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

Rafael Galupa, Elphege Pierre Nora, Rebecca Worsley-Hunt, Christel Picard, Chris Gard, Joke Gerarda van Bemmel, Nicolas Servant, Yinxiu Zhan, Fatima El Marjou, Colin Johanneau, et al.

Version of Record: https://www.sciencedirect.com/science/article/pii/S1097276519308081
Manuscript_eab5de3dffebcb72101396fa1547c23d

TITLE

A conserved noncoding locus regulates random monoallelic *Xist* expression across a topological boundary

AUTHOR LIST AND AFFILIATIONS, LEAD CONTACT FOOTNOTE, EMAIL ADDRESS CORRESPONDING AUTHOR

Rafael Galupa[1#], Elphège Pierre Nora[1‡#], Rebecca Worsley-Hunt[2‡], Christel Picard[1‡], Chris Gard[1], Joke Gerarda van Bemmel[1#], Nicolas Servant[3, 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 Heard[1,11*#]

[1] Mammalian Developmental Epigenetics Group, Genetics and Developmental Biology Unit, Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Paris, France.

[2] Berlin Institute for Medical Systems Biology, Max Delbruck Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany.

[3] Bioinformatics, Biostatistics, Epidemiology and Computational Systems Unit, Institut Curie, PSL Research University, INSERM U900, Paris, France.

[4] MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France.

[5] Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland.

[6] University of Basel, Switzerland.

[7] Transgenesis Facility, Institut Curie, Paris, France.

[8] Institut Curie Genomics of Excellence (ICGex) Platform, Institut Curie, Paris, France.

[9] Department of Developmental Biology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands.

[10] Department of Biology, Humboldt University, Berlin, Germany.

[11] Collège de France, Paris, France.

[‡] Equal contributions

[*] Lead contact and correspondence: edith.heard@embl.org

[#] Current address: European Molecular Biology Laboratory, Heidelberg, Germany (RG, EH); Gladstone Institute of Cardiovascular Diseases, San Francisco, USA (JGvB); Cardiovascular Research Institute and Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, USA (EPN).

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 *Xist* locus is regulated by sequences in the neighbouring 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 *Xist* that shares the same TAD as *Linx.* Unlike *Tsix, LinxP* is well conserved across mammals, suggesting an ancestral mechanism for random monoallelic *Xist* regulation. When introduced in the same TAD as *Xist*, *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 equalised 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) *Xist*, which is conserved across placental mammals and is expressed in female somatic cells from either the paternal or the maternal inactive X chromosome – reviewed in (Galupa and Heard, 2018). Embryonic XCI can be recapitulated *ex vivo* in differentiating mouse embryonic stem cells (mESC). These represent a powerful system to study the regulatory mechanisms of XCI, since *Xist* transcription is repressed in the pluripotent, undifferentiated state, while upon differentiation *Xist* 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 *Xist* expression in *cis*, including the elusive *X-controlling element* (*Xce*) (Cattanach and Papworth, 1981) as well as several control elements within the *X-inactivation centre* (*Xic*) – for review see (Galupa and Heard, 2015). These include *Tsix,* the antisense repressor of *Xist*, and its enhancer *Xite*; deleting either of these loci skews XCI entirely or partially, respectively, in favour 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 (Galupa 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 (~460kb), including *Xist*, *Tsix* and *Xite*, failed to induce *Xist* upregulation in differentiating female mESC (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 ~850kb (**Fig. 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 (Galupa and Heard, 2018), seems to partition two different cis-regulatory landscapes (van Bemmel et al., 2019; Nora et al., 2012). Genes within each of the two *Xic* TADs show opposite functions in the regulation of *Xist* as well as opposite transcriptional behaviours during mESC differentiation (Nora et al., 2012). The "Xist-TAD" (~550kb) contains the *Xist* promoter and some of its known positive regulators, such as *Ftx* (Furlan et al., 2018), which all become upregulated during differentiation; this domain has probably evolved as a hub of positive regulators of *Xist*. On the other hand, the "Tsix-TAD" (~300kb) 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* (**Fig. 1B**; see figure legend); this region excludes *Xite* and the *Tsix* promoter, but harbours a poorly characterised lncRNA 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 harbours cis-regulatory elements that modulate *Xist* expression and XCI choice; this *Xist*-regulatory action of *Linx* is not via the noncoding *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 harbours essential elements for endogenous *Tsix* and *Xist* regulation, we deleted a 245kb region encompassing all the loci within the Tsix-TAD except *Xite* and *Tsix* (**Fig. 1B**). This deletion does not seem to disrupt the TAD boundary nor the Xist-TAD (**Fig. S1A**). Transcriptional profiling of both control and Δ245kb 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 two days of differentiation; **Fig. 1C**). This was associated with *Xist* cloud formation in ~6% mutant male cells, which is not observed in wild type male mESC (**Fig. 1D**). Concomitantly, *Xite* and *Tsix* expression were reduced (**Fig. 1C**). Expression levels of markers for pluripotency, differentiation and proliferation were not affected (**Fig. S1B, S1C, S1D**). Therefore, the Δ245kb region contains elements that repress *Xist* and/or activate *Xite* and *Tsix*, either directly or indirectly.

To understand whether the 245kb deletion affects random XCI, we analysed heterozygous Δ245kb female ESC (**Fig. S1E**) and postimplantation embryos, derived from polymorphic mouse strains (**Fig. 1E**). Allelic ratio analyses showed that the presence of the Δ245kb region skews *Xist* expression in favour of the mutant allele (0.88 vs 0.56, $p<0.001$, **Fig. 1F** and **Fig. S1F**) and triggers preferential inactivation in *cis*, as evaluated by the expression of an X-linked gene, *Atp7a* (**Fig. 1G; Fig. S1F**). Early differentiating female mESC also displayed preferential expression of *Xist* from the Δ245kb allele (**Fig. S1E**). We conclude that this 245kb region is critical for controlling *Xist* upregulation and choice during the initiation of random XCI (see also notes in Fig. S1 legend).

We next assessed whether the Δ245kb 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 *Tsix* and *Xist* 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 245kb region on the *Xist*-GFP/*Tsix*-mCherry allele in this female mESC line (**Fig. 1H**). We found that mCherry (*Tsix*) levels were markedly reduced in Δ245kb XGTC cells compared to controls, before and after differentiation (**Fig. 1I**). The Δ245kb 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 Δ245kb allele upon differentiation (66% vs 38%; $p<0.001$) (**Fig. 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 not only contains regulators of *Tsix* but also contains elements that repress *Xist* independently of *Tsix*. This occurs despite the fact that the *Xist* promoter is located in the adjacent TAD.

Linx harbours cis-regulatory elements that modulate XCI choice independently of *Linx* transcription or RNA

Next, we set out to define the elements within the Δ245kb 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 Δ245kb 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 245kb 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 Δ245kb allele, we performed ATAC-seq (Buenrostro et al., 2013) in differentiating XX cells (d0, d1, d2) (**Fig. 2A**; **Fig. S2A**). We found strong open-chromatin sites at all known promoters within the 245kb interval (**Fig. S2A**), 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 mESC or in mice did not reveal any significant effect on *Tsix* or *Xist* expression (**Fig. S2B-D**).

None of the identified ATAC-seq peaks within the 245kb region (including *Orix*) showed significant changes during differentiation, except the promoter region of *Linx,* which showed reduced accessibility at d2 compared to d0 or d1 ($p<0.01$; **Fig. 2A**). The dynamic behaviour 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 ~2kb region centred on *Linx* TSS (ΔLinxP) in male and female mESC, as well as in mice (**Fig. 2B**; **Fig. S3A-C**; see also note in Fig. S3 legend). Differentiating (d4) ΔLinxP-heterozygous polymorphic female mESC displayed modest but significant skewing in *Xist* allelic ratios, in favour of the mutant allele (1.2-fold, $p<0.01$; **Fig. 2C**), similar to the intermediate *Xce* alleles reported to date (Galupa and Heard, 2015). Our results were consistent in both clones analysed, regardless of the strain origin of the mutated allele. We also detected preferential *Xist* cloud formation on the ΔLinxP chromosome by RNA-DNA FISH (**Fig. 2D**), implying skewed XCI choice. We observed similar results in three independent mutant clones generated in isogenic female mESC (**Fig. S3D**). Analysis of *Xist* allelic ratios in postimplantation heterozygous female embryos also revealed a slight but significant preference for *Xist* expression from the ΔLinxP allele (0.54 vs 0.48, $p<0.01$, **Fig. 2E-F**; **Fig. S3E**) and corresponding preferential *Atp7a* inactivation (0.59 vs 0.64, $p<0.01$, **Fig. 2G**; **Fig. 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 (**Fig. S2E-G** and note in Fig. S2 legend). To distinguish the contribution of the *Linx* transcript/transcription from the *LinxP* element itself, we inverted *LinxP* in mice and mESC (**Fig. 2H**), which similarly to ΔLinxP, abolished *Linx* lncRNA and transcription across the *Linx* locus (**Fig. S3F**). Unlike ΔLinxP, heterozygous *LinxP*-inv female embryos did not show bias of *Xist* or *Atp7a* allelic ratios compared to wildtype (**Fig. 2I-J, S3G**). Together, these results imply that transcription across the *Linx* locus or the *Linx* lncRNA are not mediating the effect of the *LinxP* deletion in *Xist* regulation (see also note in Fig. 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 Δ245kb mutants than in ΔLinxP 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 mESC (**Fig. 3A;** see also the first note in Fig. S4 legend). In fact, in the undifferentiated state (d0), *Tsix* is slightly upregulated in ΔLinxP mutants (**Fig. 3A**; **Fig. 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 mESC (d0), *Tsix* allelic ratios are also not affected by *LinxP* heterozygous deletion (**Fig. 3B**). However, we did detect modest but significant differences in *Xist* allelic ratios prior to differentiation (**Fig. 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 Δ245kb allele. Differences in *Xist* allelic ratios between mutant and wildtype alleles became stronger upon differentiation (**Fig. 3B**). *Tsix* allelic ratios eventually became significantly different as well (**Fig. 3B**), which may be due to silencing in *cis* by Xist RNA. To uncouple *Tsix* and *Xist* regulation, we generated heterozygous ΔLinxP mutants in the XGTC cell line (**Fig. 3C**). Cherry (*Tsix*) levels were slightly upregulated in the ΔLinxP XGTC cells compared to controls at d0 and d2 (**Fig. 3D**), consistent with the results on ΔLinxP male mESC (**Fig. 3A**), and again arguing against a role for *LinxP* as an enhancer of *Tsix*. However, the proportion of cells upregulating GFP from the ΔLinxP allele upon differentiation was slightly but significantly increased (38% vs 30%, p=0.008) (**Fig. 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 ~170kb 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 harbours three CTCF-bound sites between the regulatory elements *LinxP* and *LinxE* (**Fig. 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 intronic interval containing the CTCF sites in male ESCs (ΔLinx-int1, ~51kb) and in mice (ΔLinx-CBS, ~25kb) (**Fig. 4A**). 5C analysis of the mutant mESCs revealed disruption of local 3D organisation: increased contacts between the *Linx* 3'end region and the *Chic1* locus, which harbours CTCF sites in convergent orientation to those within the *Linx* 3'end region (**Fig. 4B**). Furthermore, the *Linx* 3'end region lost contacts with *Xite* (**Fig. 4B**) and displayed decreased basal contacts throughout the Xist-TAD (**Fig. 4C**, black arrow). The interaction frequencies were reduced between *LinxE* and the *Xist*

promoter, and unaltered between *LinxP* and the *Xist* promoter (**Fig. S6A**). However, in heterozygous female embryos, we did not observe any effect on *Xist* or *Atp7a* allelic ratios (**Fig. 4D**, **Fig. S5A-B**). This indicates that *Linx*-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 (d0, d1, d2, d4). We observed no preferential interaction peaks with *Xist* when capturing the *Linx* promoter (**Fig. 4E**) or vice-versa (**Fig. 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 (**Fig. S5C**). Together, these data do not reveal any differentiation-specific differences in the topological organisation of the *Xic* that could explain how *LinxP* regulates the *Xist* promoter during the initiation of XCI.

Finally, we wondered whether the ΔLinxP allele itself could be affecting the structural landscape of the *Xic*, and thereby influencing *Xist* expression in cis. We performed 5C on wildtype and mutant ΔLinxP male mESC, as well as LinxP-inv and ΔLinxE for comparison. Differential analysis of 5C maps comparing ΔLinxE to wild type cells revealed no obvious alterations in the structural organisation of the *Xic* TADs (**Fig. 4G; Fig. S5D**), even though ΔLinxE leads to skewing in *Xist* expression (**Fig. S2F-G**). 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 (**Fig. 4H, 4J; Fig. S5D**). Similar results were observed for the LinxP-inv allele (**Fig. 4I-J; Fig. 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 polyA cassette downstream of *LinxP*, which abolishes *Linx* transcription (**Fig. S6B-C**). 5C analysis revealed that early truncation of *Linx* transcription also led to a significant gain of contacts between the *Tsix*- and *Xist*-TADs (**Fig. S6D-E**), 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 organisation. 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 (**Fig. 2J, S3G**). The interaction frequency between *LinxP* and the *Xist* promoter in the Linx-inv allele does not seem to be significantly altered (**Fig. S6A**); this could be the reason for not seeing an effect on *Xist* regulation in this mutants, if we are to assume that the interaction frequency between *LinxP* and *Xist* is important for how *LinxP* regulates *Xist*. Our data does 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 shows that the *Linx* locus is independently involved, on the one hand, in helping to shape *Xic* folding, via its transcription or lncRNA (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 knock-ins of *LinxP* (~2kb) into the Xist-TAD, in polymorphic female cells, and we determined allelic ratios of *Xist* expression from the modified or wildtype X chromosomes. We inserted *LinxP* at two different, independent locations within the Xist-TAD: one was between *Jpx* and *Ftx* (**Fig. 5A**), ~60kb away from the *Xist* promoter and within the high-frequency contact region upstream of *Xist* (see Fig. 1A); the other was between *Ftx* and *Xpct* (**Fig. 5B**), ~170kb 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 day-0, day-2 and day-4. Our results consistently showed that the presence of *LinxP* in the Xist-TAD, regardless of its orientation or of its position, leads to preferential *Xist* expression from that chromosome at each differentiation time point (**Fig. 5A, 5B;** see also the second note in Fig. 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 neighbouring 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 (**Fig. 6A**), similarly to many lncRNA 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 (**Fig. 6B**). In particular, two conserved modules within *LinxP* show shared synteny across placental mammals, but not in the marsupial opossum (**Fig. 6C**). One of these modules coincides with binding of Nanog and Oct4 in mESC (**Fig. 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 200kb away in the neighbouring TAD. This was surprising, as current views posit that TAD boundaries prevent communication between cis-regulatory elements and genes in neighbouring 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; Groschel et al., 2014; Hnisz et al., 2016; Lupiáñez 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 others' data (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 organisation 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 (Galupa and Heard, 2015) – *LinxP* regulates *Xist* in *cis* in a manner independent of *Linx* transcripts or transcription. Thus, even though *Linx* produces an 80kb-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 neighbouring 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 about two-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 neighbouring sequences (**Figs. 4E-F**). 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 *Linx*-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 thirty 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 ~170kb 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 another gene, especially in the light of recent models of gene expression that involve the clustering of cis-regulatory elements and promoters into condensates (Plys 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 does not lead 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 ~170kb), 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 behaviour 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). 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 Mirny, 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 organisation of the locus at a given time point. Further functional studies will allow to test such hypotheses.

Besides harbouring a long-range regulator of *Xist*, the *Linx* locus is also involved in (i) regulating *Cdx4*, located ~10kb upstream of *Linx*, and (ii) in shaping the topological

organisation 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 lncRNA. 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*, *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 maybe 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 favoured 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 organisation (Galupa and Heard, 2017), suggesting that regulatory crosstalk between neighbouring 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 neighbouring TAD will certainly provide new insights into the fundamental principles of *cis*-regulatory control.

ACKNOWLEDGEMENTS

We are grateful to Katia Ancelin and Isabelle Grandjean for help and advice with animal management; to Lucile Marion-Poll for help and advice with flow cytometry experiments; to Maud Borensztein for scientific discussions as well as help with mouse genotyping; to Denis Krndija, 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 BDD team of PICT-IBiSA, the NGS Platform, the Genomics Platform (in particular David Gentien, Cécile Reyes, Audrey Rapinat and Benoit 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 Luísa Supico (1963-2019) and her inspiring mentorship, righteous indignation and precious friendship. **Funding:** This work was supported by fellowships from Région Ile-de-France (DIM Biothérapies) and Fondation pour la Recherche Médicale (FDT20160435295) to RG; NWO-ALW Rubicon (825.13.002) and Veni (863.15.016) fellowships to JvB; ERC Advanced Investigator award (ERC-2014-AdG no. 671027), Labelisation La Ligue, FRM (DEI20151234398), ANR DoseX 2017, Labex DEEP (ANR-11-LBX-0044), part of the IDEX PSL (ANR-10-IDEX-0001-02 PSL) and ABS4NGS (ANR-11-BINF-0001) to EH. High-throughput sequencing for 5C, Capture-C and RNA-seq has been performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the Canceropole Ile-de-France and by the SiRIC-Curie program - SiRIC Grant "INCa-DGOS- 4654".

AUTHOR CONTRIBUTIONS

Conceptualization: RG, EPN, LG, EH. Investigation: RG, CP, EPN, CG, PD, ALS. Methodology: RG, CP, EPN, FEM, CJ, CG, JGvB, SL, JPF, SB. Formal analysis: RG, RH, NS, YZ. Data curation: RH, NS, YZ, LG. Visualization: RG, RH, YZ. Software: RH, NS, YZ, LG. Resources: FL, JG. Supervision: RG, UO, LG, EH. Project administration: RG, EH. Funding acquisition: UO, EH. Writing, original draft: RG, EH. Writing, review & editing: RG, EPN, LG, EH.

DECLARATION OF INTERESTS

The authors declare no competing interests.

MAIN FIGURE TITLES AND LEGENDS

Fig. 1. The Tsix-TAD harbours important elements for both *Tsix* and *Xist* regulation. (**A**) Topological organisation of the *Xic*; the *Xist*/*Tsix* locus lies at the boundary between two TADs. (**B**) Targeting strategy for deleting the ~245kb region included in the transgene Tg53 but not in Tg80 (Heard et al., 1999). Tg53, but not Tg80, expresses *Tsix* in the inner cell mass of mouse blastocysts (Nora et al., 2012); both transgenes include the *Xite* element. (**C**) Gene expression analysis during differentiation. Data is normalised to wt-d0 for each gene, and represents the average of two biological replicates for each genotype. (**D**) RNA FISH for Huwe1 (X-linked gene) and Xist (exonic probe) on mESCs differentiated to d1.5. Percentage of cells with Xist RNA accumulation is indicated and represents an average from two independent clones (SD=0.07%). Scale bar: 2 μm. (**E**) Cross used for analysis of RNA allelic ratios in female hybrid embryos. Table summarises number of embryos collected. (**F**, **G**) RNA allelic ratios for Xist (**F**) and Atp7a (**G**), an X-linked gene. Each black dot corresponds to a single female embryo. Statistical analysis was performed using Mann-Whitney test (**** $p<0.0001$). Reverse cross shown in Fig. S1F. (**H**) Schematic representation of the XGTC female line (129/Cast), which harbours a double knock-in on the Cast allele: EGFP replacing *Xist* exon-1 and mCherry replacing *Tsix* exon-1. We generated Δ245kb on the Cast allele. (**I**, **J**) Cytometry profiles of mCherry (**I**) and EGFP (**J**) at d0 and d2 of differentiation. On the right, (**I**) median fluorescence intensity (FI) of mCherry (normalised to WT, d0) or (**J**) percentage of EGFP positive cells, based on illustrated threshold. Wild type data represents an average of five experimental replicates. Δ245kb data represents an average of two independent clones, five experimental replicates for each. Statistical analysis: paired two-tailed t-test (** $p<0.01$; *** $p<0.001$; **** $p<0.0001$).

Fig. 2. The *Linx* locus harbours cis-regulatory elements that control XCI choice. (**A**) ATAC-seq data for the Tsix-TAD region in differentiating XX mESC. For each time point, results of peak calling are represented by grey marks below the data. Green marks depict differential peak analysis. Identical results were found for d0 vs d2 and d1 vs d2 ($p<0.01$) within the region of interest, while no differential peaks were found for d0 vs d1. Grey box highlights the promoter of *Linx,* the only differential peak within the Δ245kb region. Normalised data shown for one replicate (second replicate in Fig. S2A); peak analysis performed on both replicates. See Methods for more details. (**B**) The *Linx* locus and its chromatin features (see Methods for sources of datasets represented). Position of introns and exons is based on (Nora et al., 2012) and mESC RNA SCRIPTURE (Guttman et al., 2010). Targeted region *LinxP* (~2kb) is indicated. (**C**) Allelic quantification of Xist RNA by pyrosequencing at day 4 of differentiation. Note that each clone harbours the deletion in a different allele and Xist RNA allelic ratios are shown from one or the other allele, depending on the mutant clone that is being compared. Data are presented as means and error bars represent SEM (six biological replicates). Statistical analysis: two-tailed paired t-test with Bonferroni's correction (** $p<0.01$). (**D**) Determining which allele is more frequently coated by Xist RNA using RNA/DNA FISH. The two alleles can be distinguished due to a TetO array present on the 129 allele (Masui et al., 2011). X chromosomes are identified by using a probe for the *Tsix/Xist* region. Data are presented as means and error bars represent standard deviation

(two biological replicates, more than 80 cells per genotype counted for each). Statistical analysis: chi-square test (* $p<0.05$). (**E, I**) Crosses used for analysis of RNA allelic ratios in female hybrid embryos. Table summarises number of embryos collected. (**F**, **G**) RNA allelic ratios for Xist (**F**) and Atp7a (**G**), an X-linked gene. Each black dot corresponds to a single female embryo. Statistical analysis: two-tailed t-test (* $p<0.05$; ** $p<0.01$). Reverse cross shown in Fig. S3E. (**H**) Inversion of the *LinxP* element. (**J**) Analysis of Xist RNA allelic ratios. Each black dot represents the ratio for a single female embryo. Statistical analysis: two-tailed t-test. Analysis of Atp7a RNA allelic ratios and reverse cross shown in Fig.S3G.

Fig. 3. The *LinxP* element is not an enhancer of *Tsix* but regulates *Xist* expression. (**A**) Gene expression analysis during differentiation. Data is normalised to wt-d0 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 Fig. 2C for more information on the clones. Data are presented as means and error bars represent SEM (six biological replicates). Statistical analysis: two-tailed paired t-test with Bonferroni's correction (** $p<0.01$). (**C**) XGTC female line (129/Cast) as in Fig. 1H. We generated ΔLinxP mutant clones on the Cast allele. (**D**, **E**) Median fluorescence intensity (FI) of mCherry (normalised to WT, d0) or percentage of EGFP positive cells (as in Fig. 1J). Wildtype data represents an average of five wildtype clones, four experimental replicates for each. ΔLinxP data represents an average of five independent clones, four experimental replicates for each. Statistical analysis: paired two-tailed t-test (** $p<0.01$; *** $p<0.001$; **** $p<0.0001$).

Fig. 4. *Linx*-related topological features are not implicated in *Xist* regulation. (**A**) The *Linx* locus, CTCF binding and orientation of CTCF motifs associated with CTCF ChIP-seq peaks. Orientation of CTCF motifs within the Tsix-TAD represented above. Targeted deletions ΔLinxCBS (~25kb) and ΔLinx-int1 (~51kb) are indicated. See Methods for sources of CTCF, DNaseI and H3K27Ac datasets. (**B**, **C**) 5C profiles: pooled data from two biological replicates for each genotype. Differential map is corrected for deletion, see Methods. Grey pixels represent either the deleted region or filtered contacts. (**D**) Left, cross used for analysis of RNA allelic ratios in female hybrid embryos. Right, Xist RNA allelic ratios; each black dot corresponds to a single female embryo. Statistical analysis: two-tailed t-test. Table summarises number of embryos collected. Analysis of Atp7a RNA allelic ratios and reverse cross shown in Fig.S5A-B. (**E, F**) Capture-C profiles for *LinxP* (**E**) and *Xist* (**F**) viewpoints, at different time points of differentiation of XX (Pgk12.1) mESC. Data represent one replicate; two or three replicates for each time point were performed and are identical to the one shown (data available in GEO). Profiles represent number of contacts for each DpnII fragment per 10,000 total contacts within a specified region (see Methods). CTCF ChIP-seq on male mESCs represented below (Nora et al. 2017). (**G, H, I**) 5C differential maps for mutant male mESC: pooled data from two biological replicates for each genotype. 5C profiles for each genotype shown in Fig.S5D. Grey pixels correspond to either deleted regions or to filtered contacts. (**J**) Quantification of 5C inter-TAD contacts (see Fig.S5E for details). Bars represent the average of the calculated proportions of four (E14, ΔLinxP) or two (ΔLinxE, LinxP-inv) independent replicates. Statistical analysis: two-tailed t-test (** $p<0.01$; *** $p<0.001$).

Fig. 5. *LinxP* enhances *Xist* expression in *cis* when knocked-in into the Xist-TAD. (A, B) (Top) Location of the two knock-in cassettes. (Bottom) Allelic quantification of Xist RNA at differentiation time points d0, d2 and d4. Note that for each clone, the cassette was knocked-in one allele only, and allelic ratios are shown for each clone relative to the knock-in allele. Data are presented as means and error bars represent SEM (three biological replicates each). Statistical analysis: two-tailed paired t-test (* $p<0.05$; ** $p<0.01$). Clones harbouring the polyA cassette alone (shades of grey) were compared to wild type (WT), while clones harbouring the *LinxP* element (shades of salmon and purple) were compared to the clones harbouring the polyA cassette alone.

Fig. 6. The *LinxP* element is conserved across placental mammals and overlaps binding site for pluripotency factors. (**A**) Sequence conservation analysis. Conservation score across placental mammals shows poor sequence conservation for *Linx* (compare 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 DNaseI-seq, represented below (same as in Fig. 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 * 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).

STAR METHODS

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 37ºC under 8% CO_2 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 37ºC (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 30ºC 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-4f88-98da-3b7d7960394c/zhang-lab-general-cloning-protocol.pdf). sgRNA constructs were amplified upon transformation of DH5α competent cells (Takara) grown at 37ºC, 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.

Engineering mESC: 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 programme 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 2.5µg of each TALEN (four constructs) or 5µg 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 1.5µg of donor plasmid. As a transfection control, 10µg 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/en/protocols/genome-engineering/816-engineering-genomic-deletions-and-inversions-in-mouse-es-cells-using-custom-designed-nucleases. 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 mice: 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 containing 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 MEGAshortscript 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 95ºC 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 × 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 infudibulum 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 hybridisation (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 10 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, 1µg/µl BSA, 10mM Vanadyl-ribonucleoside) and incubated overnight at 37°C. For RNA/DNA FISH, the coverslips were first washed three times in 2×SSC and incubated for 1h at 37 °C in 2×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×SSC (pH 7.2-7.4) and then quickly transferred to ice and washed three times in ice-cold 2×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× SSC (pH 7.2-7.4) and three times at 42–45 °C in 2× 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 wildtype 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 labelled 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 20min at 37°C, and after denaturation (75°C, 10 min). Probes were then mixed 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 (Pgk12.1, 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 (15sec each), lysates were applied directly to a QIAshredder spin column (Qiagen) and centrifuged for 3min 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 synthesised from 0.5µg 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 processed in parallel.

nCounter analysis: We used the NanoString nCounter gene expression system (Geiss et al., 2008) to systematically characterise transcriptional differences in wildtype 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: 0.1µM) in a reaction volume of 10 µL. Appropriate no-reverse transcription and no-cDNA controls were perfomed 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. Normalisation 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 37ºC 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 *LinxP* 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 15min on ice. Nuclei were resuspended in 100 µL 0.5% SDS, incubated at 62°C for 10min and quenched with 50 µL 10% Triton X-100 and 290 µL water at 37°C for 15min. Digestion was performed overnight by adding 50 µL of DpnII (Capture-C) or HindIII (5C) buffer and 10 µL of high-concentration DpnII or HindIII (NEB) and incubating samples at 37°C in a thermomixer. Before this step, an aliquot was taken from each sample as an undigested control. Digests were heat inactivated for 20 min at 65ºC and an aliquot was taken from each sample as a digested (unligated) control. Samples were cooled at room temperature for 10 min before adding the ligation cocktail. 3C libraries for Capture-C were diluted by adding 672 µL water and ligated overnight at 16°C with 8 µL T4 Ligase (30U/µl EL0013 Thermo Scientific) and 122 µL Ligation buffer in a thermomixer at 1400rpm. 3C libraries for 5C were ligated for 4 hours at 25ºC with 10U T4 ligase and ligation buffer (ThermoFisher cat 15224) in a thermomixer at 1000rpm. All ligated samples were then

centrifuged at 2000rpm, resuspended in 240 µL of 5% SDS and 1 mg Proteinase K, incubated at 55ºC for 30min, supplemented with 50 µL 5 M NaCl and incubated at 65ºC for 4 hours. DNA was then purified by adding 500 µL isopropanol, incubated at -80ºC overnight, centrifuged at 12,000 rpm at 4ºC, washed with 70% ethanol, air dried and resuspended in 100 µL water, followed by incubation with RNase A at 37ºC for one hour. 3C templates were quantified using Qubit DNA Broad-Range (ThermoFisher) and diluted to 100 ng/µL. Libraries and respective controls (undigested and digested aliquots) were verified on a gel.

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 95ºC for 5 min and incubated at 48ºC for 16-18h. 10 µL of 1X Taq ligase buffer with 5U Taq ligase were added to each annealing reaction followed by incubation at 48ºC for 4h and 65ºC for 10 min. Negative controls (no ligase, no template or no 5C oligonucleotide) 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 mL total. Cycling conditions were 95ºC for 9 min, 25 cycles of 95ºC for 30 sec, 60ºC for 30 sec, 72ºC for 30 sec followed by 72ºC 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 (KK8500, 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 13pmol of capture probes (biotinylated oligonucleotides, Integrated DNA Technologies), with the SeqCap 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 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 CTGTCTCTTATACACATCTGACGCTGCCGACGA CTGTCTCTTATACACATCTCCGAGCCCACGAGAC, clip3pAdapterMMp 0.1); (2) suppressing splice junction determination (alignIntronMax 10, alignSJoverhangMin 75, alignSJDBoverhangMin 75); (3) read pairs that represented fragments of 1500bp or less were retained (alignMatesGapMax 1500); and (4) the remaining non-default parameters were: alignEndsType Extend5pOfRead1; 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 vs 1, days 1 vs 2, and days 0 vs 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 $i \pm N$ and $j \pm N$ 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 grey 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.

SUPPLEMENTAL INFORMATION: EXCEL TABLE, FOLDER

Excel table: Table S1. List of plasmids, list of NGS data generated, list of mutant mESCs and mice and respective coordinates of deletions/inversions, and sequences of genotyping primers, TALENs, sgRNAs, nCounter probes, RNA and DNA FISH probes, Capture-C probes, qPCR primers, AQ primers, LinxP across mammals, Related to STAR Methods.

Folder: Data S1. Sequences of plasmids used for knockins, Related to Figure 5 and S6.

REFERENCES

Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. Elife *8*.

Anderson, K.M., Anderson, D.M., McAnally, J.R., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. Nature *539*, 433.

Anguera, M.C., Ma, W., Clift, D., Namekawa, S., Kelleher, R.J., and Lee, J.T. (2011). Tsx Produces a Long Noncoding RNA and Has General Functions in the Germline, Stem Cells, and Brain. PLoS Genet. *7*, e1002248.

Baniahmad, A., Muller, M., Steiner, C., and Renkawitz, R. (1987). Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. EMBO J. *6*, 2297–2303.

Bassett, A.R., Akhtar, A., Barlow, D.P., Bird, A.P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A.C., Gingeras, T.R., Haerty, W., et al. (2014). Considerations when investigating lncRNA function in vivo. Elife *3*.

van Bemmel, J.G., Galupa, R., Gard, C., Servant, N., Picard, C., Davies, J., Szempruch, A.J., Zhan, Y., Żylicz, J.J., Nora, E.P., et al. (2019). The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. Nat. Genet. *51*, 1024–1034.

Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. Mol. Cell *0*.

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.

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. *14*, 708–715.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A. V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e16.

Borensztein, M., Okamoto, I., Syx, L., Guilbaud, G., Picard, C., Ancelin, K., Galupa, R., Diabangouaya, P., Servant, N., Barillot, E., et al. (2017). Contribution of epigenetic landscapes and transcription factors to X-chromosome reactivation in the inner cell mass. Nat. Commun. *8*.

Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985). Characterization of a "silencer" in yeast: A DNA sequence with properties opposite to those of a transcriptional enhancer. Cell *41*, 41–48.

Brand, A.H., Micklem, G., and Nasmyth, K. (1987). A yeast silencer contains sequences that

can promote autonomous plasmid replication and transcriptional activation. Cell *51*, 709–719.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cao, S.X., Gutman, P.D., Dave, H.P., and Schechter, A.N. (1989). Identification of a transcriptional silencer in the 5'-flanking region of the human epsilon-globin gene. Proc. Natl. Acad. Sci. U. S. A. *86*, 5306–5309.

Cattanach, B.M., and Papworth, D. (1981). Controlling elements in the mouse. V. Linkage tests with X-linked genes. Genet. Res. *38*, 57–70.

Chodroff, R.A., Goodstadt, L., Sirey, T.M., Oliver, P.L., Davies, K.E., Green, E.D., Molnár, Z., and Ponting, C.P. (2010). Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol. *11*, R72.

Davies, J.O.J., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods *13*, 74–80.

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

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nat. Methods *14*, 629–635.

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

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doyle, H.J., Kraut, R., and Levine, M. (1989). Spatial regulation of zerknüllt: a dorsal-ventral patterning gene in Drosophila. Genes Dev. *3*, 1518–1533.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452–455.

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

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

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47*, D766–D773.

Fudenberg, G., and Mirny, L.A. (2012). Higher-order chromatin structure: bridging physics and biology. Curr. Opin. Genet. Dev. *22*, 115–124.

Furlan, G., Hernandez, N.G., Huret, C., Galupa, R., van Bemmel, J.G., Romito, A., Heard, E., Morey, C., and Rougeulle, C. (2018). The Ftx noncoding locus controls X chromosome inactivation independently of its RNA products. Mol. Cell *70*, 462–472.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science (80-.). *361*, 1341–1345.

Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. Curr. Opin. Genet. Dev. *31*, 57–66.

Galupa, R., and Heard, E. (2017). Topologically Associating Domains in Chromosome Architecture and Gene Regulatory Landscapes during Development, Disease, and Evolution. Cold Spring Harb. Symp. Quant. Biol. *82*, 267–278.

Galupa, R., and Heard, E. (2018). X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. Annu. Rev. Genet. *52*, 535–566.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat. Biotechnol. *26*, 317–325.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription. Cell *157*, 950–963.

Gray, S., and Levine, M. (1996). Transcriptional repression in development. Curr. Opin. Cell Biol. *8*, 358–364.

Greenberg, M.V.C., Glaser, J., Borsos, M., Marjou, F. El, Walter, M., Teissandier, A., and Bourc'his, D. (2017). Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. Nat. Genet. *49*, 110–118.

Groff, A.F., Barutcu, A.R., Lewandowski, J.P., and Rinn, J.L. (2018). Enhancers in the Peril lincRNA locus regulate distant but not local genes. Genome Biol. *19*, 219.

Groschel, S., Sanders, M.A., Hoogenboezem, R., De Wit, E., Bouwman, B.A.M., Erpelinck, C., Van der Velden, V.H.J., Havermans, M., Avellino, R., Van Lom, K., et al. (2014). A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. Cell *157*, 369–381.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. *28*, 503–510.

Hahne, F., and Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. (Humana Press, New York, NY), pp. 335–351.

Heard, E., Mongelard, F., Arnaud, D., and Avner, P. (1999). Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. Mol. Cell. Biol. *19*, 3156–3166.

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

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. *46*, 205–212.

Jayavelu, N.D., Jajodia, A., Mishra, A., and Hawkins, R.D. (2018). An atlas of silencer elements for the human and mouse genomes. BioRxiv 252304.

Jiang, J., Cai, H., Zhou, Q., and Levine, M. (1993). Conversion of a dorsal-dependent silencer into an enhancer: evidence for dorsal corepressors. EMBO J. *12*, 3201–3209.

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

Kent, W.J. (2002). BLAT---The BLAST-Like Alignment Tool. Genome Res. *12*, 656–664.

Kirov, N., Zhelnin, L., Shah, J., and Rushlow, C. (1993). Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in Drosophila. EMBO J. *12*, 3193–3199.

Kragesteen, B.K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A.M., Jerković, I., et al. (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. *50*, 1463–1473.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lee, J.T. (2000). Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. Cell *103*, 17–27.

Lee, J.T., and Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. Cell *99*, 47–57.

Lee, J.T., Davidow, L.S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. *21*, 400–404.

Li, L.M., and Arnosti, D.N. (2011). Long- and Short-Range Transcriptional Repressors Induce Distinct Chromatin States on Repressed Genes. Curr. Biol. *21*, 406–412.

Loos, F., Maduro, C., Loda, A., Lehmann, J., Kremers, G.-J., ten Berge, D., Grootegoed, J.A., and Gribnau, J. (2016). *Xist* and *Tsix* Transcription Dynamics Is Regulated by the X-to-Autosome Ratio and Semistable Transcriptional States. Mol. Cell. Biol. *36*, 2656–2667.

Luikenhuis, S., Wutz, A., and Jaenisch, R. (2001). Antisense Transcription through the Xist

Locus Mediates Tsix Function in Embryonic Stem Cells. Mol. Cell. Biol. *21*, 8512–8520.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell *161*, 1012–1025.

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

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., Barillot, E., Belmont, A.S., and Heard, E. (2011). Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. Cell *145*, 447–458.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and McIntosh, I. (2001). Identification of TSIX, Encoding an RNA Antisense to Human XIST, Reveals Differences from its Murine Counterpart: Implications for X Inactivation. Am. J. Hum. Genet. *69*, 951–960.

Migeon, B.R., Lee, C.H., Chowdhury, A.K., and Carpenter, H. (2002). Species Differences in TSIX/Tsix Reveal the Roles of These Genes in X-Chromosome Inactivation. Am. J. Hum. Genet. *71*, 286–293.

Minkovsky, A., Patel, S., and Plath, K. (2012). Concise review: Pluripotency and the transcriptional inactivation of the female Mammalian X chromosome. Stem Cells *30*, 48–54.

Minkovsky, A., Barakat, T.S., Sellami, N., Chin, M.H., Gunhanlar, N., Gribnau, J., and Plath, K. (2013). The pluripotency factor-bound intron 1 of Xist is dispensable for X chromosome inactivation and reactivation in vitro and in vivo. Cell Rep. *3*, 905–918.

Mutzel, V., Okamoto, I., Dunkel, I., Saitou, M., Giorgetti, L., Heard, E., and Schulz, E.G. (2019). A symmetric toggle switch explains the onset of random X inactivation in different mammals. Nat. Struct. Mol. Biol. *26*, 350–360.

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

Nakamura, N., Burt, D.W., Paul, M., and Dzau, V.J. (1989). Negative control elements and cAMP responsive sequences in the tissue-specific expression of mouse renin genes. Proc. Natl. Acad. Sci. U. S. A. *86*, 56–59.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular Coupling of Xist Regulation and Pluripotency. Science (80-.). *321*, 1693–1695.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van

Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

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

Northcott, P.A., Lee, C., Zichner, T., Stutz, A.M., Erkek, S., Kawauchi, D., Shih, D.J.H., Hovestadt, V., Zapatka, M., Sturm, D., et al. (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature *511*, 428–434.

Ogawa, Y., and Lee, J.T. (2003). Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol. Cell *11*, 731–743.

Paralkar, V.R., Taborda, C.C., Huang, P., Yao, Y., Kossenkov, A.V., Prasad, R., Luan, J., Davies, J.O.J., Hughes, J.R., Hardison, R.C., et al. (2016). Unlinking an lncRNA from Its Associated cis Element. Mol. Cell *62*, 104–110.

Peeters, S.B., Yang, C., and Brown, C.J. (2016). Have humans lost control: The elusive X-controlling element. Semin. Cell Dev. Biol. *56*, 71–77.

Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc. Natl. Acad. Sci. *108*, 13570–13575.

Plys, A.J., and Kingston, R.E. (2018). Dynamic condensates activate transcription. Science (80-.). *361*, 329–330.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ranisavljevic, N., Okamoto, I., Heard, E., and Ancelin, K. (2017). RNA FISH to Study Zygotic Genome Activation in Early Mouse Embryos. In Methods in Molecular Biology (Clifton, N.J.), pp. 133–145.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Redolfi, J., Zhan, Y., Valdes-Quezada, C., Kryzhanovska, M., Guerreiro, I., Iesmantavicius, V., Pollex, T., Grand, R.S., Mulugeta, E., Kind, J., et al. (2019). DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. Nat. Struct. Mol. Biol. *26*, 471–480.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.

Ritter, N., Ali, T., Kopitchinski, N., Schuster, P., Beisaw, A., Hendrix, D.A., Schulz, M.H., Müller-McNicoll, M., Dimmeler, S., and Grote, P. (2019). The lncRNA Locus Handsdown Regulates Cardiac Gene Programs and Is Essential for Early Mouse Development. Dev. Cell.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Sado, T., Wang, Z., Sasaki, H., and Li, E. (2001). Regulation of imprinted X-chromosome inactivation in mice by Tsix. Development *128*, 1275–1286.

Saffer, J.D., and Thurston, S.J. (1989). A negative regulatory element with properties similar to those of enhancers is contained within an Alu sequence. Mol. Cell. Biol. *9*, 355–364.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., and Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. Nat. Protoc. *7*, 171–192.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent Transcription from Active Promoters. Science (80-.). *322*, 1849–1851.

Servant, N., Lajoie, B.R., Nora, E.P., Giorgetti, L., Chen, C.-J., Heard, E., Dekker, J., and Barillot, E. (2012). HiTC: exploration of high-throughput "C" experiments. Bioinformatics *28*, 2843–2844.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.

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

Smith, E.M., Lajoie, B.R., Jain, G., and Dekker, J. (2016). Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. Am. J. Hum. Genet. *98*, 185–201.

Sousa, E.J., Stuart, H.T., Bates, L.E., Ghorbani, M., Nichols, J., Dietmann, S., and Silva, J.C.R. (2018). Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. Cell Stem Cell *22*, 919–928.

Stavropoulos, N., Lu, N., and Lee, J.T. (2001). A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. Proc. Natl. Acad. Sci. U. S. A. *98*, 10232–10237.

Studer, M., Popperl, H., Marshall, H., Kuroiwa, A., and Krumlauf, R. (1994). Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1. Science (80-.). *265*, 1728–1732.

Tsujimura, T., Klein, F.A., Langenfeld, K., Glaser, J., Huber, W., and Spitz, F. (2015). A Discrete Transition Zone Organizes the Topological and Regulatory Autonomy of the Adjacent Tfap2c and Bmp7 Genes. PLoS Genet. *11*, e1004897.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*, research0034.1.

Vicente-García, C., Villarejo-Balcells, B., Irastorza-Azcárate, I., Naranjo, S., Acemel, R.D.,

Tena, J.J., Rigby, P.W.J., Devos, D.P., Gómez-Skarmeta, J.L., and Carvajal, J.J. (2017). Regulatory landscape fusion in rhabdomyosarcoma through interactions between the PAX3 promoter and FOXO1 regulatory elements. Genome Biol. *18*, 106.

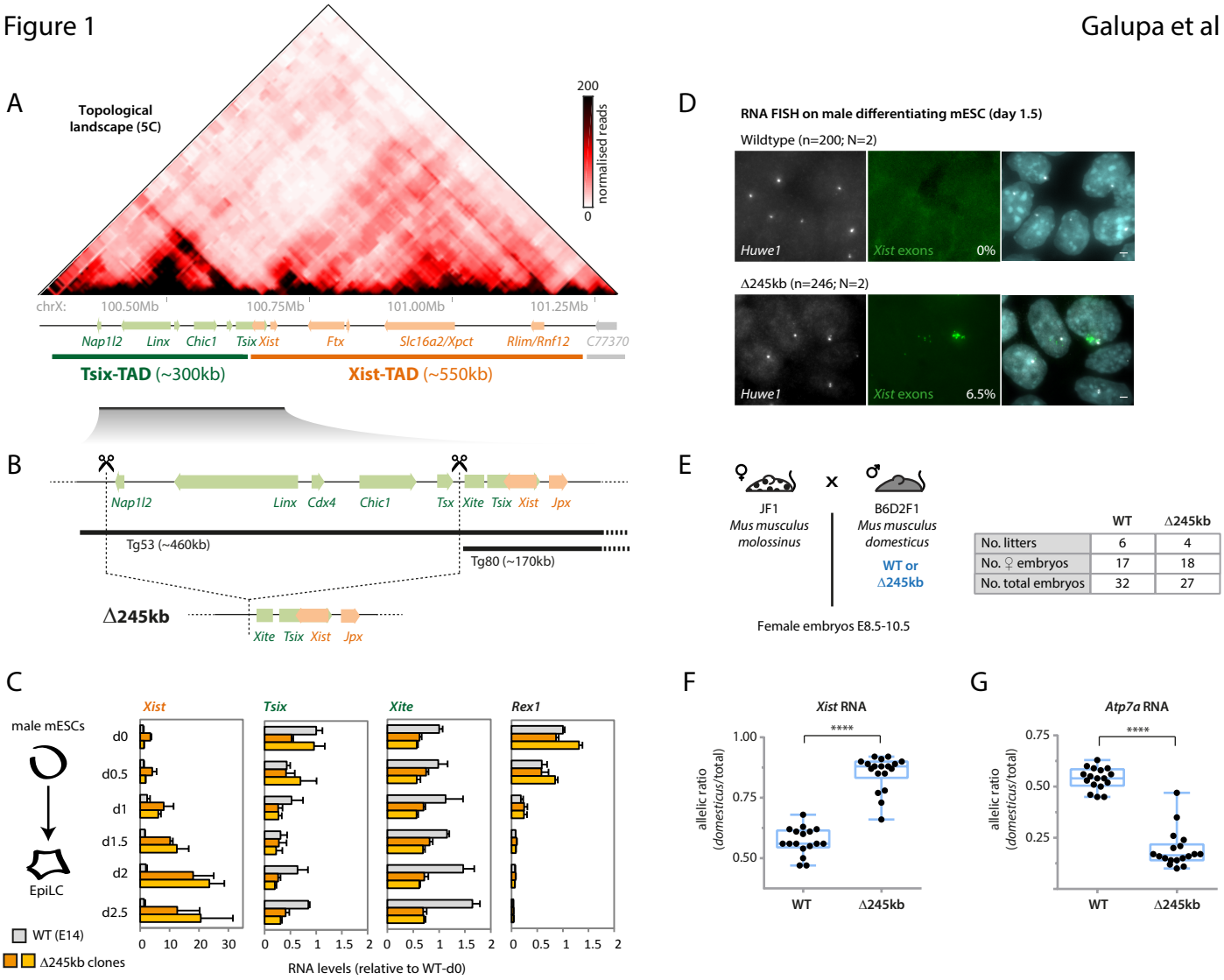
Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell *153*, 910–918.

Weintraub, S.J., Chow, K.N.B., Luo, R.X., Zhang, S.H., He, S., and Dean, D.C. (1995). Mechanism of active transcriptional repression by the retinoblastoma protein. Nature *375*, 812–816.

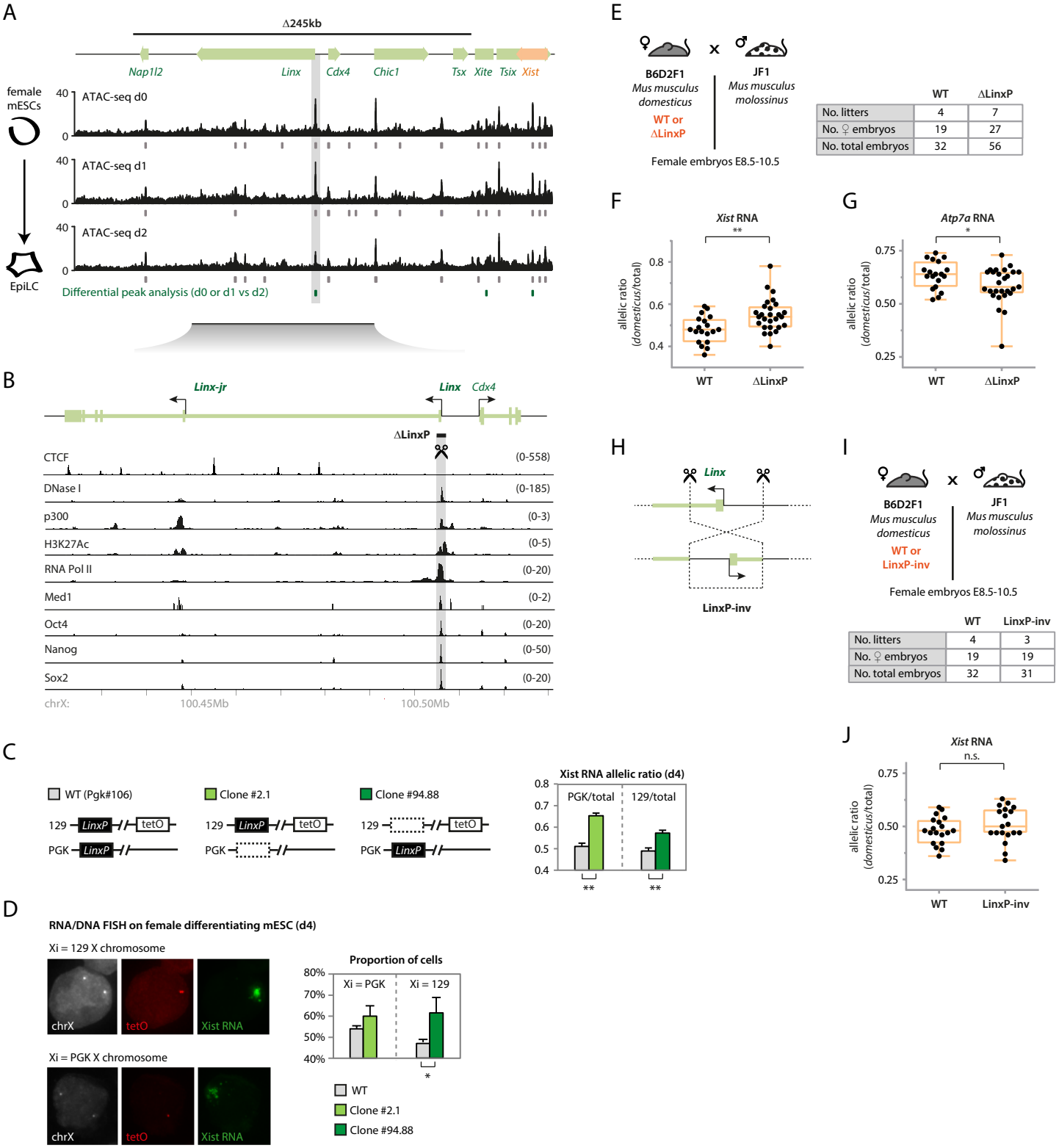
Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Blüthgen, N., Stadler, M., Tiana, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. Genome Res. *27*, 479–490.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

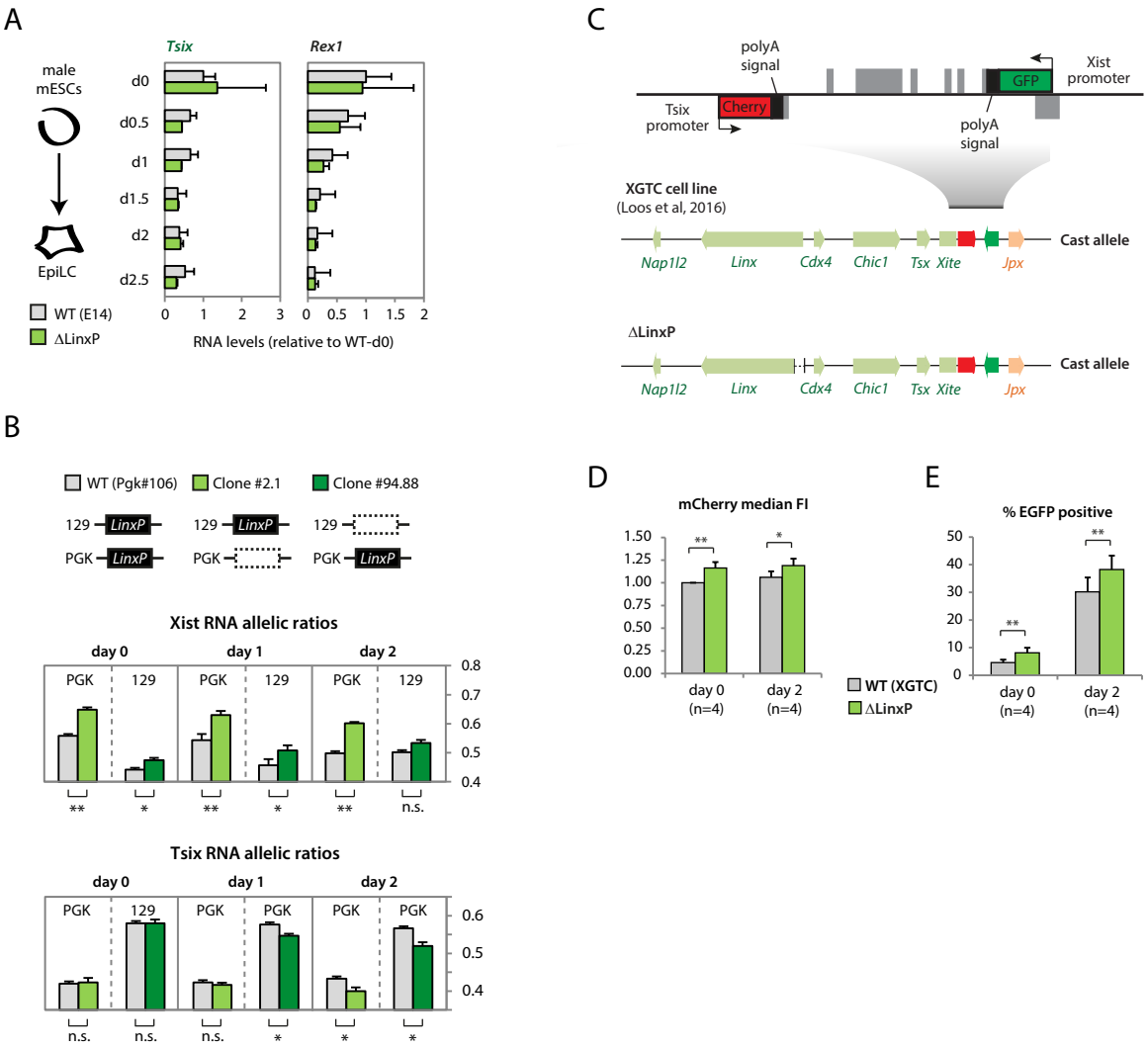


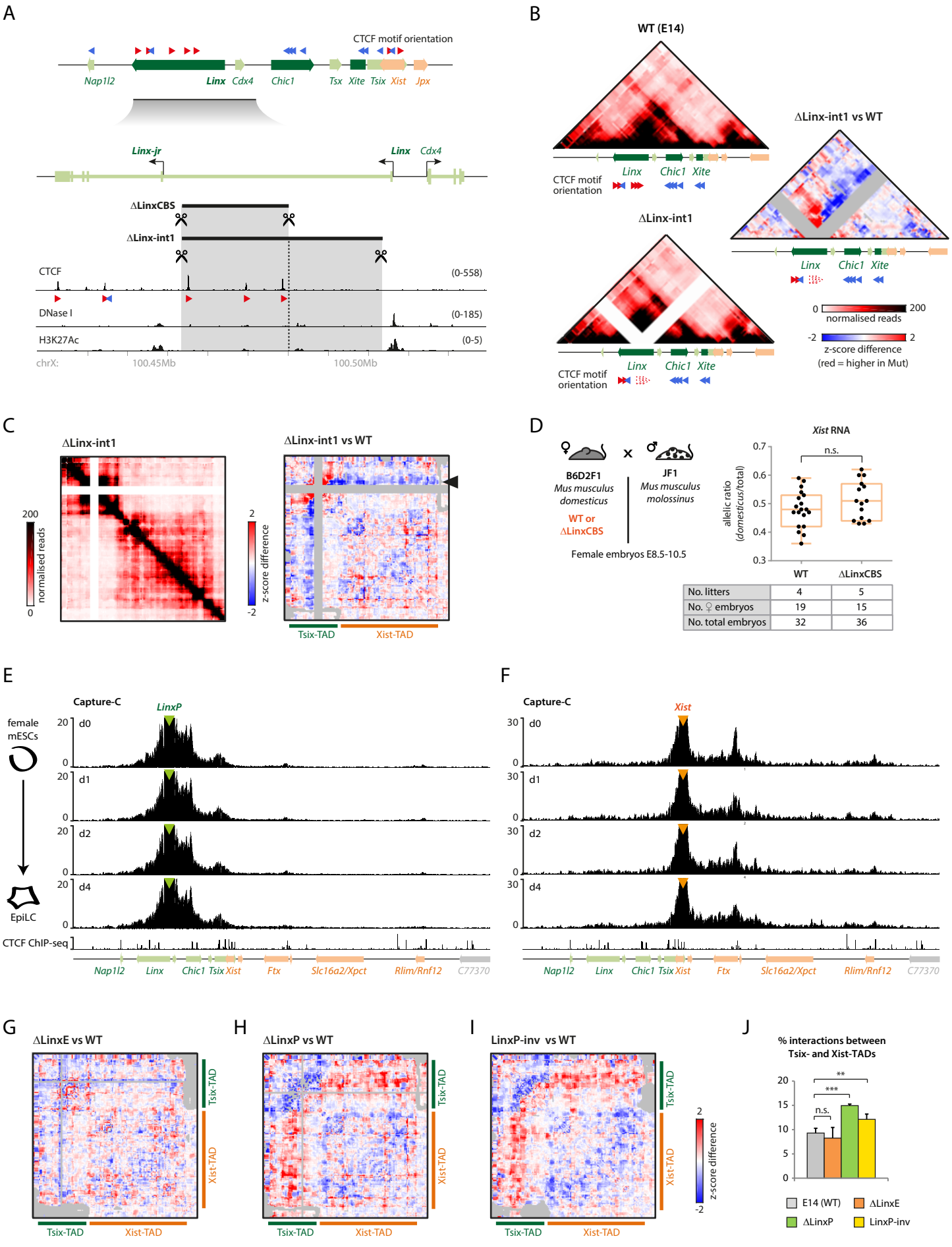
	WT	Δ245kb
No. litters	6	4
No. ♀ embryos	17	18
No. total embryos	32	27

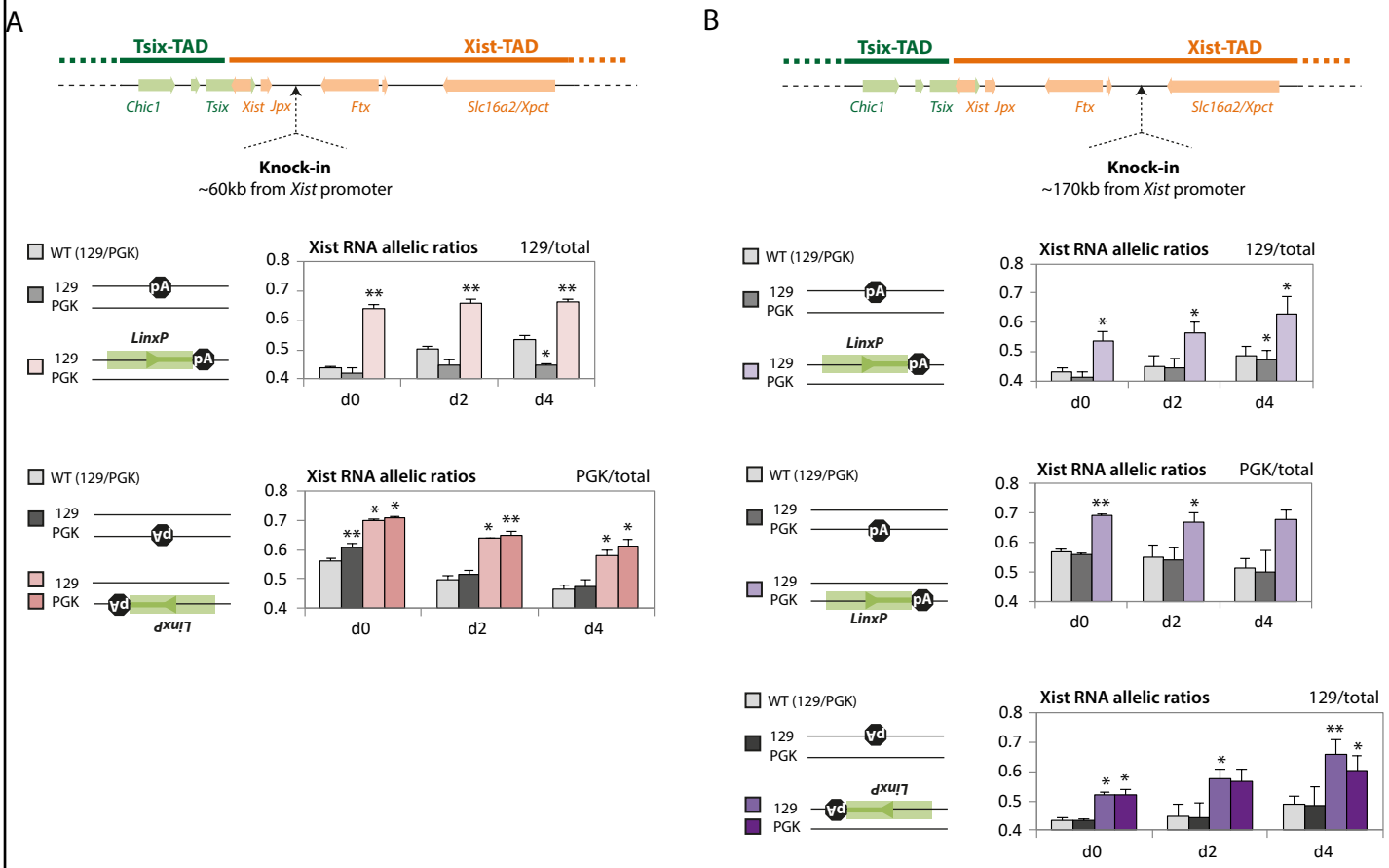


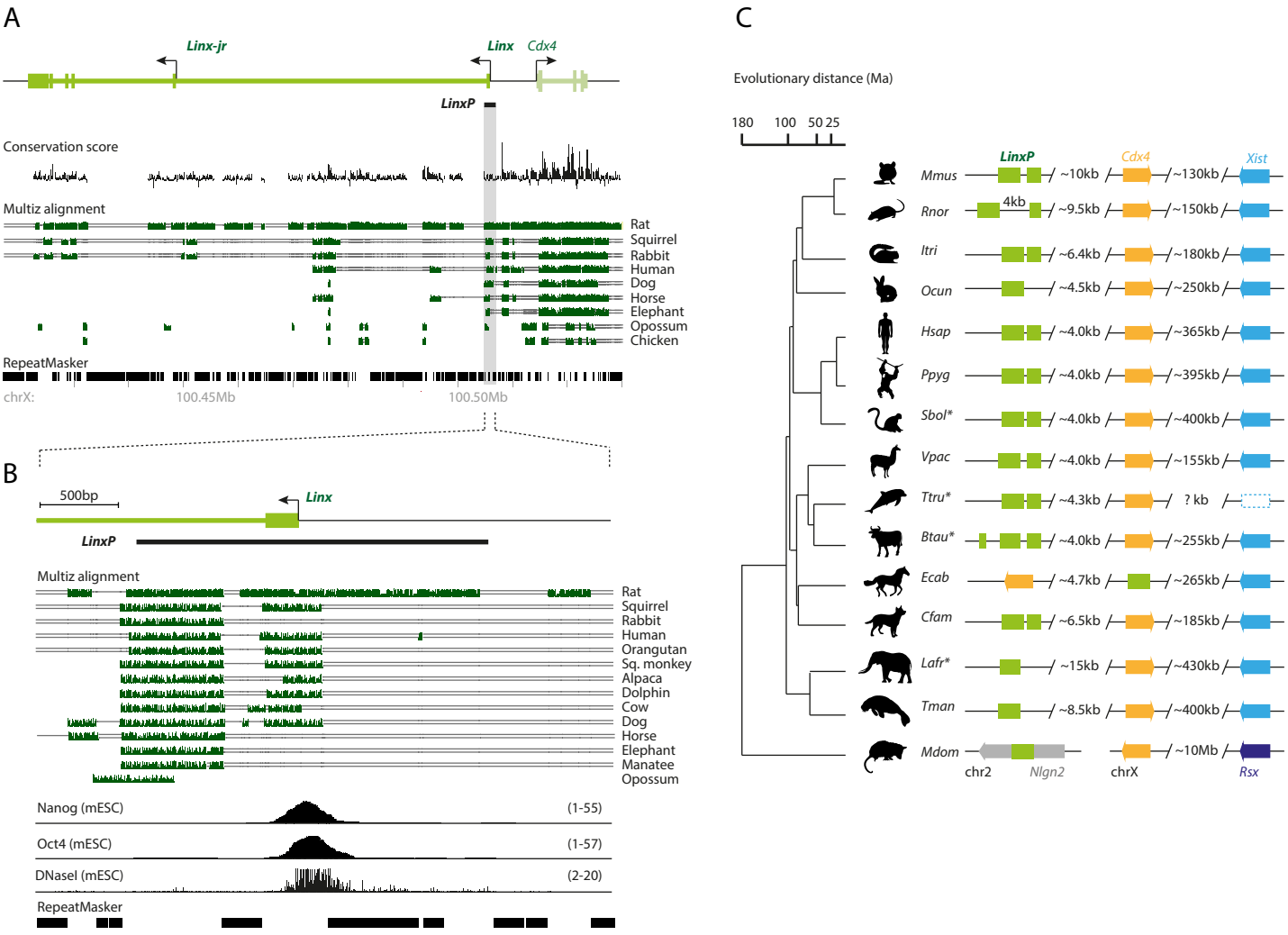
	WT	ΔLinxP
No. litters	4	7
No. ♀ embryos	19	27
No. total embryos	32	56

	WT	LinxP-inv
No. litters	4	3
No. ♀ embryos	19	19
No. total embryos	32	31

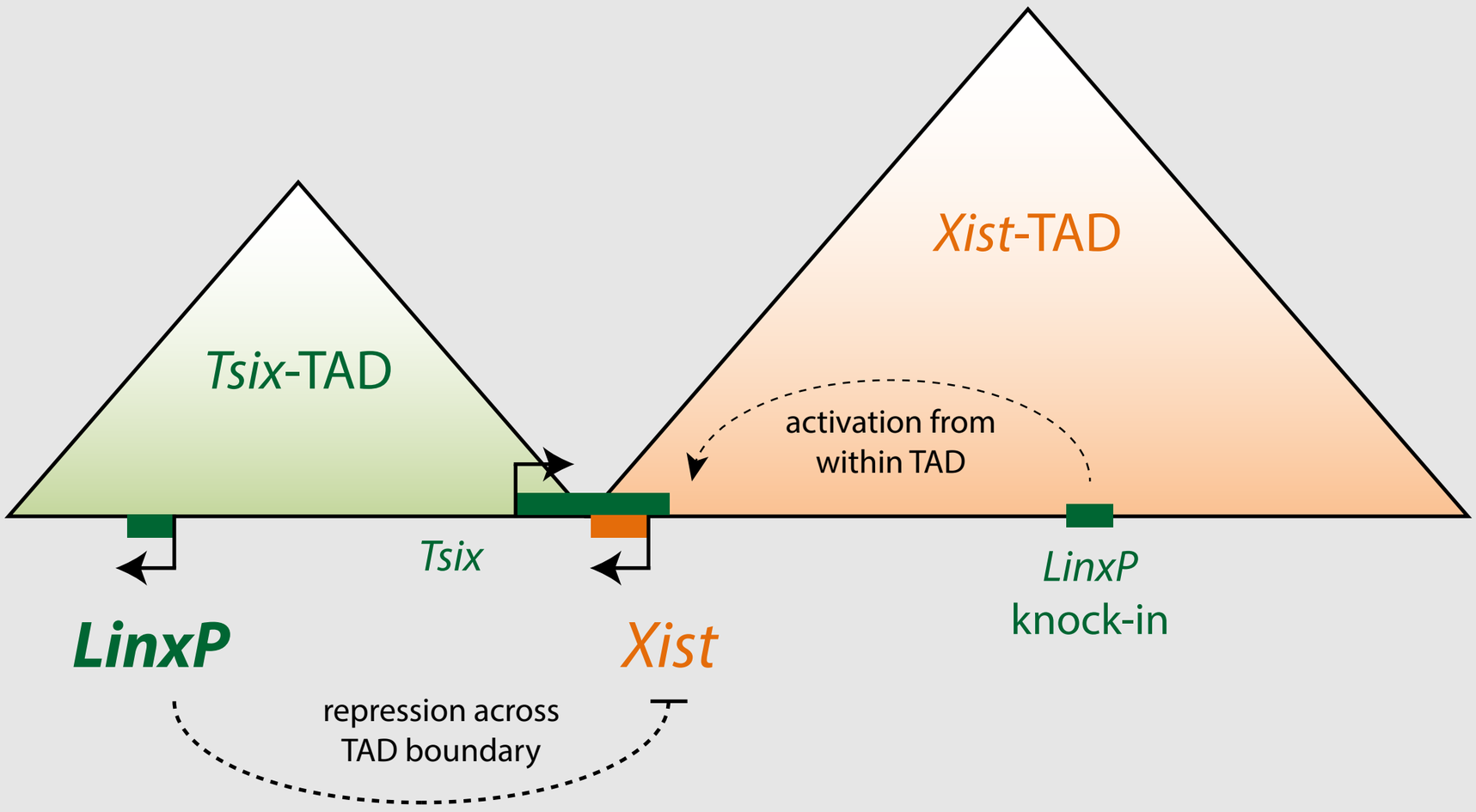

A
Tsix
Rex1
male mESCs
d0
d0.5
d1
d1.5
d2
d2.5
EpiLC
WT (E14)
ΔLinxP
RNA levels (relative to WT-d0)
B
WT (Pgk#106)
Clone #2.1
Clone #94.88
129 LinxP
PGK LinxP
PGK
Xist RNA allelic ratios
day 0
day 1
day 2
Tsix RNA allelic ratios
n.s.
C
polyA signal
Xist promoter
GFP
Tsix promoter
Cherry
XGTC cell line
(Loos et al, 2016)
Cast allele
Nap1l2
Linx
Cdx4
Chic1
Tsx
Xite
Jpx
ΔLinxP
D
mCherry median FI
day 0 (n=4)
day 2 (n=4)
WT (XGTC)
ΔLinxP
E
% EGFP positive


A
CTCF motif orientation
Nap1l2 Linx Cdx4 Chic1 Tsx Xite Tsix Xist Jpx
Linx-jr
Linx Cdx4
ΔLinxCBS
ΔLinx-int1
CTCF (0-558)
DNase I (0-185)
H3K27Ac (0-5)
chrX: 100.45Mb 100.50Mb
B
WT (E14)
ΔLinx-int1 vs WT
CTCF motif orientation
Linx Chic1 Xite
ΔLinx-int1
0 200 normalised reads
-2 2 z-score difference (red = higher in Mut)
C
ΔLinx-int1
ΔLinx-int1 vs WT
200 normalised reads 0
2 z-score difference -2
Tsix-TAD Xist-TAD
D
B6D2F1 Mus musculus domesticus
WT or ΔLinxCBS
JF1 Mus musculus molossinus
Female embryos E8.5-10.5
Xist RNA
allelic ratio (domesticus/total)
n.s.
WT ΔLinxCBS
No. litters 4 5
No. ♀ embryos 19 15
No. total embryos 32 36
E
Capture-C
LinxP
female mESCs
d0 d1 d2 d4
EpiLC
CTCF ChIP-seq
Nap1l2 Linx Chic1 Tsix Xist Ftx Slc16a2/Xpct Rlim/Rnf12 C77370
F
Capture-C
Xist
d0 d1 d2 d4
Nap1l2 Linx Chic1 Tsix Xist Ftx Slc16a2/Xpct Rlim/Rnf12 C77370
G
ΔLinxE vs WT
Tsix-TAD Xist-TAD
H
ΔLinxP vs WT
Tsix-TAD Xist-TAD
I
LinxP-inv vs WT
2 z-score difference -2
Tsix-TAD Xist-TAD
J
% interactions between Tsix- and Xist-TADs
n.s. ** ***
E14 (WT) ΔLinxE ΔLinxP LinxP-inv


A
Tsix-TAD
Xist-TAD
Chic1
Tsix
Xist
Jpx
Ftx
Slc16a2/Xpct
Knock-in
~60kb from Xist promoter
WT (129/PGK)
129 PGK
LinxP
Xist RNA allelic ratios
129/total
PGK/total
d0
d2
d4
B
Tsix-TAD
Xist-TAD
Knock-in
~170kb from Xist promoter


A
Linx-jr
Linx
Cdx4
LinxP
Conservation score
Multiz alignment
Rat
Squirrel
Rabbit
Human
Dog
Horse
Elephant
Opossum
Chicken
RepeatMasker
chrX:
100.45Mb
100.50Mb
B
500bp
Linx
LinxP
Multiz alignment
Rat
Squirrel
Rabbit
Human
Orangutan
Sq. monkey
Alpaca
Dolphin
Cow
Dog
Horse
Elephant
Manatee
Opossum
Nanog (mESC)
(1-55)
Oct4 (mESC)
(1-57)
DNaseI (mESC)
(2-20)
RepeatMasker
C
Evolutionary distance (Ma)
180
100
50 25
LinxP
Cdx4
Xist
Mmus
~10kb
~130kb
Rnor
4kb
~9.5kb
~150kb
Itri
~6.4kb
~180kb
Ocun
~4.5kb
~250kb
Hsap
~4.0kb
~365kb
Ppyg
~4.0kb
~395kb
Sbol*
~4.0kb
~400kb
Vpac
~4.0kb
~155kb
Ttru*
~4.3kb
? kb
Btau*
~4.0kb
~255kb
Ecab
~4.7kb
~265kb
Cfam
~6.5kb
~185kb
Lafr*
~15kb
~430kb
Tman
~8.5kb
~400kb
Mdom
~10Mb
chr2
Nlgn2
chrX
Rsx

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

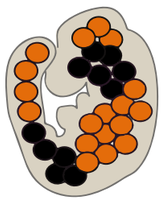


X-inactivation patterns

Wildtype

$X^{wt}X^{wt}$
50:50

ΔLinxP



$X^{wt}X^{del}$
40:60

LinxP-inv

$X^{wt}X^{inv}$
50:50

Sequence conservation across mammals

	LinxP	*Tsix*	*Xist*
Placentals			
	✔	✔	✔
	✔	✔	✔
	✔	✖	✔
	✔	✖	✔
	✔	✖	✔
Marsupials			
	✖	✖	✖