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

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.
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. Schulz1,8,*

1Otto Warburg Laboratories, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
2Development and Disease Group, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
3Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh EH4 2XU, Edinburgh, UK
4Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, 13353 Berlin, Germany
5Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
6Computational Health Center, Helmholtz Center München, 85764 Neuherberg, Germany
7These authors contributed equally
8Lead 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 Ulltrey, 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; Zylicz 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...
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 IncRNA genes, Jpx and Ftx, which activate Xist expression (Chureau et al., 2011; Furlan et al., 2018; Tio, 2010). In addition, TAD-E contains the protein-coding Rnf12/Rilm 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 (MAGEC8K test, two-sided p < 0.05 and REs (MAGEC8K test, 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 MAGEC8K (l 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 l 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-D 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.
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 MA-GeCK 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\_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\_Xic (TX\_XicR8) 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)–(I), where the cell lines shown in (A) are differentiated by 2iL withdrawal to induce Xist upregulation in the XX\_XicR8 mESCs.

(C) Overlay of DNA accessibility and histone modifications in female XX\_XicR8 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\_XicR8 and XO cell lines. Asterisks mark significant differences (p < 0.05) according to DiffBind analysis.

(H) Same as (F), but showing the XX\_XicR8 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_{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, 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 expected peak patterns and were in good agreement with native chromatin immunoprecipitation sequencing (ChIP-seq) in the parental line (Figures S2D–S2F; Table S4 (Zylcz et al., 2019). Although Xist was strongly upregulated at day 2 in XX_{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 H3K27me3 (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; Rougelle 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_{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_{Xic} than XO cells at day 0 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; Paulkin 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_{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_{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 IncRNAs (Table S6). The main TSS was located within RE93 and exhibited a chromatin state typical for enhancers, characterized by chromatin accessibility, bidirectional transcription, high H3K4me1/H3K4me3 ratio (Figures 3C and 3D), reminiscent of a previously described IncRNA class that are transcribed from enhancer elements (Gil and Ullitsky, 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 IncRNA 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,

---

**Molecular Cell**

Article

OPEN ACCESS
and Xist was initially upregulated from two alleles in a subset of cells, as previously reported for cells differentiated by 2iL withdrawal (Guyonch et al., 2014; Pacini et al., 2020; Sousa et al., 2019). We found that Xert was more frequently detected on Xert-positive than on Xert-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 Fbx 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\textsubscript{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 3t; 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 2t, 2j, and 3l), further supporting a role of these factors in Xert regulation. Re-analysis 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; Petroupolos 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 S3J), suggesting a functional conservation of enhancers in the same genomic region. Taken together, we have identified an IncRNA 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 (Heurter 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\textsubscript{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\textsubscript{Xic} and XO cells was assessed using DEseq2 (FDR < 0.05; asterisks).

(C) TT-seq (both strands) 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\textsubscript{AXic} 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 indicates an unannotated transcript and is enlarged in (C). In (B), differential gene expression between XX and XO 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 indicates an unannotated transcript and is enlarged in (C). In (B), differential gene expression between XX\textsubscript{Xic} and XO cells was assessed using DEseq2 (FDR < 0.05; asterisks).

(II) qRT-PCR quantification of Xist, Xert, Jpx, and Fbx in XX and XO cells during differentiation. The line connects the means, 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, epithelium; VE, visceral endoderm; Ect, ectoderm; End, endoderm; Mes, mesoderm; PS, primitive streak.

In (A) 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.
chromosome at day 2 in both clones (Figures 4F and 4G; Figure S4G). At day 3, this effect was lost in ΔXertPmut 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.
**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-associating 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 lies 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 ΔXertP/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 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 IncRNAs (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–6L). 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 XXXic–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 XXXic and XO cells during differentiation (Figure 7B, right; Figure S7B, right). Contact frequencies of Xist with Xert and Ftx were increased in XXXic 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
A 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 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.
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 (Makhlof 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 previously 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; Makhlof 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

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 STAR-seq. T.S. and L.R.L. and V.S. generated CRISPRa/multiguide plasmids. I.D. and V.S. generated CRISPRa/mESC lines. V.M. generated the TX1072 XXΔXicB6 line with help from R.A.F.G. R.A.F.G. analyzed mESC mutant lines with help from P.K. and T.S. RACE was performed by P.K. and T.S. and genomic features. Genome Res. 29, 1087–1099.


### Key Resources Table

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal mouse IgG</td>
<td>Cell Signaling Technology</td>
<td>Cat#5415S; Lot#9; RRID:AB_10829607</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>Active Motif</td>
<td>Cat#39635; Lot#21516012; RRID:AB_2793284</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Active Motif</td>
<td>Cat#61379; Lot#5217007; RRID:AB_2793611</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>Active Motif</td>
<td>Cat#39161; Lot#13509002; RRID:AB_2532132</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>Active Motif</td>
<td>Cat#39685; Lot#14517014; RRID:AB_2793305</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Active Motif</td>
<td>Cat#39155; Lot#31814017; RRID:AB_2561020</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>Active Motif</td>
<td>Cat#61021; Lot#23115007; RRID:AB_2614986</td>
</tr>
<tr>
<td>H2AK119Ub</td>
<td>Cell Signaling Technology</td>
<td>Cat#8240S; Lot#6; RRID:AB_10891618</td>
</tr>
<tr>
<td>Guinea Pig anti-Rabbit IgG</td>
<td>Antibodies Online</td>
<td>Cat#ABIN101961; RRID:AB_10775589</td>
</tr>
<tr>
<td>Rabbit anti-mouse IgG</td>
<td>Thermo Fisher Scientific</td>
<td>Cat#31194; Lot#TK2665056; 31194</td>
</tr>
<tr>
<td><strong>Bacterial and virus strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Shot TOP10 Chemically Competent E. coli</td>
<td>Thermo Fisher Scientific</td>
<td>Cat#C404010</td>
</tr>
<tr>
<td>MegaX DH10B T1R Electrocomp Cells</td>
<td>Thermo Fisher Scientific</td>
<td>Cat#C640003</td>
</tr>
<tr>
<td>NEB® Stable Competent E. coli (High Efficiency)</td>
<td>New England Biolabs</td>
<td>Cat#C3040H</td>
</tr>
<tr>
<td><strong>Chemicals, peptides, and recombinant proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)-Abscisic acid</td>
<td>Sigma-Aldrich</td>
<td>Cat#90769-25M</td>
</tr>
<tr>
<td>CHIR 99021</td>
<td>Axon Medchem</td>
<td>Cat#1386</td>
</tr>
<tr>
<td>PD 0325901</td>
<td>Axon Medchem</td>
<td>Cat#1408</td>
</tr>
<tr>
<td>Recombinant mouse Lif</td>
<td>Merck</td>
<td>Cat#ESG1107</td>
</tr>
<tr>
<td>Recombinant human Fibronectin</td>
<td>Merck</td>
<td>Cat#FC010</td>
</tr>
<tr>
<td>TDE1 Tagment DNA Enzyme</td>
<td>Illumina</td>
<td>Cat#15027865</td>
</tr>
<tr>
<td>TD Buffer</td>
<td>Illumina</td>
<td>Cat#15027866</td>
</tr>
<tr>
<td>pA-Tn5-3XFLAG</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Critical commercial assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PyroMark Gold Q24 reagents</td>
<td>QIAGEN</td>
<td>Cat#970802</td>
</tr>
<tr>
<td>P3 Primary Cell 4D NucleofectorTM X Kit L</td>
<td>Lonza</td>
<td>Cat#V4XP-3024</td>
</tr>
<tr>
<td>TOPO TA Cloning Kit for Subcloning with One Shot TOP10</td>
<td>Life Technologies</td>
<td>Cat#K4500J10</td>
</tr>
<tr>
<td>Qubit dsDNA HS Assay Kit</td>
<td>Life Technologies</td>
<td>Cat#Q32854</td>
</tr>
<tr>
<td>MEGA script T7 Transcription kit</td>
<td>Thermo Fisher Scientific</td>
<td>Cat#AM1334</td>
</tr>
<tr>
<td>ChiP DNA Clean &amp; Concentrator</td>
<td>Zymo</td>
<td>Cat#D5205</td>
</tr>
<tr>
<td>PrimeFlow RNA assay Kit</td>
<td>Thermo Fisher Scientific</td>
<td>Cat#88-18005-210</td>
</tr>
<tr>
<td>KAPA HiFi HotStart ReadyMix</td>
<td>Roche</td>
<td>Cat#795835001</td>
</tr>
<tr>
<td>Q5 High Fidelity DNA polymerase</td>
<td>New England Biolabs</td>
<td>Cat#MO491L</td>
</tr>
<tr>
<td>HotStarTaq Plus kit</td>
<td>QIAGEN</td>
<td>Cat#203605</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phusion High-Fidelity DNA Polymerase</td>
<td>New England Biolabs</td>
<td>Cat#M0530L</td>
</tr>
<tr>
<td>5/3 RACE kit, 2nd generation</td>
<td>Roche</td>
<td>Cat#03353621001</td>
</tr>
<tr>
<td>TruSeq® RNA Sample Preparation Kit v2</td>
<td>Illumina</td>
<td>Cat#RS-122-2001</td>
</tr>
<tr>
<td>Clontech In-Fusion HD Cloning Kit</td>
<td>Takara Bio</td>
<td>Cat#E5510S</td>
</tr>
<tr>
<td>Gibson Assembly® Cloning Kit</td>
<td>New England Biolabs</td>
<td>Cat#E5510S</td>
</tr>
</tbody>
</table>

**Original Code**
- 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: GSE6417
- 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: Zýlicz 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 XXΔXic096 (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 ΔXertP096 (B5) This study SC46
- Mouse: TX ΔXertPcasi (D5) This study SC46
- Mouse: TX ΔXertE096 (F6) This study SC66

(Continued on next page)
### Oligonucleotides

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse: TX ΔXrEP/Cast (B11)</td>
<td>This study</td>
<td>SC66</td>
</tr>
<tr>
<td>Mouse: TX ΔXrEP/ECast (F3)</td>
<td>This study</td>
<td>SC69</td>
</tr>
<tr>
<td>Mouse: TX Ftx-XrEP/Cast (C10)</td>
<td>This study</td>
<td>SC69</td>
</tr>
<tr>
<td>Mouse: TX Ftx-XrEP/Cast (F9)</td>
<td>This study</td>
<td>SC69</td>
</tr>
<tr>
<td>Human: HEK293T</td>
<td>-</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**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 570)**: LGC Biosearch Technologies Custom made (SMF-1065-5-BS): [Table S6](#).

**Ta5 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

<table>
<thead>
<tr>
<th>Plasmid: pSpCas9(BB)-2A-Puro (pX459) V2.0</th>
<th>Ran et al., 2013</th>
<th>Addgene Plasmid #42230</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid: lentiGuide-Puro</td>
<td>Sanjana et al., 2014</td>
<td>Addgene Plasmid #52963</td>
</tr>
<tr>
<td>Plasmid: SP199 (lentiGuide-Puro with optimized constant region)</td>
<td>Genolet et al., 2021</td>
<td>N/A</td>
</tr>
<tr>
<td>Plasmid: pSLQ2817</td>
<td>Gao et al., 2016</td>
<td>Addgene Plasmid #84239</td>
</tr>
<tr>
<td>Plasmid: pSLQ2818</td>
<td>Gao et al., 2016</td>
<td>Addgene Plasmid #84241</td>
</tr>
<tr>
<td>Plasmid: SP106 (pSLQ2817 with blasticidin resistance)</td>
<td>Genolet et al., 2021</td>
<td>N/A</td>
</tr>
<tr>
<td>Plasmid: SP107 (pSLQ2818 with blasticidin resistance)</td>
<td>Genolet et al., 2021</td>
<td>N/A</td>
</tr>
<tr>
<td>Plasmid: pBroad3_hyPBase_IRES_tagRFP</td>
<td>Redolfi et al., 2019</td>
<td>N/A</td>
</tr>
<tr>
<td>Plasmid: pTXB1-3xFlag-pA-Tn5-FL</td>
<td>Kaya-Okur et al., 2019</td>
<td>Addgene Plasmid #124601</td>
</tr>
<tr>
<td>Plasmid: pLP1, pLP2, VSVG</td>
<td>Thermo Fisher Scientific</td>
<td>Invitrogen K497500</td>
</tr>
<tr>
<td>Plasmid: pSTARR-seq_human</td>
<td>Arnold et al., 2013</td>
<td>Addgene Plasmid #71509</td>
</tr>
</tbody>
</table>

### Software and algorithms

| Bedtools | Quintan 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)
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 keys 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 of 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 rTA 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 Rnfl2 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-STNΔTsixP 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-KRB-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*10^6 cells/cm^2 cells were seeded for 2iL conditions, while 7*10^5 cells/cm^2 were used for differentiation. E14-STNΔTsixP 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(2BB)-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/db/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 µg/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 plasmid 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 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 Apal 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). SgRNAs (sequences are given in Table S7) were cloned following the Zhang lab protocol (https://media.addgene.org/cms/filer_public/6d/db/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 µg/ml) overnight at 37°C. Single colonies were expanded and confirmed with Sanger sequencing.

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 with 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) 2 days after transduction. RFP-positive cells were sorted 24 h after transfection and expanded as single clones under blasticidin selection (5 ng/µl, Roth).

Genome engineering

Generation of deletion mESC lines

To generate deletions, up to 4*10^6 TX1072 (for ΔXertP, ΔXertE, ΔXertP/E ΔTsx-Xert) or E14-STN (for ΔTsxP) mESCs, cultured in gelatin-coated flasks in SL medium, were nucleofected using the Lonza 4D-Nucleasefector with 2 µg of each sgRNA/Cas9 plasmid and either 3 pmol or 30 pmol repair oligo (Table S7) (ΔXertP and ΔTsxP) or with the Alt-R CRISPR-Cas9 system (ID-T) (ΔXertE, ΔXertP/E ΔTsx-Xert) using the P3 Primary Cell 4D-NucleasefectorTM kit (Lonza) and CP-106 nucleasefector program. Alt-R RNP complexes were generated according to the manufacturer’s guidelines. In short, crRNA 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<sub>ΔTasp</sub>, clone C6 was further sub-cloned and following PCR genotyping (Figures S4A and S4B), subclone E14-STN<sub>ΔTasp</sub> B2 was chosen for future experiments (here referred to as E14-STN<sub>ΔTasp</sub>).

**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<sub>2</sub>O 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<sup>6</sup> 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 digestion 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 TX<sub>A</sub>Xic<sub>0</sub> 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 cdDNA 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 using 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, cdDNA was digested with RNaseH as described before. RNaseH-treated cdDNA 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’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 LGCGenomics 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 2x SSC 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*10⁶ cells per well with Xist-specific probes, labeled with Alexa-Fluor647 (V81-14258) (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.
Assay for Transposase-Accessible Chromatin by Sequencing (ATAC-seq) was used to profile open chromatin, as described previously with adaptations (Corces et al., 2017). XXLxic and XO cells were profiled at day 0, 2 and 4 of differentiation in two biological replicates. Cells were dissociated with trypsin and 6×10⁶ cells were lysed in 50 μL cold RSB buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 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 (illuminla), 100 nM Nextera Tn5 Transposase (illuminla), 33 μL PBS, 0.01% Digitonin, 0.1% H₂O) 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 H₂O. 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*10⁶ 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 artificial chromosome (BAC) clones from the BAC PAC Resource Center of the Children’s Hospital Oakland Research Institute. E.coli 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 (Macherey-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 (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 eluates were pooled and purified using the QIAquick Gel Extraction Kit (QIAGEN). Libraries were pooled in equimolar ratios and sequenced paired-end 50 bp on the HiSeq 2500 platform yielding approximately 2.5*10⁶ fragments per sample (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*10⁶ 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*10⁶ cells were cultured under 2iL conditions and 2*10⁶ 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*10⁷ 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 sgRNA library cloning in Table S1. (Doench et al., 2016) were included as negative controls, resulting in 7358 guides in total. The sgRNA library composition is provided in Table S1.

**Lentiviral packaging**

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 H2O. 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 electroprotector (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 an annealing temperature of 63°C in the first 3 cycles and 72°C in the subsequent 11 cycles. The ampiclons were subsequently gel-purified.

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™ 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 µg/ml 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*10⁶ cells were transduced with viral supernant of the sgRNA library (MOI = 0.3). Additionally, 2*10⁶ 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/ml, 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*10⁶ 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\(^{\times}10^8\) cells were stained by Flow-FISH with an Xist-specific probe as described above. 2\(^{\times}10^7\) 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 1.1-1.5\(^{\times}10^7\) cells were recovered per fraction. After sorting, the cell pellets were snap-frozen and stored at \(-80^\circ\)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 \(\mu\)l 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\(^\circ\)C. The next day 200 \(\mu\)g of RNase A (Thermo Fisher Scientific) were added and the samples were incubated at 37\(^\circ\)C for 1 h. 100 \(\mu\)g of Proteinase K (Sigma) were subsequently added, followed by a 1 h incubation at 50\(^\circ\)C. Phenol/chloroform/isomyl 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 \(\times\) g for 10 min at RT. The aqueous phase was transferred to a new tube, 1 \(\mu\)l 100% EtOH, 90 \(\mu\)l 5 M NaCl and 1 \(\mu\)l Pellet Paint (Merck) was added to each sample, mixed, and incubated at \(-80^\circ\)C for 1 h. DNA was pelleted by centrifugation for 16,000 \(\times\) g for 15 min at 4\(^\circ\)C, pellets were washed twice with 70% EtOH, air-dried and resuspended in 50 \(\mu\)l 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 \(\mu\)g of each sample were amplified using the ReadyMix Kapa polymerase (Roche) with a total of 20 cycles and an annealing temperature of 55\(^\circ\)C. Between 0.1-2 \(\mu\)g genomic DNA was amplified per 50 \(\mu\)l 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 \(m\) in 2iL-containing medium and used as an undifferentiated control. Cells were harvested for Flow-FISH after 2 days of differentiation.

**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 Carbencilin (100 \(\mu\)g/ml) and incubated in a shaker at 37\(^\circ\)C for 4 hours. Afterward the starter culture was added to 400 mL LB medium supplemented with Carbencilin (100 \(\mu\)g/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 \(\mu\)L of 1 M IPTG was added to the culture and incubated in a cooled shaker overnight at 18\(^\circ\)C at 150 rpm. The following morning, bacteria were centrifuged in a JA-12 rotor at 10,000 rpm for 30 min at 4\(^\circ\)C. Bacterial pellets were snap frozen in liquid nitrogen and stored at \(-80^\circ\)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\(^\circ\)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\(^\circ\)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 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\(^\circ\)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 \(\times\) g for 15 min. The protein concentration was measured with the detergent compatible Bradford assay kit (Pierce), adjusted to 832 ng/\(\mu\)l 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 pH 8.0, 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 the beads, the washes were 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 0.95× 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 16,000g and 4 °C. The RNA pellet was washed once with 75% ethanol and resuspended in nuclelease free water. Residual genomic DNA was removed in solution with DNasel (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 Micro Tubes 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 nuclelease 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 MACS 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 TRITizol 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/C14) 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/C14 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 MgCl2, 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 N2.

**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 Rm2. 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 demultiplexed 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 deeptools (v2.0) (Ramirez 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) (Ramirez 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 since all 5’-RACE bands suggested, 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 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 with 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) ([Ramirez 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 -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) ([Ramirez et al., 2016]) with [bamCompare -e -bs 10 –operation ratio –normalizeUsing CPM]. Mapping statistics and quality control metrics are shown in Table S1.

**Genomic peak annotation**
For the analysis shown in Figure S2E, CUT&Tag peaks identified in undifferentiated XX\_Xic mESCs using MACS2 (see above) were assigned to gene features according to the annotation package TxDb.Musculus.UCSC.mm10.knownGene (v3.10.0) ([https://bioconductor.org/packages/release/data/annotation/html/TxDB.Musculus.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\_Xic cells were compared to native ChIP-seq data profiling the same marks in the parental TX1072 cell line ([Zylicz 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) ([Ramirez 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 [pcomp[center = TRUE, scale = TRUE]].

**CUT&Tag correlation analysis**
BAM files, excluding mitochondrial reads, were counted in 1 kb bins using deepTools2 (v3.4.1) ([Ramirez et al., 2016]) with options [multiBamSummary bins -bs 1000 -b 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\_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

**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 X Chr 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 IncRNA expression within the Xchr, we performed differential expression analysis between the XX

**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 –soon-mixed –no-discordant –phred33 -l 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 mESCs 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) (Granth 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 occurrence of the CTCF transcription factor binding motif, which was retrieved from the JASPAR database (Fornes et al., 2020) (8th 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 ≥ 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

---

**Molecular Cell**

**Article**

---

**OPEN ACCESS**

---

**Molecular Cell**

**Article**

---

**OPEN ACCESS**

---

**Molecular Cell**

**Article**

---

**OPEN ACCESS**
chrX:103,190,001-103,950,000 were considered. In this part of the enriched region, both investigated cell lines (XO and XX_{dXX}) 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.