic. We show that these two regulatory functions of Linx are genetically uncoupled from Xist regulation. Moreover, while Xist regulation does not depend on transcription across the Linx locus, regulation of Cdx4 and Xic topology are associated with Linx transcription or IncRNA. In summary, the Linx locus produces a lncRNA, and its transcription can influence TAD structure and nearby gene activity. In addition, the LinxP element at the 5’end of Linx is conserved and a regulator of Xist, which acts as a TAD context-specific modulator of Xist expression and choice making during XCI. The multifaceted Linx locus illustrates the remarkable complexity and finesse of cis-regulatory landscapes required to orchestrate appropriate gene expression during development. It also highlights the importance of careful dissection of noncoding loci (Anderson et al., 2016; Bassett et al., 2014; Engreitz et al., 2016; Paralkar et al., 2016; Ritter et al., 2019).

Finally, our study provides some important and intriguing perspectives on the mechanisms and evolution of cis-regulatory elements. Random XCI is present in all species of placental mammals examined to date, yet elements previously identified in the mouse for choice making (e.g., Tsix and Xite) do not seem conserved across most of the other species (Galupa and Heard, 2018; Migeon et al., 2002; Peeters et al., 2016). Here, we identified a novel regulator of XCI choice that is conserved across placental mammals, both in sequence and location within the Xic. Thus the Linx promoter could be the ancestral cis regulator of Xist monoallelic expression, maybe with increased relevance in species that lack Tsix. The TAD boundary that separates the Linx elements from the Xist promoter in the mouse is conserved in humans (Galupa and Heard, 2018), suggesting that this too could be an ancestral feature and may be of importance for the choice-making process during XCI. Inter-TAD regulation could be particularly relevant for such fine-tuned developmental decisions, and evolution might have favored the positioning of elements responsible for choice-making processes (such as those within the Linx locus) in a separate TAD to the promoter they control. We note that other critical developmentally associated loci also display bipartite TAD organization, as reviewed previously (Galupa and Heard, 2017), suggesting that regulatory crosstalk between neighboring TADs might be another core feature of gene regulation during development. Further dissection of mechanisms through which elements within the Tsix-TAD regulate the Xist promoter in the neighboring TAD will certainly provide new insights into the fundamental principles of cis-regulatory control.

STAR METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Tissue culture
  - Mouse experimentation
- METHOD DETAILS
  - Genomic engineering of mice and mESC
  - RNA and DNA fluorescent in situ hybridization (FISH)
  - Gene expression analysis
  - ATAC-seq (assay for transposase-accessible chromatin using sequencing)
  - Flow cytometry analysis
  - Sequence conservation and synteny analysis
  - Chromosome conformation capture techniques
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gene expression analysis
  - ATAC-seq (assay for transposase-accessible chromatin using sequencing)
  - Chromosome conformation capture techniques
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2019.10.030.

ACKNOWLEDGMENTS

We are grateful to Katia Ancelin and Isabelle Grandjean for help and advice with animal management; Lucile Marion-Poil for help and advice with flow cytometry experiments; Maud Boersenstejn for scientific discussions as well as help with mouse genotyping; and Denis Kndjia, Katia Ancelin, Inês Pinheiro,
and Simão da Rocha for critical reading of the manuscript. We thank all members of the Heard lab for advice, support, and helpful comments and discussions, in particular Catherine Corbel, Aurélie Bousard, Jan Zylicz, Laia Richart, Anne-Valérie Gendrel, Benjamin Foret, Tim Pollex, and Edda Schulz. We are also thankful to facilities at the Institut Curie, including the Mouse Facility, the Flow Cytometry Platform, the BBD team of PICT-IBISA, the NGS Platform, the Genomics Platform (in particular David Gentien, Cécile Reyes, Audrey Rapanit, and Benot Albaud), and the Bioinformatics Platform. We acknowledge the Zhang lab for sharing plasmids and the ENCODE Consortium and the Bruneau, Ren, Sharp, Stamatoyannopoulos, and Young labs for generating datasets used in this study. Finally, we wish to thank our anonymous reviewers, who provided critical comments that substantially improved the clarity and breadth of this work. R.G. would like to dedicate this article to Luísia Supico (1963–2019) and her inspiring mentorship, righteous indignation, and precious friendship.

This work was supported by fellowships from Région Ile-de-France (DIM Biothérapies) and Fondation pour la Recherche Médicale (FDT20160432595) to R.G.; NWO-ALW Rubicon (825.13.002) and Veni (863.15.018) fellowships to J.G.v.B.; and an ERC Advanced Investigator award (ERC-2014-AAdG no. 671027), Labellisation La Ligue, FRM (DEI20151234398), and ANR DoseX 2019.07.038. Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science 362, eaau1783.


Molecular Cell 77, 352–367, January 16, 2020 365


## STAR METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial and Virus Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TaKaRa E.coli DH5α Competent Cells</td>
<td>TaKaRa</td>
<td>Cat.# 9057</td>
</tr>
<tr>
<td>One Shot Stbl3 Chemically Competent E. coli</td>
<td>Invitrogen</td>
<td>C737303</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PyroMark Gold Q24</td>
<td>QIAGEN</td>
<td>Cat#970802</td>
</tr>
<tr>
<td>nCounter® Technology</td>
<td>nanoString</td>
<td>Custom-made (van Bemmel et al., 2019)</td>
</tr>
<tr>
<td>Nick Translation Kit</td>
<td>Roche</td>
<td>Cat#10976776001</td>
</tr>
<tr>
<td>P3 Primary Cell 4D-Nucleofector X Kit</td>
<td>Lonza</td>
<td>V4XP-3024</td>
</tr>
<tr>
<td>mMESSAGE mMACHINE T7 ULTRA kit</td>
<td>Life Technologies</td>
<td>AM1345</td>
</tr>
<tr>
<td>MEGAshortscript T7 kit</td>
<td>Life Technologies</td>
<td>AM1333</td>
</tr>
<tr>
<td>TALE Toolbox kit</td>
<td>Feng Zhang lab</td>
<td>Addgene kit #1000000019</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5C, ATAC-seq, Capture-C, RNA-seq</td>
<td>This Paper</td>
<td>GSE124596</td>
</tr>
<tr>
<td>CTCF ChIP-seq</td>
<td>Nora et al., 2017</td>
<td>GSE98671</td>
</tr>
<tr>
<td>RNA Pol2 ChIP-seq</td>
<td>Seila et al., 2008</td>
<td>GSE12680</td>
</tr>
<tr>
<td>Med1 ChIP-seq</td>
<td>Kagey et al., 2010</td>
<td>GSE22557</td>
</tr>
<tr>
<td>Nanog, Oct4 and Sox2 ChIP-seq</td>
<td>Marson et al., 2008</td>
<td>GSE17124</td>
</tr>
<tr>
<td>DNase-seq, and H3K27Ac and p300 ChIP-seq</td>
<td>Ren and Stamatoyannopoulos labs</td>
<td>Mouse ENCODE Consortium Yue et al., 2014</td>
</tr>
<tr>
<td><strong>Experimental Models: Cell Lines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mESC E14</td>
<td>Heard lab</td>
<td>129/Ola</td>
</tr>
<tr>
<td>mESC E14 AΔ245kb</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC E14 AΔLinxE</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC E14 AΔLinxP</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC E14 ΔOrix</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC E14 Linxp-inv</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC E14 Linx-stop</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC Pgk12.1</td>
<td>Brockdorff lab</td>
<td>129/PGK</td>
</tr>
<tr>
<td>mESC Pgk12.1 LinxP-60kb</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC Pgk12.1 LinxP-170kb</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC Pgk12.1 AΔ245kb</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC Pgk12.1 ΔCdix4P</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC Pgk#106</td>
<td>Heard lab</td>
<td>Masui et al., 2011</td>
</tr>
<tr>
<td>mESC Pgk#106 ΔLinxP</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC LF2</td>
<td>Heard lab</td>
<td>129/Ola</td>
</tr>
<tr>
<td>mESC XGTC</td>
<td>Gribnau lab</td>
<td>Loos et al., 2016</td>
</tr>
<tr>
<td>mESC XGTC ΔLinxP</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>mESC XGTC AΔ245kb</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td><strong>Experimental Models: Organisms/Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse: B6D2F1/J</td>
<td>Jackson laboratory</td>
<td>#100006</td>
</tr>
<tr>
<td>Mouse: JF1/Ms</td>
<td>Heard lab</td>
<td>-</td>
</tr>
<tr>
<td>Mouse: B6D2F1/J: AΔ245kb</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
</tbody>
</table>

(Continued on next page)
### REAGENT or RESOURCE SOURCE IDENTIFIER

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse: B6D2F1/J: ΔLinxCBS</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>Mouse: B6D2F1/J: ΔLinxE</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>Mouse: B6D2F1/J: ΔLinxP</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>Mouse: B6D2F1/J: ΔOrix</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
<tr>
<td>Mouse: B6D2F1/J: LinxP-inv</td>
<td>Heard lab</td>
<td>This Paper</td>
</tr>
</tbody>
</table>

### Oligonucleotides

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotyping primers</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>qPCR primers</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>AQ primers</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>Capture-C probes</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>Xist exons probe</td>
<td>Roche</td>
<td>Custom-made</td>
</tr>
</tbody>
</table>

### Recombinant DNA

<table>
<thead>
<tr>
<th>Recombinant DNA</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>pX459-v2</td>
<td>Feng Zang lab</td>
<td>Addgene #62988</td>
</tr>
<tr>
<td>pX459-v2: sgRNAs</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>pEN471</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>pJF1, pJF2, pJF3, pJF4</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>pFX5, pFX6, pFX7, pFX8</td>
<td>This Paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>BAC Huwe1</td>
<td>BACPAC Resources Center</td>
<td>RP24-157H12</td>
</tr>
<tr>
<td>pEN1</td>
<td>Heard lab</td>
<td>Nora et al., 2012</td>
</tr>
<tr>
<td>Linx-intron1 fosmid</td>
<td>BACPAC Resources Center</td>
<td>wi1-1985N4</td>
</tr>
<tr>
<td>pEN9</td>
<td>This paper</td>
<td>See Table S1</td>
</tr>
<tr>
<td>pSPO2/FAB/TetO</td>
<td>Heard lab</td>
<td>Masui et al., 2011</td>
</tr>
<tr>
<td>pLG10</td>
<td>Heard lab</td>
<td>Giorgetti et al., 2014</td>
</tr>
<tr>
<td>p510</td>
<td>Heard lab</td>
<td>Rougeulle et al., 1994</td>
</tr>
</tbody>
</table>

### Software and Algorithms

<table>
<thead>
<tr>
<th>Software and Algorithms</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAR mapper</td>
<td>McCarthy et al., 2012</td>
<td>v2.5.2b</td>
</tr>
<tr>
<td>edgeR package</td>
<td>McCarthy et al., 2012</td>
<td>v3.20.1</td>
</tr>
<tr>
<td>limma R package</td>
<td>Ritchie et al., 2015</td>
<td></td>
</tr>
<tr>
<td>Illumina bcl2fastq software</td>
<td><a href="https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html">https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html</a></td>
<td>v2.20.0</td>
</tr>
<tr>
<td>STAR mapper</td>
<td>Dobin et al., 2013</td>
<td>v2.4.2a</td>
</tr>
<tr>
<td>PICARD tools</td>
<td><a href="http://broadinstitute.github.io/picard">http://broadinstitute.github.io/picard</a></td>
<td>v1.90</td>
</tr>
<tr>
<td>MACS2</td>
<td>Zhang et al., 2008</td>
<td>v2.1.0</td>
</tr>
<tr>
<td>IDR</td>
<td><a href="https://github.com/nboley/idr">https://github.com/nboley/idr</a></td>
<td>v2</td>
</tr>
<tr>
<td>BEDTools</td>
<td>Quinlan and Hall, 2010</td>
<td>version 2.26.0</td>
</tr>
<tr>
<td>GVIZ</td>
<td>Hahne and Ivanek, 2016</td>
<td>v1.22.3</td>
</tr>
<tr>
<td>5C-Pro</td>
<td><a href="https://github.com/bioinfo-pf-curie">https://github.com/bioinfo-pf-curie</a> /5C-Pro</td>
<td></td>
</tr>
<tr>
<td>bowtie2</td>
<td>Langmead and Salzberg, 2012</td>
<td></td>
</tr>
<tr>
<td>HiTC BioConductor package</td>
<td>Servant et al., 2012</td>
<td></td>
</tr>
<tr>
<td>Neighborhood CV package</td>
<td><a href="https://github.com/zhanyinx/Coefficient_Variation">https://github.com/zhanyinx/Coefficient_Variation</a></td>
<td></td>
</tr>
<tr>
<td>Trim Galore!</td>
<td><a href="http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a></td>
<td></td>
</tr>
<tr>
<td>HiC-Pro</td>
<td>Servant et al., 2015</td>
<td>v2.8.0</td>
</tr>
</tbody>
</table>
LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Edith Heard (edith.heard@embl.org). There are no specific restrictions regarding the sharing of materials generated in this study.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tissue culture

Culture conditions

Feeder-independent mESC lines (E14, Pgk12.1, LF2 and clones derived from them) were grown on flasks or dishes coated with 0.1% (wt/vol) gelatin. The XGTC mESC line is feeder-dependent (Loos et al., 2016) and was grown on a mono-layer of mitomycin C-treated male MEFs. Culture media consisted in DMEM (GIBCO) except for E14, which were grown in Glasgow medium supplemented with 2mM L-Glutamine, 0.1mM nonessential amino acids and 1mM sodium pyruvate. All mESC media contained 15% FBS (GIBCO), 0.1 mM b-mercaptoethanol (Sigma) and 1000 U/mL of LIF (Chemicon). All cells were cultivated at 37°C under 8% CO₂ and passaged according to their confluency, generally every other day. Medium was refreshed daily.

Early differentiation assays

mESC were washed with 1x PBS, incubated with trypsin at 37°C (E14: 20min; Pgk12.1, LF2 and XGTC: 12 min) and resuspended in ES medium without LIF. After cell counting, desired number of cells was resuspended in differentiation medium and seeded. Differentiation medium was either “AF differentiation medium,” consisting of N2B27 medium, 20 ng/mL activin A (R&D) and 12 ng/mL FGF-basic (R&D); or “Fibro differentiation medium,” consisting of DMEM, 10% FBS, 0.1 mM b-mercaptoethanol and 100 U/mL penicillin-streptomycin. For E14 and derived clones, 8*10^5 cells per well were seeded in a fibronectin-coated (10 μg/mL, Millipore) 6-well plate in AF differentiation medium. For Pgk12.1 and derived clones, 2*10^5 cells per well were seeded in a gelatin-coated 6-well plate in AF differentiation medium. For LF2, XGTC and derived clones, 2*10^5 cells per well were seeded in a gelatin-coated 6-well plate in Fibro differentiation medium. For all differentiation assays, medium was changed daily and cells were washed in PBS before collection to remove dead cells.

Mouse experimentation

Permissions

Animal care and use for this study were performed in accordance with the recommendations of the European Community (2010/63/UE) for the care and use of laboratory animals. Experimental procedures, including genomic engineering (see below), are in compliance with international guidelines and were specifically approved by the ethics committee of the Institut Curie CEEA-IC #118 and given authorization by the French national authorities (references: APAFiS##13962-2018030717538778-v2 and APAFiS#8812-2017020611033784-v2).

Manipulation

Postimplantation embryos were collected at E8.5-10.5 stages, assuming plugging at midnight. Females with a vaginal plug were weighted every other day and only taken for dissection if a significant increase in weight was observed (~2g for B6D2F1 mice, ~1g for JF1 mice) at expected time of E8.5-E10.5 development. Extraembryonic tissues were taken for sexing the embryos. Whole embryo proper was washed three times in 1xPBS before frozen for allelic expression analysis.

METHOD DETAILS

Genomic engineering of mice and mESC

Plasmids

Deletions and inversions were generated using TALENs (mESC) or CRISPR-Cas9 (mESC and mice) technologies. We designed TALENs and sgRNAs to flank the region of interest; Table S1 contains the sequences of TALENs and sgRNAs for each engineered locus. For TALEN assembly, we used the TALE Toolbox kit (Kit # 1000000019; Addgene) and the protocol described in (Sanjana et al., 2012), except that the TALEN backbones were modified to contain a CAGGS promoter instead of the default CMV promoter. TALEN constructs were amplified upon transformation of Shot Stbl3 Chemically Competent E. coli (Life Technologies) according to manufacturer’s specifications, and sequenced for verifying correct assembly. Bacteria were grown at 30°C to minimize recombination events. For cloning sgRNAs, we used pX459-v2 (Plasmid #62988; Addgene) and protocol from the Zhang lab (https://media.addgene.org/cms/filer_public/e6/5a/e65a9ef8-c8ac-48f8-98da-3b7d7960394c/zhang-lab-general-cloning-protocol.pdf). sgRNA constructs were amplified upon transformation of DH5α competent cells (Takara) grown at 37°C, and sequenced for verifying correct cloning. Midipreps were prepared at final concentration > 1mg/mL using the NucleoBond Xtra Midi Plus kit (Macherey-Nagel). Knock-ins were generated via CRISPR/Cas9 mediated homologous recombination; Table S1 contains the sequences of sgRNAs used for each engineered locus. Donor plasmids were generated with standard cloning techniques; they are listed in Table S1 and their sequences can be found in the folder “Knockin-plasmid-sequences” accompanying this manuscript.
For knock-outs and inversions, mESC were transfected with TALEN or sgRNA constructs using the P3 Primary Cell 4D-Nucleofector X Kit (V4XP-3024) and the Amaxa 4D Nucleofector system (Lonza). We used the transfection program CG-104 for E14, LF2 and XGTC and CG-110 for Pgk12.1. Each transfection included 5 million cells resuspended in the nucleofection mix (prepared according to manufacturer’s instructions) containing 2.5 µg of each TALEN (four constructs) or 5 µg of each sgRNA (two constructs). For knock-ins, half a million cells were reverse-transfected with 3 µL of Lipofectamine-2000 (ThermoFisher) complexed with 0.5 µg of sgRNA construct and 1.5 µg of donor plasmid. As a transfection control, 10 µg of pmaxGFP (Lonza) were used, for which the nucleofection efficiency was around 90% (E14, LF2, XGTC) or 50% (Pgk12.1). For knock-outs and inversions, cells were immediately resuspended in pre-warmed culture medium after nucleofection and seeded at three serial 10x dilutions in 10-cm dishes to ensure optimal density for colony-picking. Transfected cells were selected with puromycin for 48h, and grown for 8–10 days. For knock-ins, cells were only diluted one day after transfection, and puromycin selection was started 3–4 days after dilution. Single colonies or pools of colonies were picked into 96-well plates. Genomic DNA was isolated in 96-well plates for PCR-based screening of deletions and inversions; Table S1 contains the sequences of genotyping primers for each engineered locus. The strategy was inspired on the Epigenesys protocol by Nora and Heard, 2012, described in https://www.epigenesys.eu/images/stories/protocols/pdf/20130507072445_p62.pdf. Positive clones for female cell lines were subsequently re-seeded at single-cell dilution in 96-well plates, followed by a new PCR screening, to ensure monoclonal colonies. For knock-ins, selection marker was subsequently removed by reverse lipofection with a flipase plasmid and clones were checked for puromycin sensitivity. We sequenced the PCR products from the deletion/inversion alleles to determine their exact location and, for females, the allele of the respective deletion/inversion. For knock-ins, both left and right side of the insertion were sequenced. Wild-type alleles were also sequenced, to ensure their integrity. Table S1 contains a summary of these sequencing results, including the coordinates of the deletions/inversions for each engineered locus.

**Engineering mESC**

The mouse mutant lines were generated following the strategy described in (Wang et al., 2013) with minor modifications. Cas9 mRNA was in vitro transcribed from a T7-Cas9 pCR2.1-XL plasmid (Greenberg et al., 2017) using the mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies) and purified with the RNeasy Mini kit (QIAGEN), or bought from Tebu-bio (L-7206). The sgRNAs were amplified by PCR with primers covering a 5’ T7 promoter sequence from the plasmids used for mESC transfection (Table S1). After gel purification, the T7-sgRNA PCR products were used as the template for in vitro transcription with the MEGAscript T7 kit (Life Technologies) and the products were purified using the MEGAclear kit (Life Technologies). Cas9 mRNA and the sgRNAs were eluted in DEPC-treated RNase-free water, and their quality was assessed by electrophoresis on an agarose gel after incubation at 95°C for 3min with denaturing agent provided with the in vitro transcription kits. Cas9 mRNA and sgRNAs (at 100 ng/µl and 50 ng/µl, respectively) were injected into the cytoplasm of mouse B6D2F1 zygotes from eight-week-old superovulated B6D2F1 (C57BL/6J x DBA2) females mated to stud males of the same background. Zygotes with well-recognized pronuclei were collected in M2 medium (Sigma) at E0.5. Injected embryos were cultured in M16 medium (Sigma) at 37°C under 5% CO2, until transfer at the one-cell stage the same day or at the two-cell stage the following day to the infundibulum of the oviduct of a pseudogestant CD1 female at E0.5 (25–30 embryos were transferred per female). All weaned mice (N0) were genotyped for presence of deletion or inversion alleles; Table S1 contains the sequences of genotyping primers for each engineered locus. Mice carrying engineered alleles were crossed to B6D2F1 mice and their progeny screened again for the presence of an engineered allele – in some cases, up to 6 different alleles were found from a single N0 mouse. We sequenced the PCR products of the engineered allele to determine the exact location of the deletion/inversion (Table S1 contains a summary of these results). The F1 mice were considered the “founders” and bred to B6D2F1 mice; their progeny was then intercrossed to generate homozygous mice and lines were kept in homozygosity.

**RNA and DNA fluorescent in situ hybridization (FISH)**

**On cells from tissue culture**

FISH was performed as described previously with minor modifications (Chaumeil et al., 2008). Briefly, undifferentiated or differentiating mESCs were grown on gelatin-coated coverslips or dissociated using accutase (Invitrogen) and adsorbed onto Poly-L-Lysine (Sigma) coated coverslips #1.5 (1mm) for 5 min. Cells were fixed with 3% paraformaldehyde in PBS for 5 min at room temperature and permeabilized for 5 min on ice in PBS containing 0.5% Triton X-100 and 2mM Vanadylribonucleoside complex (New England Biolabs). Coverslips were preserved in 70% EtOH at –20°C. For RNA FISH, coverslips were dehydrated through an ethanol series (80%, 95%, and 100% twice) and air-dried quickly, then lowered onto a drop of the probe/hybridization buffer mix (50% Formamide, 20% Dextran sulfate, 2x SSC, 1 µg/µl BSA, 10mM Vanadyl-ribonucleoside) and incubated overnight at 37°C. For RNA/DNA FISH, the coverslips were first washed three times in 2 x SSC and incubated for 1h at 37°C in 2 x SSC supplemented with 0.1 mg ml⁻¹ RNase A (Fermentas) and 10 U ml⁻¹ RNase H (New England Biolabs). After the RNase treatment, the coverslips were dehydrated through an ethanol series (80%, 95%, and 100% twice). Before hybridization, cells on coverslips were denatured for 38 min at 80°C in 50% formamide in 2 x SSC (pH 7.2-7.4) and then quickly transferred to ice and washed three times in ice-cold 2 x SSC. Coverslips were then lowered onto a drop of probe/hybridization buffer mix (as described for RNA FISH) and incubated overnight at 42°C. The next day, coverslips were washed three times at 42–45°C in 50% formamide in 2 x SSC (pH 7.2-7.4) and three times at 42–45°C in 2 x SSC. Nuclei were counterstained with DAPI (0.2mg/ml), coverslips were mounted (90% glycerol, 0.1X PBS, 0.1% p-phenylenediamine at pH9), and cells were imaged using a wide-field DeltaVision Core microscope (Applied Precision).
On mouse embryos

RNA FISH on mouse embryos was performed as described previously with minor modifications (Borensztein et al., 2017; Ranisavljevic et al., 2017). Embryos were recovered at E3.5-E4.5 by flushing the uterus with M2 medium (Sigma) and/or by dissection from the uterus. Zona pellucida was removed using acidic Tyrode’s solution (Sigma), and embryos were washed twice with M2 medium (Sigma). ICM was then isolated by immunosurgery, by culturing blastocysts without zona pellucida in anti-mouse red blood cell serum from rabbit (Rockland) for 30 min then in guinea pig complement serum (Sigma) for 15–30 min. For consistency, ICMs from both wild-type and homozygous knockout embryos (from separate crosses) were placed in different regions of the same coverslip before the FISH procedure (Ranisavljevic et al., 2017).

Probes

A list of RNA and DNA FISH probes used for this study can be found in Table S1. Plasmid, fosmid and bacterial artificial chromosome (BAC)-derived probes were labeled using the Nick Translation kit from Abbot and following manufacturer’s instructions. Probes were either ethanol-precipitated or vacuum-dried and resuspended in formamide with shaking at 37°C. BAC- and fosmid-derived probes were co-precipitated with mouse Cot-1 DNA (Invitrogen), and competition to block repetitive sequences was performed for at least 20 min at 37°C, and after denaturation (75°C, 10 min). Probes were then run with one volume of 2 × hybridization buffer. Probes not requiring competition were denatured at 75°C for 10 min and stored on ice until mixed with one volume of 2 × hybridization buffers.

Gene expression analysis

Time points

Cells were collected for gene expression analysis at different time points of differentiation. For XY mESC (E14 and derived clones): 0h, 12h, 24h, 36h, 48h and 60h of differentiation; for XX mESC (Pkg12.1, LF2, XGTC and derived clones): 0h, 24h, 48h, 72h and 96h of differentiation. Embryos were collected at E8.5-10.5.

Total RNA extraction for cells

Cells were lysed with Trizol (Invitrogen), and RNA was isolated using the RNAeasy Mini kit (QIAGEN), including DNase treatment. RNA samples were systematically run on an agarose gel to check their integrity.

Total RNA extraction for embryos

Embryos were lysed in RLT buffer (QIAGEN) supplemented with 0.01% 2-mercaptoethanol, and after two rounds of vortexing (15 sec each), lysates were applied directly to a QIAshredder spin column (QIAGEN) and centrifuged for 3 min at full speed. RNA was extracted using the RNAeasy Mini kit (QIAGEN), including DNase treatment, and following manufacturer’s instructions. RNA samples were systematically run on an agarose gel to check their integrity.

Reverse transcription

cDNA was synthesized from 0.5 µg of RNA using SuperScript III Reverse Transcriptase and random primers (both Invitrogen) according to the manufacturer’s recommendations. Two independent reverse transcription experiments were carried out for each sample, pooled at the end and diluted 25-fold prior to qPCR or allelic expression analysis. No-reverse transcription controls were performed in parallel.

nCounter analysis

We used the NanoString nCounter gene expression system (Geiss et al., 2008) to systematically characterize transcriptional differences in wild-type and mutant mESC, prior or during differentiation. We used 500ng of total RNA from each sample for each nCounter hybridization round. We designed a customised probe codeset to identify nearly a hundred transcripts from Xic genes, other X-linked genes, pluripotency factors, differentiation markers, proliferation markers and normalization genes (see Table S1; also published in van Bemmel et al., 2019). Standard positive controls included in the kit were used for scaling the raw data. Genes Actb, Rrm2 and Sdha were used for normalization. Differential expression was always calculated for samples run on the same nCounter hybridization.

RT-qPCR

qPCR on cDNA was performed on a ViIA7 system (Applied Biosystems) using the 2x SYBR Green Master Mix (Applied Biosystems), 2.5uL cDNA and validated primers (final concentration: 0.1 µM) in a reaction volume of 10 µL. Appropriate no-reverse transcription and no-cDNA controls were performed in parallel. All primers used were validated using standard curves (see Table S1 for a list of the primers used in this study). A threshold of 0.3 was used for determining the quantification cycle for all genes, except for Chic1, for which 0.2 was used. Normalization of gene expression levels was done using the geNorm method (Vandesompele et al., 2002) and ArpP0, Rrm2 and Gapdh used as reference genes.

Allelic expression analysis

cDNA from XX samples (cells or embryos) was PCR-amplified with biotinylated primers and pyrosequenced for allele quantification on a Pyromark Q24 system (QIAGEN). The same PCR was done on no-reverse transcription control samples to confirm absence of genomic DNA contamination. All primers used were designed using the PyroMark Assay Design software and validated on XX polymorphic genomic DNA for a ratio of 50%:50% (±4%). List of primers and SNPs used for allele quantification can be found in Table S1.

RNA-sequencing

RNA-seq libraries were prepared from 500 ng of DNase-treated total RNA (RIN = 10) using the TruSeq Stranded Total RNA kit (Illumina). Sequencing was performed using paired-end reads (PE100) in a NovaSeq System (Illumina).
ATAC-seq (assay for transposase-accessible chromatin using sequencing)

Library preparation and sequencing: ATAC-seq libraries were prepared following (Buenrostro et al., 2013) with some modifications. Fifty thousand cells were washed with cold 1xPBS twice and then resuspended directly in the transposase reaction (step with lysis buffer was omitted to reduce mitochondrial DNA content of the library). Transposase reaction was performed at 37°C for 45 minutes. DNA was purified with MinElute column (QIAGEN) and PCR amplified for 12 cycles using barcode-specific primers for each library. Total number of PCR cycles was determined by running 5 initial cycles and then monitoring the amplification of an aliquot using qPCR and the same PCR mix supplemented with 1xEvaGreen dye (Biotium) to determine additional number of PCR cycles. Amplified libraries were purified with MinElute column (QIAGEN), followed by two rounds of purification using Agencourt AMPure XP beads (A63881, Beckman Coulter) at a ratio of 1:1.6. Libraries were sequenced on a Nextseq 500 platform, with 75bp paired-end reads. Information on the sequencing reads can be found in Table S1.

Flow cytometry analysis

Single-cell suspensions in 1xPBS were prepared after accutase treatment for 5 min at 37°C. Duplets were excluded by appropriate gating. Relative fluorescence intensities were determined for EGFP and mCherry, using Blue-B-530/30 and Green-D-610/20 filters, on an LSRFortessa instrument with FACSDiva software. Subsequent analysis was performed with FlowJo.

Sequence conservation and synteny analysis

Conservation score across placental mammals – Basewise Conservation, PhyloP (Siepel et al., 2005) and Multiz alignments (Blanchette et al., 2004) were retrieved from UCSC Genome Browser (http://genome.ucsc.edu/). To determine the chromosomal position of the conserved LnxP elements, sequences for each available species were manually extracted and curated from the Multiz alignment (sequences available in Table S1) and then blasted against respective genome using BLAT in the UCSC Genome Browser (Kent, 2002).

Chromosome conformation capture techniques

3C templates

3C libraries were prepared based on previous protocols (Nora et al., 2017; Rao et al., 2014), with some modifications. Crosslinked cells (in 2% Formaldehyde; 10 million for each sample) were lysed in 10 mM Tris–HCl, pH 8, 10 mM NaCl, 0.2% NP-40, 1 × complete protease inhibitor cocktail (Roche) for 15 min on ice. Nuclei were resuspended in 100 cells (in 2% Formaldehyde; 10 million for each sample) were lysed in 10 mM Tris–HCl, pH 8, 10 mM NaCl, 0.2% NP-40, 1

5C (chromosome conformation capture carbon copy)

5C was performed as described in (Nora et al., 2017), which adopts a single-PCR strategy to construct 5C-sequencing libraries from the 3C template. Briefly, four 10 μL 5C annealing reactions were assembled in parallel, each using 500 ng of 3C template, 1 μg salmon sperm (ThermoFisher) and 10 fmol of each 5C oligonucleotide in 1X NEBuffer 4 (5C set of oligonucleotides described in Nora et al., 2012). Samples were denatured at 95°C for 5 min and incubated at 48°C for 16-18h. 10 μL of 1X Taq ligase buffer with 5U Taq ligase were added to each annealing reaction followed by incubation at 48°C for 4h and 65°C for 10 min. Negative controls (no ligase, no template or no 5C oligonucleotides) were included during each experiment to ensure the absence of contamination. To attach Illumina-compatible sequences, 5C libraries were directly PCR amplified with primers harboring 50-mer tails containing Illumina sequences that anneal to the universal T3/T7 portion of the 5C oligonucleotides (Nora et al., 2017). For this, each 5C ligation reaction was used as the template for three parallel PCRs (12 PCRs total), using per reaction 6 μL of 5C ligation with 1.125 U AmpliTaq Gold (ThermoFisher) in 1X PCR buffer II, 1.8 mM MgCl2, 0.2 mM dNTPs, 1.25 mM primers in 25 μL total. Cycling conditions were 95°C for 9 min, 25 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30 s followed by 72°C for 8 min. PCR products from the same 3C sample were pooled and run on a 2.0% agarose electrophoresis gel. 5C libraries (231 bp) were then excised and purified with the MinElute Gel Extraction kit (QIAGEN). Library concentrations were estimated using TapeStation (Agilent) and Qubit (ThermoFisher), pooled and sequenced using 12 pM for the loading on rapid flow cells using the HiSeq 2500 system (Illumina). Sequencing mode was set as 20 dark cycles followed by 80 bases in single end reads (SR80). Information on the sequencing reads can be found in Table S1.
**Capture-C**

Capture-C was performed as described in (Davies et al., 2016) with some modifications. Capture probes were designed using CapSequm (Hughes et al., 2014). To prepare Capture-C libraries, 5 μg of 3C library were sonicated using a S220 focused ultrasonicator (Covaris) to 200 bp and 2.5 μg of fragmented DNA were processed with the KAPA Hyper Prep Kit (KAPA Biosystems) according to manufacturer’s instructions. Two rounds of capture of respectively 72 and 24 hours were then performed, pooling 2 μg of each indexed library and using 13 pmol of capture probes (biotinylated oligonucleotides, Integrated DNA Technologies), with the Seq-Cap EZ system (#06953212001, Roche/NimbleGen). This capture was performed according to manufacturer’s instructions, except for the first round when the volume of reagents was multiplied by the number of pooled libraries. Library size was confirmed using LabChip GXII Touch HT (Perkin Elmer) with a DNA High Sensitivity chip, and DNA concentrations were estimated using Qubit (Thermo Fisher Scientific). Capture-C libraries were sequenced on a MiSeq instrument (Illumina) using 75bp paired end reads and 5% PhiX.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**Gene expression analysis**

**RNA FISH, RT-qPCR, nCounter, allelic expression analysis**

All statistical details of experiments can be found in the figure legends, figures and/or Results, including the statistical tests used, exact value of n and what n represents.

**RNA-sequencing**

RNA sequencing reads have been aligned on the mouse reference genome (mm9) using the STAR mapper (v2.5.2b) (McCarthy et al., 2012), with the following parameters: outFilterMultimapNmax 20; outFilterMismatchNmax 999; outFilterMismatchNoverLmax 0.04; outSAMprimaryFlag OneBestScore; outMultimapperOrder Random. Read counts per gene were also generated with STAR and combined across samples to generate the raw counts table. Gene counts were filtered to be > 1 in at least one sample and normalized out by the trimmed mean of M values (TMM) using the edgeR package (McCarthy et al., 2012; Robinson et al., 2010). Differential expression was determined using the limma R package (Ritchie et al., 2015). Information on the sequencing reads can be found in Table S1.

**ATAC-seq (assay for transposase-accessible chromatin using sequencing)**

**Mapping and filters**

Demultiplexing was performed with the Illumina bcl2fastq software, version 2.20.0 (https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html). The reads were mapped with STAR 2.4.2a (Dobin et al., 2013) to the mm9 genome. A 75bp index was built using STAR’s generate_genome command and GENCODE mouse annotation, version M1 (Frankish et al., 2019). STAR parameters were as follows: (1) trimming the Nextera Transposase Adapters (clip3pAdapterSeq CTGTCTCTTATACA CATCTGACGCTGCCGACGA CTGTCTCTTATACATCCTGGAGGCCACGAGAC, clip3pAdapterMMp 0.1); (2) suppressing splice junction determination (alignIntronMax 10, alignSJoverhangMin 75, alignSJBoverhangMin 75); (3) read pairs that represented fragments of 1500bp or less were retained (alignMatesGapMax 1500); and (4) the remaining non-default parameters were: alignEndsType Extend5pOffRead1; outSAMattributes NH HI AS nM MD NM; outFilterMismatchNoverReadLmax 0.04; outFilterMismatchNoverLmax 1. After mapping, the reads were subject to further filtering. First we collapsed read duplicates with PICARD tools v1.90 (http://broadinstitute.github.io/picard), and selected only uniquely mapping reads using the flag “NH:i:1.” Then we removed chrM and any non-reference chromosomes, and retained only concordant read pairs that represented fragments >= 38bp and <= 1500bp. As a quality check, we assessed for low read duplication and a low percentage of reads mapping to chrM. We also verified that the ratio of short reads to long (> 150bp) reads was consistent with published ATAC-seq datasets for both mouse and human (i.e., approximately 1:1).

**Peak calling and reproducibility**

To identify potential open chromatin regions within the Xic region, ATAC-seq peaks were called using MACS2 v.2.1.0 (Zhang et al., 2008) and submitted to IDR (version 2; https://github.com/nboley/idr) to determine the subset of reproducible peaks. MACS2 was used to generate two types of peak lists for IDR: (1) a statistical cut-off of q-value < 0.01 was used on the pooled replicates of each time point, to generate an “oracle” peak list; and (2) for each individual replicate a “relaxed” list of true and false positives was created using a cut-off of p value < 0.1. The remaining arguments to the MACS2 callpeak command were as follows: gsize mm, nomodel, shift 100, extsize 200, keep-dup all. The blacklist regions reported by ENCODE for mm9 were removed from the MACS2 peak files using BEDTools intersect (version 2.26.0) (Quinlan and Hall, 2010). The final list of ATAC-seq peaks was determined with IDR: for a given time point, the oracle list from MACS2 and the top scoring 125,000 peaks from each replicate’s relaxed MACS2 list were input to IDR. The remaining IDR parameters were: input-file-type narrowPeak, rank p.value, idr-threshold 0.05. The subset of regions that passed the IDR threshold were used for downstream analysis.

**Differential peak analysis**

EdgeR, version 3.20.1 (McCarthy et al., 2012) was used to call differential ATAC-seq peaks between time points: days 0 versus 1, days 1 versus 2, and days 0 versus 2. To create a list of regions-of-interest for EdgeR, the IDR peaks from all time points were merged using BEDTools (version 2.26.0) merge command (Quinlan and Hall, 2010). The regions-of-interest and the ATAC-seq bam files were input to EdgeR with default parameters and an FDR of 0.01.
**Data visualization**

The processed data was visualized using the R package GVIZ, version 1.22.3 (Hahne and Ivanek, 2016), and the bam files for each time point were normalized using DeepTools bamCoverage, with parameters: normalizeUsingRPKM, binSize 20, smoothLength 60.

**Chromosome conformation capture techniques**

**5C (chromosome conformation capture carbon copy)**

Sequencing data was processed using our custom pipeline, 5C-Pro, available at https://github.com/bioinfo-pf-curie/5C-Pro. Briefly, single-end sequencing reads were first trimmed to remove Illumina adapters and aligned on an in silico reference of all pairs of forward and reverse primers using the bowtie2 software (Langmead and Salzberg, 2012). Aligned reads were then directly used to infer the number of contacts between pairs of forward and reverse primers, thus providing a 5C map at the primer resolution. Based on our previous experiments, inefficient primers were discarded from downstream analysis. Quality controls of the experiments were then performed using the HiTC BioConductor package (Servant et al., 2012). Data from biological replicates were pooled (summed) and binned using a running median (window = 30kb, final resolution = 6kb). We normalized 5C contacts for the total number of reads and filtered out outlier probes and singletons, as previously described (Hnisz et al., 2016; Nora et al., 2012; Smith et al., 2016). We also developed a novel method to exclude noisy contacts in the 5C maps, called “neighbourhood coefficient of variation,” available at https://github.com/zhanyinx/Coefficient_Variation. Considering that the chromatin fiber behaves as a polymer, the contact frequency of a given pair of genomic loci (e.g., i and j) cannot be very different from those of fragments i ± N and j ± N if N is smaller (or in the order of) than the persistence length of the chromatin fiber. Hence, a given pixel in the 5C map (which is proportional to the contact frequency between the two corresponding loci) can be defined as noisy if its numerical value is too different from those corresponding to neighboring interaction frequencies. To operatively assess the similarity of a given interaction with neighboring contacts, we calculated the coefficient of variation (CV) of contacts (pixels in the 5C map) in a 10x10 square centered on every contact. We then set out to discard pixels for which the corresponding coefficient of variation was bigger than a threshold. Given that the distribution of the coefficient of variation of all 5C samples in this study is bimodal around CV = 1, we set the CV threshold to 1. Discarded contacts appear as gray pixels in the differential 5C maps. For differential analysis between two samples of interest (generally wild-type versus mutant), we calculated the difference between Z-scores determined for each individual map (Smith et al., 2016). Samples corresponding to inversions of genomic regions were mapped to a virtually inverted map before analysis. Samples corresponding to deletions were corrected for the new distance between genomic elements; this distance-adjustment was performed along with the Z-score calculation.

**Capture-C**

Raw reads were first trimmed using the Trim Galore! pipeline (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and then processed using the HiC-Pro pipeline, v2.8.0 (Servant et al., 2015), until the detection of valid interaction products. Interaction products including the viewpoint of choice were selected using the make Viewpoint HiC-Pro utility. For plotting, interaction frequencies were normalized to the number of contacts per DpnII fragment per 10,000 total contacts within the analyzed region (chrX:100214149-101420149), followed by a running mean with a window size of 7 DpnII fragments.

**DATA AND CODE AVAILABILITY**

All next-generation sequencing data generated in this study has been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE124596, as indicated in the Key Resources Table. Codes used in this study and their availability are also indicated in the Key Resources Table.