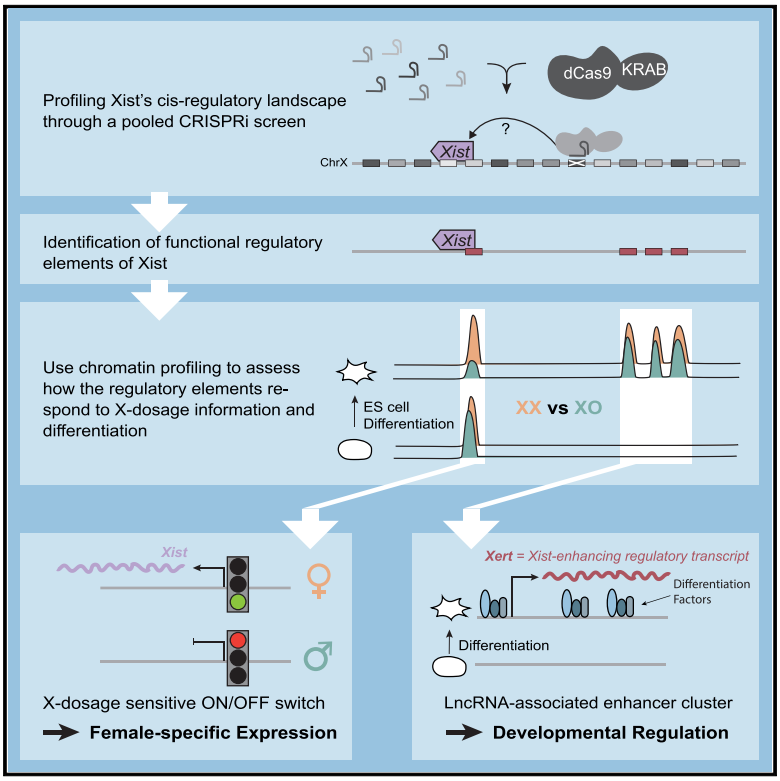
Molecular Cell

Article

Distal and proximal *cis*-regulatory elements sense X chromosome dosage and developmental state at the *Xist* locus

Graphical abstract



Highlights

- Promoter-proximal elements control female-specific Xist expression in a binary fashion
- Long-range enhancer elements regulate developmental timing of Xist upregulation
- Several distal enhancers are associated with a previously unannotated lncRNA, Xert
- Xert is upregulated concomitantly with Xist and activates Xist in *cis*

Authors

Rutger A.F. Gjaltema, Till Schwämmle, Pauline Kautz, ..., Annalisa Marsico, Stefan Mundlos, Edda G. Schulz

Correspondence

edda.schulz@molgen.mpg.de

In brief

Gjaltema et al. show that developmental upregulation of Xist, the master regulator of X inactivation, is controlled by several distal enhancer elements, which are associated with a *cis*-acting lncRNA called Xert. However, they drive Xist expression only in female cells, as those maintain the promoter-proximal region in a responsive state.

Gjaltema et al., 2022, Molecular Cell *82*, 190–208
January 6, 2022
https://doi.org/10.1016/j.molcel.2021.11.023

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

Distal and proximal *cis*-regulatory elements sense X chromosome dosage and developmental state at the *Xist* locus

Rutger A.F. Gjaltema,[1,7] Till Schwämmle,[1,7] Pauline Kautz,[1] Michael Robson,[2,3] Robert Schöpflin,[2,4,5] Liat Ravid Lustig,[1] Lennart Brandenburg,[1] Ilona Dunkel,[1] Carolina Vechiatto,[1] Evgenia Ntini,[1] Verena Mutzel,[1] Vera Schmiedel,[1] Annalisa Marsico,[6] Stefan Mundlos,[2,4] and Edda G. Schulz[1,8,*]

[1]Otto Warburg Laboratories, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[2]Development and Disease Group, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[3]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh EH4 2XU, Edinburgh, UK
[4]Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, 13353 Berlin, Germany
[5]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[6]Computational Health Center, Helmholtz Center München, 85764 Neuherberg, Germany
[7]These authors contributed equally
[8]Lead contact
*Correspondence: edda.schulz@molgen.mpg.de
https://doi.org/10.1016/j.molcel.2021.11.023

SUMMARY

Developmental genes such as *Xist*, which initiates X chromosome inactivation, are controlled by complex *cis*-regulatory landscapes, which decode multiple signals to establish specific spatiotemporal expression patterns. *Xist* integrates information on X chromosome dosage and developmental stage to trigger X inactivation in the epiblast specifically in female embryos. Through a pooled CRISPR screen in differentiating mouse embryonic stem cells, we identify functional enhancer elements of *Xist* at the onset of random X inactivation. Chromatin profiling reveals that X-dosage controls the promoter-proximal region, while differentiation cues activate several distal enhancers. The strongest distal element lies in an enhancer cluster associated with a previously unannotated *Xist*-enhancing regulatory transcript, which we named *Xert*. Developmental cues and X-dosage are thus decoded by distinct regulatory regions, which cooperate to ensure female-specific *Xist* upregulation at the correct developmental time. With this study, we start to disentangle how multiple, functionally distinct regulatory elements interact to generate complex expression patterns in mammals.

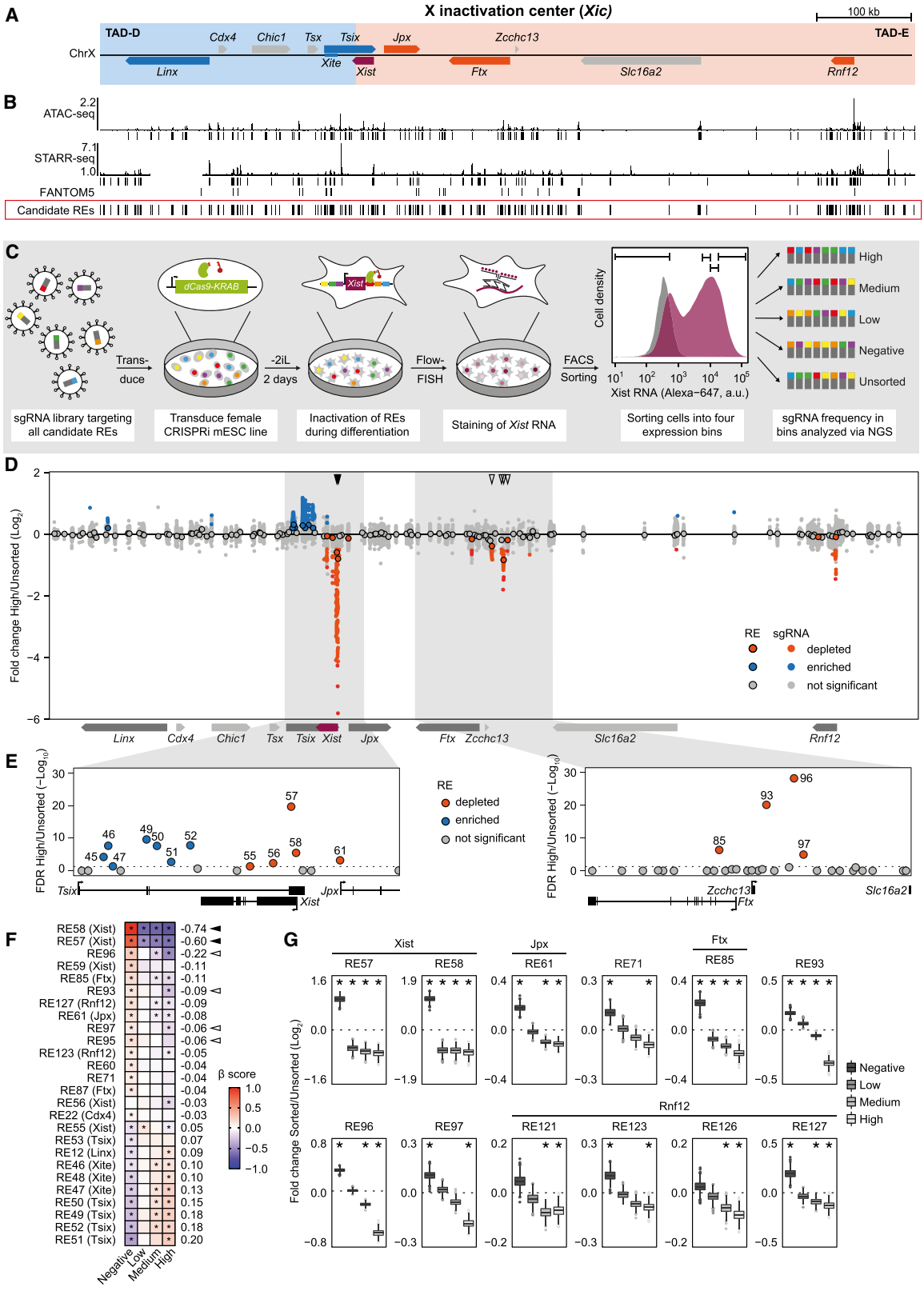
INTRODUCTION

During embryonic development, correct spatiotemporal gene expression is controlled by complex *cis*-regulatory landscapes (Bolt and Duboule, 2020). Multiple *trans*-acting signals in the form of sequence-specific transcription factors bind to *cis*-acting proximal and distal regulatory elements (REs) and control transcription from a gene's core promoter, to precisely tune tissue- and stage-specific gene expression (Long et al., 2016; Spitz and Furlong, 2012). Another layer of regulation is composed of long non-coding RNAs (lncRNAs) that regulate neighboring genes in *cis* and are often transcribed from or through enhancer elements (Gil and Ulitsky, 2018). Although *cis*-regulatory landscapes have been mapped for a number of genes (Fulco et al., 2016, 2019; Klann et al., 2017), it remains poorly understood how they decode complex information to precisely tune gene expression during development.

Here we use the murine *Xist* locus as a model to study information processing by *cis*-regulatory landscapes. *Xist* is an essential developmental regulator, which initiates X chromosome inactivation (XCI) in females to ensure X chromosome dosage compensation between the sexes (Brown et al., 1991; Penny et al., 1996). *Xist* is upregulated during early embryonic development from one of two X chromosomes in females in an X-dosage-dependent manner and induces chromosome-wide gene silencing (Galupa and Heard, 2018; Mutzel and Schulz, 2020; Żylicz and Heard, 2020). The *Xist* locus must thus integrate differentiation cues and X-dosage information to establish the correct expression pattern.

In mice, XCI occurs in two waves. Shortly after fertilization, the paternal X chromosome (Xp) is inactivated in an imprinted form of XCI, which is maintained in the extraembryonic tissues (Mak et al., 2004; Okamoto et al., 2004). In the inner cell mass of the blastocyst, which will give rise to the embryo, the Xp becomes reactivated again. Shortly after, at the epiblast stage, random XCI is initiated, causing each cell to inactivate either the paternal or the maternal X. Random XCI, which is thought to occur in all

CellPress
OPEN ACCESS



(legend on next page)

placental mammals, can be recapitulated in cell culture by inducing differentiation of pluripotent mouse embryonic stem cells (mESCs) (Monk, 1981).

The regulatory landscape of *Xist*, called the X-inactivation center (*Xic*), is thought to encompass a region of ~800 kb surrounding the *Xist* gene (Figure 1A). The *Xic* is structured into two topologically associating domains (TADs), TAD-D and TAD-E, with the *Xist* gene being transcribed across their boundary (Nora et al., 2012). TAD-D contains several *Xist* repressors, including *Xist*'s non-coding antisense transcript *Tsix*, the *Tsix* enhancer region *Xite*, and the more distal *Linx* locus (Galupa et al., 2020; Lee and Lu, 1999; Lee et al., 1999; Luikenhuis et al., 2001; Nora et al., 2012; Ogawa and Lee, 2003). TAD-E comprises the *Xist* promoter and multiple positive regulators, including two more lncRNA genes, *Jpx* and *Ftx*, which activate *Xist* expression (Chureau et al., 2011; Furlan et al., 2018; Tian et al., 2010). In addition, TAD-E contains the protein-coding *Rnf12*/*Rlim* gene, which contributes to X-dosage-dependent *Xist* upregulation (Jonkers et al., 2009). Although a series of *cis*- and *trans*-acting *Xist* activators have thus been identified, to our knowledge, no classical enhancer elements have been described to date.

X-dosage information is in part transmitted through a double dose of RNF12 in female cells, which targets the *Xist* repressor REX1/ZFP42 for degradation (Gontan et al., 2012, 2018; Jonkers et al., 2009). REX1 is thought to repress *Xist* indirectly by enhancing *Tsix* transcription (Navarro et al., 2010) and directly through binding *Xist*'s transcription start site (TSS) and a CpG island ~1.5 kb downstream of the TSS, where it competes for binding with the ubiquitous *Xist* activator YY1 (Chapman et al., 2014; Gontan et al., 2012; Makhlouf et al., 2014). Developmental regulation of *Xist* has been attributed to pluripotency factors (Donohoe et al., 2009; Navarro et al., 2008, 2010; Payer et al., 2013). They repress *Xist* in pluripotent cells, while their downregulation following differentiation triggers *Xist* upregulation. This pluripotency factor-induced repression is thought to be mediated by a pluripotency factor binding site within intron 1 of *Xist*, together with transcriptional activation of *Tsix* (Donohoe et al., 2009; Navarro et al., 2008, 2010). However, neither deletion of the intronic binding site nor of the *Tsix* promoter results in *Xist* upregulation prior to differentiation (Barakat et al., 2011; Lee and Lu, 1999; Minkovsky et al., 2013; Nesterova et al., 2011). It thus remains an open question how the developmental state is sensed by the *Xist* locus and whether developmental regulation is indeed ensured through pluripotency factor repression alone or whether differentiation cues might also activate *Xist*.

To understand how the complex *cis*-regulatory landscape of *Xist* integrates information on X-dosage and development, we comprehensively mapped *cis*-regulatory elements that control *Xist* in mESCs. We then profiled how their activity is modulated by X-dosage and differentiation. In this way, we identified an enhancer cluster that controls developmental *Xist* upregulation and is associated with a previously unannotated transcript we named *Xert*. We found that the *Xert* locus activates *Xist* transcription in *cis* and interacts with the *Xist* promoter in three-dimensional (3D) space. Overall, our data show that differentiation cues are integrated by a series of distal REs. However, they can stimulate *Xist* transcription only in female cells, for which double X-dosage acts to prevent repression of the promoter-proximal region.

RESULTS

Identification of *Xist*-controlling REs through a pooled CRISPR screen

To understand how information is processed by *Xist*'s regulatory landscape, we comprehensively identified REs that control *Xist* upregulation at the onset of random XCI. We performed a pooled CRISPR interference (CRISPRi) screen (Klein et al., 2018), where catalytically dead Cas9 (dCas9) fused to a KRAB repressor domain is targeted to putative REs in a pooled fashion to inactivate one RE per cell. Subsequent enrichment of cells with high or low *Xist* expression allows identification of functional REs by comparing single guide RNA (sgRNA) abundance among cell populations.

To establish a set of candidate REs to be tested in the screen, we profiled DNA accessibility using assay for transposase-accessible chromatin using sequencing (ATAC-seq) and enhancer activity in a massively parallel reporter assay, called STARR-seq, both in mESCs and upon differentiation (Figure 1B; Figure S1A; see STAR Methods for details) (Arnold et al., 2013; Buenrostro et al., 2013). After integrating these datasets with enhancer regions reported by the FANTOM5 consortium (Lizio et al., 2015), we defined a total of 138 candidate REs with a

Figure 1. Identification of *Xist*-regulating genomic elements through a pooled CRISPR screen

(A) Schematic representation of the *Xic*; known *Xist* regulators are red (activators) and blue (repressors).

(B) ATAC-seq, STARR-seq, and FANTOM5 data used to identify candidate REs (red box). Vertical bars below the tracks represent peaks identified by MACS2 (false discovery rate [FDR] < 0.1). A region within *Linx* was missing from the STARR-seq library.

(C) Schematic outline of the pooled CRISPRi screen used to identify functional *Xist* REs in (D)–(G). After lentiviral transduction with the sgRNA library, female mESCs stably expressing a CRISPRi system were differentiated for 2 days by 2iL withdrawal and stained for *Xist* RNA using Flow-FISH (purple; gray, undifferentiated cells). SgRNA distributions were analyzed in four expression bins, each containing 15% of cells as indicated.

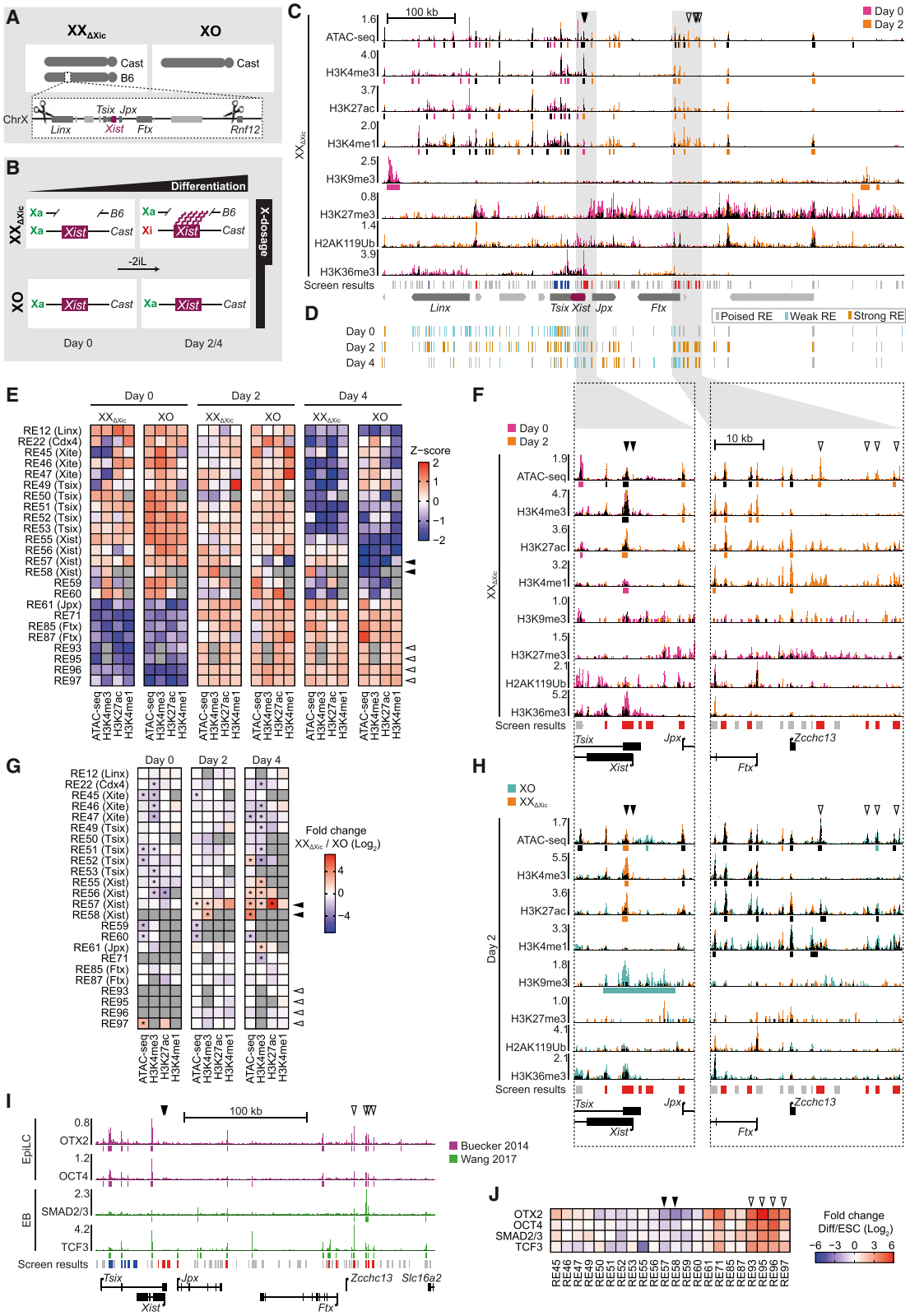
(D and E) Comparison of sgRNA abundance in the Xist-high fraction compared with the unsorted population. Small dots in (D) show individual sgRNAs, and rimmed circles in (D) and (E) represent REs. Significantly enriched and depleted sgRNAs (MAGeCK test, two-sided $p < 0.05$) and REs (MAGeCK mle, Wald FDR < 0.05) are colored blue and red, respectively. In (E), significantly enriched or depleted REs are annotated with their number. The dashed line represents an FDR of 0.05.

(F) Heatmap showing effect size estimated using MAGeCK mle (β score) for each sorted fraction. All candidate REs significantly enriched or depleted in at least one fraction (FDR < 0.05; asterisks) are shown and sorted by their mean β score across all fractions (indicated on the right), with the score for the negative fraction negated.

(G) Fold change of sorted and unsorted populations for 1,000 bootstrap samples of 50 randomly selected sgRNAs for each RE. REs in TAD-E with an empirical FDR < 0.01 (asterisks) in at least two populations are shown.

Arrowheads in (D) and (F) indicate the promoter-proximal elements RE57–58 (filled) and the newly identified distal enhancers RE93, 95–97 (open). See also Figure S1 and Tables S1 and S2.

CellPress
OPEN ACCESS



(legend on next page)

median length of 991 bp (Figure 1B; Figures S1B and S1C; see STAR Methods for details). SgRNAs targeting those candidate REs were cloned into a lentiviral vector, resulting in a library with 7,358 sgRNAs and a median of 43 sgRNAs per RE (Figures S1D and S1E; Table S1).

The screen was performed in a female mESC line (TX-SP107) stably expressing an abscisic acid (ABA)-inducible CRISPRi system (Gao et al., 2016). The parental TX1072 line has been derived by crossing two distantly related mouse strains, C57BL/6 (B6) and Cast/EiJ (Cast) (Schulz et al., 2014). Although TX1072 mESCs carry a doxycycline-inducible TetO element upstream of the *Xist* TSS at the B6 chromosome and express the rtTa transactivator, *Xist* is upregulated normally upon differentiation in the absence of doxycycline (Pacini et al., 2021). TX-SP107 mESCs were transduced with the sgRNA library, differentiated for 2 days by 2i/LIF (2iL) withdrawal and stained for *Xist* RNA using Flow-FISH (fluorescent *in situ* hybridization) (Figure 1C; Figures S1F and S1G). After sorting four cell populations (Xist negative, low, medium, and high), sgRNA frequency within the unsorted and sorted populations was determined using deep sequencing (Figure 1C; Figures S1H–S1J).

To identify REs controlling *Xist*, we compared sgRNA abundance between sorted and unsorted populations using the MAGeCK tool suite (Figures 1D and 1E; Figures S1K and S1L; Table S2) (Li et al., 2014, 2015). All regions previously described to activate *Xist* were depleted from the Xist-high fraction and enriched in the negative population, while known repressive elements showed the opposite pattern (see Table S3). Among others, the screen identified the *Xist* promoter (RE58), the promoter-proximal CpG island (RE57) (Johnston et al., 1998; Newall et al., 2001), the *Jpx* promoter (RE61) (Tian et al., 2010), the *Ftx* promoter region (RE85, RE87) (Chureau et al., 2011; Furlan et al., 2018; Soma et al., 2014), and multiple elements within *Tsix* (RE46–50) (Cohen et al., 2007; Lee and Lu, 1999; Ogawa and Lee, 2003; Vigneau et al., 2006) (Figures 1D and 1E; Figures S1K and S1L). Several of the identified elements overlap with promoters of known *Xist* regulators. Their effect on *Xist* might be mediated either by the linked transcripts or by an enhancer-like function of promoters, as previously reported (Dao et al., 2017; Engreitz et al., 2016).

In addition to known elements, the screen also identified several regions that, to our knowledge, have not yet been shown to regulate *Xist*. Multiple intronic elements within *Tsix* (RE51–53) had repressive effects, and an element downstream of *Rnf12* (RE123), which might act as an *Rnf12* enhancer, activated *Xist* expression (Figures 1D and 1E; Figures S1K and S1L). The most prominent region identified was a cluster of activating REs (RE93, 95–97) ~150–170 kb telomeric to *Xist*, which were all enriched in the Xist-negative population, and all except RE95 were also depleted from Xist-high cells (Figures 1D and 1E; Figures S1K and S1L, open arrowheads).

We next ranked all REs according to their contribution to *Xist* regulation using two different approaches (Table S2; see STAR Methods for details). As expected, the strongest activating regions were located around the *Xist* promoter, most notably at the TSS (RE58) and the promoter-proximal CpG island (RE57) (Figure 1F, filled arrowheads). Among the distal elements, the newly discovered RE96 showed the strongest effect, followed by a region in *Ftx* (RE85) and another previously unknown element, RE93 (Figure 1F). Interestingly, we observed distinct enrichment patterns among elements, across the different Xist-positive populations. Although promoter-proximal REs (RE57, 58) were depleted to a similar extent across populations, most distal elements, in particular the newly identified RE93–97 region, showed a gradual increase in depletion from the Xist-low to Xist-high populations (Figure 1G; Figure S1M). The promoter-proximal elements thus appear to control *Xist* in a binary fashion, constituting an on/off switch, while distal elements modulate expression levels.

Proximal and distal elements integrate X-dosage information and differentiation cues

In the next step, we investigated how activity of the identified *Xist*-controlling REs was modulated by differentiation and X chromosome dosage. We profiled histone modifications and DNA accessibility in mESCs and at day 2 (when *Xist* is strongly upregulated) or day 4 (when gene silencing is established) of differentiation (Pacini et al., 2021). To assess the chromatin state of the inactive X (Xi), which upregulates *Xist*, we used a female mESC line ($XX_{\Delta Xic}$) with a heterozygous ~800 kb deletion around *Xist* (Figures 2A and 2B) (Pacini et al., 2021). In this line, *Xist* is

Figure 2. Differentiation cues and X-dosage control distal and proximal REs, respectively

(A) Cell lines used to assess X-dosage effects, where an *Xic* deletion in female $XX_{\Delta Xic}$ ($TX\Delta Xic_{B6}$) mESCs allows profiling of the Xi, while an XO line is used to profile the Xa.

(B) Schematic representation of the experimental setup used in (C)–(H), where the cell lines shown in (A) are differentiated by 2iL withdrawal to induce *Xist* upregulation in the $XX_{\Delta Xic}$ mESCs.

(C) Overlay of DNA accessibility and histone modifications in female $XX_{\Delta Xic}$ mESCs profiled using ATAC-seq and CUT&Tag at days 0 and 2. Peaks called in at least one time point are shown below some tracks and are colored if significantly different between time points across biological replicates (FDR < 0.05).

(D) Chromatin segmentation using ChromHMM.

(E) Quantification of the data shown in (C) at *Xist*-regulating REs. REs with insufficient coverage are colored gray.

(F) Zoom-ins of (C).

(G) Heatmap as in (E), showing the fold change between the $XX_{\Delta Xic}$ and XO cell lines. Asterisks mark significant differences ($p < 0.05$) according to DiffBind analysis.

(H) Same as (F), but showing the $XX_{\Delta Xic}$ and XO cell lines at day 2.

(I) Published ChIP-seq tracks (Buecker et al., 2014; Wang et al., 2017) for OTX2, OCT4, SMAD2/3, and TCF3 in epiblast-like cells (EpiLCs) or embryoid bodies (EBs).

(J) Quantification of (I) and corresponding data in ESCs within *Xist*-regulating REs.

Arrowheads in (C)–(J) indicate the promoter-proximal elements RE57–58 (filled) and the distal enhancers RE93, 95–97 (empty). The screen results (Figure 1) are shown below the tracks, and inhibiting (blue) or activating (red) REs are colored. See also Figure S2 and Table S4.

upregulated exclusively from the wild-type allele and signals detected within the deleted region will thus originate from the Xi. *Xist* expression was only slightly reduced in $XX_{\Delta Xic}$ cells compared with the parental TX1072 line, with >70% of cells expressing *Xist* at days 2–4 of differentiation (Figures S2A and S2B). To profile the active X (Xa) in a cellular context with single X chromosome dosage, we used an XO subclone of the parental mESC line, which, similar to male mESCs, does not upregulate *Xist* (Figure S2C).

We profiled seven histone modifications (H3K4me3, H3K27ac, H3K4me1, H3K9me3, H3K27me3, H2AK119ub, and H3K36me3), using CUT&Tag (Kaya-Okur et al., 2019). The data showed the expected peak patterns and were in good agreement with native chromatin immunoprecipitation sequencing (ChIP-seq) in the parental line (Figures S2D–S2F; Table S4) (Żylicz et al., 2019). Although *Xist* was strongly upregulated at day 2 in $XX_{\Delta Xic}$ cells (Figure S2C), few changes occurred at its promoter-proximal region. It was devoid of repressive marks already in undifferentiated cells, exhibited DNA accessibility and was decorated by active histone modifications, such as H3K4me3, H3K4me1, and H3K27ac (Figures 2C–2F). Only a small but significant increase in H3K27ac and loss of the H3K4me1 mark was observed upon differentiation, together with a reduction of H3K36me3 (Figures 2E and 2F; Figure S2G). The latter likely reflects *Tsix* downregulation, which is thought to repress *Xist* by co-transcriptional deposition of this mark (Loos et al., 2015; Ohhata et al., 2015). The *Xist* promoter thus resides in a "poised" state already prior to differentiation.

In contrast to promoter-proximal elements, the distal REs identified in our screen (Ftx, RE93–97) were largely inactive in undifferentiated cells and gained active chromatin marks and DNA accessibility only during differentiation (Figures 2E and 2F; Figure S2G). This observation was confirmed by chromatin segmentation with ChromHMM (Figure 2D; Figure S2H) (Ernst and Kellis, 2012). Moreover, the distal elements were covered by a broad H3K27me3 domain in undifferentiated cells (Figure 2C; Figure S2G), which corresponds to a previously described "H3K27me3 hotspot" (Marks et al., 2009; Rougeulle et al., 2004). As previously reported, the hotspot disappeared during differentiation (Marks et al., 2009), potentially contributing to the observed activation of the Ftx-RE93–97 region (Figures 2E and 2F; Figure S2G). These results suggest that *Xist* upregulation during differentiation is driven primarily by distal regulatory elements.

When comparing distal REs between $XX_{\Delta Xic}$ and XO cells, we found that they gained active marks and lost H3K27me3 in a similar manner in both cell lines, suggesting a largely X-dosage-independent regulation (Figures 2G and 2H; Figures S2I and S2J). The only regions with higher activity in $XX_{\Delta Xic}$ than XO cells at day 2 were the promoter-proximal elements. Although they appeared mostly active in both cell lines at day 0, they lost activity in XO cells during differentiation. Concomitantly, a broad ~16-kb-wide H3K9me3 domain, covering the *Xist* promoter region, appeared only in XO cells at days 2 and 4 (Figure 2H; Figure S2J) and was similarly observed in differentiating male ESCs in a published dataset (Figure S2K) (Bleckwehl et al., 2021). X chromosome dosage thus appears to control mainly the promoter-proximal region, where it counteracts active repression by H3K9me3 during differentiation. Developmental cues, on the other hand, are sensed primarily by distal REs.

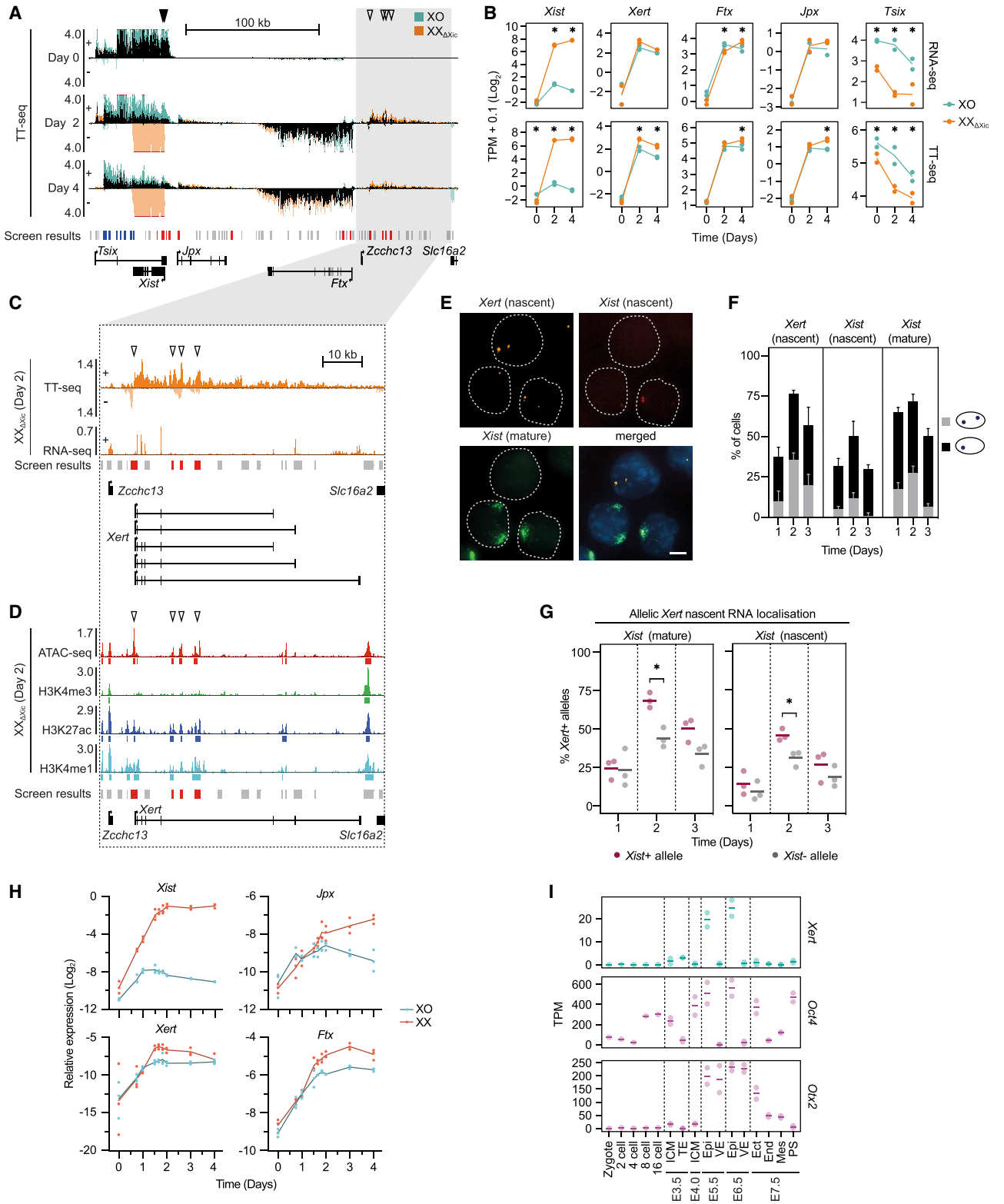
To investigate how distal REs, in particular RE93–97, are activated, we screened for transcription factors that might regulate these elements. We identified factors enriched at RE93, 95, 96, and 97 using the Cistrome database (Zheng et al., 2019), which contains a large collection of published ChIP-seq experiments in different cell types and tissues (Figures 2I and 2J; Figure S2L). All four REs were bound by OTX2 in epiblast-like cells (EpiLCs), a factor that regulates epiblast differentiation (Acampora et al., 2013; Yang et al., 2014) and induces repositioning of OCT4 (Buecker et al., 2014), which is also recruited to RE93–97 in a differentiation-dependent manner. Moreover, we detected binding of two other regulators of ESC differentiation, SMAD2/3 and TCF3 (Guo et al., 2011; Pauklin and Vallier, 2015), specifically in differentiating cells (Wang et al., 2017). Of note, pluripotency factors, which have previously been implicated in developmental regulation of *Xist* (Donohoe et al., 2009; Navarro et al., 2008, 2010; Payer et al., 2013), did not bind to RE93–97 (Figure S2M) (Chronis et al., 2017; Gontan et al., 2012; Tu et al., 2016).

In sum, the *Xist* promoter is already in a mostly active chromatin configuration prior to differentiation, while distal enhancers are inaccessible and covered by a broad repressive H3K27me3 domain. These distal elements are then activated by several differentiation-associated transcription factors in $XX_{\Delta Xic}$ and XO cells, but *Xist* upregulation appears to be prevented in XO cells through H3K9me3 deposition at the *Xist* promoter.

A lncRNA named *Xert* is transcribed through RE95–97 concomitantly with *Xist* upregulation

To investigate transcriptional activity at *Xist*-controlling REs, we profiled nascent transcription and mature RNA expression in our $XX_{\Delta Xic}$-XO model by transient transcriptome sequencing (TT-seq; Schwalb et al., 2016) and RNA sequencing (RNA-seq), respectively (Figures 3A and 3B; Figure S3A; Table S5). We detected an unannotated transcript, which overlapped with RE93–97 and was expressed upon differentiation in both cell lines (Figure 3A, gray box). Through polyA-enriched RNA-seq as well as 3′- and 5′-RACE, we identified several relatively short (~400–800 bp), spliced and poly-adenylated transcripts originating from a ~50 kb genomic region (Figure 3C; Figures S3B and S3C). They showed limited protein-coding potential, supporting a classification as lncRNAs (Table S6). The main TSS was located within RE93 and exhibited a chromatin state typical for enhancers, characterized by chromatin accessibility, bidirectional transcription, H3K27ac, and a high H3K4me1/H3K4me3 ratio (Figures 3C and 3D), reminiscent of a previously described lncRNA class that are transcribed from enhancer elements (Gil and Ulitsky, 2018; Marques et al., 2013; Tan et al., 2020). As the promoter of this unknown transcription unit was identified as an *Xist* enhancer element in our screen (RE93), we hypothesized that the lncRNA might activate *Xist* transcription, similar to *Jpx* and *Ftx*. We thus named the locus *Xist*-enhancing regulatory transcript (*Xert*).

To further characterize *Xert*, we assessed its expression dynamics both *in vitro* and *in vivo*. We performed RNA-FISH for *Xert* and *Xist* in differentiating female mESCs (Figures 3E and 3F). *Xert* transcription foci were observed in 40%–70% of cells,



(legend on next page)

and *Xist* was initially upregulated from two alleles in a subset of cells, as previously reported for cells differentiated by 2iL withdrawal (Guyochin et al., 2014; Pacini et al., 2020; Sousa et al., 2018). We found that *Xert* was more frequently detected on Xist-positive than on Xist-negative alleles, suggesting that it might activate *Xist* in *cis* (Figure 3G). We then performed a high-resolution time-course experiment, which revealed that *Xert* was upregulated concomitantly with *Xist*, *Jpx*, and *Ftx* at the onset of differentiation (Figure 3H). In contrast to *Ftx* and *Jpx*, which maintained high expression throughout the time course, *Xert* levels started to decrease after day 2, suggesting a role in initial *Xist* upregulation. *Xert* reached ~4-fold higher levels in wild-type XX mESCs compared with XO cells, which is more than expected from the 2-fold copy number difference for the *Xert* gene between the two cell lines. Therefore, not only differentiation cues but also X chromosome dosage appeared to modulate *Xert* expression, further supported by slightly higher expression in $XX_{\Delta Xic}$ compared with XO cells (Figure 3B).

We next reanalyzed published datasets to characterize activity of the *Xert* region *in vivo*. RNA-seq data from sex-mixed embryos and adult tissues (Bauer et al., 2021; Deng et al., 2014; Söllner et al., 2017; Wang et al., 2019; Zhang et al., 2018) revealed that *Xert* was specifically expressed at the onset of random XCI, which occurs in the epiblast at embryonic days (E) 5.5 and 6.5, but not in somatic cells (Figure 3I; Figures S3D and S3E) (Mak et al., 2004; Shiura and Abe, 2019). During early embryogenesis, the *Xert* expression pattern largely mirrored co-expression of *Xert* binding factors *Otx2* and *Oct4* (Figures 2I, 2J, and 3I), further supporting a role of these factors in *Xert* regulation. Reanalysis of ChIP-seq data from post-implantation embryos (Yang et al., 2019) showed that the *Xert* promoter and the RE95–97 enhancer region located in its longest intron were marked with an active enhancer signature (H3K4me1 and H3K27ac) in the E6.5 epiblast with levels decreasing at E7.5 (Figure S3F).

To assess conservation of the *Xert* locus in humans, we reanalyzed several published datasets of human preimplantation embryos and ESCs (Guo et al., 2017; Li et al., 2019; Petropoulos et al., 2016; Rada-Iglesias et al., 2011; Wang et al., 2015). In agreement with poor sequence conservation, we could not detect a longer transcript resembling *Xert* (Figures S3G–S3I). However, we did identify an active enhancer element in human ESCs, marked by open chromatin, H3K27ac, H3K4me1, and p300, which was bound by SMAD2/3 similarly to the mouse *Xert* locus (Figure S3I), suggesting a functional conservation of enhancers in the same genomic region.

Taken together, we have identified a lncRNA within the *Xic*, which is associated with a series of functional *Xist*-activating elements. As it is specifically expressed at the onset of random XCI and is positively correlated with *Xist* transcription, it might function as an early *cis*-acting *Xist* activator.

Xert activates *Xist* in *cis*

To test a functional role of *Xert* in *Xist* regulation, we perturbed *Xert* transcription using multiple approaches (Figure 4A). First, we attenuated *Xert* promoter (XertP) activity in female cells using CRISPRi (Figures 4B and 4C). A ~20-fold reduction of *Xert* levels resulted in ~2-fold reduced *Xist* expression at day 2 (Figure 4B, right). Flow-FISH revealed a 10%–20% decrease in *Xist*-expressing cells with a 25% reduction of *Xist* levels within the positive population (Figure 4C). Next, we overexpressed *Xert* in male cells using the SunTag CRISPR activation (CRISPRa) system (Heurtier et al., 2019; Tanenbaum et al., 2014) in a mESC line carrying a *Tsix* mutation to facilitate ectopic *Xist* upregulation Figures S4, resulting in a significant increase in *Xist* expression both before and during differentiation (Figure 4D; Figures S4A and S4B). Although *Xert* could be induced more strongly in undifferentiated cells (because of low basal expression), the effect on *Xist* was more pronounced in differentiating cells. The *Xert* promoter region thus appears to promote *Xist* expression, in particular in differentiating mESCs.

To test whether *Xert* regulates *Xist* in *cis* or in *trans*, we deleted the XertP region on one allele in female TX1072 mESCs and assessed the effect on *Xist* expression. We deleted a ~1.5 kb region around the *Xert* TSS either on the Cast or B6 allele (ΔXertP; Figure 4A; Figures S4C–S4F). Monoallelic transcription of *Xert* was confirmed using RNA-FISH and pyrosequencing, which performs quantitative sequencing over single SNPs on cDNA (Figure 4E; Figures S4G and S4H). Between 65% and 80% of *Xist* RNA originated from the wild-type allele in ΔXertP cells, compared with 50% in the parental cell line (Figure 4E). Moreover, the deletion led to a shift from biallelic to monoallelic *Xist* expression, with preferential upregulation from the wild-type X

Figure 3. An unannotated enhancer-associated transcript is upregulated concomitantly with *Xist* at the onset of XCI

(A and B) TT-seq (A and B) and RNA-seq (B) in $XX_{\Delta Xic}$ and XO cells. Regions in (A) where the signal extends beyond the depicted range are marked in red; shaded area indicates an unannotated transcript and is enlarged in (C). In (B), differential gene expression between $XX_{\Delta Xic}$ and XO cells was assessed using DEseq2 (FDR < 0.05; asterisks).

(C) TT-seq (both stands) and pA-RNA-seq (+ strand only) tracks of the region shaded in gray in (A) with five isoforms of the newly identified *Xert* transcript (bottom).

(D) DNA accessibility and histone modifications in differentiating $XX_{\Delta Xic}$ cells at the *Xert* locus.

(E–G) RNA-FISH in differentiating XX mESCs (TX1072). An example image (E) is shown for day 2, where nuclei are denoted by a dashed outline and the scale bar marks 5 μm. Quantification (F) and the frequency of *Xert* detection at Xist− or Xist+ alleles (G) are shown for three biological replicates. Asterisks indicate $p < 0.05$ using an unpaired two-tailed t test.

(H) qRT-PCR quantification of *Xist*, *Xert*, *Jpx*, and *Ftx* in XX and XO cells during differentiation. The line connects the mean, dots represent individual replicates (n = 3).

(I) *Xert*, *Otx2*, and *Oct4 (Pou5f1)* RNA expression during early mouse development in sex-mixed embryos (Deng et al., 2014; Zhang et al., 2018). ICM, inner cell mass; TE, trophectoderm; Epi, epiblast; VE, visceral endoderm; Ect, ectoderm; End, endoderm; Mes, mesoderm; PS, primitive streak.

In (G) and (I), the horizontal bar indicates the mean, and dots indicate individual replicates. In (F), mean and SD of three biological replicates are shown. One hundred cells were quantified in each sample. Arrowheads in (A), (C), and (D) indicate RE57–58 (filled) and RE93, 95–97 (open). See also Figure S3 and Tables S5 and S6.

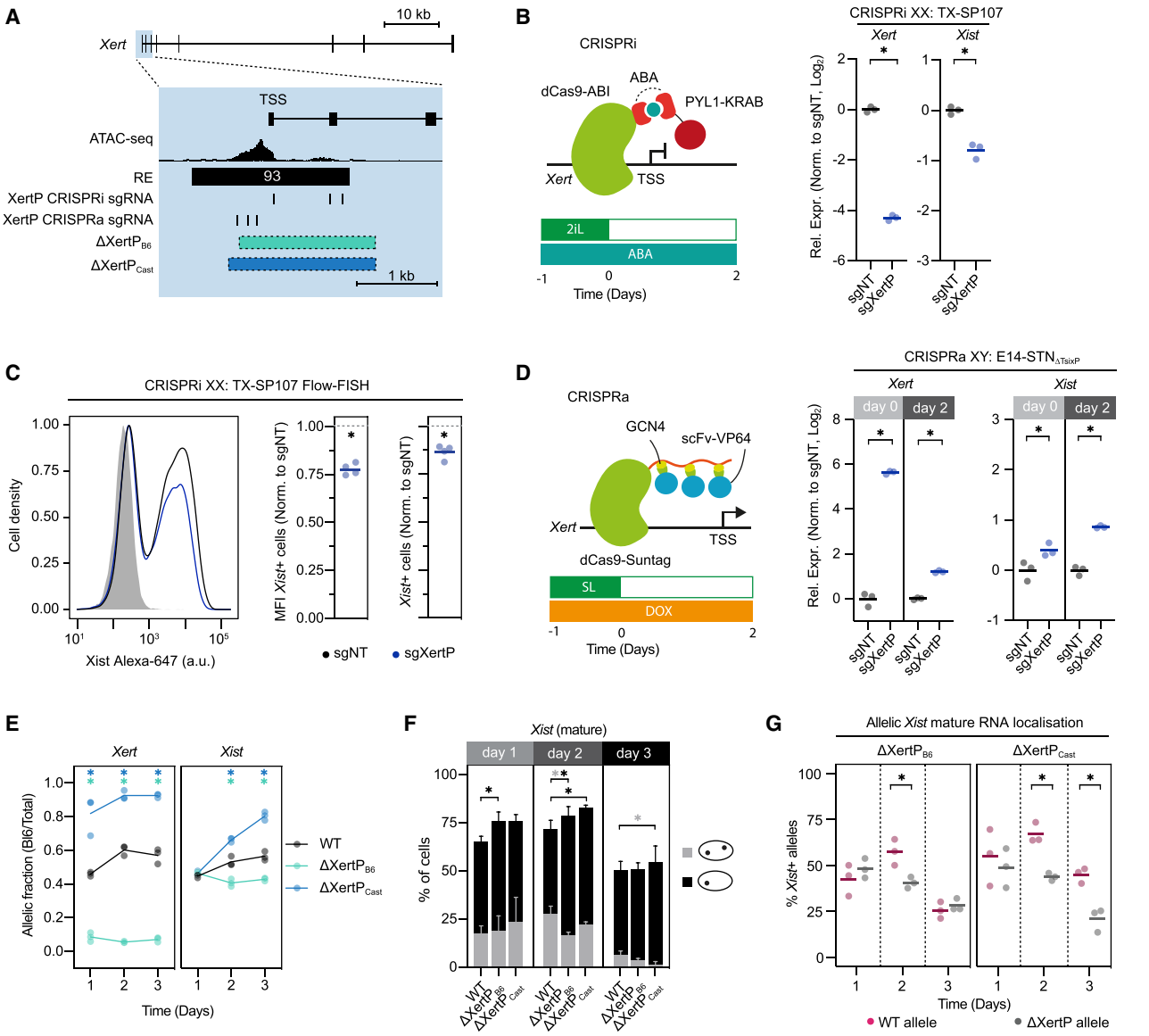


Figure 4. *Xert* enhances *Xist* expression in *cis*

(A) Schematic of XertP perturbations used in (B)–(G).

(B–D) Repression of XertP through an ABA-inducible CRISPRi system in female TX-SP107 mESCs (B and C) and XertP activation through a doxycycline (Dox)-inducible CRISPRa system in male E14-STN$_{\Delta TsixP}$ cells (D). Cells were transduced with multiguide expression vectors of three sgRNAs against XertP as indicated in (A) or with non-targeting (NT) controls. qRT-PCR (B and D) and Flow-FISH (C) of three or four biological replicates after 2 days of differentiation are shown. In (C), the sample shaded in gray denotes undifferentiated (2iL) TX-SP107 cells.

(E–G) Differentiation time course of two heterozygous ΔXertP lines (position shown in A) and the parental wild-type (WT) line assessed using pyrosequencing (E) and RNA-FISH (F and G). For (F) and (G), 100 cells were quantified per replicate. To assess the frequency of *Xist* upregulation from the wild-type (Xert+) or the deleted allele (Xert−) in (G), only cells with a single *Xert* signal (14–66 cells per replicate) were included in the analysis.

Horizontal bars (B–D and G) or lines (E) denote the mean of three or four biological replicates; dots represent individual measurements. In (F), mean and SD of three biological replicates are shown. Asterisks indicate significance of $p < 0.05$ using an unpaired two-tailed t test or for (C) of a one-sample two-tailed t test. Colored asterisks in (G) denote comparison of the respective mutant with the wild-type control. See also Figure S4 and Table S7.

chromosome at day 2 in both clones (Figures 4F and 4G; Figure S4G). At day 3, this effect was lost in ΔXertP$_{B6}$ cells, probably because of preferential *Xist* upregulation from the B6 allele in the parental line (Pacini et al., 2020). These results show that the XertP region enhances *Xist* transcription in *cis*.

In summary, we could show that *Xert* is a *cis*-acting *Xist* activator, as deleting its promoter reduces *Xist* upregulation from the mutated allele. As three *Xist* enhancer elements (RE95–97) are located in an intron of the *Xert* gene, *Xert* might at least in part function by regulating activity of this enhancer cluster.

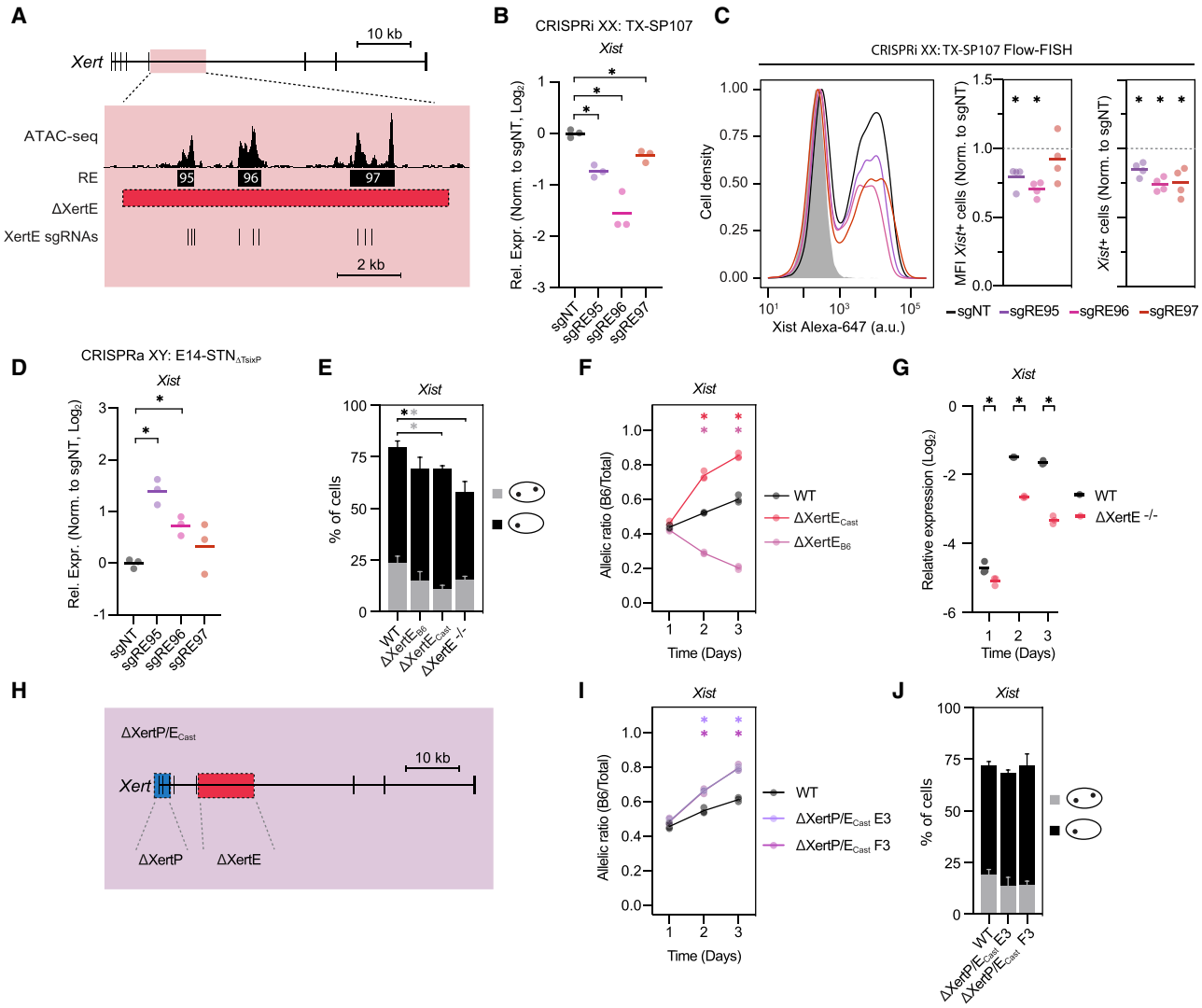


Figure 5. An intronic enhancer cluster within *Xert* activates *Xist* expression in *cis*

(A) Schematic of *Xert* enhancer cluster (XertE) perturbations used in (B)–(G).

(B–D) *Xist* expression assessed using qRT-PCR (B and D) and Flow-FISH (C) at day 2 of differentiation upon targeting of XertE with CRISPRi (B and C) and CRISPRa (D) in female (B and C) and male (D; ΔTsix) mESCs as indicated. In (C), the sample shaded in gray denotes undifferentiated (2iL) TX-SP107 cells.

(E–G) Characterization of female mESC lines carrying a heterozygous (E and F) or homozygous deletion of XertE. *Xist* expression was quantified using RNA-FISH (E), pyrosequencing (F), and qRT-PCR (G).

(H–J) Characterization of heterozygous ΔXertP/E double-mutant mESCs using pyrosequencing (I) and RNA-FISH (J).

Horizontal bars (B–D and G) or lines (F and I) denote the mean of three or four biological replicates (dots). In (E) and (J), mean and SD of n = 3 biological replicates are shown. Asterisks indicate significance of $p < 0.05$ using an unpaired two-tailed t test or for (C) of a one-sample two-tailed t test. See also Figure S5 and Table S7.

Xert-associated enhancer elements control *Xist* upregulation

RE95–97 were identified as *Xist*-activating regions in our screen, with RE96 being the most potent distal element in the *Xic* (Figure 1F). We termed this region *Xert*-associated enhancer cluster (XertE). To confirm a functional role of XertE in *Xist* upregulation, we targeted each RE individually with CRISPRi, which reduced *Xist* expression up to 3-fold after 2 days of differentiation (Figures 5A–5C). Next, we activated XertE through CRISPRa, which led to increased *Xist* expression in male and female cells (Figure 5D; Figure S5A), confirming a functional role of XertE in *Xist* upregulation. To further characterize XertE, we deleted a ~10 kb region containing all three REs either on one or on both X chromosomes in TX1072 mESCs (ΔXertE; Figure 5A; Figures S5B–S5D). We observed a slight, not always significant, reduction of *Xist*-expressing cells at day 2 in all clones (Figure 5E). A strong skewing toward the wild-type allele, which produced 80%–85% of *Xist* RNA in both heterozygous lines, pointed to impaired *Xist* upregulation from the mutant chromosome (Figure 5F). Homozygous XertE deletion led to ~4-fold reduced *Xist* levels (Figure 5G),

thus demonstrating that XertE was not strictly required for *Xist* upregulation. This observation supports the results from our CRISPRi screen that the distal enhancers in XertE do not primarily control frequency of *Xist* upregulation, but rather expression levels (Figure 1G).

As XertE is located within an intron of the *Xert* gene, we next asked whether *Xert* and XertE function independently or in the same pathway. To this end, we deleted XertE in the $\Delta XertP_{Cast}$ line (Figures 5H–5J; Figures S5E and S5F). *Xist* expression was skewed toward the wild-type allele in the double mutant to a similar extent as seen in the XertP and XertE single-mutant lines (compare Figure 5I with Figures 4E and 5F). Thus, the XertP and XertE elements do not function additively but seem to rather lie in the same regulatory axis.

Through a series of perturbations of the XertE region, we could confirm that it functions as an enhancer cluster controlling *Xist* transcription in *cis*. Moreover, XertP and XertE appear to function in the same pathway, supporting the idea that *Xert* might regulate activity of XertE by transcribing through the enhancer cluster.

Feedback and feedforward loops might amplify *Xert* enhancer activity

As XertE lies ~10 kb downstream of the *Xert* promoter, we asked whether it might affect *Xert* transcription in addition to regulating *Xist*. We thus analyzed *Xert* expression upon XertE perturbation (Figures 6A–6D; Figures S6A and S6B). Overall *Xert* levels were reduced upon XertE repression by CRISPRi (Figure 6A). In the ΔXertE mutant mESCs, *Xert* expression was impaired at the deleted allele, which was accompanied by a shift toward more monoallelic expression (Figures 6C and 6D). Conversely, *Xert* expression was increased upon XertE activation with CRISPRa (Figures S6A and S6B). These findings show that XertE also functions as an enhancer of *Xert* transcription itself. If *Xert* would increase activity of XertE by transcribing through the enhancer cluster, as shown for other lncRNAs (Anderson et al., 2016), such mutual activation could constitute a positive feedback loop between XertE and XertP to amplify *Xist* activation.

Next we asked if and how *Xert* might cooperate with other *Xist* activators in TAD-E. We thus analyzed how perturbation of XertE or XertP affected *Jpx*, *Ftx*, and *Rnf12*. *Ftx* expression was reduced when targeting XertE or XertP with CRISPRi and increased upon ectopic activation by CRISPRa (Figures 6A and 6B; Figures S6A–S6C). Similarly, *Ftx* showed a clear skewing toward the non-deleted allele in ΔXertP and ΔXertE cells (Figures 6D and 6E). These findings show that in addition to activating *Xist*, the *Xert* elements also promote *Ftx* expression in *cis*.

As *Ftx* is a well-characterized *cis*-acting *Xist* activator (Furlan et al., 2018), we asked how *Xert* and *Ftx* might cooperate. We therefore generated cell lines with a larger deletion, encompassing the *Ftx* promoter, XertP and XertE (Figure 6F; Figures S6D–S6F). We observed a clear reduction of biallelic *Xist* expression in all clones, a strong skewing of *Xist* of up to 98% toward the wild-type allele in the heterozygous lines and a 3.5-fold reduction of *Xist* levels in the homozygous clone (Figures 6G–6I). The phenotype of the heterozygous deletion was thus clearly more pronounced than that of the XertE deletion alone (compare Figures 6H and 5F). As deletion of the *Ftx* promoter region alone has been reported to induce ~70% skewing (Furlan et al., 2018), which is significantly weaker than the ΔFtx-Xert phenotype, *Ftx* and *Xert* appear to activate *Xist* expression at least in part independently of each other.

Increased contacts between *Xert* and *Xist* during *Xist* upregulation

To further corroborate a direct role of *Xert* in *Xist* regulation, we investigated whether the *Xert* region would spatially interact with the *Xist* promoter and how such interactions might change during *Xist* upregulation. We performed capture Hi-C (cHi-C) within our $XX_{\Delta Xic}$-XO cell model at days 0 and 2 of differentiation, where the *Xic* and surrounding sequences were enriched from a Hi-C library by affinity capture (Figure 7A). In undifferentiated cells, we observed the characteristic split of the *Xic* into TAD-D and TAD-E (Nora et al., 2012) in both cell lines (Figure 7A, top). During differentiation, a sub-TAD formed within TAD-E, which stretched from the *Xist* promoter to a CTCF-site ~20 kb downstream of XertE, thus covering the entire ~200 kb activating region upstream of *Xist* (Figure 7A, bottom; Figure S7A). A comparison between days 0 and 2 revealed an increase in the contact frequency between *Xert* and *Xist* upon differentiation (Figure 7B, left; Figure S7B, left). To investigate whether the identified contact patterns were specific for the inactive X, we compared the contact maps between $XX_{\Delta Xic}$ and XO cells during differentiation (Figure 7B, right; Figure S7B, right). Contact frequencies of *Xist* with *Xert* and *Ftx* were increased in $XX_{\Delta Xic}$ compared with XO cells, which might be either a cause or a consequence of *Xist* expression. In summary, we show that the *Xert* region interacts with *Xist*, supporting its role as an *Xist* enhancer. Moreover, their contact frequency is modulated by differentiation cues and X chromosome dosage. Changes in chromatin conformation of the locus might thus contribute to female-specific and monoallelic *Xist* upregulation at the onset of differentiation.

DISCUSSION

In the present study, we show how an important developmental locus decodes complex input signals to precisely control gene expression. We identify REs that regulate *Xist* during the onset of random XCI. We then categorize them through chromatin profiling in a cell model that allows dissection of X-dosage sensing and developmental regulation. Hereby we show that only the *Xist* promoter-proximal region responds to X-dosage, while developmental cues activate a ~200 kb region upstream of *Xist*, containing *Jpx*, *Ftx*, and the newly identified *Xert* region. Through a series of (epi)genome editing approaches, we show that the *Xert* promoter and a cluster of intronic enhancers within *Xert*'s gene body (XertE) activate *Xist* expression in *cis* and form a regulatory hub with *Ftx*. We can now draw a detailed picture of how distinct transcription factors controlled by X-dosage and differentiation activate specific regulatory regions within the *Xic* to ensure *Xist* upregulation at the epiblast stage in a female-specific manner (Figure 7C).

We discovered a strong distal enhancer cluster of *Xist*, associated with a previously unknown transcript, which we named *Xert*. It had long been suspected that long-range REs must exist in that region, as a ~450 kb single-copy transgene containing *Xist* and

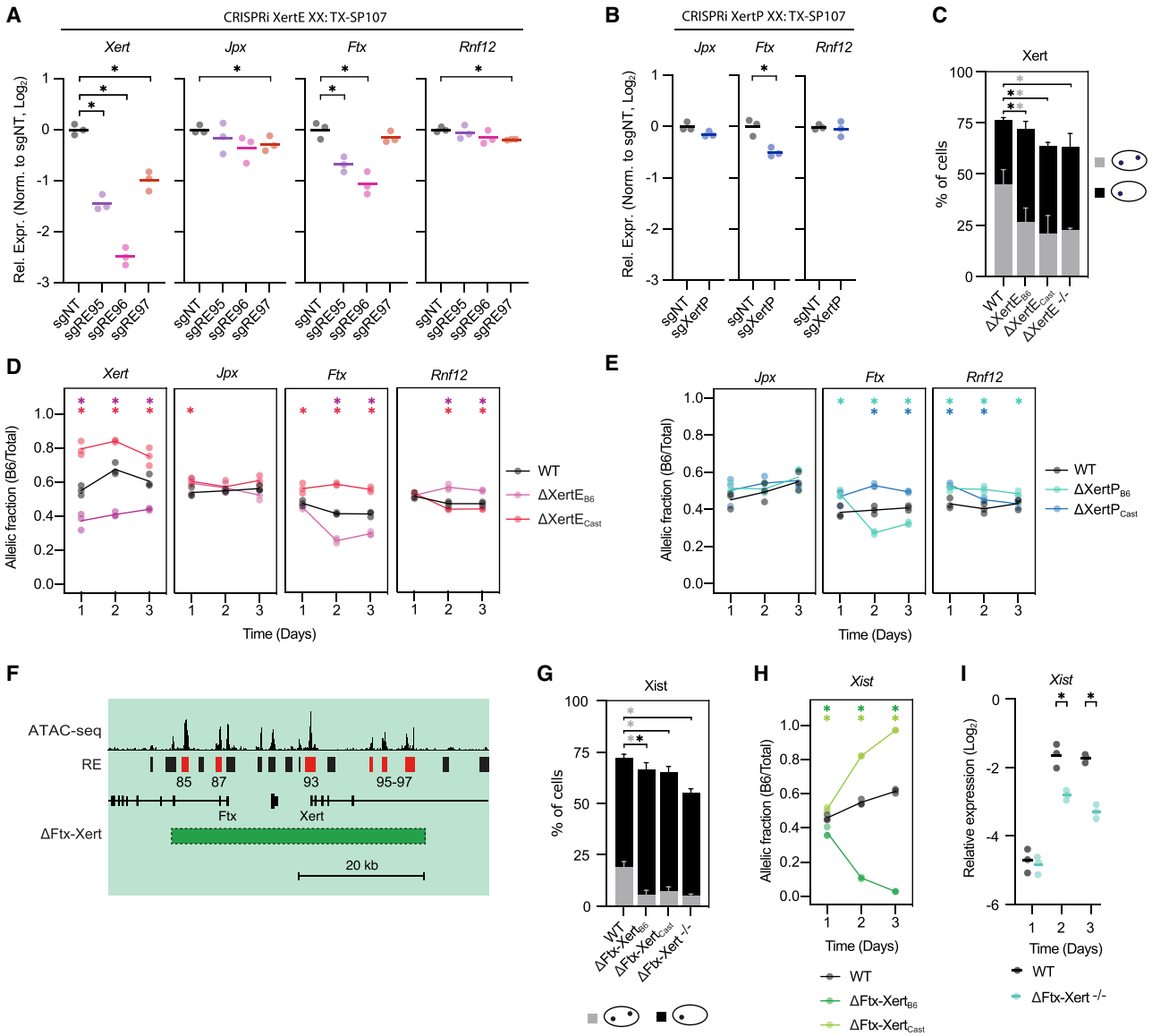


Figure 6. *Xert* and *Ftx* form a regulatory hub during XCI initiation

(A and B) Quantification of *Xist* regulators upon CRISPRi repression of individual REs at XertE (A) and of XertP (B) as in Figures 4B and 5B.

(C–E) Analysis of *Xist* regulators in ΔXertE (C and D) and ΔXertP (E) mutant lines using RNA-FISH (C) and pyrosequencing (D and E).

(F–I) *Xist* expression in lines carrying a heterozygous (G and H) or homozygous (I) Ftx-Xert deletion, assessed using RNA-FISH (G), pyrosequencing (H), and qRT-PCR (I). In (F) the screen results are shown below the ATAC-seq (day 2) tracks.

Horizontal bars (A, B, and I) or lines (D, E, and H) denote the mean of three biological replicates (dots). In (C) and (G) mean and SD of three biological replicates are shown. Asterisks indicate significance of $p < 0.05$ using an unpaired two-tailed t test. Colored asterisks in (D), (E), and (H) denote comparison of the respective mutant line with the wild-type control. See also Figure S6.

~100 kb of upstream sequence, which includes *Jpx*, but not *Xert* and the *Ftx* promoter, cannot drive *Xist* upregulation in tissues undergoing random XCI *in vivo* or *in vitro* (Heard et al., 1996, 1999). We show that *Ftx* and *Xert* cooperate to form a regulatory hub, wherein their transcripts and enhancer elements promote each other's activity to jointly allow strong *Xist* upregulation upon differentiation. For both *Xert* and *Ftx*, their strongest REs (RE85/96) lie within their major transcripts. At both loci, transcription might help activate transcript-embedded enhancers, as shown previously at the *Hand2* locus (Anderson et al., 2016). As nascent transcription can block H3K27me3 deposition (Hosogane et al., 2016; Kaneko et al., 2014; Laugesen et al., 2019), transcription might also accelerate removal of the repressive H3K27me3 hotspot, which covers the entire region before differentiation. Although *Ftx* is expressed rather ubiquitously (Chureau et al., 2011), *Xert* transcription appears to be restricted to a short period when random XCI is initiated. As *Xert* seems to primarily boost *Xist* expression levels, as revealed by the binned

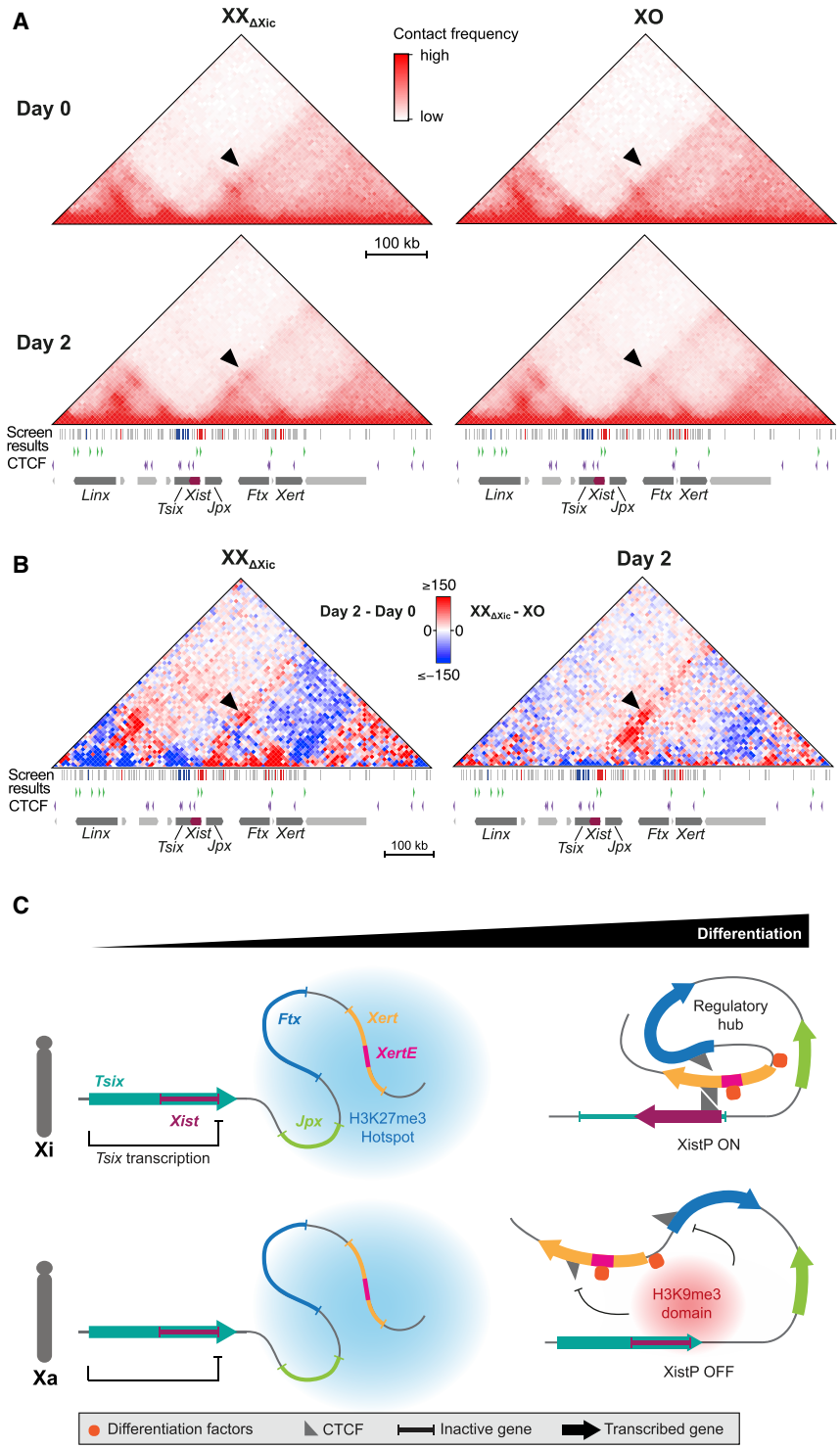


Figure 7. Increased contacts of *Xert* with the *Xist* promoter during XCI initiation

(A) Capture Hi-C in XX$_{\Delta Xic}$ (left) and XO cells (right) at day 0 (top) and day 2 of differentiation (bottom).

(B) Subtraction heatmap of the data shown in (A), comparing days 0 and 2 in XX$_{\Delta Xic}$ cells (left) and XX$_{\Delta Xic}$ and XO cells at day 2 (right). In (A) and (B), two replicates were merged for all samples except XO day 0, for which only a single replicate is available. Arrowheads indicate contacts between *Xert* and *Xist*. Screen results and CTCF binding sites and their orientation are indicated below the heatmaps.

(C) Schematic of *Xist* regulation where *Xist* activators *Jpx*, *Ftx*, and *Xert* are repressed in undifferentiated cells, in part by a H3K27me3 hotspot (blue cloud), and *Xist* is repressed by antisense transcription through *Tsix*. Following differentiation, *Xist* activators are upregulated, and their associated enhancers are activated by differentiation factors. The *Xist* promoter (XistP) acquires a repressive H3K9me3 domain (red cloud) at the Xa, which reduces *Xist* contacts with *Xert* and *Ftx*. At the Xi, XistP is in an active configuration allowing CTCF recruitment and contacts with distal activating elements, thereby inducing *Xist* upregulation. See also Figure S7 and Table S5.

sorting strategy we used in our CRISPR screen, its activation at the onset of XCI might be important to pass a previously postulated activation threshold (Monkhorst et al., 2008; Mutzel and Schulz, 2020; Mutzel et al., 2019). Subsequent *Xert* downregulation might help prevent spurious *Xist* upregulation from the Xa, while *Ftx* and *Jpx* maintain *Xist* expression on the Xi in somatic cells.

Our results finally answer the long-standing question of how developmental regulation of *Xist* is ensured. We show that in addition to downregulation of the repressors *Tsix* and REX1, *Xist* upregulation requires activation of a series of distal enhancer elements, which appear to be controlled by primed pluripotency factors. Among these are SMAD2/3, which are activated by the TGF-β/activin pathway. The activin receptor has previously been identified as XCI activator in two different short hairpin RNA (shRNA) screens, further supporting a role of this pathway in *Xist* regulation (Bhatnagar et al., 2014; Sripathy et al., 2017). Intriguingly, the TGFβ pathway is also regulated by RNF12, which enhances SMAD2/3 signaling via degradation of inhibitory SMAD7 (Zhang et al., 2012). This might be the reason why *Xert* is transcribed slightly more than double in cells with two X chromosomes, which might also contribute to X-dosage-dependent *Xist* regulation. Nevertheless, the distal enhancer elements in the *Ftx-Xert* region were strongly activated both in XX and XO cells upon differentiation, showing that they mainly sense developmental progression.

X-dosage, by contrast, acts primarily on *Xist*'s promoter-proximal region, including a CpG island ~1.5 kb downstream

of the TSS and a region encoding the repeat A of the *Xist* RNA, both of which have previously been implicated in *Xist* regulation (Hoki et al., 2009; McDonald et al., 1998; Norris et al., 1994; Royce-Tolland et al., 2010) (see Table S3). The region is bound by CTCF, YY1, and REX1 (Makhlouf et al., 2014; Navarro et al., 2006), with REX1 being targeted for degradation in an X-dosage-dependent manner (Gontan et al., 2018), further supporting a role of this region in X-dosage sensing. YY1 and also CTCF, both of which have been implicated in long-range chromatin interactions (Nora et al., 2017; Weintraub et al., 2017), bind this region preferably on the Xi in somatic cells, with binding to the Xa likely being inhibited by DNA methylation (Calabrese et al., 2012; Chapman et al., 2014; Makhlouf et al., 2014; Norris et al., 1994; Tarjan et al., 2019). Differential CTCF and YY1 binding between the alleles might underlie the increase in long-range contacts that we observe on the *Xist*-expressing chromosome. At the same time when the *Xist* promoter is activated on the future Xi, we observe active repression at the Xa through deposition of H3K9me3. This might be mediated by TRIM28/KAP1, which has been reported to bind the region on the Xa (Enervald et al., 2021) and recruits H3K9-specific histone methyl-transferases (Ecco et al., 2017). How KAP1 is targeted to the region however remains an open question.

Overall, our analyses reveal that the *Xic* assumes at least three distinct states (Figure 7C). In undifferentiated mESCs, the *Xist* promoter is accessible, but transcription is repressed by *Tsix* and REX1, while distal enhancers are repressed by the H3K27me3 hotspot. Upon differentiation, distal enhancers are derepressed and activated by primed pluripotency factors, resulting in upregulation of *Jpx*, *Ftx*, and *Xert*. Those distal regions will then drive *Xist* upregulation, but only if the promoter-proximal region is maintained in an active configuration by X-dosage-dependent mechanisms, thereby restricting *Xist* upregulation to females. In males, and presumably also on the future Xa in females, the promoter region assumes a heterochromatic state. Activation by distal enhancers and active repression thus appear to be two competing processes at the *Xist* promoter, and their relative dynamics must be tightly tuned in an X-dosage-dependent manner.

Taken together, we have uncovered a regulatory hierarchy at the *Xic*, which allows coincidence detection of two signals that inform the locus on sex and developmental stage of the cell. Similar to other developmental genes, multiple distal elements function as tissue-specific enhancers. The promoter-proximal region by contrast acts as a binary switch, which, when turned off, renders the core promoter unresponsive to long-range regulation. In this way, two signals controlling distal and proximal elements, respectively, are integrated with an AND logic. Our findings are thus the first step toward understanding how logical operations are performed by *cis*-regulatory landscapes to generate the complex expression patterns of developmental genes in mammals.

Limitations of the study

The fact that our CRISPR screen has identified nearly all known REs of *Xist* suggests that the approach is sufficiently sensitive to comprehensively profile *Xist*'s regulatory landscape. However, the screen was performed at an early stage of differentiation, and the relative importance of the various elements might vary in a time- and tissue-specific manner. Moreover, genomic resolution of the screen is limited by the ability of H3K9me3 to spread over several kilobases (Thakore et al., 2015). The relative importance of closely positioned REs, such as RE95–97 within XertE, can thus not be fully resolved. Another open question is whether transcription of *Xert* through XertE indeed modulates activity of the enhancer cluster. Unfortunately, we did not succeed in efficiently terminating *Xert* transcription upstream of XertE, when integrating tandem poly-adenylation signals without additional sequence. The mechanistic interplay between XertP and XertE thus remains to be addressed in the future.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - mESC culture and differentiation
- METHOD DETAILS
  - Molecular cloning
  - Piggybac transposition
  - Lentiviral transduction
  - Genome engineering
  - NGS karyotyping
  - RNA extraction, reverse transcription, qPCR
  - 3′- and 5′ RACE
  - Pyrosequencing
  - RNA FISH
  - Flow-FISH
  - Poly-adenylated RNA-seq and *de novo* transcriptome assembly
  - ATAC-seq
  - STARR-seq
  - CRISPRi screen
  - CUT&Tag of histone modifications
  - TT-seq and RNA-seq
  - Capture Hi-C
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of Flow-FISH data
  - Statistical analysis for qPCR, pyrosequencing and RNA-FISH
  - Statistical analysis of CRISPRi/CRISPRa experiments
  - NGS karyotyping analysis
  - NGS data processing
  - Xert isoforms detection and analysis
  - ATAC-seq data processing
  - STARR-seq data processing
  - CRISPRi screen analysis
  - CUT&Tag analysis

- ChromHMM analysis
- Quantitative analysis of ATAC-seq and CUT&Tag data
- TT-seq analysis
- Analysis of published data
- cHi-C analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2021.11.023.

ACKNOWLEDGMENTS

We want to thank Abhishek Sampath Kumar for help generating the E14-STN$_{\Delta TsixP}$ line, Steven Henikoff for support in setting up CUT&Tag, and Kristina Zumer for support in setting up TT-seq. We thank Tyler Klann and Charles Gersbach for support in setting up the CRISPR screen, Laura Glaser and Melissa Bothe for support in establishing ATAC-seq, and Pablo Navarro for sharing the E14 SunTag cell line. We thank Rafael Galupa for sharing the pFX5 plasmid and critical reading of the manuscript. We thank the Max Planck Institute for Molecular Genetics Seqcore and fluorescence-activated cell sorting (FACS) facilities. This work was supported by the Max-Planck Research Group Leader program, E:bio Module III—Xnet grant (BMBF 031L0072), and Human Frontiers Science Program (CDA-00064/2018) to E.G.S. T.S. is supported by Deutsche Forschungsgemeinschaft (DFG) (IRTG2403, Regulatory Genome). V.M. is supported by DFG (GRK1772, Computational Systems Biology) and M.R. by the Wellcome Trust (206475/Z/17/Z).

AUTHOR CONTRIBUTIONS

R.A.F.G., T.S., and E.G.S. conceived the project and designed the experiments. R.A.F.G. performed CUT&Tag, TT-seq, and CRISPRi/a experiments. T.S. performed ATAC-seq, CRISPRi screen, and CRISPRi Flow-FISH. L.R.L. established the Flow-FISH screening assay. L.B. performed STARR-seq. L.R.L. and V.S. generated CRISPRa/i multiguide plasmids. I.D. and V.S. generated CRISPRi/a mESC lines. V.M. generated the TX1072 XXΔXic$_{B6}$ line with help from R.A.F.G. R.A.F.G. generated and analyzed mESC mutant lines with help from P.K. and T.S. RACE was performed by P.K. and *de novo* transcript assembly by E.N. M.R. performed cHi-C, and R.S. analyzed the data. C.V. performed statistical analysis for CUT&Tag. T.S. performed all other computational analyses. R.A.F.G., T.S., and E.G.S. wrote the manuscript with input from all authors. Funding acquisition was performed by A.M., S.M., and E.G.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 17, 2021
Revised: November 22, 2021
Accepted: November 23, 2021
Published: December 20, 2021

REFERENCES

Acampora, D., Di Giovannantonio, L.G., and Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. Development *140*, 43–55.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. *44* (W1), W3–W10.

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*, 433–436.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202–8.

Barakat, T.S., Gunhanlar, N., Pardo, C.G., Achame, E.M., Ghazvini, M., Boers, R., Kenter, A., Rentmeester, E., Grootegoed, J.A., and Gribnau, J. (2011). RNF12 activates Xist and is essential for X chromosome inactivation. PLoS Genet. *7*, e1002001.

Barakat, T.S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. Mol. Cell *53*, 965–978.

Barros de Andrade E Sousa, L., Jonkers, I., Syx, L., Dunkel, I., Chaumeil, J., Picard, C., Foret, B., Chen, C.J., Lis, J.T., Heard, E., et al. (2019). Kinetics of *Xist*-induced gene silencing can be predicted from combinations of epigenetic and genomic features. Genome Res. *29*, 1087–1099.

Bauer, M., Vidal, E., Zorita, E., Üresin, N., Pinter, S.F., Filion, G.J., and Payer, B. (2021). Chromosome compartments on the inactive X guide TAD formation independently of transcription during X-reactivation. Nat. Commun. *12*, 3499.

Bhatnagar, S., Zhu, X., Ou, J., Lin, L., Chamberlain, L., Zhu, L.J., Wajapeyee, N., and Green, M.R. (2014). Genetic and pharmacological reactivation of the mammalian inactive X chromosome. Proc. Natl. Acad. Sci. USA *111*, 12591–12598.

Bleckwehl, T., Crispatzu, G., Schaaf, K., Respuela, P., Bartusel, M., Benson, L., Clark, S.J., Dorighi, K.M., Barral, A., Laugsch, M., et al. (2021). Enhancer-associated H3K4 methylation safeguards in vitro germline competence. Nat. Commun. *12*, 5771.

Bolt, C.C., and Duboule, D. (2020). The regulatory landscapes of developmental genes. Development *147*, dev171736.

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

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

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.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

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

Chapman, A.G., Cotton, A.M., Kelsey, A.D., and Brown, C.J. (2014). Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. BMC Genet. *15*, 89.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. Cell *168*, 442–459.e20.

Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and Rougeulle, C. (2011). Ftx is a non-coding RNA which affects Xist expression

and chromatin structure within the X-inactivation center region. Hum. Mol. Genet. *20*, 705–718.

Cohen, D.E., Davidow, L.S., Erwin, J.A., Xu, N., Warshawsky, D., and Lee, J.T. (2007). The DXPas34 repeat regulates random and imprinted X inactivation. Dev. Cell *12*, 57–71.

Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Dao, L.T.M., Galindo-Albarrán, A.O., Castro-Mondragon, J.A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., et al. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. Nat. Genet. *49*, 1073–1081.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

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.

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.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature *460*, 128–132.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98.

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719–2729.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Enervald, E., Powell, L.M., Boteva, L., Foti, R., Blanes Ruiz, N., Kibar, G., Piszczek, A., Cavaleri, F., Vingron, M., Cerase, A., and Buonomo, S.B.C. (2021). RIF1 and KAP1 differentially regulate the choice of inactive versus active X chromosomes. EMBO J. Published online November 17, 2021. https://doi.org/10.15252/embj.2020105862.

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.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Finak, G., Frelinger, J., Jiang, W., Newell, E.W., Ramey, J., Davis, M.M., Kalams, S.A., De Rosa, S.C., and Gottardo, R. (2014). OpenCyto: an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis. PLoS Comput. Biol. *10*, e1003806.

Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *48* (D1), D87–D92.

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* (D1), D766–D773.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science *354*, 769–773.

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669.

Furlan, G., Gutierrez Hernandez, N., 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.e8.

Galupa, R., and Heard, E. (2018). X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. Annu. Rev. Genet. *52*, 535–566.

Galupa, R., Nora, E.P., Worsley-Hunt, R., Picard, C., Gard, C., van Bemmel, J.G., Servant, N., Zhan, Y., El Marjou, F., Johanneau, C., et al. (2020). A conserved noncoding locus regulates random monoallelic Xist expression across a topological boundary. Mol. Cell *77*, 352–367.e8.

Gao, Y., Xiong, X., Wong, S., Charles, E.J., Lim, W.A., and Qi, L.S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. Nat. Methods *13*, 1043–1049.

Genolet, O., Monaco, A.A., Dunkel, I., Boettcher, M., and Schulz, E.G. (2021). Identification of X-chromosomal genes that drive sex differences in embryonic stem cells through a hierarchical CRISPR screening approach. Genome Biol. *22*, 110.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome Res. *15*, 1451–1455.

Gil, N., and Ulitsky, I. (2018). Production of spliced long noncoding RNAs specifies regions with increased enhancer activity. Cell Syst. *7*, 537–547.e3.

Girardot, C., Scholtalbers, J., Sauer, S., Su, S.-Y., and Furlong, E.E.M. (2016). Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. BMC Bioinformatics *17*, 419.

Gontan, C., Achame, E.M., Demmers, J., Barakat, T.S., Rentmeester, E., van IJcken, W., Grootegoed, J.A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature *485*, 386–390.

Gontan, C., Mira-Bontenbal, H., Magaraki, A., Dupont, C., Barakat, T.S., Rentmeester, E., Demmers, J., and Gribnau, J. (2018). REX1 is the critical target of RNF12 in imprinted X chromosome inactivation in mice. Nat. Commun. *9*, 4752.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Guo, G., Huang, Y., Humphreys, P., Wang, X., and Smith, A. (2011). A PiggyBac-based recessive screening method to identify pluripotency regulators. PLoS ONE *6*, e18189.

Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., et al. (2017). Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. Cell Stem Cell *21*, 533–546.e6.

Guyochin, A., Maenner, S., Chu, E.T.-J., Hentati, A., Attia, M., Avner, P., and Clerc, P. (2014). Live cell imaging of the nascent inactive X chromosome during the early differentiation process of naive ES cells towards epiblast stem cells. PLoS ONE *9*, e116109.

Hahne, F., LeMeur, N., Brinkman, R.R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J., Strain, E., and Gentleman, R. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics *10*, 106.

Hansen, P., Ali, S., Blau, H., Danis, D., Hecht, J., Kornak, U., Lupiáñez, D.G., Mundlos, S., Steinhaus, R., and Robinson, P.N. (2019). GOPHER: generator of probes for capture Hi-C experiments at high resolution. BMC Genomics *20*, 40.

Heard, E., Kress, C., Mongelard, F., Courtier, B., Rougeulle, C., Ashworth, A., Vourc'h, C., Babinet, C., and Avner, P. (1996). Transgenic mice carrying an Xist-containing YAC. Hum. Mol. Genet. *5*, 441–450.

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.

Heurtier, V., Owens, N., Gonzalez, I., Mueller, F., Proux, C., Mornico, D., Clerc, P., Dubois, A., and Navarro, P. (2019). The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. Nat. Commun. *10*, 1109.

Hoki, Y., Kimura, N., Kanbayashi, M., Amakawa, Y., Ohhata, T., Sasaki, H., and Sado, T. (2009). A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. Development *136*, 139–146.

Hosogane, M., Funayama, R., Shirota, M., and Nakayama, K. (2016). Lack of transcription triggers h3k27me3 accumulation in the gene body. Cell Rep. *16*, 696–706.

Johnston, C.M., Nesterova, T.B., Formstone, E.J., Newall, A.E., Duthie, S.M., Sheardown, S.A., and Brockdorff, N. (1998). Developmentally regulated Xist promoter switch mediates initiation of X inactivation. Cell *94*, 809–817.

Jonkers, I., Barakat, T.S., Achame, E.M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J.A., and Gribnau, J. (2009). RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. Cell *139*, 999–1011.

Kaneko, S., Son, J., Bonasio, R., Shen, S.S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev. *28*, 1983–1988.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat. Commun. *10*, 1930.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. *35*, 561–568.

Klein, J.C., Chen, W., Gasperini, M., and Shendure, J. (2018). Identifying novel enhancer elements with CRISPR-based screens. ACS Chem. Biol. *13*, 326–332.

Knight, P.A., and Ruiz, D. (2013). A fast algorithm for matrix balancing. IMA J. Numer. Anal. *33*, 1029–1047.

Krueger, F., and Andrews, S.R. (2016). SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. F1000Res. *5*, 1479.

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

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. Mol. Cell *74*, 8–18.

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, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554.

Li, W., Köster, J., Xu, H., Chen, C.-H., Xiao, T., Liu, J.S., Brown, M., and Liu, X.S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome Biol. *16*, 281.

Li, Q.V., Dixon, G., Verma, N., Rosen, B.P., Gordillo, M., Luo, R., Xu, C., Wang, Q., Soh, C.-L., Yang, D., et al. (2019). Genome-scale screens identify JNK-JUN signaling as a barrier for pluripotency exit and endoderm differentiation. Nat. Genet. *51*, 999–1010.

Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. *47*, e47.

Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., et al.; FANTOM consortium (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. *16*, 22.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell *167*, 1170–1187.

Loos, F., Loda, A., van Wijk, L., Grootegoed, J.A., and Gribnau, J. (2015). Chromatin-mediated reversible silencing of sense-antisense gene pairs in embryonic stem cells is consolidated upon differentiation. Mol. Cell. Biol. *35*, 2436–2447.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

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.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science *303*, 666–669.

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014). A prominent and conserved role for YY1 in Xist transcriptional activation. Nat. Commun. *5*, 4878.

Marks, H., Chow, J.C., Denissov, S., Françoijs, K.-J., Brockdorff, N., Heard, E., and Stunnenberg, H.G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res. *19*, 1361–1373.

Marques, A.C., Hughes, J., Graham, B., Kowalczyk, M.S., Higgs, D.R., and Ponting, C.P. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. Genome Biol. *14*, R131.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j *17*, 10.

McDonald, L.E., Paterson, C.A., and Kay, G.F. (1998). Bisulfite genomic sequencing-derived methylation profile of the xist gene throughout early mouse development. Genomics *54*, 379–386.

Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L., et al. (2017). Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res *45*, D658–D662.

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.

Monk, M. (1981). A stem-line model for cellular and chromosomal differentiation in early mouse-development. Differentiation *19*, 71–76.

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F., and Gribnau, J. (2008). X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. Cell *132*, 410–421.

Mutzel, V., and Schulz, E.G. (2020). Dosage sensing, threshold responses, and epigenetic memory: a systems biology perspective on random X-chromosome inactivation. BioEssays *42*, e1900163.

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.

Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Dev. *20*, 2787–2792.

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 *321*, 1693–1695.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.

Nesterova, T.B., Senner, C.E., Schneider, J., Alcayna-Stevens, T., Tattermusch, A., Hemberger, M., and Brockdorff, N. (2011). Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. Epigenetics Chromatin *4*, 17.

Newall, A.E., Duthie, S., Formstone, E., Nesterova, T., Alexiou, M., Johnston, C., Caparros, M.L., and Brockdorff, N. (2001). Primary non-random X inactivation associated with disruption of Xist promoter regulation. Hum. Mol. Genet. *10*, 581–589.

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.

Norris, D.P., Patel, D., Kay, G.F., Penny, G.D., Brockdorff, N., Sheardown, S.A., and Rastan, S. (1994). Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell *77*, 41–51.

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

Ohhata, T., Matsumoto, M., Leeb, M., Shibata, S., Sakai, S., Kitagawa, K., Niida, H., Kitagawa, M., and Wutz, A. (2015). Histone H3 lysine 36 trimethylation is established over the Xist promoter by antisense Tsix transcription and contributes to repressing Xist expression. Mol. Cell. Biol. *35*, 3909–3920.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644–649.

Pacini, G., Dunkel, I., Mages, N., Mutzel, V., Timmermann, B., Marsico, A., and Schulz, E.G. (2020). Integrated analysis of Xist upregulation and gene silencing at the onset of random X-chromosome inactivation at high temporal and allelic resolution. BioRxiv.

Pacini, G., Dunkel, I., Mages, N., Mutzel, V., Timmermann, B., Marsico, A., and Schulz, E.G. (2021). Integrated analysis of Xist upregulation and X-chromosome inactivation with single-cell and single-allele resolution. Nat. Commun. *12*, 3638.

Pauklin, S., and Vallier, L. (2015). Activin/Nodal signalling in stem cells. Development *142*, 607–619.

Payer, B., Rosenberg, M., Yamaji, M., Yabuta, Y., Koyanagi-Aoi, M., Hayashi, K., Yamanaka, S., Saitou, M., and Lee, J.T. (2013). Tsix RNA and the germline factor, PRDM14, link X reactivation and stem cell reprogramming. Mol. Cell *52*, 805–818.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

Perez, A.R., Pritykin, Y., Vidigal, J.A., Chhangawala, S., Zamparo, L., Leslie, C.S., and Ventura, A. (2017). GuideScan software for improved single and paired CRISPR guide RNA design. Nat. Biotechnol. *35*, 347–349.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *165*, 1012–1026.

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

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (W1), W160–W165.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

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., and Aiden, E.L. (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.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

Rougeulle, C., Chaumeil, J., Sarma, K., Allis, C.D., Reinberg, D., Avner, P., and Heard, E. (2004). Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. Mol. Cell. Biol. *24*, 5475–5484.

Royce-Tolland, M.E., Andersen, A.A., Koyfman, H.R., Talbot, D.J., Wutz, A., Tonks, I.D., Kay, G.F., and Panning, B. (2010). The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. Nat. Struct. Mol. Biol. *17*, 948–954.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell *14*, 203–216.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shen, L., Shao, N.-Y., Liu, X., Maze, I., Feng, J., and Nestler, E.J. (2013). diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. PLoS ONE *8*, e65598.

Shiura, H., and Abe, K. (2019). Xist/Tsix expression dynamics during mouse peri-implantation development revealed by whole-mount 3D RNA-FISH. Sci. Rep. *9*, 3637.

Söllner, J.F., Leparc, G., Hildebrandt, T., Klein, H., Thomas, L., Stupka, E., and Simon, E. (2017). An RNA-Seq atlas of gene expression in mouse and rat normal tissues. Sci. Data *4*, 170185.

Soma, M., Fujihara, Y., Okabe, M., Ishino, F., and Kobayashi, S. (2014). Ftx is dispensable for imprinted X-chromosome inactivation in preimplantation mouse embryos. Sci. Rep. *4*, 5181.

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.e6.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Sripathy, S., Leko, V., Adrianse, R.L., Loe, T., Foss, E.J., Dalrymple, E., Lao, U., Gatbonton-Schwager, T., Carter, K.T., Payer, B., et al. (2017). Screen for reactivation of MeCP2 on the inactive X chromosome identifies the BMP/TGF-β superfamily as a regulator of XIST expression. Proc. Natl. Acad. Sci. U S A *114*, 1619–1624.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Tan, J.Y., Biasini, A., Young, R.S., and Marques, A.C. (2020). Splicing of enhancer-associated lincRNAs contributes to enhancer activity. Life Sci. Alliance *3*, e202000663.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Tarjan, D.R., Flavahan, W.A., and Bernstein, B.E. (2019). Epigenome editing strategies for the functional annotation of CTCF insulators. Nat. Commun. *10*, 4258.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.

Tian, D., Sun, S., and Lee, J.T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell *143*, 390–403.

Tu, S., Narendra, V., Yamaji, M., Vidal, S.E., Rojas, L.A., Wang, X., Kim, S.Y., Garcia, B.A., Tuschl, T., Stadtfeld, M., and Reinberg, D. (2016). Co-repressor CBFA2T2 regulates pluripotency and germline development. Nature *534*, 387–390.

Tycko, J., Wainberg, M., Marinov, G.K., Ursu, O., Hess, G.T., Ego, B.K., Aradhana, Li, A., Truong, A., Trevino, A.E., et al. (2019). Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. Nat. Commun. *10*, 4063.

Vigneau, S., Augui, S., Navarro, P., Avner, P., and Clerc, P. (2006). An essential role for the DXPas34 tandem repeat and Tsix transcription in the counting process of X chromosome inactivation. Proc. Natl. Acad. Sci. U S A *103*, 7390–7395.

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell *16*, 386–399.

Wang, Q., Zou, Y., Nowotschin, S., Kim, S.Y., Li, Q.V., Soh, C.-L., Su, J., Zhang, C., Shu, W., Xi, Q., et al. (2017). The p53 family coordinates wnt and nodal inputs in mesendodermal differentiation of embryonic stem cells. Cell Stem Cell *20*, 70–86.

Wang, C.-Y., Colognori, D., Sunwoo, H., Wang, D., and Lee, J.T. (2019). PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. Nat. Commun. *10*, 2950.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. Cell *171*, 1573–1588.e28.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. JOSS *4*, 1686.

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. *4*, 1310.

Yang, S.-H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A.D. (2014). Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. Cell Rep. *7*, 1968–1981.

Yang, X., Hu, B., Liao, J., Qiao, Y., Chen, Y., Qian, Y., Feng, S., Yu, F., Dong, J., Hou, Y., et al. (2019). Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development. Cell Res. *29*, 911–926.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.

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

Zhang, L., Huang, H., Zhou, F., Schimmel, J., Pardo, C.G., Zhang, T., Barakat, T.S., Sheppard, K.-A., Mickanin, C., Porter, J.A., et al. (2012). RNF12 controls embryonic stem cell fate and morphogenesis in zebrafish embryos by targeting Smad7 for degradation. Mol. Cell *46*, 650–661.

Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z.-A., et al. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nat. Genet. *50*, 96–105.

Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.-H., Brown, M., Zhang, X., Meyer, C.A., and Liu, X.S. (2019). Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. *47* (D1), D729–D735.

Żylicz, J.J., and Heard, E. (2020). Molecular mechanisms of facultative heterochromatin formation: an X-chromosome perspective. Annu. Rev. Biochem. *89*, 255–282.

Żylicz, J.J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D., et al. (2019). The implication of early chromatin changes in X chromosome inactivation. Cell *176*, 182–197.e23.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
normal mouse IgG	Cell Signaling Technology	Cat#5415S; Lot#9; RRID:AB_10829607
H3K4me1	Active Motif	Cat#39635; Lot#21516012; RRID:AB_2793284
H3K4me3	Active Motif	Cat#61379; Lot#5217007; RRID:AB_2793611
H3K9me3	Active Motif	Cat#39161; Lot#13509002; RRID:AB_2532132
H3K27ac	Active Motif	Cat#39685; Lot#14517014; RRID:AB_2793305
H3K27me3	Active Motif	Cat#39155; Lot#31814017; RRID:AB_2561020
H3K36me3	Active Motif	Cat#61021; Lot#23115007; RRID:AB_2614986
H2AK119Ub	Cell Signaling Technology	Cat#8240S; Lot#6; RRID:AB_10891618
Guinea Pig anti-Rabbit IgG	Antibodies Online	Cat#ABIN101961; RRID:AB_10775589
Rabbit anti-mouse IgG	Thermo Fisher Scientific	Cat#31194; Lot#TK2665056; 31194
Bacterial and virus strains		
One Shot TOP10 Chemically Competent *E. coli*	Thermo Fisher Scientific	Cat#C404010
MegaX DH10B T1R Electrocomp Cells	Thermo Fisher Scientific	Cat#C640003
NEB® Stable Competent *E. coli* (High Efficiency)	New England Biolabs	Cat#C3040H
Chemicals, peptides, and recombinant proteins		
(+)-Abscisic acid	Sigma-Aldrich	Cat#90769-25M
CHIR 99021	Axon Medchem	Cat#1386
PD 0325901	Axon Medchem	Cat#1408
Recombinant mouse Lif	Merck	Cat#ESG1107
Recombinant human Fibronectin	Merck	Cat#FC010
TDE1 Tagment DNA Enzyme	Illumina	Cat#15027865
TD Buffer	Illumina	Cat#15027866
pA-Tn5-3XFLAG	This study	N/A
Critical commercial assays		
PyroMark Gold Q24 reagents	QIAGEN	Cat#970802
P3 Primary Cell 4D NucleofectorTM X Kit L	Lonza	Cat#V4XP-3024
TOPO TA Cloning Kit for Subcloning with One Shot TOP10	Life Technologies	Cat#K4500J10
Qubit dsDNA HS Assay Kit	Life Technologies	Cat#Q32854
MEGA script T7 Transcription kit	Thermo Fisher Scientific	Cat#AM1334
ChIP DNA Clean & Concentrator	Zymo	Cat#D5205
PrimeFlow RNA assay Kit	Thermo Fisher Scientific	Cat#88-18005-210
KAPA HiFi HotStart ReadyMix	Roche	Cat#7958935001
Q5 High Fidelity DNA polymerase	New England Biolabs	Cat#M0491L
HotStarTaq Plus kit	QIAGEN	Cat#203605

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Phusion High-Fidelity DNA Polymerase	New England Biolabs	Cat#M0530L
5′/3′RACE kit, 2nd generation	Roche	Cat#03353621001
TruSeq® RNA Sample Preparation Kit v2	Illumina	Cat#RS-122-2001
Clontech In-Fusion HD Cloning Kit	Takara Bio	Cat#638920
Gibson Assembly® Cloning Kit	New England Biolabs	Cat#E5510S
Deposited data		
Original Code	This study	https://doi.org/10.5281/zenodo.5706318
ATAC-seq	This study	GEO: GSE167350
Capture Hi-C	This study	GEO: GSE167351
CRISPRi Screen	This study	GEO: GSE167352
CUT&Tag	This study	GEO: GSE167353
PolyA-enriched RNA-seq	This study	GEO: GSE167354
STARR-seq	This study	GEO: GSE167355
TT-seq	This study	GEO: GSE167356
RNA-seq in NPCs	Bauer et al., 2021	GEO: GSE157448
ChIP-seq for histone modifications in XY ESCs	Bleckwehl et al., 2021	GEO: GSE155089
ChIP-seq for OTX2 and OCT4	Buecker et al., 2014	GEO: GSE56098
ChIP-seq for pluripotency factors	Chronis et al., 2017	GEO: GSE90893
scRNA-seq in mouse embryos	Deng et al., 2014	GEO: GSE45719
ChIP-seq for REX1	Gontan et al., 2012	GEO: GSE36417
ATAC-seq in human ESCs	Guo et al., 2017	GEO: GSE92280
ChIP-seq for SMAD2/3 in human ESCs	Li et al., 2019	GEO: GSE109524
scRNA-seq in human embryos	Petropoulos et al., 2016	Arrayexpress: E-MTAB-3929
ChIP-seq for histone modifications in human ESCs	Rada-Iglesias et al., 2011	GEO: GSE28874
RNA-seq in adult mouse tissues	Söllner et al., 2017	Arrayexpress: E-MTAB-6081
ChIP-seq for CTCF	Stadler et al., 2011	GEO: GSE30206
ChIP-seq for PRDM14	Tu et al., 2016	GEO: GSE71675
GRO-seq in human ESCs	Wang et al., 2015	GEO: GSE54471
ChIP-seq for SMAD2/3 and TCF3	Wang et al., 2017	GEO: GSE70486
RNA-seq in MEFs	Wang et al., 2019	GEO: GSE116413
ChIP-seq for histone modifications in mouse embryos	Yang et al., 2019	GEO: GSE98101
RNA-seq in mouse embryos	Zhang et al., 2018	GEO: GSE76505
ChIP-seq for histone modifications in TX1072 cell line	Żylicz et al., 2019	GEO: GSE116990
Experimental models: Cell lines		
Mouse: TX1072 (A3)	Schulz et al., 2014	SC02
Mouse: TX1072 XO (B7)	Pacini et al., 2021	SC13
Mouse: TX XXΔXic_{B6} (A1)	Pacini et al., 2021	SC34
Mouse: TX SP106 (D5)	Genolet et al., 2021	SC47
Mouse: TX SP107 (B6)	Genolet et al., 2021	SC37
Mouse: E14-STN	Heurtier et al., 2019	SC40
Mouse: E14-STNΔTsixP (B2)	This study	SC41
Mouse: TX Δ$XertP_{B6}$ (B5)	This study	SC46
Mouse: TX Δ$XertP_{Cast}$ (D5)	This study	SC46
Mouse: TX Δ$XertE_{B6}$ (F6)	This study	SC66

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: TX ΔXertE$_{Cast}$ (B11)	This study	SC66
Mouse: TX ΔXertE−/− (A10)	This study	SC66
Mouse: TX ΔXertP/E$_{Cast}$ (E3)	This study	SC69
Mouse: TX ΔXertP/E$_{Cast}$ (F3)	This study	SC69
Mouse: TX ΔFtx-Xert$_{B6}$ (A10)	This study	SC67
Mouse: TX ΔFtx-Xert$_{Cast}$ (C10)	This study	SC67
Mouse: TX ΔFtx-Xert−/− (F9)	This study	SC67
Human: HEK293T	-	N/A
Oligonucleotides		
RT-qPCR primers	This study	Table S6
3′/5′-RACE primers	This study	Table S6
sgRNA oligos	This study	Table S6
Cloning primers	This study	Table S6
Genotyping primers	This study	Table S6
NGS primers	This study	Table S6
Capture HiC probes	Illumina	Custom made: Table S6
Xert Stellaris probes (Quasar 570)	LGC Biosearch Technologies	Custom made (SMF-1065-5-BS): Table S6
Xist Mature Stellaris probes (Fluorescein)	LGC Biosearch Technologies	Custom made (SMF-1065-5-BS): Table S6
Xist Mature Stellaris probes (Quasar 670)	LGC Biosearch Technologies	Custom made (SMF-1065-5-BS): Table S6
Xist Nascent Stellaris probes (Quasar 670)	LGC Biosearch Technologies	Custom made (SMF-1065-5-BS): Table S6
Huwe1 Nascent Stellaris probes (Quasar 570)	LGC Biosearch Technologies	Custom made (SMF-1065-5-BS): Table S6
Tsix Nascent Stellaris probes (Fluorescein)	LGC Biosearch Technologies	Custom made (SMF-1025-5-BS): Table S6
Xist Flow-FISH probes	Thermo Fisher Scientific	Cat#VB1-14258
Recombinant DNA		
Plasmid: pSpCas9(BB)-2A-Puro (pX459) V2.0	Ran et al., 2013	Addgene Plasmid #42230
Plasmid: lentiGuide-Puro	Sanjana et al., 2014	Addgene Plasmid #52963
Plasmid: SP199 (lentiGuide-Puro with optimized constant region)	Genolet et al., 2021	N/A
Plasmid: pSLQ2817	Gao et al., 2016	Addgene Plasmid #84239
Plasmid: pSLQ2818	Gao et al., 2016	Addgene Plasmid #84241
Plasmid: SP106 (pSLQ2817 with blasticidin resistance)	Genolet et al., 2021	N/A
Plasmid: SP107 (pSLQ2818 with blasticidin resistance)	Genolet et al., 2021	N/A
Plasmid: pBroad3_hyPBase_IRES_tagRFP	Redolfi et al., 2019	N/A
Plasmid: pTXB1-3xFlag-pA-Tn5-FL	Kaya-Okur et al., 2019	Addgene Plasmid #124601
Plasmid: pLP1, pLP2, VSVG	Thermo Fisher Scientific	Invitrogen K497500
Plasmid: pSTARR-seq_human	Arnold et al., 2013	Addgene Plasmid #71509
BACs for STARR-seq	BAC PAC	RP23-106C4, RP23-11P22, RP23-423B1, RP23-273N4, RP23-71K8
Software and algorithms		
Bedtools	Quinlan and Hall 2010	2.29.2
Bowtie	Langmead et al., 2009	1.2.2
bowtie2	Langmead and Salzberg 2012	2.3.5.1
BWA	Li and Durbin 2009	0.7.17
ChIPseeker	Yu et al., 2015	1.22.1
ChromHMM	Ernst and Kellis 2012	1.119

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cistrome DB Toolkit	Mei et al., 2017; Zheng et al., 2019	N/A
Cutadapt	Martin, 2011	1.15
deepTools	Ramírez et al., 2016	2
DESeq2	Love et al., 2014	1.26.0
DiffBind	Ross-Innes et al., 2012	2.6.6
diffReps	Shen et al., 2013	1.55.6
fasterq-dump	https://rnnh.github.io/bioinfo-notebook/docs/fasterq-dump.html	2.9.4
flowCore	Hahne et al., 2009	1.52.1
Galaxy	Afgan et al., 2016; Giardine et al., 2005	N/A
GOPHER	Hansen et al., 2019	N/A
Graphpad PRISM	https://www.graphpad.com	9
GuideScan	Perez et al., 2017	N/A
HiCUP pipeline	Wingett et al., 2015	0.7.04
Htsjdk	https://samtools.github.io/htsjdk/	2.12.0
IGV	https://software.broadinstitute.org/software/igv/home	2.3.94
JASPAR	Fornes et al., 2020	8th release
Je suite	Girardot et al., 2016	1.2.1
Juicer tools	Durand et al., 2016	1.19.02
MACS2	Zhang et al., 2008	2.1.2
MAGeCK	Li et al., 2014; Li et al., 2015	0.5.9.3
MEME suite	Bailey et al., 2009; Grant et al., 2011	5.2.0
openCyto	Finak et al., 2014	1.24.0
Picard	http://broadinstitute.github.io/picard	2.5
Rstudio	https://www.rstudio.com/	3.6.3
Rsubread	Liao et al., 2019	2.0.1
Samtools	Li et al., 2009	1.1
SNPsplit	Krueger and Andrews 2016	0.3.4
STAR	Dobin et al., 2013	2.7.5a
Tidyverse	Wickham et al., 2019	1.3.0
Trim Galore	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	0.6.4
TxDb.Mmusculus.UCSC.mm10.knownGene	https://bioconductor.org/packages/release/data/annotation/html/TxDb.Mmusculus.UCSC.mm10.knownGene.html	3.10.0
UCSC Genome browser	Kent et al., 2002	N/A
UpSetR	Conway et al., 2017	1.4.0
Other		
μMACS Separator	Miltenyi Biotec	cat#130-042-602
Covaris S2 Focused-ultrasonicators	Covaris	cat#500217

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Edda Schulz (Edda.Schulz@molgen.mpg.de)

Materials availability

All cell lines and other materials generated within this study will be made available by the lead author upon request.

Data and code availability

- All NGS data generated within this study has been deposited at GEO and is publicly available at the time of publication (SuperSeries GEO: GSE167358). Individual accession numbers are listed in the key resources table. Microscopy data reported in this paper will be shared by the lead contact upon request. *Xert* transcription variant information has been deposited at GenBank and will be accessible under the following accession numbers: OK239717, OK239718, OK239719, OK239720 and OK239721.
- All original code has been deposited at https://github.com/EddaSchulz/Xert_paper/ and is publicly available as of the date of publication. The DOI through Zenodo is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

The female TX1072 cell line (clone A3) is a F1 hybrid ESC line derived from a cross between the C57BL/6 (B6) and CAST/EiJ (Cast) mouse strains that carries a doxycycline-responsive promoter in front of the *Xist* gene on the B6 chromosome and an rtTA insertion in the Rosa26 locus (Schulz et al., 2014). TXΔXicB6 (clone TXdXic_A1, here referred to as XXΔXic) carries a 773 kb deletion around the *Xist* locus on the B6 allele (chrX:103,182,701-103,955,531, mm10) (Pacini et al., 2021). Only the *Rnf12* gene at the distal end of TAD-E remains intact in that line to not preclude *Xist* upregulation from the wild-type allele (Barakat et al., 2014). The TX1072 XO line (clone B7) has lost the B6 X chromosome and is trisomic for chromosome 16. Female 1.8 XX mESCs carry a homozygous insertion of 7xMS2 repeats in *Xist* exon 7 and are a gift from the Gribnau lab (Schulz et al., 2014). The female TXΔXertP (Clone B5 and D5), TXΔXertE (Clone A10, B11 and F6) and TXΔFtx-Xert (Clone A10, C10 and F9) cell lines were generated by introducing heterozygous and/or homozygous deletions in TX1072 mESCs. The female TXΔXertP/E line (clone E3 and F3) was generated by introducing a heterozygous XertE deletion in the TXΔXertP D5 line. The B6 chromosome is modified in TXΔXertP B5, TXΔXertE F6, TXΔFtx-Xert A10 lines, and the Cast allele carries the deletion in TXΔXertP D5, TXΔXertE B11, TXΔFtx-Xert C10 and both TXΔXertP/E lines (E3 and F3). The cell lines were generated using CRISPR-Cas9 mediated genome editing (see below) and the deleted regions are specified in Table S7. TXΔXertP B5 carries duplications of parts of Chr 10 and TXΔXertP D5 and both TXΔXertP/E lines are trisomic for Chr 8 (Figure S4F).

The male E14-STN$_{\Delta TsixP}$ mESC cell line expresses the CRISPRa Sun-Tag system (Tanenbaum et al., 2014) under a doxycycline-inducible promoter and carries a 4.2 kb deletion around the major *Tsix* promoter (ChrX:103445995-103450163, mm10, Table S7). The cell line was generated by introducing the *Tsix* deletion in E14-STN mESCs (Heurtier et al., 2019) (a kind gift from Navarro lab) and NGS karyotyping (see below) detected duplications of parts of Chr 2.

The TX-SP106 (Clone D5) mESC line stably expresses PYL1-VPR-IRES-Blast and ABI-tagBFP-SpdCas9, constituting a two-component CRISPRa system, where dCas9 and the VPR activating domain are fused to ABI and PYL1 proteins, respectively, which dimerize upon treatment with abscisic acid (ABA). The TX-SP107 (Clone B6) mESC line stably expresses PYL1-KRAB-IRES-Blast and ABI-tagBFP-SpdCas9, constituting a two-component CRISPRi system, where dCas9 and the KRAB repressor domain are fused to ABI and PYL1 proteins, respectively, which dimerize upon ABA treatment. Both cell lines were generated through piggybac transposition (see below). Correct karyotype was confirmed for TX-SP106 (Clone D5) and TX-SP107 (Clone B6) by NGS (Figure S4F). Since repression in TX-SP107 cells transduced with sgRNAs was often observed already without ABA treatment, we could not make use of the inducibility of the system. Instead, TX-SP107 cells were always treated with ABA (100 μM) 24 h before differentiation and effects were compared to NTC sgRNAs.

mESC culture and differentiation

TX1072 mESCs, TX1072 derived mutant cell lines and 1.8 cells were grown on 0.1% gelatin-coated flasks in serum-containing medium supplemented with 2i and LIF (2iL) (DMEM (Sigma), 15% ESC-grade FBS (GIBCO), 0.1 mM β-mercaptoethanol, 1000 U/ml leukemia inhibitory factor (LIF, Millipore), 3 μM Gsk3 inhibitor CT-99021, 1 μM MEK inhibitor PD0325901, Axon). Differentiation was induced by 2iL withdrawal in DMEM supplemented with 10% FBS and 0.1 mM β-mercaptoethanol at a density of $1.6*10^4$ cells/cm^2 on fibronectin-coated (10 μg/ml) tissue culture plates, if not stated otherwise. During the pooled CRISPR screen and CRISPRi experiments, cells were differentiated at a density of $3.6*10^4$ cells/cm^2. For STARR-seq, $1*10^5$ cells/cm^2 cells were seeded for 2iL conditions, while $7*10^4$ cells/cm^2 were used for differentiation. E14-STN$_{\Delta TsixP}$ mESC cells were grown on 0.1% gelatin-coated flasks in serum-containing medium (DMEM (Sigma), 15% ESC-grade FBS (GIBCO), 0.1 mM β-mercaptoethanol), supplemented with 1000 U/ml LIF (SL). Differentiation was induced by LIF withdrawal in DMEM supplemented with 10% FBS and 0.1 mM β-mercaptoethanol at a density of $5.2*10^4$ cells/cm^2 in fibronectin-coated (10 μg/ml) tissue culture plates. Before each experiment and for each cell line generated, the presence of two X chromosomes was verified by RNA FISH for two 2 X-linked genes, either *Huwe1* and *Tsix* or *Huwe1* and *Xist*. The results are provided in Table S7.

METHOD DETAILS

Molecular cloning

Cloning sgRNA plasmids

For genomic deletion of XertP and the Tsix promoter, sgRNAs were designed to target the 5′ and 3′ end of the region of interest and cloned into pSpCas9(BB)-2A-Puro (pX459) V2.0 (Ran et al., 2013). pX459 was a kind gift from Feng Zhang (Addgene plasmid # 42230). SgRNAs (sequences are given in Table S7) were cloned following the Zhang lab protocol (https://media.addgene.org/cms/filer_public/6d/d8/6dd83407-3b07-47db-8adb-4fada30bde8a/zhang-lab-general-cloning-protocol-target-sequencing_1.pdf). In short, two complementary oligos containing the guide sequence and a BbsI recognition site were annealed and ligated with the BbsI (New England Biolabs) digested target plasmid. The ligation mixes were heat shock transformed into NEB Stable competent cells (New England Biolabs) and grown as single colonies on LB-Agar plates (supplemented with Ampicillin 100 ug/ml) overnight at 37°C. Single colonies were expanded and confirmed with Sanger sequencing.

Cloning of sgRNAs in multiguide expression system

For CRISPRa and CRISPRi experiments three different sgRNAs targeting the same RE (Table S7) were cloned into a single sgRNA expression plasmid with Golden Gate cloning as described previously (Genolet et al., 2021). Each sgRNA is controlled by a different Pol III promoter (mU6, hH1 or hU6) and fused to the optimized sgRNA constant region described in Chen et al. (Chen et al., 2013). To this end, the sgRNA constant region of the lentiGuide-puro sgRNA expression plasmid (Sanjana et al., 2014) (Addgene 52963) was exchanged for the optimized sgRNA constant region, thus generating the vector SP199. The vector was digested with BsmBI (New England Biolabs) overnight at 37°C and gel-purified. Two fragments were synthesized as gene blocks (IDT) containing the optimized sgRNA constant region coupled to the mU6 or hH1 promoter sequences. These fragments were then amplified with primers that contained part of the sgRNA sequence and a BsmBI restriction site (primer sequences can be found in Table S7) and purified using the gel and PCR purification kit (Macherey & Nagel). The vector (100 ng) and two fragments were ligated in an equimolar ratio in a Golden Gate reaction with T4 ligase and the BsmBI isoschizomer Esp3I for 20 cycles (5 min 37°C, 5 min 20°C) with a final denaturation step at 65°C for 20 min. Vectors were transformed into NEB Stable competent E.coli. Successful assembly was verified by ApaI digest and Sanger sequencing.

Piggybac transposition

TX-SP106 and TX-SP107 lines were generated by piggybac transposition. To this end the puromycin resistance cassette in the piggybac CRISPRa and CRISPRi expression plasmid (pSLQ2817 and pSLQ2818) was exchanged for a blasticidin resistance, resulting in plasmid SP106 and SP107 respectively. pSLQ2817 and pSLQ2818 were gifts from Stanley Qi (Gao et al., 2016) (Addgene plasmids #84239 and #84241). The respective plasmid was then transfected with Lipofectamin 2000 (Thermo Fisher Scientific) into female TX1072 mESCs in a 5-to-1 transposase-to-target ratio with the hyperactive transposase (pBroad3_hyPBase_IRES_tagRFP) (Redolfi et al., 2019). RFP-positive cells were sorted 24 h after transfection and expanded as single clones under blasticidin selection (5 ng/μl, Roth).

Lentiviral transduction

To package lentiviral vectors into lentiviral particles, $1*10^6$ HEK293T cells were seeded into one well of a 6-well plate and transfected the next day with the lentiviral packaging vectors: 1.2 μg pLP1, 0.6 μg pLP2 and 0.4 μg pVSVG (Thermo Fisher Scientific), together with 2 μg of the desired construct using Lipofectamine 2000 (Thermo Fisher Scientific). HEK293T supernatant containing the viral particles was harvested after 48 h. $0.2*10^6$ mESCs were seeded per well in a 12-well plate with 2iL (for TX-SP106 and TX-SP107) or SL medium (for E14-STN$_{\Delta TsixP}$) and transduced the next day with 1ml of 5:1 concentrated (lenti-X, Clontech) and filtered viral supernatant with 8 ng/μl polybrene (Sigma Aldrich). Puromycin selection (1 ng/μl, Sigma Aldrich) was started two days after transduction and kept for at least 2 passages.

Genome engineering

Generation of deletion mESC lines

To generate deletions, up to $4*10^6$ TX1072 (for ΔXertP, ΔXertE, ΔXertP/E ΔFtx-Xert) or E14-STN (for ΔTsixP) mESCs, cultured in gelatin-coated flasks in SL medium, were nucleofected using the Lonza 4D-Nucleofector with 2 μg of each sgRNA/Cas9 plasmid and either 3 pmol or 30 pmol repair oligo (Table S7) (ΔXertP and ΔTsixP) or with the Alt-R CRISPR-Cas9 system (ID-T) (ΔXertE, ΔXertP/E ΔFtx-Xert) using the P3 Primary Cell 4D-NucleofectorTM kit (Lonza) and CP-106 nucleofection program. Alt-R RNP complexes were generated according to the manufacturer's guidelines. In short, cRNA and tracrRNA were diluted to 200mM with duplex buffer and duplexed in a 1:1 ratio at 95°C for 5 min. To generate RNP complexes 1.2 μl of a 1:1 mix of both duplexes (targeting the 3′ and 5′ of the deletion region) were incubated with 1.7 μl Alt-R® S.p. HiFi Cas9 Nuclease V3 and 1.1 μl PBS for 15 min at RT and used for nucleofection together with 1 μl electroporation enhancer. Afterward the cells were plated on gelatin-coated 10 cm plates with SL medium. Between 18 and 24 h following nucleofection, cells nucleofected with sgRNA/Cas9 plasmid were selected in SL medium supplemented with puromycin (1 ng/ml) for 24 h. Two to 3 days later, the cells were trypsinized and seeded at low densities in gelatin-coated 10 cm plates in SL medium wherein they were cultured until single colonies were visible (up to 12 days).

Genotyping of engineered clones

Semi-confluent 96-well plates with clones were split into 2 low density and 1 high density gelatin-coated 96-well plates with SL medium. Up to 2 days later gDNA was isolated from the high density plate. The cells were washed with PBS and lysed in the 96-wells plate with 50 μl Bradley lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS, 10 mM NaCl, 1 mg/ml Proteinase K (Invitrogen)). The plate was incubated overnight at 55°C in a humidified chamber. To precipitate gDNA, 150 μl ice-cold 75 mM NaCl in 99% EtOH was added per well and the plate was incubated for 30 min at RT. The plate was centrifuged for 15 min at 4000 rpm and 4°C. The pellet was washed once with 70% EtOH and centrifuged for 15 min at 4000 rpm and air-dried at 45°C for 10 min. The gDNA was resuspended in 150 μl TE buffer (10 mM Tris pH 8.0 and 1 mM EDTA, pH 7.5) for 1 h at 37°C. The clones were initially characterized by PCR using either QIAGEN HotStarTaq Plus kit (QIAGEN) or Q5 High Fidelity DNA polymerase (New England Biolabs) following the manufacturer's guidelines, and primer combinations that distinguish between WT and deletions, insertions, or inversions (Table S7). A small number of positive clones were expanded from low density plates. PCR genotyping was repeated on gDNA isolated using the DNeasy Blood and Tissue Kit (QIAGEN). To identify the targeted allele, amplicons containing SNPs were gel-purified and sequenced. Primers and SNP positions are given in Table S7. Few clones were selected and adapted to 2iL medium for at least 4 passages prior to subsequent experiments. For E14-STN$_{\Delta TsixP}$, clone C6 was further sub-cloned and following PCR genotyping (Figures S4A and S4B), subclone E14-STN$_{\Delta TsixP}$ B2 was chosen for future experiments (here referred to as E14-STN$_{\Delta TsixP}$).

NGS karyotyping

Cell lines were karyotyped via double digest genotyping-by-sequencing (ddGBS), a reduced representation genotyping method, as described previously (Genolet et al., 2021). Briefly, the forward and reverse strands of a barcode adaptor and common adaptor were diluted and annealed, after which they were pipetted into each well of a 96-well PCR plate together with 1 μg of each sample and dried overnight (oligo sequences are listed in Table S7). The following day, the samples were digested with 20 μL of a NlaIII and PstI enzyme mix (New England Biolabs) in NEB Cutsmart Buffer at 37°C for 2 h. After the digest, a 30 μL mix with 1.6 μL of T4 DNA ligase (New England Biolabs) was added to each well and placed on a thermocycler (16°C 60 min followed by 80°C 30 min for enzyme inactivation). By doing this, barcode and common adapters with ends complementary to those generated by the two restriction enzymes were ligated to the genomic DNA.

Samples were cleaned with CleanNGS beads (CleanNA) using 90 μL of beads for each well and following manufacturer's instructions. Samples were eluted in 25 μL ddH_2O and DNA was quantified using a dsDNA HS Qubit assay (Thermofisher). Samples were pooled in an equimolar fashion, size-selected (300-450bp) by loading 400 ng of each pooled sample on an agarose gel followed by a cleaning step using the Nucleospin Gel and PCR Cleanup kit (Macherey-Nagel). Samples were PCR amplified using the Phusion High-Fidelity DNA Polymerase (New England Biolabs) and an annealing temperature of 68°C over 15 amplification cycles (OG218/OG219). Resulting amplicons were cleaned with CleanNGS beads in a 1:1.2 ratio (sample:beads) and sequenced with 2x75bp on the Miseq platform or 1x150bp on the NextSeq platform (12 pM loading concentration), yielding 0.2-4*10^6 fragments per sample. The read counts per sample are provided in Table S7 (Cell lines).

RNA extraction, reverse transcription, qPCR

Cells were lysed directly in the plate by adding up to 1 mL of Trizol (Invitrogen). RNA was isolated using the Direct-Zol RNA Miniprep Kit (Zymo Research) following the manufacturer's instructions with on-column DNase digestion. If cDNA was subsequently analyzed by pyrosequencing, DNase digest was performed using Turbo DNA free kit (Ambion). Up to 1 μg RNA was reverse transcribed using Superscript III Reverse Transcriptase (Invitrogen) with random hexamer primers and expression levels were quantified in the QuantStudio 7 Flex Real-Time PCR machine (Thermo Fisher Scientific) using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific). Primers used are listed in Table S7 (RT-qPCR Primers).

3′- and 5′ RACE

To identify transcript isoforms as well as exact stop and start sites of *Xert*, 3′- and 5′RACE were performed. First, RNA was isolated from 2 day-differentiated TXΔXic$_{B6}$ cells using the Direct-Zol RNA Miniprep Kit (Zymo Research). To remove any remaining gDNA, RNA samples were rigorously treated with DNase for 20 min at 37°C using the TURBO DNA-freeTM Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Poly-adenylated RNAs were purified from 5 μg total RNA with the Dynabeads®Oligo (dT)25 Kit (Thermo Fisher Scientific) following the manufacturer's instructions.

For 3′-RACE cDNA was synthesized as described before, instead using 50 ng purified polyadenylated RNA and the oligo(dT)-anchor primer from the 5′/3′RACE kit, by following the manufacturer's guidelines. To remove DNA:RNA duplexes, 25 ng of cDNA was digested with 0.5 μl 1:40 diluted RNaseH (New England Biolabs) for 20 min at 37°C. To specifically amplify the 3′ end of the transcript for 3′RACE, RNaseH-treated cDNA was PCR-amplified by Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions using the gene-specific forward primer PK1 and the anchor primer PK9. PCR products were analyzed on agarose gels and purified using QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions. To increase specificity the isolated PCR product was PCR amplified with the nested gene-specific forward primer PK4 and the anchor primer PK9. Additionally, a PCR targeting putative exon 2 and exon 6, was performed using the gene-specific PK4 and PK17 primers.

For 5′-RACE, 50 ng of purified poly-adenylated RNA was reverse transcribed using the gene-specific reverse primer PK35. To remove DNA:RNA duplexes, cDNA was digested with RNaseH as described before. RNaseH-treated cDNA was purified using QIA-

quick PCR purification Kit according to the manufacturer's instructions. Subsequently, 5′ pA tailing of the product was performed with the 5′/3′RACE kit, 2nd generation (Roche) according to the manufacturer's guidelines. The anchor sequence was added to the 5′ end of the transcript by PCR amplification using the gene-specific reverse primer PK13 and the oligo(dT)-anchor primer and Phusion High-Fidelity DNA Polymerase (New England Biolabs). PCR products were analyzed on agarose gels and purified using the QIAquick PCR purification Kit according to the manufacturer's instructions. To increase specificity, the cleaned PCR product was amplified in a nested PCR using the nested gene-specific reverse primer PK34 and the anchor primer PK9. All primer sequences are given in Table S7 (3′5'RACE Primers).

TOPO TA cloning and Sanger sequencing

Blunt-end PCR amplicons underwent A-tailing using HotStarTaq DNA Polymerase (New England Biolabs). The PCR products from the nested 3′/5′RACE were cleaned-up using QIAquick PCR purification Kit and then mixed with 5 μl 10x DNA polymerase reaction buffer, 10 μl of 1 mM dATP, 0.2 μl of HotStarTaq DNA polymerase, filled up to 50 μl with nuclease free water and incubated for 20 min at 72°C. The A-tailed PCR products were separated on agarose gels, and bands were individually isolated from agarose gels (Figure S3B) using the QIAquick Gel Extraction Kit (QIAGEN) and cloned into TOPO vector pCR2.1 using the Topo TA cloning kit (Invitrogen) according to the manufacturer's instructions. For the ligation, 1 μl 10x T4 DNA ligase buffer, 1.5 μl pCR2.1 vector, 10 ng A-tailed gel-isolated PCR product and 1 μl T4 DNA ligase (New England Biolabs) were mixed in a total reaction volume of 10 μl and incubated for 15 min at room temperature. One Shot® TOP10 chemically competent *E. coli* (Thermo Fisher Scientific) were heat shock transformed and plated on LB-agar plates supplemented with Ampicillin 100 μg/ml, 100 μl 20 mg/ml X-gal (Sigma Aldrich). Plates were incubated overnight at 37°C, the following day colonies were assessed by blue/white screening. Five white colonies were picked per plate, inoculated in 5 mL LB medium (supplemented with Ampicillin 100 μg/ml) and shaken overnight at 37°C. Plasmids were purified from the bacterial cultures using Plasmid Mini Prep (PeqLab) according to the manufacturer's instructions and analyzed by Sanger sequencing via LGC Genomics GmbH, PK11 was used as sequencing primer. The obtained sequence data between the gene-specific forward primer (for 3′RACE) or gene-specific reverse primer (for 5′-RACE) and the anchor primer was extracted and aligned to the mouse genome (mm10) via basic local alignment search tool (BLAST) and visualized using the UCSC genome browser (Figure S3C). After analyzing splice isoforms in pA-RNA-seq data, we detected two additional unidentified *Xert* isoforms, which were confirmed by conventional PCR with primers PK4+PK17 on RNaseH treated cDNA as described above. Sanger sequencing of isolated bands A1-4 indeed revealed these 2 additional *Xert* isoforms (Figure S3C).

Pyrosequencing

To quantify relative allelic expression for individual genes, an amplicon containing a SNP at the Cast allele was amplified by PCR from cDNA using Hot Start Taq (QIAGEN) for 38 cycles. The PCR product was sequenced using the Pyromark Q24 system (QIAGEN). Assay details are given in Table S7 (PyroSeq Assay).

RNA FISH

RNA FISH was performed using Stellaris FISH probes (Biosearch Technologies). Probe details can be found in Table S7 (FISH Probes). Cells were dissociated using Accutase (Invitrogen) and adsorbed onto coverslips (#1.5, 1 mm) coated with Poly-L-Lysine (Sigma) for 5 min. Cells were fixed with 3% paraformaldehyde in PBS for 10 min at RT (18–24°C) and permeabilized for 5 min on ice in PBS containing 0.5% Triton X-100 and 2 mM Ribonucleoside Vanadyl complex (New England Biolabs). Coverslips were preserved in 70% EtOH at −20°C. Prior to FISH, coverslips were incubated for 5 minutes in wash buffer containing 2x SSC and 10% formamide, followed by hybridization for 6 hours to overnight at 37°C with 250 nM of each FISH probe in 50 μL Stellaris RNA FISH Hybridization Buffer (Biosearch Technologies) containing 10% formamide. Coverslips were washed twice for 30 min at 37°C with 2x SSC/10% formamide with 0.2 mg/ml Dapi being added to the second wash. Prior to mounting with Vectashield, mounting medium coverslips were washed with 2xSSC at RT for 5 minutes. Images were acquired using a widefield Z1 Observer microscope (Zeiss) using a 100x objective.

Flow-FISH

For Flow-FISH the PrimeFlow RNA assay (Thermo Fisher Scientific) was used according to the manufacturer's recommendations. Specifically, the assay was performed in conical 96-well plates with $5*10^6$ cells per well with Xist-specific probes, labeled with Alexa-Fluor647 (VB1-14258) (Thermo Fisher Scientific). Samples were resuspended in PrimeFlow RNA Storage Buffer before flow cytometry. Cells were analyzed or sorted using the BD FACSAria™ II or BD FACSAria Fusion flow cytometers. The sideward and forward scatter areas were used for live cell gating, whereas the height and width of the sideward and forward scatter were used for doublet discrimination. At least 20,000 cells were measured per replicate.

Poly-adenylated RNA-seq and *de novo* transcriptome assembly

Total RNA (100ng) from 2 days differentiated TX1072 mESCs was subjected to strand-specific RNA-seq library preparation with the TruSeq® RNA Sample Preparation Kit v2 (Illumina), which included polyadenylated RNA enrichment using oligo-dT magnetic beads, by following the manufacturers guidelines. The libraries were subjected to Illumina NGS PE50 on the HiSeq 4000 platform to obtain approximately 65 mio fragments.

ATAC-seq

Assay for Transposase-Accessible Chromatin by Sequencing (ATAC-seq) was used to profile open chromatin, as described previously with adaptations (Corces et al., 2017). $XX_{\Delta Xic}$ and XO cells were profiled at day 0, 2 and 4 of differentiation in two biological replicates. Cells were dissociated with trypsin and $6*10^4$ cells were lysed in 50 μL cold RSB buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20) supplemented with 0.1% Igepal CA-630 and 0.01% Digitonin. The lysis buffer was washed out using 1 mL of cold RSB buffer. Nuclei were then pelleted by centrifugation (500 x g, 10 min, 4°C) and the supernatant aspirated. Subsequently, they were resuspended in 50 μL Transposase Mix (1x TD buffer (Illumina), 100 nM Nextera Tn5 Transposase (Illumina), 33 μL PBS, 0.01% Digitonin, 0.1% H_2O) and incubated at 37°C for 30 minutes at 1000 rpm. The reaction was stopped by adding 2.5 μL 10% SDS and purified using the DNA Clean & Concentrator Kit (Zymo). 20 μL of the transposed DNA was then amplified using the NEBNext High-Fidelity 2x PCR Master mix with i5 and i7 Nextera barcoded primers for 12 cycles (see Table S7 for primer sequences). The PCR product was size-selected using NGS Clean beads (CleanNA), by adding them at first at a 70%-ratio and transferring the supernatant. Afterward, the beads were added once more at a 180%-ratio and the PCR product was eluted from the beads in 20 μL H_2O. The success of the transposition was verified with the BioAnalyzer High Sensitivity DNA system (Agilent Technologies). Sequencing libraries were pooled in equimolar ratios and sequenced paired-end 75 bp on the HiSeq 4000 platform yielding approximately $2.5*10^7$ fragments per sample (Table S1).

STARR-seq

STARR-seq is a massively parallel reporter assay, where a large number of genomic fragments are tested for enhancer activity in an episomal context (Arnold et al., 2013).

STARR-seq library cloning

A STARR-seq library covering the *Xic* was cloned as described previously (Arnold et al., 2013) with modifications. The Bacterial Artificial Chromosome (BAC) clones RP23-106C4, RP23-11P22, RP23-423B1, RP23-273N4, RP23-71K8 were purchased as bacterial stabs from the BAC PAC Resource Center of the Children's Hospital Oakland Research Institute. *E.coli* BAC clones were grown in 200 mL LB medium (10 g/l NaCl, 10 g/l Bacto Tryptone, 5 g/l Yeast extract, 1 mM NaOH) supplemented with 12.5 μg/ml Chloramphenicol (Sigma) in a shaking incubator at 30°C for 20 h. The BAC DNA was isolated using the NucleoBond BAC 100 kit (Machery-Nagel). BAC DNA was pooled (2.5 μg each) and split into four tubes, which were filled with TE buffer to a total volume of 100 μl. The DNA was sheared by sonication (Bioruptor Plus, low intensity, 3 cycles with 32 s 'on'/ 28 s 'off'), size-selected on a 1% agarose gel and extracted with the QIAquick Gel Extraction Kit (QIAGEN). The eluates were pooled and purified using the QIAquick PCR Purification Kit (QIAGEN). The purified fragments were end-repaired, dA-tailed and ligated to adapter_STARR1/adapter_STARR2 to be compatible with Illumina sequencing according to the NEBnext DNA library prep master mix set for Illumina (New England Biolabs, oligonucleotide sequences shown in Table S7). The ligated fragments were purified using Agencourt AMPure XP beads (Beckman Coulter) and eluted in 25 μL elution buffer. Four PCR reactions were then carried out with 1 μL of the purified DNA using the KAPA HotStart HiFi Ready Mix (KAPA Biosystems) for 10 cycles inserting a 15 nt homology sequence for the subsequent cloning step (IF_fwd/IF_rev). The PCR products were size-selected on a 1% agarose gel and purified using the MinElute PCR Purification Kit (QIAGEN). The pSTARR-seq_human vector (kindly provided by Alexander Stark) was digested with AgeI-HF and SalI-HF for 3.5 h at 37°C, size-selected on a 1% agarose gel and purified using the MinElute PCR Purification Kit (QIAGEN). The DNA library was then cloned into the vector using 4 In-Fusion cloning reactions according to the manufacturer's instructions (Clontech In-Fusion HD). The reactions were pooled, ethanol-precipitated and transformed into MegaX DH10BTM T1R Electrocompetent Cells (Invitrogen) according to the manufacturer's instructions in a total of eight separate reactions. Plasmid DNA was amplified overnight and isolated using the NucleoBond Xtra Midi Plus Kit (Macherey-Nagel). The plasmid library was sequenced paired-end 50 bp on the HiSeq 2500 platform yielding approximately $1.1*10^7$ fragments (Table S1). Due to a partial deletion in one of the BAC clones used, a ~55 kb region within the *Linx* gene was not covered by the STARR-seq library.

Transfection and sequencing

$5.0*10^6$ 1.8 XX and 1.8 XO cells were transfected with 2.5 μg of the STARR-seq library using Lipofectamine LTX (Thermo) according to manufacturer's instructions in three biological replicates. $3*10^6$ cells were cultured under 2iL conditions and $2*10^6$ under differentiation conditions for 48 h. RNA was isolated using the Direct-Zol RNA Miniprep Kit (Zymo Research). The mRNA fraction was recovered from the total RNA using Dynabeads Oligo-dT25 (Invitrogen) with 1 mg beads per 50 μg of total RNA. The RNA was reverse transcribed into cDNA using Superscript III (Invitrogen) with a gene specific primer (STARR_GSP). Three reactions were performed for each sample. The reactions were then treated with RNaseI (Thermo) for 60 min at 37°C and cleaned using the MinElute PCR Purification Kit (QIAGEN). Subsequently, a junction PCR was performed using an intron-spanning primer pair (Junction_fwd and Junction_rev) and 8 μL cDNA with the KAPA HotStart HiFi Ready Mix (KAPA Biosystems) for 15 cycles. Three reactions each were pooled using the QIAquick PCR Purification Kit (QIAGEN). Sequencing adapters were added in a second PCR using three reactions with 10 μL of the purified junction PCR product with the KAPA HotStart HiFi Ready Mix (KAPA Biosystems) for 12 cycles. Lastly, the samples were isolated via agarose gel extraction and the QIAquick Gel Extraction Kit (QIAGEN) and purified once more using the QIAquick PCR Purification and MinElute PCR Purification Kits (QIAGEN). Libraries were pooled in equimolar ratios and sequenced paired-end 50 bp on the HiSeq 2500 platform yielding approximately $1.0*10^7$ fragments per sample (Table S1). All primer sequences are provided in Table S7.

CRISPRi screen

sgRNA library design

BAM files of all conditions and replicates of the ATAC-seq and STARR-seq data (see above) were merged. Peaks were called using *MACS2* (v2.1.2) with options [callpeak -f BAMPE -g mm -q 0.1] (Quinlan and Hall, 2010; Zhang et al., 2008). The resulting narrowPeak files were then filtered for peaks in the *Xic* (chrX:103198658-104058961). In addition, a list of candidate enhancer elements across different mouse tissues in the region, identified by the FANTOM5 consortium based on Cap Analysis of Gene Expression (CAGE), was used (Lizio et al., 2015). Afterward, candidate regions from ATAC-seq, STARR-seq and FANTOM5 data were combined using *bedtools* (v2.29.2) with option [merge]. Regions longer than 2000 bp were split manually according to visual inspection of the ATAC-seq data and adjacent REs with a total combined length below 2000 bp (including the distance between them) were merged, resulting in a list of 138 candidate REs. Since the efficiency of targeting REs with CRISPRi is known to be highly variable (Klann et al., 2017), the candidate REs were saturated with sgRNA sequences generated from the *GuideScan* webtool (Perez et al., 2017) with a specificity score of > 0.2 (Tycko et al., 2019). 300 randomly chosen non-targeting guides from the mouse CRISPR Brie lentiviral pooled library (Doench et al., 2016) were included as negative controls, resulting in 7358 guides in total. The sgRNA library composition is provided in Table S1.

sgRNA library cloning

The sgRNA library was cloned into the lentiGuide-puro sgRNA expression plasmid (Addgene 52963, (Sanjana et al., 2014)). The vector was digested with BsmBI (New England Biolabs) at 55°C for 1 h and gel-purified. sgRNA sequences were synthesized by Genscript flanked with OligoL (TGGAAAGGACGAAACACCG) and OligoR (GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC) sequences. For the amplification of the library, 7 PCR reactions with primers OG113/OG114 and approx. 5ng of the synthesized oligo pool were carried out using the Phusion Hot Start Flex DNA Polymerase (New England Biolabs), with a total of 14 cycles and an annealing temperature of 63°C in the first 3 cycles and 72°C in the subsequent 11 cycles. The amplicons were subsequently gel-purified.

Amplified sgRNAs were ligated into the vector through Gibson assembly (New England Biolabs). Three 20 μl Gibson reactions were carried out using 7 ng of the gel-purified insert and 100 ng of the vector. The reactions were pooled, EtOH-precipitated to remove excess salts which might impair bacterial transformation and resuspended in 12.5 μl H_2O. 9 μl of the eluted DNA were transformed into 20 μl of electrocompetent cells (MegaX DH10B, Thermo Fisher Scientific) according to the manufacturer's protocol using the ECM 399 electroporator (BTX). After a short incubation period (1h, 37°C 250 rpm) in 1 mL SOC medium, 9 mL of LB medium with Ampicillin (0.1 mg/ml, Sigma) were added to the mixture and dilutions were plated in Agar plates (1:100, 1:1000 and 1:10000) to determine the coverage of the sgRNA library (526-fold). 500 mL of LB media with Ampicillin were inoculated with the rest of the mixture and incubated overnight for subsequent plasmid purification using the NucleoBond Xtra Maxi Plus kit (Macherey & Nagel) following the manufacturer's instructions. To assess library composition by deep-sequencing, a PCR reaction was carried out to add illumina adaptors by using the KAPA HiFi HotStart ReadyMix (Roche), with an annealing temperature of 60°C and 12 cycles (OG125/OG126). The PCR amplicon was gel-purified by using the Nucleospin Gel and PCR Clean-up kit (Macherey & Nagel) following the manufacturer's instructions. The library was sequenced paired-end 75 bp on the HiSeq 4000 Platform yielding approximately 7.5 million fragments. Read alignment statistics found in Table S2. A $\log_2$-distribution width of 1.7 for the plasmid library showed that sufficient coverage was attained during library cloning (Figures S1E and S1J). Only one sgRNA was missing from the library (gRNA_6494). All primer sequences are given in Table S7.

Lentiviral packaging

HEK293T cells were cultured in DMEM supplemented with 10% FBS and passaged every 2 to 3 days. For lentiviral packaging, 20 10cm plates with HEK293T cells were transfected at 90% confluency, each with 6.3 μg pPL1, 3.1 μg pLP2 and 2.1 μg VSVG vectors (Thermo Fisher Scientific) together with 10.5 μg of the cloned sgRNA library. Plasmids and 60 μl Lipofectamine 2000 reagent (Thermo Fisher Scientific) were each diluted in 1 mL of OptiMEM, incubated separately for 5 min and then together for 20 min. The mix was added dropwise to the HEK293T cells and the medium was changed 6 h after transfection. After 48 h the medium was collected and viral supernatant was concentrated 10-fold using the lenti-X^{TM} Concentrator (Takara Bio) following the manufacturer's instructions and subsequently stored at −80°C.

To estimate the viral titer, serial 10-fold dilutions were prepared from the viral stock and used to transduce mESCs in a 6-well plate (Mock plus 10^{-2} to 10^{-6}) together with 8 ng/μl polybrene (Merck) in duplicates. Selection with puromycin (1 ng/μl, Sigma) was started two days after transduction and colonies were counted in each well after 8 days. The estimated titer was $0.68*10^5$ transducing units (TU) per ml.

Lentiviral transduction

The TX-SP107 mESC line, carrying an ABA-inducible dCas9-KRAB system, was grown for at least two passages in SL medium prior to transduction. Transduction was carried out in SL medium, as X chromosome loss was sometimes observed upon transduction in 2iL medium. A total of $6*10^6$ cells were transduced with viral supernatant of the sgRNA library (MOI = 0.3). Additionally, $2*10^5$ cells each were transduced with either an empty pLenti vector or an sgRNA targeting the *Xist* TSS (Table S7, sgRNA targets). Both controls were taken along for the rest of the experiment and confirmed CRISPRi efficiency (Figures S1F and S1G). Puromycin selection (1 ng/μl, Sigma) was started two days after transduction and kept for the rest of the experiment. At the next passage, the cells were transferred into 2iL medium. After two more passages, cells were differentiated by 2iL-withdrawal. Recruitment of dCas9-KRAB to target sites was induced using ABA (100 μM) one day before differentiation and kept throughout the rest of the protocol. $1*10^6$ cells were kept

in 2iL-containing medium and used as an undifferentiated control. Cells were harvested for Flow-FISH after 2 days of differentiation.

Flow-FISH and cell sorting

$2*10^8$ cells were stained by Flow-FISH with an Xist-specific probe as described above. $2*10^7$ cells were snap-frozen after the two fixation steps to be used as the unsorted fraction. Four different populations were sorted, where 15% cells with the lowest signal were termed Xist-negative, while 45% cells with the strongest signal were sorted into 3 positive populations (0%–15% = High, 15%–30% = Medium, 30%–45% = Low). Around $1.1\text{-}1.5*10^7$ cells were recovered per fraction. After sorting, the cell pellets were snap-frozen and stored at −80°C for further analysis.

Preparation of sequencing libraries and sequencing

Sequencing libraries were prepared from all sorted cell populations and the unsorted cells for each of the two independent screen replicates. DNA from frozen cell pellets was isolated through phenol/chloroform extraction since it yields significantly more DNA than DNA isolation kits based on silica columns. Cell pellets were thawed and resuspended in 250 μl of lysis buffer (1% SDS (Thermo Fisher Scientific), 0.2 M NaCl and 5 mM DTT (Roth) in TE Buffer) and incubated overnight at 65°C. The next day 200 μg of RNase A (Thermo Fisher Scientific) were added and the samples were incubated at 37°C for 1 h. 100 μg of Proteinase K (Sigma) were subsequently added, followed by a 1 h incubation at 50°C. Phenol/chloroform/isoamyl alcohol (Roth) was added to each sample in a 1:1 ratio, the mixture was vortexed for 1 min and subsequently centrifuged at 16,000 x g for 10 min at RT. The aqueous phase was transferred to a new tube, 1 mL 100% EtOH, 90 μl 5 M NaCl and 1 μl Pellet Paint (Merck) was added to each sample, mixed, and incubated at −80°C for 1 h. DNA was pelleted by centrifugation for 16,000 x g for 15 min at 4°C, pellets were washed twice with 70% EtOH, air-dried and resuspended in 50 μl water.

The genomically integrated sgRNA cassette was amplified in two successive PCR reactions as described previously (Shalem et al., 2014) with minor modifications. To ensure sufficient library coverage (> 300x), 14.5 μg of each sample were amplified using the ReadyMix Kapa polymerase (Roche) with a total of 20 cycles and an annealing temperature of 55°C (OG115/OG116). Between 0.1-2 μg genomic DNA was amplified per 50 μl PCR reaction. In particular, in samples stained with Flow-FISH PCR amplification was inhibited at higher DNA concentrations such that up to 145 PCR reactions had to be performed per sample. Successful amplification was verified on a 1% agarose gel and the reactions were pooled. The PCR product was isolated and concentrated using the Zymo DNA Clean and Concentrator Kit. A second nested PCR was performed to attach sequencing adaptors and sample barcodes using 2.5-50 μl of the product from the first PCR as template, with a total of 11 cycles and an annealing temperature of 55°C (OG125/OG170-OG180). Resulting amplicons were loaded on a 1% agarose gel and purified using the Nucleospin Gel and PCR clean-up kit (Macherey-Nagel). Libraries were sequenced paired-end 75bp on the NextSeq 500 platform yielding approximately $4*10^6$ fragments per sample (Table S2). All primer sequences are given in Table S7 (NGS Oligos).

CUT&Tag of histone modifications

Cleavage Under Targets and Tagmentation (CUT&Tag) makes use of Tn5 transposition at protein A (pA) bound antibody recognition sites and was performed as described previously with minor modifications (Kaya-Okur et al., 2019).

Purification of 3xFLAG-pA-Tn5

The 3xFlag-pA-Tn5 protein was purified from *E. coli* containing pTXB1-3xFlag-pA-Tn5-FL (Addgene, #124601), a kind gift from Steven Henikoff (Kaya-Okur et al., 2019). From an overnight streak LB agar plate, a single colony was selected for a liquid starter culture in LB medium supplemented with Carbenicilin (100 μg/ml) and incubated in a shaker at 37°C for 4 hours. Afterward the starter culture was added to 400 mL LB medium supplemented with Carbenicilin (100 μg/ml) and incubated until it reached an OD_{600} of 0.6 (roughly three hours) and was directly cooled on ice. After 30 min on ice, 100 μL of 1 M IPTG was added to the culture and incubated in a cooled shaker overnight at 18°C at 150 rpm. The following morning, bacteria were centrifuged in a JA-12 rotor at 10,000 rpm for 30 min at 4°C. Bacterial pellets were snap frozen in liquid nitrogen and stored at −80°C. The pellets were thawed on ice and resuspended in 40 mL HEGX buffer (20 mM HEPES-KOH pH 7.2, 0.8 M NaCl, 1 mM EDTA pH8.0, 10% Glycerol, 0.2% Triton X-100). Following this the cell suspension was divided over two 50 mL tubes and lysed with a Branson tip sonicator on ice by using a 1064 (10-150 ml) tip with the following settings: 20 s on, 20 s off, 50% duty cycle for 9 min total. Afterward the lysate was centrifuged in a JA-12 rotor at 10,000 rpm for 30 min at 4°C and the supernatant was collected. Two 20 mL columns (Biorad were each packed with 2.5 mL Chitin resin slurry (New England BioLabs) and washed once with 20ml HEGX buffer. To each column 20 mL supernatant was added, locked on both openings and incubated overnight with rotation. The following morning the columns were washed four times with 20ml pre-cooled HEGX buffer supplemented with protease inhibitor cocktail tablets (Roche). Afterward the chitin resin holding the 3xFLAG-pA-Tn5 was collected in a total of 10 mL elution buffer (20 mM HEPES-KOH pH7.2, 0.8 M NaCl, 1 mM EDTA, 10% Glycerol, 0.2% Triton X-100, 100 mM DTT), transferred to a 15ml falcon tube and extracted for 48 h on a rotator (15 rpm) at 4°C. Afterward the resin was allowed to settle to the bottom over 40 min on ice followed by centrifugation for two min at 300 rpm at 4°C to collect all chitin resin. The supernatant holding the 3xFLAG-pA-Tn5 was dialysed in 800 mL cold dialysis buffer (100 mM HEPES-KOH pH7.2, 0.1 M NaCl, 0.2 mM EDTA, 2 mM DTT, 0.2% Triton X-100, 20% Glycerol) using a Slide-A-lyzer 30K dialysis cassette for 24 h at 4°C with magnetic stirring at 300 rpm, with the buffer being refreshed after the initial 12 h. The dialysed protein extract (~5.5ml) was concentrated 6-fold using an Amicon Ultra 4 30K 15ml falcon filtration system with successive rounds of centrifugation in a swing bucket centrifuge at 3000 xg at 4°C for 15 min. The protein concentration was measured with the detergent compatible Bradford assay kit (Pierce), adjusted to 832 ng/μl with dialysis buffer and diluted

1:1 volume with 100% glycerol (= 5.5 μM). The 3xFLAG-pA-Tn5 fusion protein was confirmed on a GelCode Blue (Thermo Fisher Scientific) stained SDS-PAGE. Aliquots of 100 μL of 5.5 μM 3xFLAG-pA-Tn5 fusion protein were loaded with mosaic end adapters. For this, 10 μL ME-A with 10 μL ME-reverse and 10 μL ME-B with 10 μL ME-reverse 200 μM oligos (Table S7) (dissolved in 10 mM Tris-HCl pH8.0, 50 mM NaCl, 1 mM EDTA) were annealed in separate reactions on a thermocycler for 5 minutes at 95°C with a ramp down to 21°C over 30 min and mixed together afterward. Aliquots of 100 μL containing 5.5 μM 3xFLAG-pA-Tn5 fusion protein were mixed with 16 μl of adaptor mix and incubated for 1 h at RT with rotation and stored at −20°C.

Cell preparations and CUT&Tag

Cells were washed with PBS and dissociated with accutase. For each CUT&Tag reaction 1*10^5 cells were collected and washed once with Wash buffer (20 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 10 mM Sodium butyrate, 1 mM PMSF). 10 μL Concanavalin A (Bangs Laboratories) beads were equilibrated with 100 μL Binding buffer (20 mM HEPES-KOH, pH 7.5, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$) and afterward concentrated in 10 μL binding buffer. The cells were bound to the Concanavalin A beads by incubating for 10 min at RT with rotation. Following this, the beads were separated on a magnet and resuspended in 100 μL chilled Antibody buffer (Wash buffer with 0.05% Digitonin and 2 mM EDTA). Subsequently 1 μl of primary antibody (antibodies can be found in Table S7) was added and incubated on a rotator for 3 hours at 4°C. After magnetic separation the beads were resuspended in 100 μL chilled Dig-wash buffer (Wash buffer with 0.05% Digitonin) containing 1 μl of matching secondary antibody and were incubated for 1 h at 4°C with rotation. The beads were washed three times with ice cold Dig-wash buffer and resuspended in chilled Dig-300 buffer (20 mM HEPES-KOH, pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, 0.01% Digitonin, 10 mM Sodium butyrate, 1 mM PMSF) with 1:250 diluted 3xFLAG-pA-Tn5 preloaded with Mosaic-end adapters. After incubation for 1 h at 4°C with rotation, the beads were washed four times with chilled Dig-300 buffer and resuspended in 50 μL Tagmentation buffer (Dig-300 buffer 10 Mm $MgCl_2$). Tagmentation was performed for 1 h at 37°C and subsequently stopped by adding 2.25 μL 0.5 M EDTA, 2.75 μL 10% SDS and 0.5 μL 20 mg/mL Proteinase K and vortexing for 5 s. DNA fragments were solubilized overnight at 55°C followed by 30 min at 70°C to inactivate residual Proteinase K. DNA fragments were purified with the ChIP DNA Clean & Concentrator kit (Zymo Research) and eluted with 25 μL elution buffer according to the manufacturer's guidelines.

Library preparation and sequencing

NGS libraries were generated by amplifying the CUT&Tag DNA fragments with i5 and i7 barcoded HPLC-grade primers (Buenrostro et al., 2015) (Table S7, NGS Oligos) with NEBNext® HiFi 2x PCR Master Mix (New England BioLabs) on a thermocycler with the following program: 72°C for 5 min, 98°C for 30 s, 98°C for 10 s, 63°C for 10 s (14-15 Cycles for step 3-4) and 72°C for 1 min. Post PCR cleanup was performed with Ampure XP beads (Beckman Coulter). For this 0.95x volume of Ampure XP beads were mixed with the NGS libraries and incubated at RT for 10 min. After magnetic separation, the beads were washed three times on the magnet with 80% ethanol and the libraries were eluted with Tris-HCl, pH 8.0. The quality of the purified NGS libraries was assessed with the BioAnalyzer High Sensitivity DNA system (Agilent Technologies). Sequencing libraries were pooled in equimolar ratios, cleaned again with 1.2x volume of Ampure XP beads and eluted in 20 μL Tris-HCl, pH 8.0. The sequencing library pool quality was assessed with the BioAnalyzer High Sensitivity DNA system (Agilent Technologies) and subjected to Illumina PE75 next generation sequencing on the NextSeq500 platform totalling approximately 5 mio fragments per library.

TT-seq and RNA-seq

S4U metabolic labeling of nascent RNA

Transient transcriptome sequencing (TT-seq), which is based on enrichment of S4U-labeled nascent RNA after a short, 5 min labeling pulse (Schwalb et al., 2016), was performed to profile genome-wide nascent RNA levels. To this end the cells were cultured in 10 cm plates with 2iL or differentiated for 2 or 4 days by 2iL withdrawal. Cells were metabolically labeled with culture medium containing 750 μM 4-Thiouridine (S4U) (Sigma Aldrich) for 5 min at 37°C and 5% CO_2. Directly afterward, the cells were washed with PBS and lysed on ice with TRIzol (Ambion).

RNA isolation

The lysates were pre-cleared by centrifugation and per 1*10^7 cells supplemented with 2.4 ng equimolar mix of *in vitro* transcribed S4U-labeled and unlabeled ERCC spike-ins, as previously described (Schwalb et al., 2016). Total RNA was extracted from TRIzol with chloroform. In short, 200 μL chloroform was mixed per ml lysate and phase separated by centrifugation in phase-lock tubes. RNA was precipitated from the aqueous phase with isopropanol supplemented with 0.1 mM DTT and centrifugation for 10 min at 16.000x g and 4°C. The RNA pellet was washed once with 75% ethanol and resuspended in nuclease free water. Residual genomic DNA was removed in solution with DNaseI (QIAGEN) following the manufacturer's guidelines. The total RNA was purified for a second round with the Direct-zol RNA Miniprep Plus kit (Zymo Research) by following the manufacturer's guidelines but in addition including 100 mM DTT in all wash buffers to prevent oxidation of S4U-labeled RNA.

RNA fragmentation and biotinylation

For each TT-seq reaction 300 μg of purified RNA was divided over 2 Covaris MicroTubes and fragmented on the Covaris S2 platform for 10 s, 1% duty cycle, intensity 2, 1 cycle, 200 cycles/burst. Corresponding samples were pooled and 3 μg taken aside for quality control (detailed below). The remaining S4U-treated RNA (~260 μl) was biotinylated by adding 240 μL nuclease free water, 100 μL 10x Biotinylation buffer (100 mM Tris-HCl pH7.4, 10 mM EDTA), 200 μL DMSO and 200 μL Biotin-HPDP (1 μg/μl in DMSO) and incubated for 1.5 h on a thermoblock at 37°C with 750 rpm agitation. To remove unreacted Biotin-

HPDP, the biotinylated RNA was extracted with phenol:chloroform (PCI) 5:1, pH 4.5) (Ambion). For this an equal volume of PCI was mixed with the biotinylated RNA and phase separated by centrifuging in phase-lock tubes at 12,000 x *g* for 5 min. The aqueous phase was collected, mixed with an equal volume of isopropanol and 1:10 volume of 5 M NaCl, and centrifuged at 16,000 x *g* for 15 min 4°C to precipitate the biotinylated RNA. The RNA pellet was washed twice with 500 μL 75% ethanol and dissolved in 50 μL RNase-free water.

Nascent RNA enrichment

Biotinylated RNA was denatured at 65°C for 10 min, directly followed by cooling on ice. To capture the biotinylated (nascent) RNA, 100 μL μMACS Streptavidin Microbeads (Miltenyi) were added and incubated on a heat block at 24°C for 15 min with 750 rpm agitation. Bead mixture was loaded on pre-equilibrated MACS μColumn while attached to a μMACS separator. The initial flow-through was collected and loaded one more time on the MACS μColumn. The columns were washed 3 times with 900 μL heated (65°C) wash buffer (100 mM Tris-HCl pH 7.4, 10 mM EDTA, 100 mM NaCl, 0.1% Tween-20) and 3x with RT wash buffer. To elute the enriched nascent RNA, the columns were loaded twice with 100 μL 100 mM DTT and collected. The nascent RNA was purified by adding 3 volumes of TRIzol and processed with the Direct-zol RNA Miniprep kit (Zymo Research) with addition of 1/100 volume of 100 mM DTT to each supplied wash buffer.

To confirm the quality of the total input and nascent RNA, the samples were analyzed with the Agilent RNA 600 pico kit on the Bioanalyzer platform (Agilent). Furthermore, enrichment of labeled (nascent) RNA over unlabeled RNA was assessed by RT-qPCR. For this, 1 μl of eluted nascent RNA and 500 ng fragmented total RNA were reverse transcribed (as described above) and enrichment of labeled ERCC spike-ins over unlabelled spike-ins (included during the first RNA isolation steps) was assessed by qRT-PCR (Figure S3A) with primers specific for each spike-in sequence (Table S7).

Library preparation and sequencing

Total RNA and nascent RNA samples were subjected to strand-specific RNA-seq library preparation with the KAPA RNA HyperPrep kit with RiboErase (Illumina), which included 1st and 2nd strand synthesis and ribosomal RNA depletion, by following the manufacturer's guidelines. The libraries were sequenced PE75 (for $XX_{\Delta Xic}$) on the Illumina HiSeq 4000 platform or PE100 (for XO) on the NovaSeq 6000 platform with approximately 25 mio fragments for total RNA and 100 mio fragments for nascent RNA.

Capture Hi-C

Nuclei preparation

$XX_{\Delta Xic}$ and XO cultured in 2iL (day 0) or after 2 days differentiation (2iL withdrawal) were disassociated with 0.1% (w/v) accutase for 7 min at 37 °C. Cells were counted and $2*10^6$ cells were transferred in a 50 mL falcon tube through a 40 μm cell strainer and complemented with 10% FBS in PBS. 37% formaldehyde (Sigma-Aldrich) was added to a final concentration of 2% to fix the cells for 10 min at RT. Crosslinking was quenched by adding glycine to a final concentration of 125 mM. Fixed cells were washed twice with cold PBS and lysed using fresh lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA with protease inhibitor cocktail tablets (Roche) to isolate nuclei. After 10 min incubation in ice, cell lysis was assessed microscopically. Nuclei were centrifuged for 5 min at 480 x *g*, washed once with PBS and snap frozen in liquid N_2.

Chromosome conformation capture library preparation and sequencing

3C libraries were prepared from fixed nuclei as described previously (Despang et al., 2019). In summary, nuclei pellets were thawed on ice and subjected to DpnII digestion, ligation and decrosslinking. Re-ligated products were sheared using a Covaris sonicator (duty cycle: 10%, intensity: 5, cycles per burst: 200, time: 2 cycles of 60 sec each, set mode: frequency sweeping, temperature: 4–7 °C). Adapters were then added to the sheared DNA and amplified according to Agilent instructions for Illumina sequencing. The library was hybridized to the custom-designed SureSelect library and indexed for sequencing (100 bp, paired end) following manufacturer's instructions. The custom-designed SureSelect library was described to capture informative GATC fragments within chrX:102238718-105214261 (mm10) using *GOPHER*, as described previously (Hansen et al., 2019). Except for XO mESCs at day 0, where the library preparation failed, capture Hi-C experiments were performed in duplicate, which displayed strong replicate correlation (Figure S7B).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless stated otherwise, data processing and visualization was conducted using Rstudio with *tidyverse* (v1.9) (Wickham et al., 2019).

Analysis of Flow-FISH data

FCS files were gated using RStudio with the *flowCore* (v1.52.1) and *openCyto* packages (v1.24.0) (Finak et al., 2014; Hahne et al., 2009). All cells that showed a fluorescent intensity above the 99th-percentile of the 2iL-control were marked as Xist-positive. These cells were then used to calculate the mean fluorescent intensity in the Xist-positive fraction after background correction by subtracting the mean intensity of the 2iL-control. Both, the percentage of Xist-positive cells and the mean fluorescent intensity of the Xist-positive fraction were plotted as a ratio to the non-targeting control.

Statistical analysis for qPCR, pyrosequencing and RNA-FISH

Statistical analysis of RT-qPCR, pyrosequencing, RNA-FISH and Flow-FISH experiments was conducted using *Graphpad PRISM* (v9). RT-qPCR data was normalized to the geometric mean of Arpo and Rrm2. Significance was defined as a two-sided p value < 0.05 using an unpaired two-tailed t test (one-sample t test for Flow-FISH experiments).

Statistical analysis of CRISPRi/CRISPRa experiments

To ensure robust normalization, 2-4 multiguide sgNT plasmids were employed in each experiment. Measurements of each targeting guide were then normalized to the geometric mean of those different multiguide NT plasmids.

NGS karyotyping analysis

Data processing and statistical analysis was performed on the public Galaxy server usegalaxy.eu. Fastq files were uploaded and demultiplex using the "Je-demultiplex" tool (v1.2.1) (Girardot et al., 2016) . Reads of the karyotyping analysis were mapped to the mouse genome (mm10) using BWA (Li and Durbin, 2009). The reads for each chromosome were then counted using deeptols (v2.0) (Ramírez et al., 2016) on the useGalaxy platform (Afgan et al., 2016; Giardine et al., 2005) with option [multiBamSummary]. The counts per chromosome were divided by the sum of all counts per sample. The relative counts were then normalized to the wild-type and visualized as a heatmap (Figure S4F).

NGS data processing

Genome preparation

For all alignment of data generated within the TX1072 cell lines, all SNPs in the mouse genome (mm10) were N-masked (Barros de Andrade E Sousa et al., 2019; Pacini et al., 2021) using *SNPsplit* (v0.3.2) (Krueger and Andrews, 2016) for high-confidence SNPs between present in the TX1072 cell line as described previously (Barros de Andrade E Sousa et al., 2019; Pacini et al., 2020). For all other data (STARR-seq, published data) data was aligned to the reference genome.

Read filtering

Following alignment, sequencing data was filtered for mapped and, for paired-end data, properly paired reads using *samtools* (v1.10) (Li et al., 2009) with options [view -f 2 -q 20] for ATAC-seq, CUT&Tag and paired ChIP-seq, [view -F 4 -q 20] for unpaired ChIP-seq, [view -f 2 -q 10] for STARR-seq and [view -q 7 -f 3] for RNA-seq and TT-seq data. Afterward, the BAM files were sorted using *samtools* with [sort]. Blacklisted regions for mm10 (ENCODE Project Consortium, 2012) were then removed using *bedtools* (Quinlan and Hall, 2010) (v2.29.2) with options [intersect -v] . Unless stated otherwise, duplicates were marked and removed using *Picard* (v2.18.25) with options [MarkDuplicates VALIDATION_STRINGENCY = LENIENT REMOVE_DUPLICATES = TRUE] (http://broadinstitute.github.io/picard). For analysis, BAM files of individual replicates were merged using samtools with [merge] .

Generation of coverage tracks

BIGWIG coverage tracks for sequencing data were created using *deeptools2* (v3.4.1) (Ramírez et al., 2016) merged replicates, if available. For TT-seq and polyadenylated RNA-seq, BAM files were split depending on the strand prior to track generation. Normalization was performed using the total number of autosomal reads. RNA-seq and unpaired ChIP-seq data was processed with the options [bamCoverage -bs 10 –normalizeUsing CPM -ignore chrX chrY]. For paired and unspliced data types (ATAC-seq, CUT&Tag, ChIP-seq & TT-seq) reads were additionally extended using[-e]. The tracks were visualized using the UCSC genome browser (Kent et al., 2002).

Peak calling

Unless stated otherwise, peaks were called using *MACS2* (Zhang et al., 2008) (v2.1.2) with standard options [callpeak -f BAMPE/BAM -g mm -q 0.05] on individual replicates. For ChIP-seq, input samples were included for normalization using [-c]. Only peaks detected in all replicates were retained by merging replicates using *bedtools* (v2.29.2) (Quinlan and Hall, 2010) with [intersect] .

Xert isoforms detection and analysis

3′ and 5′RACE had identified multiple isoforms of *Xert* with a length of 398-767bp (Table S6). To verify the *Xert* isoforms as detected by 3′/5′-RACE, we used polyadenylated RNA-seq data from TX1072 cells differentiated for 2 days. Reads were aligned using *STAR* (v2.7.5a) *(Dobin et al., 2013)* with options [–outSAMattributes NH HI NM MD]. For *de novo* transcript assembly, the sorted bam file was analyzed in *Cufflinks* (v2.2.1) with the parameter [–library-type fr-firststrand]. Mapping statistics can be found in Table S5. We generated a Sashimi plot in IGV (v2.3.94) and analyzed the transcripts predicted by the *de novo* transcript assembly (Figure S3C).

Our 5′-RACE data suggested that *Xert* contains 2 TSSs. However, since we detected band D6 (Figure S3C) only once from all 5′-RACE colonies analyzed, and this far upstream TSS was neither detected in RNA-seq nor in TT-seq data (Figure 3A), we considered the TSS starting at ChrX:103637012 (5′-RACE bands D1-5 in Figure S3C) the only TSS driving *Xert* transcription in our mESCs.

To detect any open reading frames (ORFs) that could potentially code for protein, DNA sequences from all processed *Xert* transcript isoforms were loaded into Geneious (v10.2.6) and assessed with the *Find ORF* option for a minimum size of 150 bp in 5′-3′ direction. Six ORFs with a length between 153bp and 234bp were identified (Table S6).

ATAC-seq data processing

Read filtering was performed as described in the section "NGS data processing." Sequences were trimmed using *Trim Galore* (v0.6.4) with options [–paired –nextera] (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Afterward, the trimmed

FASTQ files were aligned using *bowtie2* (v2.3.5.1) (Langmead and Salzberg, 2012) with the options [–local –very-sensitive-local -X 2000]. Mitochondrial reads were removed using a custom Python script. Mapping statistics can be found in Table S1.

STARR-seq data processing

The data was processed as described previously (Arnold et al., 2013). FASTQ files were mapped using *bowtie* (Langmead et al., 2009) (v1.2.2) with options [-S -t -v 3 -m 1 -l 250 -X 2000]. As the amount of reads per sample was very low after deduplication (~99% duplicates) and the samples were similar between conditions, all samples were then merged using *samtools* (v1.10) (Li et al., 2009) with options [merge] for further analysis. For visualization, BIGWIG tracks normalized to the cloned library were created using *deepTools2* (v3.4.1) (Ramírez et al., 2016) with [bamCompare -e -bs 10 –operation ratio –normalizeUsing CPM]. Mapping statistics and quality control metrics are shown in Table S1.

CRISPRi screen analysis

Data processing and statistical analysis was performed using the *MAGeCK* CRISPR screen analysis tools (Li et al., 2014, 2015) (v0.5.9.3). Alignment and read counting was performed with options [count –norm-method control] for all samples together. At least $3.25*10^6$ mapped reads were obtained per sample. Correlation between the two replicates was computed as a Pearson correlation coefficient on the normalized counts (Figure S1H). The NTC distribution width was similar across samples, suggesting that sufficient library coverage was maintained during all steps (Figure S1I).

Statistical analysis was performed on the RE levels with options [mle –norm-method control –max-sgrnapergene-permutation 350] and on the sgRNA level [test –norm-method control] by comparing each sorted fraction to the unsorted control. In tracks showing the screen results, all REs that were significantly enriched or depleted (FDR < 0.05) in either the high or the negative fraction are colored. In order to rank REs based on their effect on *Xist* expression, we averaged their beta score, a measure of the effect size estimated by the *MAGeCK mle* tool, across populations for each RE that exhibited an FDR < 0.05 in at least one bin, with inverting the sign in the negative bin. To ensure robustness of the ranking and to exclude an analysis bias associated with the variable number of sgRNAs per RE, we implemented an alternative strategy focusing on those REs that were targeted by > 50 guides. First, normalized counts were averaged across replicates for each sgRNA. For 1000 bootstrap samples, each containing 50 sgRNAs randomly selected with replacement, the fold change between sorted and unsorted fractions was calculated and averaged. Ranking REs according to the mean of those fold-change distributions led to nearly identical results as the beta-score based approach. An empirical p value was calculated from the resulting distribution and Benjamini-Hochberg corrected. Alignment statistics, normalized counts, gene hit summary files and RE ranking is provided in Table S2.

CUT&Tag analysis

Data processing

Read filtering was performed as described in the section "NGS data processing." Read sequences were trimmed using *Trim Galore* (0.6.4) with options [–paired –nextera] (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Afterward, the trimmed FASTQ files were aligned according to (Kaya-Okur et al., 2019) with modifications using *bowtie2* (v2.3.5.1) with the options [–local –very-sensitive-local –no-mixed –no-discordant –phred33 -I 10 -X 2000] (Langmead and Salzberg, 2012). As the percentage of duplicate reads was very low, duplicated reads were kept for analysis. Mapping statistics and quality control metrics can be found in Table S4.

Genomic peak annotation

For the analysis shown in Figure S2E, CUT&Tag peaks identified in undifferentiated $XX_{\Delta Xic}$ mESCs using *MACS2* (see above) were assigned to gene features according to the annotation package *TxDb.Mmusculus.UCSC.mm10.knownGene* (v3.10.0) (https://bioconductor.org/packages/release/data/annotation/html/TxDb.Mmusculus.UCSC.mm10.knownGene.html) using *ChIPseeker* (v1.22.1) (Yu et al., 2015).

Comparison of CUT&Tag with native ChIP-seq data

The H3K4me1, H3K4me3, H3K27ac, H3K27me3 and H2AK119ub histone marks profiled via CUT&Tag in undifferentiated $XX_{\Delta Xic}$ cells were compared to native ChIP-seq data profiling the same marks in the parental TX1072 cell line (Żylicz et al., 2019). FASTQ files were retrieved from the GEO Accession Viewer (GSE116990) using *fasterq-dump* (v2.9.4) (https://rnnh.github.io/bioinfo-notebook/docs/fasterq-dump.html). In order to keep the data comparable, processing was done analogous to the CUT&Tag data, as described above. Subsequently, reads were quantified in 1 kb bins using *deepTools2* (v3.4.1) (Ramírez et al., 2016) with the options [multiBamSummary -bs 1000] on merged replicates. Afterward, a PCA analysis was conducted using the base R package *stats* (v3.6.3) with [prcomp(center = TRUE, scale = TRUE)].

CUT&Tag correlation analysis

BAM files, excluding mitochondrial reads, were counted in 1 kb bins using *deepTools2* (v3.4.1) (Ramírez et al., 2016) with options [multiBamSummary bins -bs 1000 -bl chrM.bed]. The Pearson correlation coefficient between different histone marks, conditions or replicates was then computed using *deepTools2* (v3.4.1) with options [plotCorrelation -c pearson]. The resulting values were hierarchically clustered and plotted as a heatmap.

ChromHMM analysis

Chromatin segmentation was performed using *ChromHMM (Ernst and Kellis, 2012)* (v1.19) on ATAC-seq and CUT&Tag data (H3Kme1, H3Kme3, H3K27ac, H3K27me3) for the $XX_{\Delta Xic}$ cell line at all three time points. The model was learned for 10 to 15 emission

states. After visual inspection of the resulting BED files, 12 emission states were chosen for further analysis. Chromatin states were then assigned as 'no RE', 'poised RE', 'weak RE' or 'strong RE' states depending on the enrichment of the different chromatin marks (Figure S2H). Only REs are shown in Figure 2D.

Quantitative analysis of ATAC-seq and CUT&Tag data

ATAC-seq, H3K4me3, H3K27ac and H3K4me1 reads were quantified from the replicate BAM files at the candidate REs using *Rsubread* (v2.0.1) (Liao et al., 2019) with the options [featureCounts(isPairedEnd = TRUE)]. Counts per Million (CPM) were then calculated for all samples. To compare between different conditions, we computed a z-score for the individual REs (Figure 2E). In Figures 2E and 2G, comparisons in which all of the conditions failed to reach 5 raw reads in both replicates were colored in dark gray.

Differential peaks were identified for ATAC-seq, H3K4me3, H3K4me1 and H3K27ac with the *DiffBind* bioconductor package (v2.6.6) (Ross-Innes et al., 2012). The analysis was performed either for all peaks that were identified with *MACS2* (see above) in all replicates of at least one condition or for all candidate REs from the CRISPR screen (Figure 2G). All peaks on the X chromosome outside of the deleted region in the $XX_{\Delta Xic}$ cell line (chrX: 103182701-103955531) were excluded from the analysis to remove any potential bias due to the different number of X chromosomes between the cell lines. Differential peaks were then analyzed between time points and cell lines with the options [dba.analyze(method = DBA_ALL_METHODS)]. The results were exported using [dba.report(method = DBA_ALL_METHODS, th = 0.05, bUsepVal = FALSE)].

In order to find differentially enriched regions for the broader H3K9me3 mark, we used *diffReps (Shen et al., 2013)* (v1.55.6). Here the number of reads mapping to 5 kb intervals was compared between time points or cell lines using a sliding window approach with options [–window 5000 –step 1000]. To identify consensus peaks on the X chromosome present in both conditions, *diffreps* was used to compare each condition with a modified version of itself, where all X chromosome reads had been removed. Peaks identified in both conditions, which did not overlap with a differential peak were defined as consensus peaks.

TT-seq analysis

Total and nascent RNA data was processed according to (Schwalb et al., 2016). Reads were aligned using *STAR* (v2.7.5a) with options [–outSAMattributes NH HI NM MD] (Dobin et al., 2013). Mapping statistics and quality control can be found in Table S5.

Gene quantification

To quantify gene expression, the GENCODE M25 annotation (Frankish et al., 2019) was supplemented with the *Xert* coordinates (Table S6). *Rsubread* (v2.0.1) (Liao et al., 2019) was used with the options [featureCounts(isPairedEnd = TRUE, GTF.featureType = "gene," strandSpecific = 2, allowMultiOverlap = TRUE)] to count reads over the entire gene for TT-seq or with [featureCounts(isPairedEnd = TRUE, GTF.featureType = "exon," strandSpecific = 2, allowMultiOverlap = FALSE)] to only count exonic reads for RNA-seq. In order to detect statistical differences in the expression of lncRNA expression within the *Xic*, we performed differential expression analysis between the $XX_{\Delta Xic}$ and XO cell lines using *DESeq2 (Love et al., 2014)* (v1.26.0). Comparisons with an adjusted p value < 0.05 were marked as significant (Figure 3B). TPM values and the results of the differential expression analysis from our TT-seq and RNA-seq data can be found in Table S5.

Analysis of published data

FASTQ files or processed WIG tracks of published NGS data were retrieved from the GEO Accession Viewer using *fasterq-dump* (v2.9.4) (https://rnnh.github.io/bioinfo-notebook/docs/fasterq-dump.html) or from the EMBL-EBI Array Express (additional information on the datasets is detailed in Table S5).

Processing of ChIP-seq data

Reads were trimmed for adaptor fragments using *Trim Galore* (v0.6.4) with options [–illumina] (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were aligned using *bowtie2* (v2.3.5.1) (Langmead and Salzberg, 2012) with the options [–local –very-sensitive-local –no-mixed –no-discordant –phred33 -I 10 -X 2000] for paired-end and [–very-sensitive] for single-end data (Langmead and Salzberg, 2012)

Visualization of CTCF binding sites within the Xic

CTCF binding sites (CBS's) in mESC's were detected using published CTCF ChIP-seq data (Stadler et al., 2011) using the FIMO program within the MEME suite web tool (v5.2.0) (Grant et al., 2011). To this end, we generated a FASTA file containing the sequences within the CTCF peaks using bedtools (v2.29.2) (Quinlan and Hall, 2010) with options [getfasta]. Then we scanned the peaks for the occurence of the CTCF transcription factor binding motif, which was retrieved from the JASPAR database (Fornes et al., 2020) (8^{th} release). Lastly, the direction of the CBS's were annotated by the strandedness of the binding motif.

cHi-C analysis

Mapping, filtering and deduplication of short reads were performed with the HiCUP pipeline (Wingett et al., 2015) (v0.7.4) [no size selection, Nofill: 1, Format: Sanger]. The pipeline employed bowtie2 (v2.3.5.1) (Langmead and Salzberg, 2012) for mapping short reads to the N-masked reference genome mm10. For merging replicates, the corresponding bam files with valid and unique read pairs from the *HiCUP* pipeline were combined. Juicer tools (v1.19.02) (Durand et al., 2016) was used to generate binned contact maps from valid and unique read pairs with MAPQ $\geq$ 30 and to normalize contact maps by Knight and Ruiz (KR) matrix balancing (Knight and Ruiz, 2013; Rao et al., 2014). For the generation of contact maps, only read pairs mapping to the genomic region

chrX:103,190,001-103,950,000 were considered. In this part of the enriched region, both investigated cell lines (XO and $XX_{\Delta Xic}$) have only one allele. Afterward, KR-normalized maps were exported at 10 kb bin size.

To compare contact maps between different cell lines or time points, subtraction maps were generated from KR-normalized maps, which were normalized in a pairwise manner before subtraction. To account for differences between two maps in their distance-dependent signal decay, the maps were scaled jointly across their sub-diagonals. Therefore, the values of each sub-diagonal of one map were divided by the sum of this sub-diagonal and multiplied by the average of these sums from both maps. Afterward, the maps were scaled by 10^6/total sum.

cHi-C maps were visualized as heatmaps with linear scale and with values above the 0.92-quantile being truncated to improve visualization. In the heatmaps of subtraction maps, values were truncated at thresholds, which are indicated in the color bar.