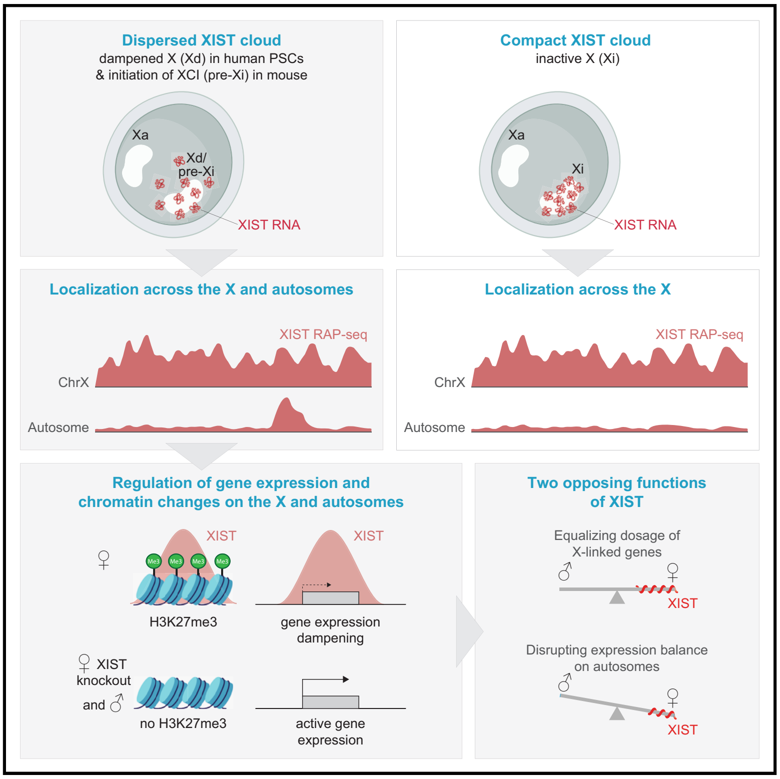
Cell

Article

XIST directly regulates X-linked and autosomal genes in naive human pluripotent cells

Graphical abstract



Highlights

- XIST localizes across the X and spreads to autosomal regions in female naive hPSCs
- XIST dampens the expression of X-linked and autosomal genes in female naive hPSCs
- XIST mediates chromatin changes at target regions on the X and autosomes
- Xist spreads to autosomal regions during XCI initiation in differentiating mouse PSCs

Authors

Iris Dror, Tsotne Chitiashvili, Shawn Y.X. Tan, ..., Douglas L. Black, Fangyuan Ding, Kathrin Plath

Correspondence

kplath@mednet.ucla.edu

In brief

During X chromosome dampening, when XIST expression is uncoupled from complete gene silencing and displays a non-typical dispersed configuration, XIST spreads beyond the X chromosome to downregulate gene expression on autosomes.

Dror et al., 2024, Cell *187*, 110–129
January 4, 2024
https://doi.org/10.1016/j.cell.2023.11.033

CellPress

Cell

Article

XIST directly regulates X-linked and autosomal genes in naive human pluripotent cells

Iris Dror,[1,7] Tsotne Chitiashvili,[1,7] Shawn Y.X. Tan,[1,8] Clara T. Cano,[1,8] Anna Sahakyan,[1,8] Yolanda Markaki,[1,2] Constantinos Chronis,[1,3] Amanda J. Collier,[1] Weixian Deng,[1] Guohao Liang,[4] Yu Sun,[1] Anna Afasizheva,[1] Jarrett Miller,[1] Wen Xiao,[5] Douglas L. Black,[5] Fangyuan Ding,[4,6] and Kathrin Plath[1,9,*]

[1]Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Institute for Structural and Chemical Biology & Department of Molecular and Cell Biology, University of Leicester, Leicester LE1 7RH, UK
[3]Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA
[4]Department of Biomedical Engineering, University of California Irvine, Irvine, CA 92697, USA
[5]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA
[6]Department of Developmental and Cell Biology, Department of Pharmaceutical Sciences, University of California Irvine, Irvine, CA 92697, USA
[7]These authors contributed equally
[8]These authors contributed equally
[9]Lead contact
*Correspondence: kplath@mednet.ucla.edu
https://doi.org/10.1016/j.cell.2023.11.033

SUMMARY

X chromosome inactivation (XCI) serves as a paradigm for RNA-mediated regulation of gene expression, wherein the long non-coding RNA XIST spreads across the X chromosome in *cis* to mediate gene silencing chromosome-wide. In female naive human pluripotent stem cells (hPSCs), XIST is in a dispersed configuration, and XCI does not occur, raising questions about XIST's function. We found that XIST spreads across the X chromosome and induces dampening of X-linked gene expression in naive hPSCs. Surprisingly, XIST also targets specific autosomal regions, where it induces repressive chromatin changes and gene expression dampening. Thereby, XIST equalizes X-linked gene dosage between male and female cells while inducing differences in autosomes. The dispersed Xist configuration and autosomal localization also occur transiently during XCI initiation in mouse PSCs. Together, our study identifies XIST as the regulator of X chromosome dampening, uncovers an evolutionarily conserved *trans*-acting role of XIST/Xist, and reveals a correlation between XIST/Xist dispersal and autosomal targeting.

INTRODUCTION

Embryonic development requires the unique regulation of X-linked gene expression due to the different number of X chromosomes between males and females. In placental mammals, X chromosome dosage compensation occurs by X chromosome inactivation (XCI) in females, which transcriptionally silences most genes on one of the two X chromosomes, equalizing X-linked gene expression with males.[1–8] The long non-coding RNA (lncRNA) Xist is necessary and sufficient for the induction of XCI[9–15] and therefore critical for establishing gene expression balance between the sexes.

Xist spreads from its active transcription locus on the X chromosome across the entire X to mediate gene silencing and heterochromatin formation.[16–20] *Cis*-limited spreading and silencing also occur when Xist is expressed ectopically from autosomes.[21–26] Therefore, the prevailing view is that Xist localizes and acts exclusively on the chromosome it is expressed from.

On the inactive X chromosome (Xi), Xist is localized within ~50 foci, each containing two RNA molecules.[27–29] The cluster of Xist foci is known as Xist cloud.[30] Recent reports described a non-typical, dispersed configuration of XIST in female human and cynomolgus monkey pre-implantation embryos,[31–33] naive human pluripotent stem cells (hPSCs),[34,35] human primordial germ cells (hPGCs),[36] mouse and human immune cells,[37–39] and cancer cells,[40] wherein XIST foci are detected throughout a large portion of the nucleus. The significance of this dispersed configuration is currently unknown. The dispersed XIST distribution in human pre-implantation embryos, naive hPSCs, and hPGCs is accompanied by a distinctive form of X chromosome dosage compensation coined "X chromosome dampening" (XCD).[33,34,36,41] Genes on the dampened X chromosome (Xd) are downregulated, yet not completely silenced. Thus, XIST expression, which is normally accompanied by XCI,[1–8] is uncoupled from complete gene silencing in cells with XCD. The mechanisms underlying XCD and whether XIST has any regulatory role in XCD are unknown. XIST dispersal may represent the

CellPress
OPEN ACCESS

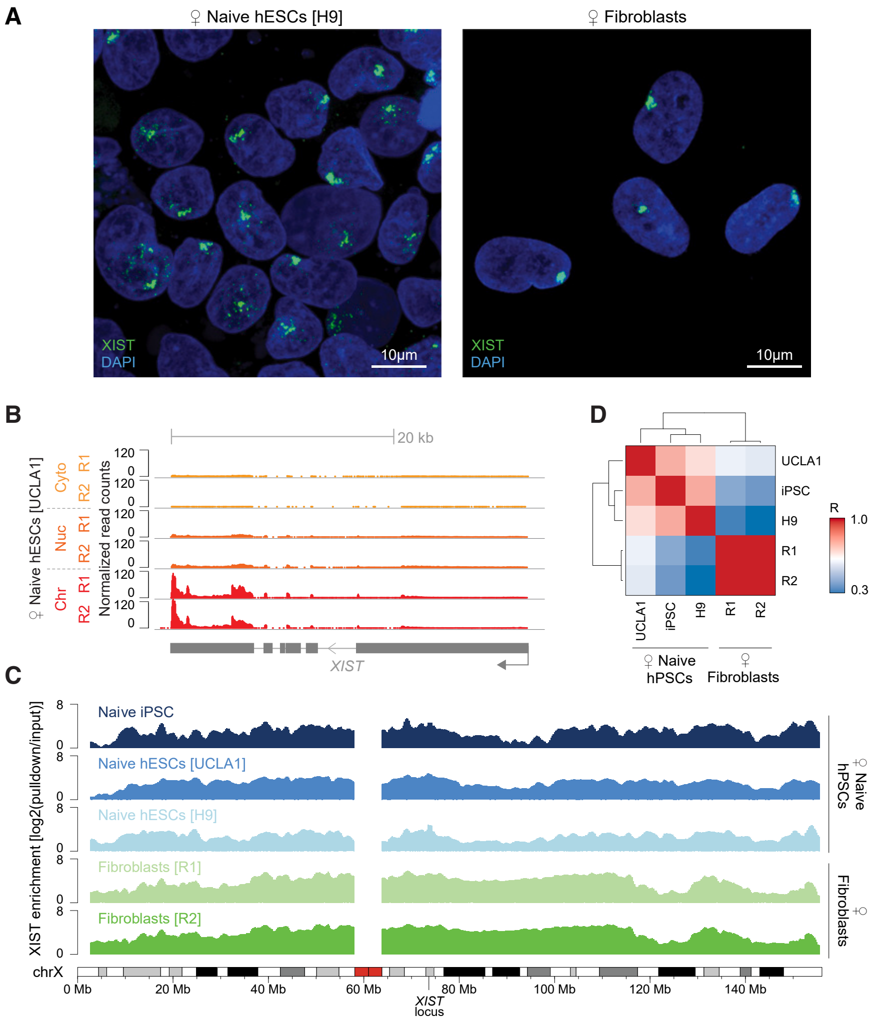


Figure 1. XIST localizes across the X chromosome in female naive hPSCs

(A) RNA-FISH images for XIST in indicated DAPI-stained cell lines.

(B) Genome browser tracks of the *XIST* locus showing normalized read counts from poly(A) + RNA isolated from the chromatin (Chr), nucleoplasmic (Nuc), and cytoplasmic (Cyto) fractions of naive human embryonic stem cells (hESCs) (UCLA1) from two replicates (R1/R2).

(C) XIST enrichment (based on RAP-seq) along chrX in indicated cell lines and replicates (R1/R2); centromeric region (red) is masked.

(D) Pearson correlation (R) of XIST enrichment (100 kb windows, every 25 kb) along chrX for cell lines from (C); dendrogram represents Euclidean distances.

See also Figure S1.

dissociation of the RNA from chromatin, which normally only happens during mitosis,[30] association of the RNA with an X chromosome in an extended conformation, or spreading beyond the X-territory to autosomal regions. The latter would be intriguing as, in principle, Xist is able to regulate autosomal genes.[16,21–26]

In this study, we explored the significance of the dispersed XIST configuration and XIST's role in XCD by examining XIST's localization and function in female naive hPSCs. We found that despite its widespread distribution, XIST localizes across the X chromosome and mediates gene expression dampening and chromatin changes on the Xd, establishing XIST as the regulator of XCD. Additionally, we discovered a remarkable spreading of XIST to specific autosomal regions, where XIST alters the chromatin state and dampens target gene expression, inducing an imbalance of autosomal gene expression between female and male cells. Xist's dispersed distribution and autosomal targeting also occur transiently during XCI initiation in mouse PSCs. Overall, our study shifts the paradigm from XIST/Xist being a *cis*-limited regulator to showing that it can also operate in *trans*, which may be linked to its dispersed configuration.

RESULTS

XIST spreads across the X in female naive hPSCs

The dispersed localization of XIST in female naive hPSCs detected by RNA fluorescent *in situ* hybridization (FISH)[34,35] differs significantly from the dense XIST cloud in somatic cells like fibroblasts (Figure 1A). To investigate this unusual XIST distribution, we first asked whether the RNA associates with chromatin in female naive hPSCs. Biochemical fractionation into cytoplasm, nucleoplasm, and chromatin established that spliced and polyadenylated XIST predominantly partitions to chromatin (Figures 1B and S1A).

To map at high resolution where on chromatin XIST localizes, we applied XIST RNA antisense purification (RAP)-seq[16] to two female naive human embryonic stem cell (hESC) lines (H9 and UCLA1) and one female naive human induced pluripotent stem cell (hiPSC) line (Tables S1 and S2), all of which exhibit the dispersed XIST distribution (Figures S1B and S1C). For comparison, we also generated replicate XIST RAP-seq data for female human fibroblasts with a compact XIST cloud on the Xi (Figure 1A). The *XIST* locus with its accumulating nascent transcripts was among the most enriched genomic regions (Figure S1D), confirming the specificity of the approach.

We found that XIST accumulates across the X in naive hPSCs and fibroblasts, with a ≥2-fold enrichment in ≥94% of 100 kb windows (Figures 1C and S1E). However, the enrichment of XIST is higher in fibroblasts than in naive hPSCs (~44% vs. ~4% of 100 kb windows enriched ≥20-fold), which we confirmed with a quantitative PCR measurement (Figures S1E and S1F). Moreover, the distribution along the X reproducibly differs between hPSCs and fibroblasts (Figures 1C, 1D, and S1G), which is likely explained by the differential folding of the chromosome.[42,43] Overall, these results demonstrate that XIST localizes across the X chromosome in

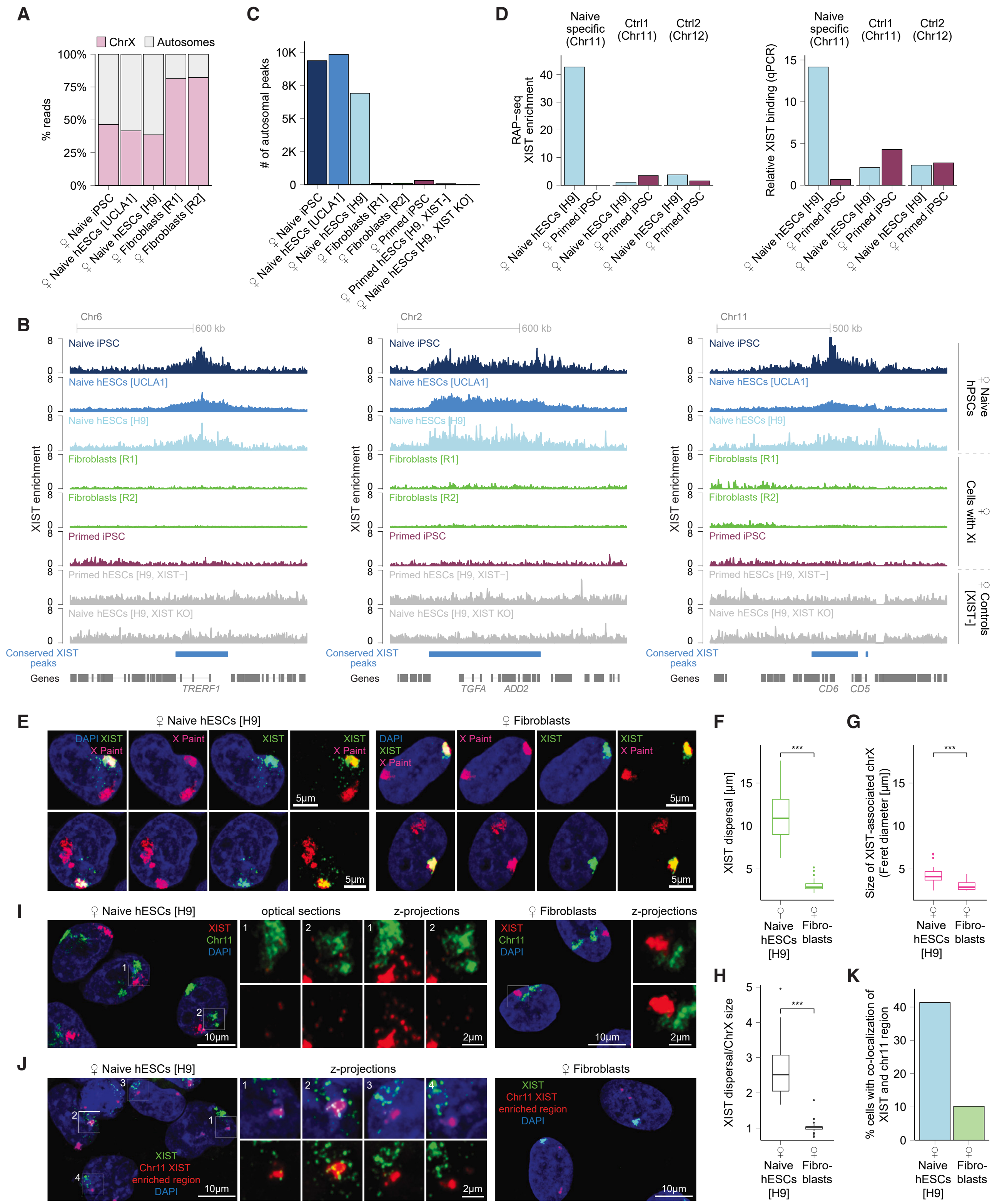


Figure 2. XIST spreading to specific autosomal regions in female naive hPSCs

(A) Proportion of XIST RAP-seq reads aligning to chrX or autosomes for indicated cell lines.

(B) Examples of autosomal XIST enrichment in indicated cell lines with conserved peaks marked.

(legend continued on next page)

naive hPSCs, despite its dispersed appearance and the lack of XCI.

XIST spreads to autosomal regions in naive hPSCs

We next explored whether the dispersed XIST distribution represents spreading to autosomes. We found that a higher proportion of RAP-seq reads aligns to autosomes in naive hPSCs than in fibroblasts (Figure 2A) and that autosomal regions with a >2-fold XIST enrichment are nearly exclusively present in naive hPSCs (on average 397 100 kb windows across naive hPSC lines compared with 1 across fibroblast replicates; Figure S1H). Next, we applied peak calling to accurately define XIST-enriched regions and identified thousands of autosomal peaks in each naive hPSC line (6,903 MACS2 peaks in H9, 9,845 in UCLA1, and 9,347 in iPSCs; see STAR Methods) and ~100 in the two fibroblast replicates (79 and 100 MACS2 peaks in R1 and R2; often located close to centromeres and telomeres) (Figures 2B, 2C, and S1I). These findings indicate that autosomal regions are enriched for XIST specifically in naive hPSCs. The lack of autosomal XIST accumulation in fibroblasts is in line with the widely accepted *cis*-limited function of XIST on the Xi.

To validate that we are capturing autosomal XIST interactions, we performed additional experiments. (1) To ensure XIST-specific hybridization of the RAP probes, we applied RAP-seq to female naive XIST knockout (KO) H9 hESCs (see below) and detected virtually no enrichment on the X or autosomes (Figures 2B, 2C, S1J, and S1K). We made a similar observation with primed hESCs that do not express XIST due to Xi erosion[44] (XIST-negative H9; Figures 2B, 2C, S1J, and S1K), confirming the specificity of the RAP-seq approach to XIST. (2) We additionally performed XIST RAP-seq in primed hiPSCs with an XIST-coated Xi derived from the same fibroblast population that gave rise to the naive hiPSC line with strong autosomal XIST binding. We found only a few autosomal XIST peaks in these cells (Figures 2B, 2C, S1J, and S1K), therefore excluding the possibility of genetic background influencing the detection of autosomal XIST. (3) We confirmed the naive hPSCs-specific accumulation of XIST at an autosomal region with an independent measurement by combining XIST RAP with a quantitative PCR readout (Figure 2D).

Next, we used imaging approaches to confirm the autosomal spreading of XIST in individual cells. Combining XIST RNA FISH with the detection of the X chromosome by DNA FISH, we detected XIST throughout the X in female naive hESCs and observed many RNA foci beyond the X territory, whereas the RNA is largely restricted to the Xi-territory in fibroblasts (Figure 2E). Quantifications show that the dispersal of XIST occurs over a region larger than the X chromosome specifically in naive hPSCs (Figures 2F–2H). Combining XIST RNA FISH with paint for chromosome (chr) 11, which encompasses many XIST-enriched regions, we found that XIST foci frequently neighbor or immerse within the chr11-territory in naive hESCs and not in fibroblasts (Figure 2I). Moreover, XIST foci overlap with a specific highly XIST-enriched autosomal region in 40% of the naive hESCs vs. only 10% of the fibroblasts (Figures 2J and 2K).

Together, these results identify the association of endogenously expressed XIST with chromosomes in *trans*. Moreover, autosomal targeting of XIST is uniquely associated with the naive pluripotent state as it is not observed in primed hPSCs and fibroblasts, suggesting that the spreading of XIST to autosomes is cell-type specific and linked to the dispersed XIST configuration of naive hPSCs.

XIST targets developmental genes on autosomes

To characterize the autosomal localization of XIST, we assembled a reference XIST peak set encompassing all peaks from naive and primed hPSCs and fibroblasts, wherein individual peaks are assigned to one or more samples (Table S3; see STAR Methods). XIST peaks are distributed across all autosomes in each naive hPSC line (Figures 3A and S2A). 724 peaks, referred to as conserved peaks, are detected in all three naive hPSC lines (Figure 3B), which are more than expected by random chance ($p < 0.001$; permutation testing). These conserved peaks are present on all autosomes except chr13, most abundant on chr11 (101 peaks, 3.7% of chr), chr1 (98 peaks, 3.1%), and chr20 (50 peaks, 2.4%) (Figures 3C, S2B, and S2C; Table S3), more significant and wider than non-conserved peaks (Figures 3D and S2D), and encompass 434 protein-coding genes (Figure S2E; Table S4). Targeted genes are involved in developmental processes based on gene ontology analysis ($p < 5 \times 10^{-6}$; Figure 3E), include several transcription factors (TFs) such as *TEAD1* and *ETV5*,[45,46] and are significantly upregulated during capacitation for multi-lineage differentiation (55% of genes, hypergeometric test $-\log_{10}(p) > 16$; Figure 3F) or the naive-to-primed pluripotency transition (Figures 3G and 3H).[47,48] Thus, XIST associates with autosomal genes that are regulated during early human development.

XIST preferentially targets spatially close and polycomb-marked autosomal regions

To characterize XIST's autosomal spreading, we compared XIST enrichment in naive hPSCs with genomic features and found that it best correlates with the density of mammalian-wide

(C) Number of autosomal XIST peaks (MACS2 peaks) in indicated cell lines. k = ×1,000.
(D) XIST enrichment in indicated cell lines at a naive hPSC-specific XIST-enriched region and two control regions with low XIST (Ctrl1/2) based on RAP-seq (left) and RAP-qPCR (right).
(E) RNA FISH for XIST and chrX-paint in indicated cell lines (z-projections, two examples each).
(F) Quantification of XIST dispersal in indicated cell lines based on (E). Wilcoxon p value: ***$p < 0.001$.
(G) Size quantification of the XIST-associated chrX based on (E). Wilcoxon p value: ***$p < 0.001$.
(H) Ratio of XIST dispersal from (F) and chrX size from (G). Wilcoxon p value: ***$p < 0.001$.
(I) RNA FISH for XIST and chr11-paint in indicated cell lines; zoom-in's show optical sections and z-projections for boxed regions.
(J) As in (I), except for XIST RNA-FISH and DNA FISH for an XIST-enriched region on chr11.
(K) Percentage of cells with the XIST-enriched region on chr11 from (J) overlapping with XIST foci.
See also Figure S1.

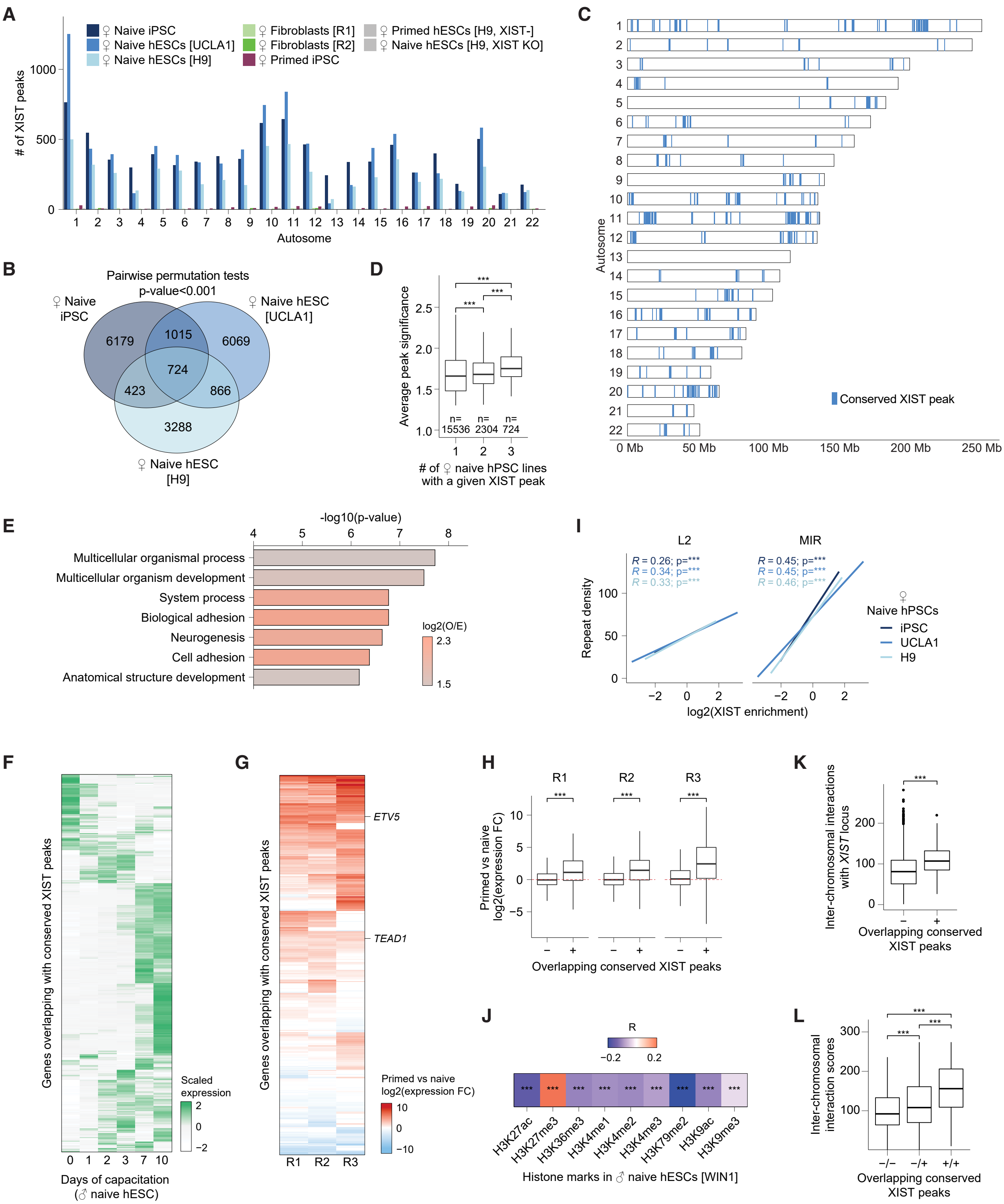


Figure 3. XIST preferentially targets autosomal regions marked by H3K27me3 and enriched for developmental genes

(A) Number of XIST peaks per autosome in indicated cells lines.

(B) Overlap of naive hPSCs-specific autosomal XIST peaks between indicated naive hPSC lines.

(C) Location of the 724 conserved XIST peaks on autosomes.

(legend continued on next page)

interspersed repeats (MIRs) and L2 long interspersed retrotransposable elements (LINEs) (Figure 3I; Table S5). Since XIST enrichment on the X chromosome also best correlates with these features in addition to gene density and short interspersed retrotransposable elements (SINEs) (Figure S2F), related regulatory mechanisms may govern the association of XIST with the X and autosomes.

To uncover additional regulatory features, we mined ENCODE chromatin immunoprecipitation sequencing (ChIP-seq) data from diverse cell types and found that binding sites of the polycomb-repressive complex 2 (PRC2) subunits SUZ12 and EZH2 are enriched in conserved XIST peaks (Figure S2G; Table S5). Consistent with this finding, genomic regions with conserved peaks are more lowly expressed than non-targets (Figure S2H) and within the B1 sub-compartment of the nucleus (Figure S2I), formed by spatial interactions of facultative heterochromatin regions carrying high levels of the PRC2 mark H3K27me3.[49] To explore if autosomal XIST target regions are pre-marked with particular histone modifications, we performed ChIP-seq for nine histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me3, H3K27me3, H3K27ac, H3K36me3, and H3K79me2) in male naive hESCs (WIN1[50]), which do not express XIST (Figures S1B, S1C, and S3A), and found that autosomal XIST targeting correlates best with H3K27me3 levels (Figure 3J), suggesting that XIST preferentially targets polycomb protein-regulated autosomal regions.

We previously showed that during the initiation of XCI, Xist first spreads to X-linked sites that are spatially close to the *Xist* transcription locus.[16] We therefore explored in high-resolution Hi-C data from primed hESCs[51] whether XIST-occupied autosomal regions are closer to the *XIST* locus in 3D space than non-target regions and observed significantly stronger inter-chromosomal interactions of XIST-bound autosomal regions (Figure 3K). Autosomal target regions also more strongly interact with each other than with non-target regions (Figure 3L), perhaps due to the fact that genomic regions with similar chromatin characteristics colocalize.[49,52] Together, these results suggest that XIST spreads to spatially close autosomal sites and that nuclear organization and specific chromatin states guide its localization.

The distinctive localization of XIST across the X chromosome and at specific autosomal sites in female naive hPSCs raised the possibility that XIST may control gene expression and chromatin state on the X chromosome and autosomes. Alternatively, XIST may not be active in naive hPSCs, particularly given that XCI does not occur. In this case, XCD would be regulated through an XIST-independent mechanism. Another possibility is that XIST has different roles on the X and autosomes. To distinguish between these possibilities, we first explored the gene expression state on the XIST-associated X chromosome.

Genes on the XIST-associated X are dampened to varying degrees

Leveraging the fact that only one X chromosome in female naive hPSCs is coated with XIST[34] (Figures 2E, S1B, and S1C) and that XIST stays associated with the X it is transcribed from (Figures S3B and S3C), we could ask whether gene expression on the XIST-associated X is dampened compared with the X without XIST. We applied quantitative RNA FISH to measure the nascent transcript levels of five X-linked genes and the intensity of the X-encoded lncRNA XACT[53] in several female naive hPSCs (UCLA1, H9, UCLA9, and HNES3), in combination with XIST RNA FISH. We found that all genes are biallelically expressed, which confirmed the lack of XCI, although the proportion of cells with biallelic expression varies (Figures S3D–S3F; STAR Methods). In cells with biallelic transcription, we consistently observed lower transcript signals on the XIST-associated X compared with the X lacking XIST (Figure S3G). Since the examined naive hPSCs were derived from primed pluripotent cells (UCLA1, UCLA9, and H9[34,44]) or the blastocyst (HNES3[54]) and cultured in different media (5iLAF [H9, UCLA1, and UCLA9] and t2iL+Gö [HNES3]), the reduction of nascent transcription on the XIST-coated X chromosome occurs regardless of the culture condition and origin of the cells. Moreover, the reduced expression on the XIST-associated X is observed in naive hPSCs derived from primed hESCs with two active X chromosomes (UCLA9[44]) or directly from the blastocyst (HNES3[54]), indicating that dampening on the XIST-associated X is not a remnant of a prior Xi (STAR Methods).

(D) Significance (−log(q)) of naive hPSCs-specific autosomal XIST peaks present in one, two, or three hPSC line(s). Wilcoxon p value: $^{***}p < 0.001$; n = number of peaks; outliers omitted for clearer visualization.

(E) Most highly enriched ontologies and their significance (Fisher's exact test $-\log_{10}(p)$) for genes associated with conserved autosomal XIST peaks, colored by $\log_2$(observed/expected [O/E]).

(F) Scaled gene expression during capacitation time course,[47] for genes overlapping with conserved autosomal XIST peaks.

(G) Log_2 gene expression fold change (FC) between the primed and naive pluripotent states for male HNES1 hESCs[47] (replicate R1) and female H9 hESCs from two publications (R2[47] and R3[48]), for genes overlapping with conserved autosomal XIST peaks (45%, 47%, and 66% of genes, respectively, display significant upregulation).

(H) Log_2 expression FC between primed vs. naive pluripotent cells from (G) for genes overlapping (+) conserved autosomal XIST peaks or not (−). Dashed line represents no change. Wilcoxon p value: $^{***}p < 0.001$.

(I) Linear regression for RAP-seq-based XIST enrichment along autosomes (1 Mb every 250 kb) vs. the density of L2 and MIR elements for indicated cell lines. Pearson correlation (R) and p value ($^{***}p < 0.001$) are given.

(J) Pearson correlation (R) and p value ($^{***}p < 0.001$) for comparison of autosomal XIST enrichment (average of three female naive hPSCs; 100 kb every 25 kb) and histone marks derived by ChIP-seq for male naive hESCs (WIN1).

(K) Inter-chromosomal interactions of the *XIST* locus with autosomal windows overlapping (+) or not (−) with conserved XIST peaks. Wilcoxon p value: $^{***}p < 0.001$.

(L) Inter-chromosomal interactions between pairs of autosomal windows where either both windows overlap with conserved XIST peaks (+/+), one (+/−), or none (−/−). Wilcoxon p value: $^{***}p < 0.001$. Outliers omitted for clearer visualization.

See also Figure S2.

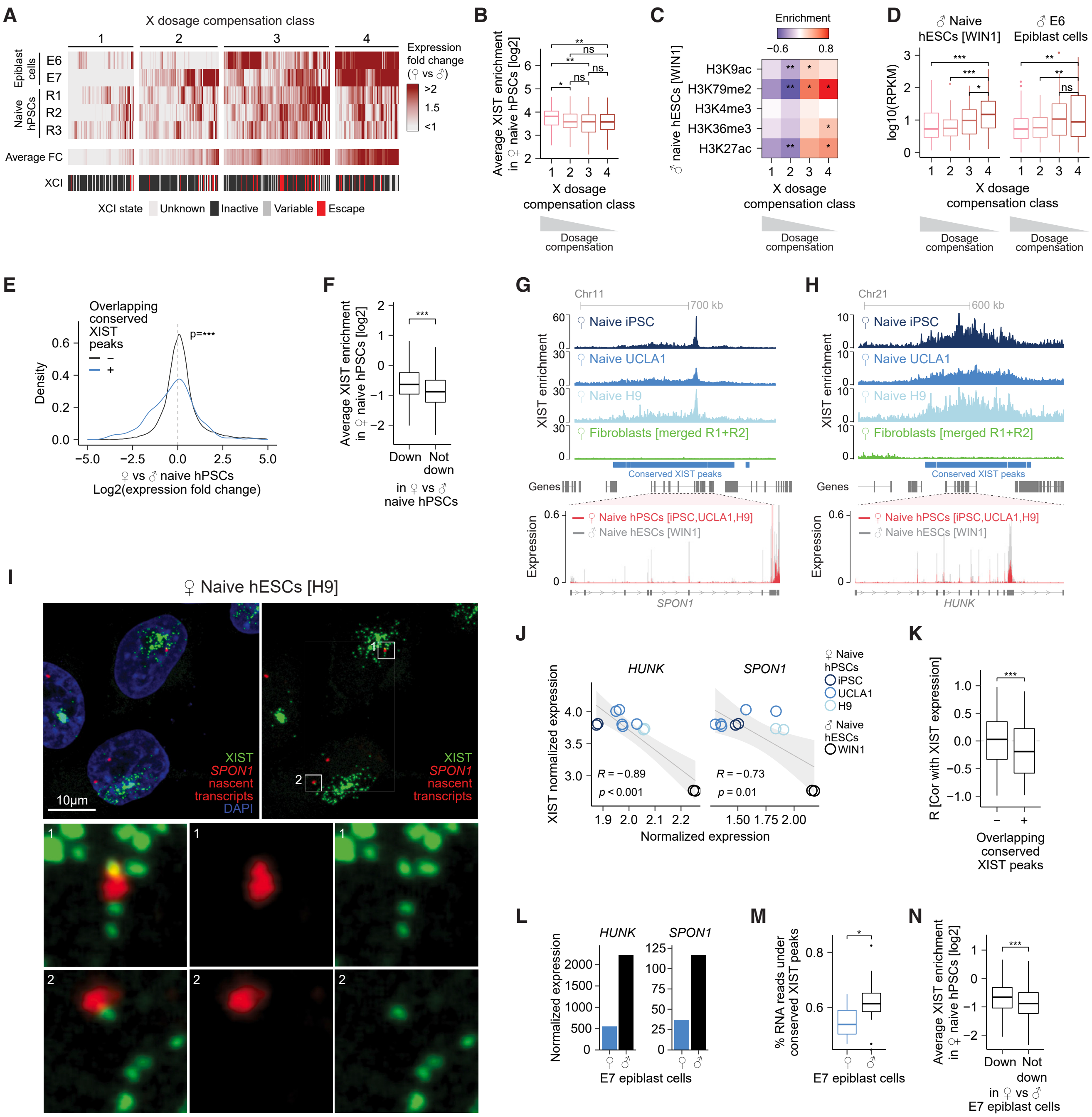


Figure 4. XIST-targeted X-linked and autosomal genes are downregulated in naive female hPSCs and female epiblast cells

(A) X-linked gene expression differences between female and male epiblast cells from E6 and E7 pre-implantation embryos[33], and between different female and male naive hPSC lines (R1: female naive H9 vs. male HNES1 hESCs[47]; R2: female naive iPSCs, UCLA1 and H9 hPSCs vs. male WIN1 hESCs; R3: female naive HNES3 vs. male HNES1 hESCs); clustered into four X-dosage compensation classes. Average FC and XCI state are given.

(B) Average XIST enrichment at the transcriptional start site of genes from each X-dosage compensation class. Wilcoxon p values: ns $\geq$ 0.05, *p < 0.05, **p < 0.01. The degree of female/male dosage compensation from (A) is indicated with the triangle.

(C) Enrichment of histone marks (from ChIP-seq in male cells) in each X-dosage compensation class from (A). Wilcoxon p values: *p < 0.05, **p < 0.01.

(D) Expression level for genes in each X-dosage compensation class in naive male hESCs and male E6 pre-implantation epiblast cells.[33] Wilcoxon p values: ns $\geq$ 0.05, *p < 0.05, **p < 0.01, ***p < 0.001.

(E) Density plot of expression differences between female naive iPSCs, UCLA1, and H9 hPSCs vs. male naive WIN1 hESCs, for autosomal genes overlapping (+) conserved autosomal XIST peaks or not (–). Wilcoxon p value: ***p < 0.001 comparing (+) vs. (–) genes. Dashed line = no difference.

(legend continued on next page)

Overall, these data uncover the correlation of XIST expression with gene expression dampening on the X chromosome in *cis* and suggest that XIST mediates XCD in female naive hPSCs. Consistent with this, single-cell RNA sequencing (scRNA-seq) and bulk RNA-seq data show a decrease in X-linked relative to the autosomal gene transcripts with increasing XIST levels in female naive hPSCs, consistent with a prior report,[34] and allelic analysis confirmed reduced gene expression on the XIST-positive compared with the XIST-negative X (Figures S3H–S3M; Table S2).

To explore whether all X-linked genes are similarly repressed, we compared transcript levels between female and male naive hPSCs or epiblast cells from pre-implantation embryos. We found that X-linked genes have diverse degrees of dosage compensation, captured by classification into four dosage compensation classes (Figures 4A, S4A, and S4B). The most dosage-compensated genes (class 1) display a more similar expression level between female and male cells and have the highest accumulation of XIST (Figure 4B), whereas less dosage-compensated genes (classes 2–4) exhibit an increasingly female-biased expression. More genes in classes 3 and 4 than in classes 1 and 2 escape XCI in somatic cells (Figure 4A). Class 3 and 4 genes are characterized by a higher level of active histone modifications and higher expression in male cells *in vitro* and *in vivo* (Figures 4C and 4D), suggesting that the active state protects X-linked genes from being strongly dampened. These data suggest that genes are differentially regulated by XIST.

XIST-targeted autosomal genes show male expression bias

Next, we investigated the possibility that XIST regulates autosomal gene expression in female cells. Comparing male and female expression levels, we found that XIST-targeted autosomal genes have significantly lower expression in female vs. male naive hPSCs than genes not overlapping with conserved peaks (Figures 4E and S4C). Similarly, autosomal genes that are more lowly expressed in female vs. male naive hPSCs display a higher enrichment of XIST than all other autosomal genes (Figures 4F and S4D). As seen on the X, on autosomes only a subset of XIST target genes is strongly regulated. Specifically, of the 246 genes associated with conserved XIST peaks and expressed in hPSCs, 42 genes ($p < 0.001$; Fisher's exact test) are significantly more lowly expressed in female compared with male hPSCs (Table S6). Female-repressed genes include the most highly XIST-enriched genes *SPON1* ($\log_2$(female/male expression) = -3.5, chr11) known to promote neuronal differentiation[55] and *HUNK* ($\log_2$(female/male expression) = -1.66, chr21) known to control proliferation and differentiation[56] (Figures 4G and 4H). RNA FISH for nascent transcripts of *SPON1* revealed a higher proportion of monoallelic expression in female compared with male naive hPSCs (Figures S4E and S4F) and the co-localization of the *SPON1* transcription site with XIST foci (Figure 4I). We also found a negative correlation between the transcript levels of *HUNK* or *SPON1* with XIST levels across naive hPSC lines (Figure 4J). On average, this negative correlation applies to all autosomal genes bound by XIST, whereas autosomal genes not targeted by XIST show no correlation with XIST expression (Figure 4K). These data suggest that XIST deposition at autosomal loci mediates their downregulation in female naive hPSCs.

Examining scRNA-seq data,[33] we also found that epiblast cells of the pre-implantation embryo have male-biased expression of *SPON1* and *HUNK* and that XIST-targeted autosomal genes display slightly lower, albeit significant, expression in female compared with male epiblast cells (Figures 4L and 4M). Similarly, autosomal genes significantly downregulated in female compared with male cells display higher XIST enrichment than genes not downregulated (Figure 4N). These findings show that the repression of XIST-targeted autosomal genes extends to the human embryo.

XIST mediates dampening of X-linked and autosomal gene expression

To directly explore the function of XIST in female naive hPSCs, we excised the first ~2 kb of the *XIST* gene, a region critical for the expression and function of mouse Xist,[22,57] in female primed H9 hESCs (Figures 5A and S5A–S5D) and converted two KO clones (C7/C18) and wild-type (WT) H9 hESCs to the naive pluripotent state (Figure 5A). RNA FISH and bulk RNA-seq confirmed the successful establishment of the naive state and deletion of XIST (Figures 5B–5E).

Bulk and scRNA-seq data showed that the X-to-autosome (X/A) transcript ratio is significantly increased in KO cells due to the global upregulation of X-linked relative to autosomal genes (Figures 5F, 5G, and S5E–S5J). On the X, genes of X-dosage compensation class 1 are more strongly upregulated in the absence of XIST than less dosage-compensated genes

(F) Average XIST enrichment from three female hPSC lines (Figure 1C) at autosomal genes significantly downregulated, or not, in naive female iPSCs, UCLA1, and H9 hPSCs vs. male WIN1 hESCs. Wilcoxon p value: ***$p < 0.001$.

(G) Top: XIST enrichment across the chr11-region containing *SPON1* in indicated cell lines. Bottom: normalized RNA-seq reads across the zoomed-in *SPON1* locus (thick lines = exons; arrowheads = transcription direction) in indicated male and female naive hPSC lines.

(H) As in (G), except for the chr21-region containing *HUNK*.

(I) RNA FISH for *SPON1* nascent transcripts and XIST with zoom-in for boxed regions.

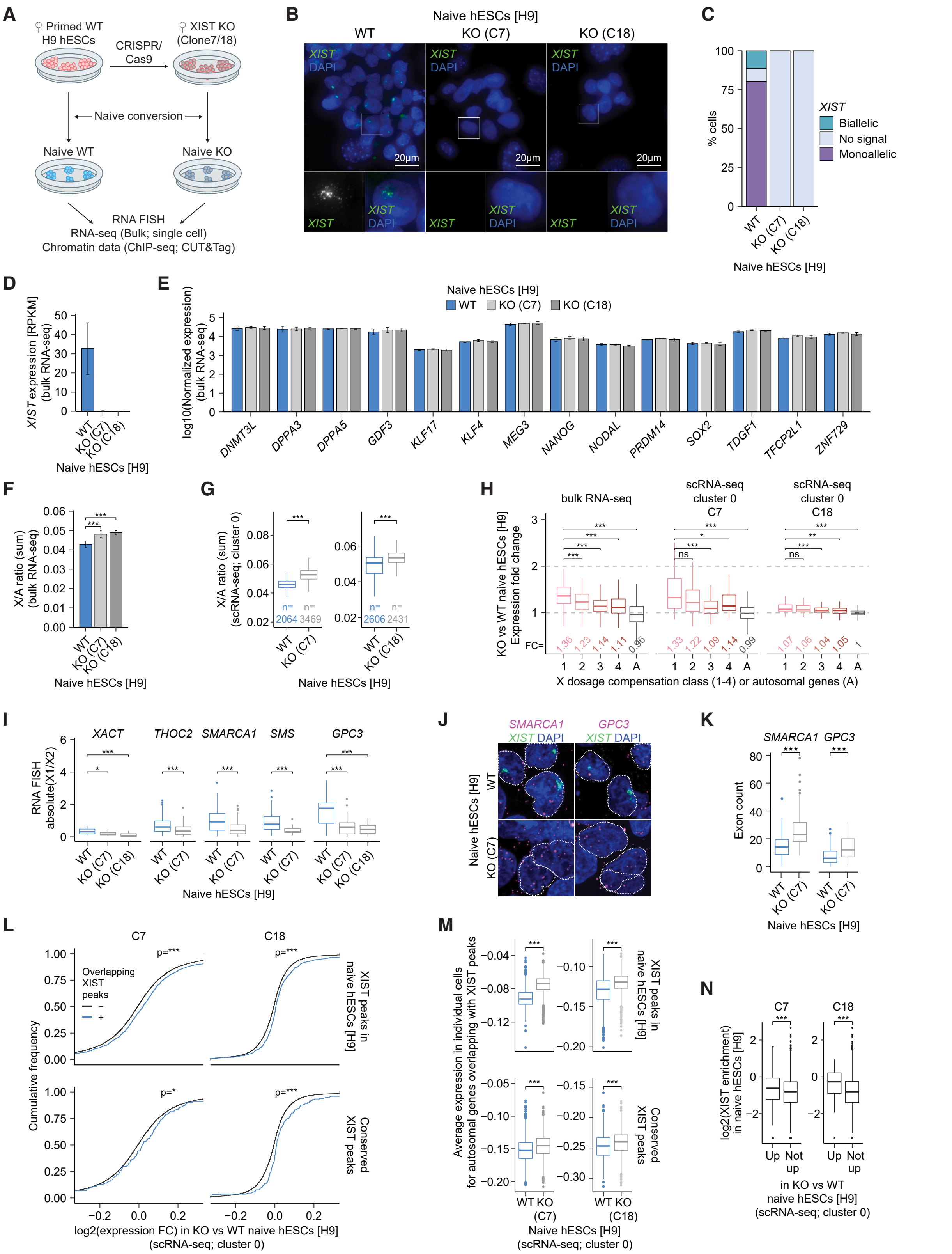
(J) Normalized transcript levels of *HUNK* or *SPON1* vs. XIST levels in indicated naive hPSC lines based on bulk RNA-seq data. Pearson correlation (R), p value and linear regression line (95% confidence interval) are given.

(K) Boxplot of Pearson correlation coefficients between transcript levels of XIST and genes overlapping (+) or not (−) with conserved autosomal XIST peaks. Wilcoxon p value: ***$p < 0.001$.

(L) Normalized expression for *HUNK* and *SPON1* in female and male E7 epiblast cells based on scRNA-seq.[33]

(M) Proportion of reads for genes overlapping with conserved autosomal XIST peaks (normalized by total reads) in female and male E7 epiblast cells based on scRNA-seq.[33] Wilcoxon p value: *$p < 0.05$.

(N) Average XIST enrichment for autosomal genes significantly downregulated or not in female vs. male E7 epiblast cells.[33] Wilcoxon p value: ***$p < 0.001$.
See also Figures S3 and S4.



(legend on next page)

(Figure 5H). Furthermore, RNA FISH for XACT and nascent transcripts of X-linked genes revealed that the difference in signal intensities between the two X chromosomes is significantly reduced in the absence of XIST (Figure 5I). We also detected an increase in the total transcripts for the X-linked genes *GPC3* and *SMARCA1* in individual KO cells by single-molecule RNA FISH (Figures 5J and 5K). Together, these findings uncover XIST as a regulator of XCD.

To explore whether XIST also mediates the regulation of autosomal genes in female cells, we examined scRNA-seq data, focusing on the cells with the highest expression of naive pluripotency markers (cluster 0; Figure S5I; STAR Methods). Deletion of XIST leads to a small, yet significant upregulation of XIST-targeted autosomal genes compared with non-targeted genes (Figures 5L and 5M). Consistent with this result, the proportion of biallelically SPON1-expressing cells is increased in XIST KO cells (Figure S4F), and genes upregulated upon XIST KO are associated with higher XIST enrichment in WT cells (Figure 5N).

Together, these data demonstrate that XIST mediates the dampening of X-linked and autosomal genes, which yields a similar expression of X-linked between the sexes and male-biased expression of autosomal.

SPEN contributes to the dampening of XIST targets in naive hPSCs

We next explored how XIST mediates gene expression dampening. The repressor SPEN is essential for gene silencing by Xist on the Xi.[27,58–62] To investigate the role of SPEN in the regulation of XIST targets in naive hPSCs, we knocked down *SPEN* by RNAi in female WT naive hESCs and, additionally, in male naive hESCs and female XIST KO hESCs, allowing us to distinguish XIST-dependent and -independent SPEN-mediated regulation (Figures S6A–S6C).

SPEN knockdown resulted in the significant upregulation of X-linked genes relative to autosomal genes specifically in female WT cells, without affecting the steady-state level of XIST or naive pluripotency markers (Figures S6D–S6J), indicating that XIST exploits SPEN for XCD. In contrast to the larger effect of XIST deletion on strongly dosage-compensated X-linked genes (class 1) (Figure 5H), X-linked genes are affected similarly by *SPEN* knockdown, regardless of which X-dosage compensation class they belong to (Figure S6G), suggesting that additional regulatory mechanisms are required for strong dosage compensation by XIST.

Autosomal XIST target genes become weakly, yet significantly, upregulated upon SPEN depletion in female WT hPSCs, and autosomal genes significantly upregulated upon SPEN knockdown accumulate more XIST than others (Figures S6K and S6L). However, SPEN also regulates autosomal XIST-target genes in male hESCs and female XIST KO cells (Figures S6K–S6N), indicating that they are dependent on SPEN, regardless of XIST presence. These data suggest that XIST exploits SPEN to mediate gene dampening in naive hPSCs and that SPEN dependency on the X and autosomes differs.

The Xd is less compacted than the Xi

Previous work has shown that XIST-induced chromatin compaction is critical for XCI.[27,63–67] To explore whether XCD is associated with compaction, we determined the compaction state of the XIST-coated Xd relative to the active X chromosome (Xa) in naive hPSCs and the Xi and Xa in somatic cells by applying XIST RNA FISH and X chromosome painting. In naive hESCs, the Xd is slightly more compact than the Xa, and each of these chromosomes is less compacted than the Xi and Xa, respectively, in somatic cells (Figure S6O). DNA FISH with probes targeting specific genomic regions on the X confirmed these results, indicating that they are not due to experimental limitations associated with chromosome painting (Figures S6P and S6Q). Together, these findings suggest that the limited compaction of the Xd may be functionally linked to the occurrence of XCD instead of XCI.

Figure 5. XIST is required for X-linked and autosomal gene dampening in female naive hPSCs

(A) XIST deletion approach: female primed H9 WT hESCs and two derivative XIST KO clones (C7/C18) were converted to the naive state and subsequently analyzed.

(B) XIST RNA FISH in female naive WT and XIST KO hESCs, with marked cells magnified.

(C) Percentage of cells from (B) with no, biallelic or monoallelic XIST accumulation.

(D) XIST transcript level in indicated cell lines based on bulk RNA-seq data. Each bar represents the mean expression from replicate data sets and error bars give the standard deviation (SD).

(E) As in (D), for normalized expression of naive pluripotency markers.

(F) As in (D), for X/A ratios (read count sum). Wilcoxon p value: ***$p < 0.001$. Each bar represents the mean ratio from replicate data sets and error bars give the SD.

(G) X/A ratio (read count sum) for individual cells of indicated cell lines based on scRNA-seq data. Wilcoxon p value: ***$p < 0.001$; n = number of cells. Outliers omitted for clearer visualization. Each KO clone was independently converted and analyzed along-side WT cells.

(H) Expression FC between naive KO and WT cells for X-linked (split by X-dosage compensation classes) and autosomal genes (A), derived from bulk- or scRNA-seq data as marked. Dashed lines mark no difference (y = 1) and double dosage in KO cells (y = 2). Median FC values are given. Wilcoxon p values: ns ≥ 0.05, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(I) Absolute signal intensity difference of nascent transcription foci of indicated genes between the two X chromosomes in naive WT and KO cells. Wilcoxon p values: *$p < 0.05$, ***$p < 0.001$.

(J) Single-molecule RNA FISH for *SMARCA1* or *GPC3* transcripts and XIST in naive WT and KO hESCs.

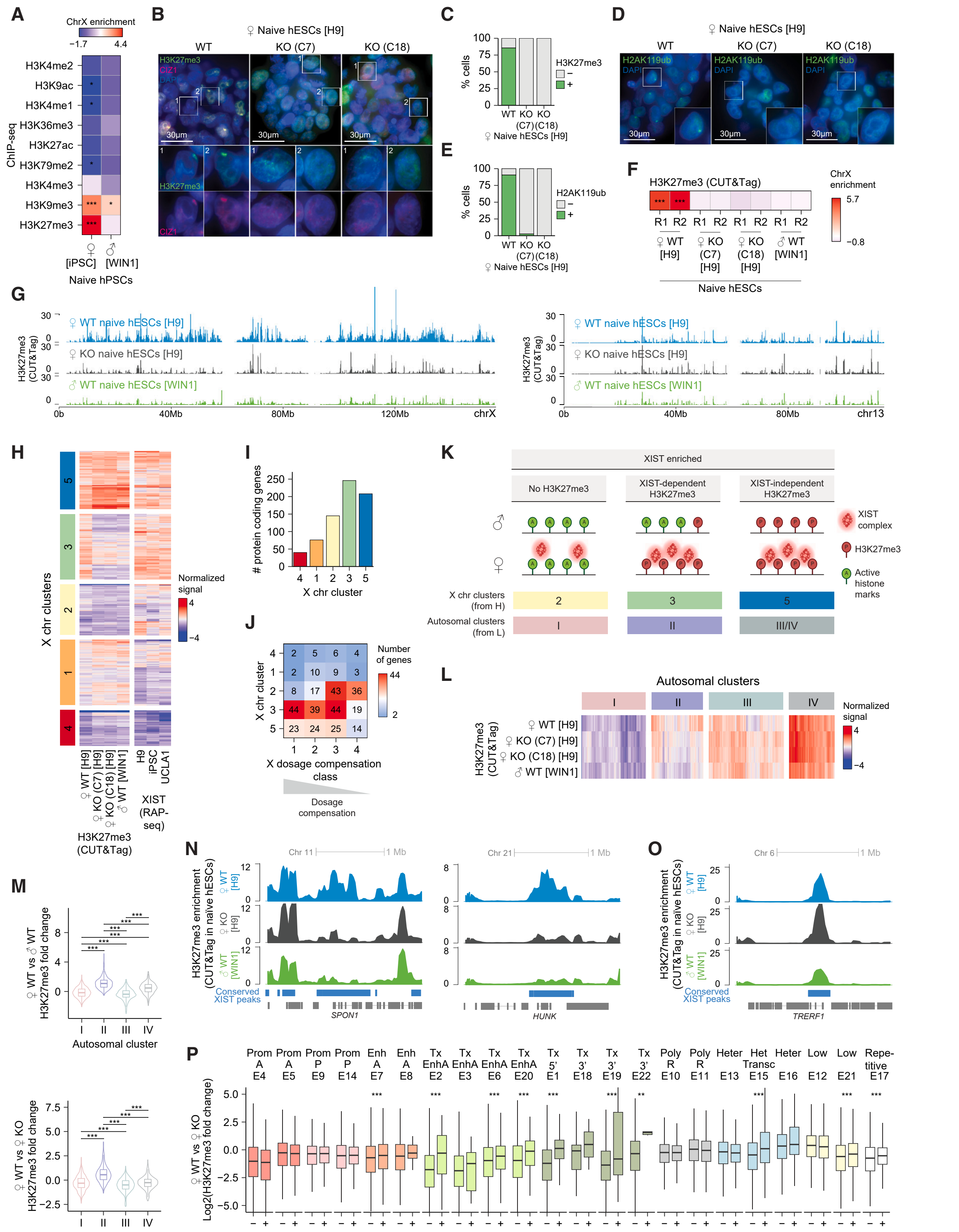
(K) Quantification of total cell transcripts from (J). Wilcoxon p value: ***$p < 0.001$.

(L) Cumulative distribution plot of expression FC between naive KO and WT cells for autosomal genes overlapping (+) or not (−) with H9 hESC-derived XIST peaks (top) or with conserved autosomal XIST peaks (bottom). For clearer visualization, x axis range was limited. Wilcoxon p values indicate that genes under XIST peaks are upregulated in KO cells: *$p < 0.05$, ***$p < 0.001$.

(M) Normalized average expression of autosomal genes overlapping H9 hESC-derived (top) or conserved (bottom) autosomal XIST peaks in individual cells. Wilcoxon p value: ***$p < 0.001$.

(N) XIST enrichment at autosomal genes significantly upregulated or not in naive KO vs. WT hESCs. Wilcoxon p value: ***$p < 0.001$.

See also Figures S5 and S6.



(legend on next page)

XIST mediates chromatin changes on the Xd and autosomes

Given the known role of XIST in altering the chromatin state on the Xi,[68–72] we next asked whether XIST also mediates chromatin changes in naive hPSCs. We generated ChIP-seq data for nine histone modifications in female naive hiPSCs, complementing the male hPSC data (Figure 3J).

On the X chromosome, we found that the repressive H3K27me3 and H3K9me3 modifications are significantly enriched in female compared with male cells (Figure 6A). Immunostaining revealed a strong accumulation of H3K27me3 on the XIST-associated Xd, together with CIZ1, a known Xist-interacting protein,[27,62,73] specifically in WT and not in KO cells (Figures 6B, 6C, and S7A–S7D; STAR Methods), indicating that the accumulation of H3K27me3 on the Xd is XIST dependent. A similar result was obtained for H2AK119ub, another polycomb-mediated histone modification occurring on the Xi[74] (Figures 6D and 6E). Genomic analysis of H3K27me3 by CUT&Tag demonstrated a strong H3K27me3 enrichment across the entire X chromosome in female cells, which is absent in XIST KO and male WT hESCs (Figures 6F and 6G). Together, these data indicate that XIST induces dramatic chromatin changes across the Xd.

To characterize the relationship between chromatin changes, XIST coating, and the degree of X-linked dosage compensation, we stratified genomic regions on the X chromosome into five clusters based on XIST enrichment in naive female WT hPSC lines and H3K27me3 levels in naive female WT and XIST KO hESCs and male hESCs, allowing us to define XIST-dependent H3K27me3 changes (Figure 6H; Table S7), and explored if the clusters are characterized by unique features (Figures 6I, 6J, and S7E–S7N). This analysis revealed that X-linked regions with the highest XIST enrichment either gain H3K27me3 in an XIST-dependent manner (cluster 3; Figure 6H) or are pre-marked with H3K27me3 (cluster 5; Figure 6H) and mostly contain strongly dosage-compensated genes (Figure 6J), indicating a link between XIST-mediated dampening and polycomb-mediated regulation. Consistent with this, PRC2 inhibition in female naive hESCs[75] induces the upregulation of the most dosage-compensated X-linked genes (Figure S7O). Regions with intermediate XIST enrichment and constitutive lack of H3K27me3 (cluster 2, Figure 6H) contain X-linked genes with weaker dosage compensation (Figure 6J), in agreement with their active chromatin state and high expression level (Figures 4C and 4D). Heterochromatic regions of clusters 1 and 4 are minimally targeted by XIST and contain only a few genes (Figures 6H–6J and S7F–S7L). Overall, this analysis uncovers distinct pathways of X-linked gene regulation downstream to XIST binding (Figure 6K).

We also explored whether XIST induces H3K27me3 gains at autosomal target regions. Similar to X chromosome cluster 3, we identified autosomal target regions with an XIST-dependent H3K27me3 accumulation (autosomal cluster II; Figures 6L and 6M), which include the most XIST-enriched target genes *HUNK1* and *SPON1* (Figure 6N). Other regions are either lacking H3K27me3 (autosomal cluster I) or constitutively enriched for H3K27me3 (autosomal clusters III and IV) (Figures 6L, 6M, 6O, and S7P), mirroring clusters 2 and 5, respectively, on the X chromosome (Figure 6H). Exploiting chromatin state annotations (Figure S7F), we found that the XIST-mediated H3K27me3 gain on autosomes occurs predominantly at enhancers and actively transcribed gene bodies (Figure 6P).

Overall, these findings demonstrate that the relationship between XIST binding and H3K27me3 on the X chromosome and autosomes is similar, with a subset of XIST-targeted regions gaining H3K27me3 (Figure 6K), and that XIST induces a repressive chromatin state at active *cis*-regulatory sites and gene bodies on autosomes.

Figure 6. XIST mediates chromatin changes on the X chromosome and at autosomal target regions

(A) ChrX enrichment of histone marks based on ChIP-seq in male and female naive hPSCs. *Z* score p values: *p < 0.05, ***p < 0.001.

(B) H3K27me3 and CIZ1 immunostainings in naive WT and XIST KO hESCs, with magnification of marked cells.

(C) Percentage of cells with H3K27me3 accumulation from (B).

(D) H2AK119ub immunostaining in naive WT and XIST KO hESCs.

(E) Percentage of cells with H2AK119ub accumulation from (D).

(F) As in (A), except for H3K27me3 obtained by CUT&Tag in indicated cell lines (R, replicate).

(G) H3K27me3 enrichment along chrX (left) and chr13 (right; chr13 lacks conserved XIST peaks) based on merged replicate data from (F);additionally, C7 and C18 data were averaged for the XIST KO data.

(H) Heatmap of the normalized H3K27me3 signal for chrX bins (100 kb, every 25 kb) and XIST enrichment from indicated naive hPSC lines, grouped into five clusters.

(I) Number of protein-coding genes in each chrX cluster from (H).

(J) Number of genes from each X-dosage compensation class (Figure 4A) within the chrX clusters from (H).

(K) Scheme summarizing the three H3K27me3 states at XIST-enriched regions on chrX from (H) and autosomes from (L) and correspondence between X and autosome results. Left: no H3K27me3 in male and female cells; middle: XIST-dependent H3K27me3 deposition in female cells; right: XIST-independent H3K27me3 in male and female cells.

(L) Heatmap of the normalized H3K27me3 level from indicated naive hESC lines in autosomal windows overlapping conserved XIST peaks, clustered into four groups.

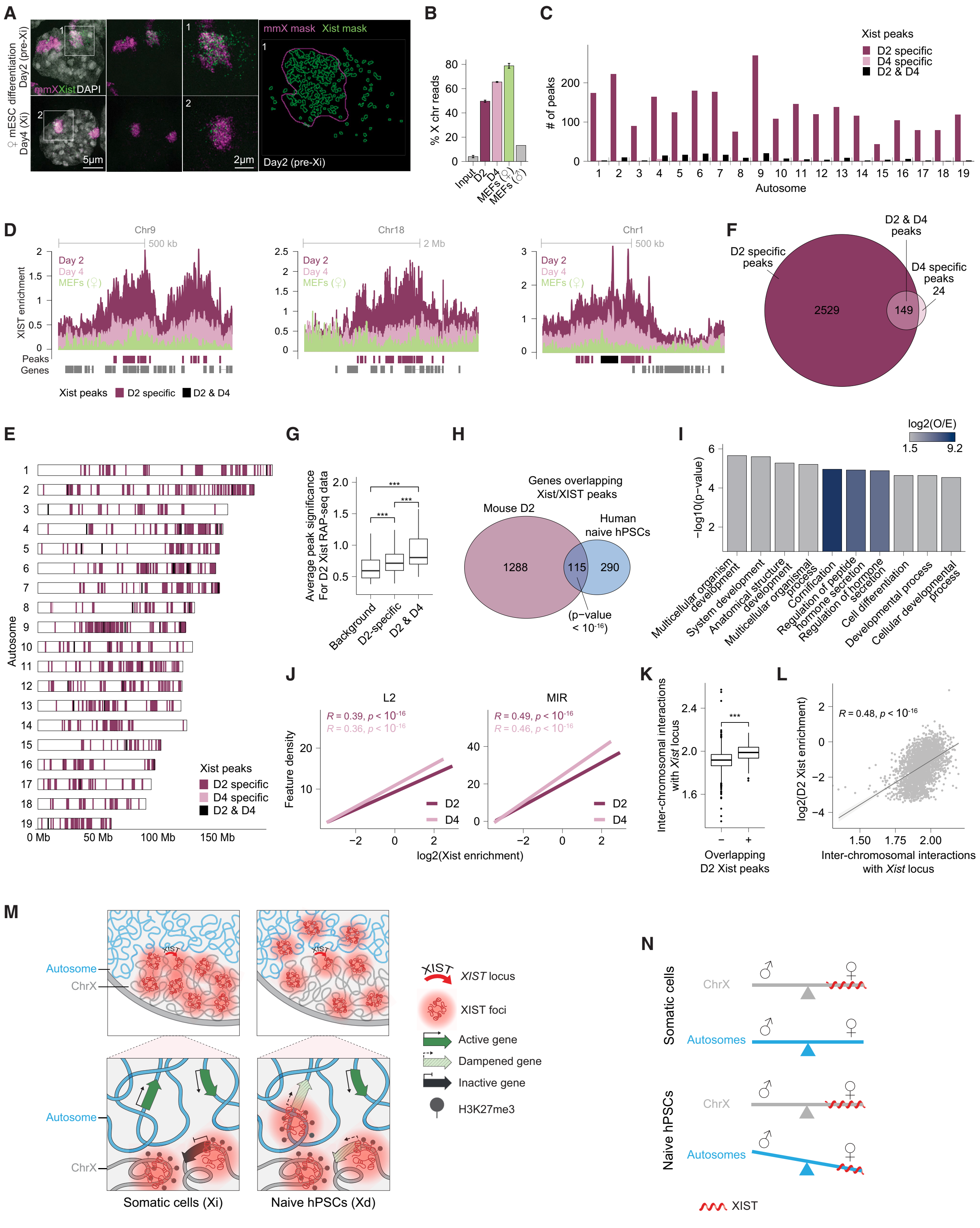
(M) Difference in H3K27me3 level between naive female (H9) and male (WIN1) hESCs (top) or naive female WT and KO H9 hESCs (bottom) for autosomal groups from (L). Wilcoxon p value: ***p < 0.001. Outliers omitted for clearer visualization.

(N) Examples of H3K27me3 enrichment in indicated naive hESC lines across genomic regions containing *SPON1* and *HUNK* loci and overlapping conserved XIST peaks.

(O) As in (N), except for *TRERF1*.

(P) H3K27me3 FC between female naive WT and KO H9 hESCs for autosomal regions overlapping conserved XIST peaks (+) or not (–), split according to chromatin state in male naive hESCs explained in Figure S7F. Wilcoxon p values indicate that XIST targets gain H3K27me3 in WT cells compared with non-targets: *p < 0.05, **p < 0.01, ***p < 0.001.

See also Figure S7.



(legend on next page)

XIST transiently spreads to autosomal regions during the initiation of mouse XCI

To explore whether the dispersed XIST configuration and autosomal spreading occur in other contexts, we investigated Xist's localization during the initiation of XCI in differentiating mouse ESCs (mESCs). We previously showed that Xist coats the X chromosome by day 2 of differentiation; however, gene silencing is limited at day 2 and only completed by day 4.[27] Applying Xist RNA FISH and X chromosome painting, we confirmed that Xist covers the pre-Xi at day 2 and the Xi at day 4 (Figure 7A). Additionally, we found that Xist is in a dispersed configuration and spreads beyond the X-territory at day 2, whereas it is largely restricted to the X at day 4 (Figure 7A). Thus, Xist dispersal happens transiently during XCI initiation when X-linked gene silencing is at most partial.

We next investigated the localization of Xist at day 2 and day 4 by RAP-seq (two replicates per time point). In addition to the enrichment across the X chromosome at both time points (Figure 7B),[27] we found a remarkable accumulation of Xist at specific regions on every autosome on day 2, which is largely lost at day 4 (Figures 7C–7E). We identified 2,529 day 2-specific autosomal Xist peaks, 24 day 4-specific peaks, and 149 day 2 peaks that are maintained at day 4 and display the highest significance scores (Figures 7C–7G; STAR Methods). In female mouse embryonic fibroblasts (MEFs), which represent a later developmental stage with a fully formed Xi, autosomes are even less enriched for Xist than at day 4, and the absence of Xist peaks in male MEFs confirmed the specificity of the RAP-seq approach (Figures 7B, 7D, and S7Q). Together, these data show that Xist transiently spreads to autosomes during the initiation of mouse XCI and that autosomal localization of Xist/XIST is evolutionarily conserved.

Autosomal day 2 Xist peaks encompass 1,646 autosomal genes including 1,403 with known human orthologs, of which a significant number (115) displays XIST binding in naive female hPSCs (Figure 7H; Table S4). Genes with conserved XIST/Xist binding are associated with developmental gene ontology terms (Figure 7I) and include the TFs *ESRRB*, *ARNTL*, *RBPJ*, and *ETV5*. Interestingly, the triple KO of *ETV5, RBPJ,* and *TCF3* traps mESCs in the naive pluripotent state,[46] supporting the idea that autosomal targeting of XIST/Xist plays a role in the regulation of the naive-to-primed pluripotency transition. Autosomal enrichment of Xist in mouse cells best correlates with the presence of MIR and L2 repetitive elements (Figure 7J; Table S5), and autosomal Xist targets have high inter-chromosomal interactions with the *Xist* locus (Figures 7K and 7L), as seen for human XIST in naive hPSCs. Thus, autosomal Xist/XIST localization follows similar guiding principles in early mouse and human development.

Interestingly, autosomal targeting in mouse cells occurs when gene repression on the X is inclomplete,[27] Xist is in a dispersed configuration (Figure 7A), and the pre-Xi is not compacted,[27] which are all features of the Xd in hPSCs. However, XIST/Xist dispersal, autosomal binding of XIST/Xist, limited gene repression, and lack of compaction of the X are stably maintained in naive hPSCs and transient during the initiation of XCI in differentiating mESCs.

DISCUSSION

Confinement to the X chromosome in *cis* is a hallmark of XIST/Xist action. We uncovered the spreading of XIST/Xist to specific autosomal regions adding an unexpected "*trans*" dimension to XIST's actions (Figure 7M). Our work also shows that XIST alters the chromatin state of its autosomal targets and induces their downregulation (Figure 7M). Autosomal targets are in close spatial proximity to the genomic *XIST/Xist* locus, suggesting

Figure 7. Autosomal spreading of Xist occurs during mouse XCI initiation

(A) 3D-structured illumination microscopy (SIM) projections of RNA-FISH staining for Xist and chrX paint (mmX) at days 2 (pre-Xi) and 4 (Xi) of mESC differentiation (left). Middle images show only the Xist and chrX signals and right image a magnification of the Xi with chrX and XIST masks.

(B) Percentage of genomic reads aligned to chrX for input genomic DNA and Xist RAP-seq pull-down in indicated samples. Each bar represents the mean percentage from replicate data sets and error bars the SD. D2 = day 2, D4 = day 4.

(C) Number of Xist peaks per autosome labeled according to sample specificity (D2 and D4 = day 2 peaks maintained at day 4)

(D) Xist enrichment and peaks at three autosomal regions from indicated samples.

(E) Location of Xist peaks from (C).

(F) Number of Xist peaks from (C).

(G) Average significance ($-\log_{10}(q)$) of autosomal Xist peaks for day 2 RAP-seq replicates; classified into background, day 2 specific peaks (D2), and day 2 peaks maintained at day 4 (D2 and D4) (STAR Methods). Wilcoxon p value: ***$p < 0.001$.

(H) Overlap of genes associated with conserved autosomal XIST peaks in naive hPSCs and genes associated with Xist peaks at day 2 of mESC differentiation. Fisher's exact p value is given.

(I) Most highly enriched ontologies and their significance (Fisher's exact test $-\log_{10}(p)$) for the 115 genes associated with autosomal XIST/Xist binding in both human and mouse cells from (H), colored by $\log_2$(observed/expected).

(J) Linear regression line comparing RAP-seq-based Xist enrichment along autosomes (1 Mb every 250 kb) and the density of L2 and MIR elements, for day 2 and day 4. Pearson correlation (R) and p values are given.

(K) Inter-chromosomal interaction scores between the *Xist* locus and 1 Mb autosomal windows overlapping (+) or not (−) with Xist peaks from day 2. Wilcoxon p value: ***$p < 0.001$.

(L) Inter-chromosomal interactions between the *Xist* locus and all 1 Mb autosomal regions vs. day 2 Xist enrichment. Pearson correlation (R) and p value are given.

(M) Summary of XIST localization and function. Top: nuclear area containing chrX and neighboring autosomes in female somatic cells with an Xi (left) and female naive hPSCs with an Xd (right). Bottom left: On the Xi, XIST only spreads in *cis*, silencing gene expression. Autosomal genes are not targeted and regulated by XIST. Bottom right: In naive hPSCs, XIST spreads in *cis* across the Xd and in *trans* to specific autosomal regions and induces gene dampening and chromatin changes at these sites.

(N) Model of the dosage compensation function of XIST. On the Xi/Xd in female somatic cells or naive hPSCs, XIST acts to normalize gene expression dosage between male and female cells. Additionally, XIST disrupts autosomal dosage equality between the sexes in naive hPSCs.

See also Figure S7.

that the autosomal localization is instructed by nuclear organization, similar to the spreading of Xist across the Xi.[16] Although it is well established that XIST/Xist can downregulate autosomal genes in the context of X/A translocations or in *cis* upon transgenic expression from an autosomal site,[14,21,23–26,76] our study reveals that XIST, endogenously expressed from the X chromosome, can downregulate autosomal genes in *trans*.

Our study reveals a coupling between Xist/XIST dispersal and autosomal localization in both naive hPSCs and mouse cells initiating XCI. Epiblast cells of female human pre-implantation embryos display dispersed XIST configuration[32,33] and male-biased expression of autosomal XIST targets (Figures 4L–4N), suggesting that XIST also regulates autosomal genes during early human embryonic development. Intriguingly, the dispersed configuration of XIST/Xist has also been described for cynomolgus monkey pre-implantation embryos,[31] hPGCs,[36] different immune cells including natural killer, dendritic, and T and B cells,[37–39] Alveolar type 2 lung cells,[77] and cancer cells.[40] It is therefore conceivable that autosomal XIST spreading occurs in all these cell types and may have a broad role in development, the regulation of specific cellular functions and disease progression. Intriguingly, in the fruit fly, the X chromosome dosage-compensating roX RNA/MSL complex also targets autosomal genes and regulates their expression, which is required for proper fly development.[78] Further studies are necessary to define the biological consequences of autosomal regulation by XIST/Xist; however, it has been suggested that male embryos develop more rapidly than female embryos.[79–81]

We show that XIST mediates XCD in female naive hPSCs and that X-linked genes have different susceptibility to XIST-mediated dampening. X-linked genes less regulated by XIST display strong female-biased expression and those more strongly regulated have similar transcript levels in female and male naive hPSCs and human pre-implantation embryos. Our data suggest that different regulatory features and chromatin regulation underlie the diverse susceptibility to XIST-mediated dampening. XCD is also observed in female human and non-human primate pre-implantation embryos[31–33] and hPGCs,[36] suggesting that XIST also mediates XCD in these instances.

Our study shows that XIST can generate two functionally different outputs—dampening (XCD) and silencing (XCI)—and identifies multiple features that are similar between them: XIST isoforms (Figure S5A), localization over the entire X, protein effectors (CIZ1 and SPEN), downstream H3K27me3 and H2AK119ub deposition, and the variable degree of X-linked gene regulation. One noticeable difference is that the Xd is less compact than the Xi. The same is true for the pre-Xi at day 2 of mESC differentiation.[27] During mouse XCI initiation, compaction only occurs at day 4, as gene silencing completes and Xist becomes confined to the Xi territory, and requires the action of polycomb complexes and SMCHD1.[27,65–67,82] Interestingly, differentiation is required to achieve targeting of SMCHD1 to and maximal silencing on the Xi.[63] Based on these observations, we suggest that the absence of differentiation-induced compaction on the Xd in naive hPSCs and the day 2 pre-Xi may be responsible for the inability of XIST/Xist to mediate complete gene silencing in these states. Additionally, lower levels of XIST on the Xd may contribute to less efficient gene repression. Unraveling the mechanistic differences between XCD and XCI presents an important goal for future studies.

As the master regulator of XCI and XCD, XIST balances X-linked gene dosage between females and males. Conversely, XIST-mediated repression of autosomal genes leads to their male-biased expression (Figure 7N). Thus, XIST has opposing functions, equalizing X-linked while unbalancing autosomal gene expression. Another intriguing finding is that XIST induces sexually dimorphic H3K27me3 profiles at a subset of its autosomal targets (Figure 6K). The female-specific autosomal H3K27me3 deposition may carry an epigenetic memory and have a lasting impact on development, akin to the role of H3K27me3 in non-canonical imprinting and inter-generational inheritance.[83]

Limitations of the study

One limitation of this study is that the inter- and intra-chromosomal interactions were not extracted from naive hPSCs due to the need for high-resolution contact maps. Further studies are needed to directly explore the role of 3D structure in XIST spreading to autosomes. Additionally, although female 5iLAF-cultured hPSCs model the pre-implantation pluripotent state, a notable difference is that female naive hPSCs express XIST only from one X chromosome, unlike the embryo with two XIST-expression X chromosomes. Nevertheless, the monoallelic expression of XIST in hPSCs allowed us to explore whether the XIST-associated chromosome is specifically regulated. Furthermore, although our work uncovers the XIST-mediated dampening of X-linked genes, we cannot exclude the possibility that X chromosome upregulation[4,84] additionally occurs on the XIST-negative X chromosome.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.11.033.

ACKNOWLEDGMENTS

We thank the Broad Stem Cell Research Center (BSCRC) and TCGB of the Jonsson Comprehensive Cancer Center (JCCC) at UCLA for support. This work was supported by: T.C. (Boehringer Ingelheim PhD and UCLA Dissertation Year Fellowships), S.Y.X.T. (A*STAR National Science Scholarship, BSCRC PhD Fellowship), C.T.C. (NRSA F32 Fellowship GM007185, HHMI Gilliam, and UCLA Whitcome Fellowships), A.S. (NRSA F31 Fellowship GM115122, Mangasar M. Mangasarian Scholarship, UCLA Dissertation Year Fellowship), Y.M. (NIH R03HD095086), C.C. (CIRM Postdoctoral Training Grant, LLR Visiting Fellowship, NIH (P01HL160469)), A.J.C. (UCLA BSCRC Postdoctoral Fellowship), W.D. (CSST and UCLA Whitcome Fellowships), G.L. and F.D. (NIH Director's New Innovator Award, DP2GM149554), A.A. (UCLA Whitcome Fellowship, NRSA F30 Fellowship HD102190), D.L.B. (NIH (R35GM136426) and BSCRC Innovation Award), and K.P. (BSCRC Innovation Award, Iris Cantor-UCLA Women's Health Center Award (CTSI grant number UL1TR000124), the BSCRC, David Geffen School of Medicine and JCCC at UCLA, NIH (R01HD098387), and a HHMI Faculty Scholar grant).

AUTHOR CONTRIBUTIONS

Conceptualization, A.S., I.D., and K.P.; methodology, Y.M. (high-/super-resolution microscopy), A.J.C. (CUT&Tag and naive hPSC cultures), and D.L.B. (smRNA FISH); investigation, T.C. (cell lines generation, imaging, bulk and scRNA-seq, XIST KO, and SPEN KD), S.Y.X.T. (RAP-seq), C.T.C. (RAP-seq and qPCR), A.S. (cell lines generation, imaging, bulk RNA-seq, and RAP-seq), Y.M. (imaging), C.C. (ChIP-seq and ChromHMM modules), A.J.C. (cell lines generation, scRNA-seq), W.D. (CUT&Tag), Y.S. (qPCRs and immunostainings), A.A. (RNA fractionation and qPCR), J.M. (sequential RNA/DNA FISH), and W.X. (smRNA FISH); formal analysis, G.L., F.D. (quantitative FISH), Y.M. (RNA/DNA FISH), and I.D. (all computational analyses); visualization, I.D.; writing, supervision, and project administration, I.D. and K.P.; funding acquisition, K.P.

DECLARATION OF INTERESTS

K.P. is a member of *Cell*'s advisory board.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: December 21, 2020
Revised: April 1, 2023
Accepted: November 28, 2023
Published: January 4, 2024

REFERENCES

1. Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. *12*, 429–442.
2. Avner, P., and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. Nat. Rev. Genet. *2*, 59–67.
3. Brockdorff, N., Bowness, J.S., and Wei, G. (2020). Progress toward understanding chromosome silencing by Xist RNA. Genes Dev. *34*, 733–744.
4. Deng, X., Berletch, J.B., Nguyen, D.K., and Disteche, C.M. (2014). X chromosome regulation: diverse patterns in development, tissues and disease. Nat. Rev. Genet. *15*, 367–378.
5. Gendrel, A.-V., and Heard, E. (2014). Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu. Rev. Cell Dev. Biol. *30*, 561–580.
6. Minkovsky, A., Patel, S., and Plath, K. (2012). Concise review: pluripotency and the transcriptional inactivation of the female mammalian X chromosome. Stem Cells *30*, 48–54.
7. Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. *42*, 733–772.
8. Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. *36*, 233–278.
9. Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.B., Vassilev, I., Galupa, R., Servant, N., et al. (2017). Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. *24*, 226–233.
10. Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*, 329–331.
11. Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.
12. Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. *11*, 156–166.
13. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.
14. Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705.
15. Yang, L., Kirby, J.E., Sunwoo, H., and Lee, J.T. (2016). Female mice lacking Xist RNA show partial dosage compensation and survive to term. Genes Dev. *30*, 1747–1760.
16. Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

17. Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465–469.

18. Brockdorff, N. (2019). Localized accumulation of Xist RNA in X chromosome inactivation. Open Biol. *9*, 190213.

19. Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M., and Brockdorff, N. (1999). Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. Hum. Mol. Genet. *8*, 195–204.

20. Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

21. Popova, B.C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T.B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. Proc. Natl. Acad. Sci. USA *103*, 7706–7711.

22. Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. *30*, 167–174.

23. Kelsey, A.D., Yang, C., Leung, D., Minks, J., Dixon-McDougall, T., Baldry, S.E.L., Bogutz, A.B., Lefebvre, L., and Brown, C.J. (2015). Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. Genome Biol. *16*, 208.

24. Lee, J.T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275–279.

25. Loda, A., Brandsma, J.H., Vassilev, I., Servant, N., Loos, F., Amirnasr, A., Splinter, E., Barillot, E., Poot, R.A., Heard, E., et al. (2017). Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. Nat. Commun. *8*, 690.

26. Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., and Lawrence, J.B. (2002). An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. USA *99*, 8677–8682.

27. Markaki, Y., Gan Chong, J., Wang, Y., Jacobson, E.C., Luong, C., Tan, S.Y.X., Jachowicz, J.W., Strehle, M., Maestrini, D., Banerjee, A.K., et al. (2021). Xist nucleates local protein gradients to propagate silencing across the X chromosome. Cell *184*, 6174–6192.e32.

28. Smeets, D., Markaki, Y., Schmid, V.J., Kraus, F., Tattermusch, A., Cerase, A., Sterr, M., Fiedler, S., Demmerle, J., Popken, J., et al. (2014). Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. Epigenetics Chromatin *7*, 8.

29. Rodermund, L., Coker, H., Oldenkamp, R., Wei, G., Bowness, J., Rajkumar, B., Nesterova, T., Susano Pinto, D.M., Schermelleh, L., and Brockdorff, N. (2021). Time-resolved structured illumination microscopy reveals key principles of Xist RNA spreading. Science *372*, eabe7500.

30. Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

31. Okamoto, I., Nakamura, T., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Nakamura, S.I., Ema, M., Yamamoto, T., and Saitou, M. (2021). The X chromosome dosage compensation program during the development of cynomolgus monkeys. Science *374*, eabd8887.

32. Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

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

34. Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., et al. (2017). Human naive pluripotent stem cells model X chromosome dampening and X inactivation. Cell Stem Cell *20*, 87–101.

35. Vallot, C., Patrat, C., Collier, A.J., Huret, C., Casanova, M., Liyakat Ali, T.M.L., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J., et al. (2017). XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. Cell Stem Cell *20*, 102–111.

36. Chitiashvili, T., Dror, I., Kim, R., Hsu, F.M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K., et al. (2020). Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. Nat. Cell Biol. *22*, 1436–1446.

37. Syrett, C.M., Sindhava, V., Sierra, I., Dubin, A.H., Atchison, M., and Anguera, M.C. (2018). Diversity of epigenetic features of the inactive X-chromosome in NK cells, dendritic cells, and macrophages. Front. Immunol. *9*, 3087.

38. Syrett, C.M., Paneru, B., Sandoval-Heglund, D., Wang, J., Banerjee, S., Sindhava, V., Behrens, E.M., Atchison, M., and Anguera, M.C. (2019). Altered X-chromosome inactivation in T cells may promote sex-biased autoimmune diseases. JCI Insight *4*, e126751.

39. Wang, J., Syrett, C.M., Kramer, M.C., Basu, A., Atchison, M.L., and Anguera, M.C. (2016). Unusual maintenance of X chromosome inactivation predisposes female lymphocytes for increased expression from the inactive X. Proc. Natl. Acad. Sci. USA *113*, E2029–E2038.

40. Chaligné, R., Popova, T., Mendoza-Parra, M.A., Saleem, M.A., Gentien, D., Ban, K., Piolot, T., Leroy, O., Mariani, O., Gronemeyer, H., et al. (2015). The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. Genome Res. *25*, 488–503.

41. Sahakyan, A., and Plath, K. (2016). Transcriptome encyclopedia of early human development. Cell *165*, 777–779.

42. Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.-C., Sanborn, A.L., Machol, I., Shamim, M., et al. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. USA *113*, E4504–E4512.

43. Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J.B., Blau, C.A., Shendure, J., Duan, Z., et al. (2015). Bipartite structure of the inactive mouse X chromosome. Genome Biol. *16*, 152.

44. Patel, S., Bonora, G., Sahakyan, A., Kim, R., Chronis, C., Langerman, J., Fitz-Gibbon, S., Rubbi, L., Skelton, R.J.P., Ardehali, R., et al. (2017). Human embryonic stem cells do not change their X-inactivation status during differentiation. Cell Rep. *18*, 54–67.

45. Currey, L., Thor, S., and Piper, M. (2021). TEAD family transcription factors in development and disease. Development *148*, dev196675.

46. Kalkan, T., Bornelöv, S., Mulas, C., Diamanti, E., Lohoff, T., Ralser, M., Middelkamp, S., Lombard, P., Nichols, J., and Smith, A. (2019). Complementary activity of ETV5, RBPJ, and TCF3 drives formative transition from naive pluripotency. Cell Stem Cell *24*, 785–801.e7.

47. Rostovskaya, M., Stirparo, G.G., and Smith, A. (2019). Capacitation of human naïve pluripotent stem cells for multi-lineage differentiation. Development *146*, dev172916.

48. Collier, A.J., Panula, S.P., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., et al. (2017). Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. Cell Stem Cell *20*, 874–890.e7.

49. Rao, S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al.

(2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

50. Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

51. Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388.

52. Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602–616.

53. Vallot, C., Huret, C., Lesecque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A., Duret, L., and Rougeulle, C. (2013). XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. Nat. Genet. *45*, 239–241.

54. Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. Stem Cell Rep. *6*, 437–446.

55. Gyllborg, D., Ahmed, M., Toledo, E.M., Theofilopoulos, S., Yang, S., ffrench-Constant, C., and Arenas, E. (2018). The matricellular protein R-spondin 2 promotes midbrain dopaminergic neurogenesis and differentiation. Stem Cell Rep. *11*, 651–664.

56. Gardner, H.P., Belka, G.K., Wertheim, G.B.W., Hartman, J.L., Ha, S.I., Gimotty, P.A., Marquis, S.T., and Chodosh, L.A. (2000). Developmental role of the SNF1-related kinase Hunk in pregnancy-induced changes in the mammary gland. Development *127*, 4493–4509.

57. Gjaltema, R.A.F., Schwämmle, T., Kautz, P., Robson, M., Schöpflin, R., Ravid Lustig, L., Brandenburg, L., Dunkel, I., Vechiatto, C., Ntini, E., et al. (2022). Distal and proximal cis-regulatory elements sense X chromosome dosage and developmental state at the Xist locus. Mol. Cell *82*, 190–208.e17.

58. Dossin, F., Pinheiro, I., Żylicz, J.J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y., et al. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. Nature *578*, 455–460.

59. McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

60. Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. Cell Rep. *12*, 554–561.

61. Nesterova, T.B., Wei, G., Coker, H., Pintacuda, G., Bowness, J.S., Zhang, T., Almeida, M., Bloechl, B., Moindrot, B., Carter, E.J., et al. (2019). Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. Nat. Commun. *10*, 3129.

62. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

63. Bowness, J.S., Nesterova, T.B., Wei, G., Rodermund, L., Almeida, M., Coker, H., Carter, E.J., Kadaster, A., and Brockdorff, N. (2022). Xist-mediated silencing requires additive functions of SPEN and Polycomb together with differentiation-dependent recruitment of SmcHD1. Cell Rep. *39*, 110830.

64. Pandya-Jones, A., and Plath, K. (2016). The "lnc" between 3D chromatin structure and X chromosome inactivation. Semin. Cell Dev. Biol. *56*, 35–47.

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

66. Nozawa, R.-S., Nagao, K., Igami, K.-T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nat. Struct. Mol. Biol. *20*, 566–573.

67. Wang, C.-Y., Jégu, T., Chu, H.-P., Oh, H.J., and Lee, J.T. (2018). SMCHD1 merges chromosome compartments and assists formation of super-structures on the inactive X. Cell *174*, 406–421.e25.

68. Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

69. Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H.F.M., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell *4*, 481–495.

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

71. Boeren, J., and Gribnau, J. (2021). Xist-mediated chromatin changes that establish silencing of an entire X chromosome in mammals. Curr. Opin. Cell Biol. *70*, 44–50.

72. Wutz, A. (2007). Xist function: bridging chromatin and stem cells. Trends Genet. *23*, 457–464.

73. Ridings-Figueroa, R., Stewart, E.R., Nesterova, T.B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., et al. (2017). The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. Genes Dev. *31*, 876–888.

74. Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664–678.

75. Kumar, B., Navarro, C., Winblad, N., Schell, J.P., Zhao, C., Weltner, J., Baqué-Vidal, L., Salazar Mantero, A., Petropoulos, S., Lanner, F., et al. (2022). Polycomb repressive complex 2 shields naïve human pluripotent cells from trophectoderm differentiation. Nat. Cell Biol. *24*, 845–857.

76. Russell, L.B. (1963). Mammalian X-chromosome action: inactivation limited in spread and in region of origin. Science *140*, 976–978.

77. Sierra, I., Pyfrom, S., Weiner, A., Zhao, G., Driscoll, A., Yu, X., Gregory, B.D., Vaughan, A.E., and Anguera, M.C. (2023). Unusual X chromosome inactivation maintenance in female alveolar type 2 cells is correlated with increased numbers of X-linked escape genes and sex-biased gene expression. Stem Cell Rep. *18*, 489–502.

78. Valsecchi, C.I.K., Basilicata, M.F., Semplicio, G., Georgiev, P., Gutierrez, N.M., and Akhtar, A. (2018). Facultative dosage compensation of developmental genes on autosomes in Drosophila and mouse embryonic stem cells. Nat. Commun. *9*, 3626.

79. Gardner, D.K., Larman, M.G., and Thouas, G.A. (2010). Sex-related physiology of the preimplantation embryo. Mol. Hum. Reprod. *16*, 539–547.

80. Kawase, Y., Tachibe, T., Kamada, N., Jishage, K.-I., Watanabe, H., and Suzuki, H. (2021). Male advantage observed for in vitro fertilization mouse embryos exhibiting early cleavage. Reprod. Med. Biol. *20*, 83–87.

81. Ray, P.F., Conaghan, J., Winston, R.M., and Handyside, A.H. (1995). Increased number of cells and metabolic activity in male human preimplantation embryos following in vitro fertilization. J. Reprod. Fertil. *104*, 165–171.

82. Blewitt, M.E., Gendrel, A.-V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat. Genet. *40*, 663–669.

83. Inoue, A. (2023). Noncanonical imprinting: intergenerational epigenetic inheritance mediated by Polycomb complexes. Curr. Opin. Genet. Dev. *78*, 102015.

84. Larsson, A.J.M., Coucoravas, C., Sandberg, R., and Reinius, B. (2019). X-chromosome upregulation is driven by increased burst frequency. Nat. Struct. Mol. Biol. *26*, 963–969.

85. Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., et al. (2017). Landscape of X chromosome inactivation across human tissues. Nature *550*, 244–248.

86. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. Cell *174*, 744–757.e24.

87. Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., and Clark, A.T. (2012). Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. Hum. Mol. Genet. *21*, 751–764.

88. Panning, B., Dausman, J., and Jaenisch, R. (1997). X chromosome inactivation is mediated by Xist RNA stabilization. Cell *90*, 907–916.

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

90. R Core Team (2020). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

91. Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2012). FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

92. Krueger, F., James, F., Ewels, P., Afyounian, E., and Schuster-Boeckler, B. (2021). FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo. Version 0.6.7 (Zenodo).

93. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915.

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

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

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

97. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

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

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

100. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

101. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

102. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

103. McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186.

104. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007.

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

106. Picard Tools - By Broad Institute http://broadinstitute.github.io/picard/.

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

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

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

110. Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics *32*, 289–291.

111. Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics *32*, 587–589.

112. Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *22*, 1600–1607.

113. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

114. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101.

115. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

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

117. Akalin, A., Franke, V., Vlahoviček, K., Mason, C.E., and Schübeler, D. (2015). genomation: a toolkit to summarize, annotate and visualize genomic intervals. Bioinformatics *31*, 1127–1129.

118. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

119. Hayashi, K., and Saitou, M. (2013). Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. Nat. Protoc. *8*, 1513–1524.

120. Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896–1908.

121. Yeom, K.-H., Pan, Z., Lin, C.-H., Lim, H.Y., Xiao, W., Xing, Y., and Black, D.L. (2021). Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring. Genome Res. *31*, 1106–1119.

122. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of lincRNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell *44*, 667–678.

123. An, C., Feng, G., Zhang, J., Cao, S., Wang, Y., Wang, N., Lu, F., Zhou, Q., and Wang, H. (2020). Overcoming autocrine FGF signaling-induced heterogeneity in naive human ESCs enables modeling of random X chromosome inactivation. Cell Stem Cell *27*, 482–497.e4.

124. Teklenburg, G., Weimar, C.H.E., Fauser, B.C.J.M., Macklon, N., Geijsen, N., Heijnen, C.J., Chuva de Sousa Lopes, S.M.C.de S., and Kuijk, E.W. (2012). Cell lineage specific distribution of H3K27 trimethylation accumulation in an in vitro model for human implantation. PLoS One *7*, e32701.

125. Zijlmans, D.W., Talon, I., Verhelst, S., Bendall, A., Van Nerum, K., Javali, A., Malcolm, A.A., van Knippenberg, S.S.F.A., Biggins, L., To, S.K., et al. (2022). Integrated multi-omics reveal polycomb repressive complex 2 restricts human trophoblast induction. Nat. Cell Biol. *24*, 858–871.

126. Kupai, A., Vaughan, R.M., Dickson, B.M., and Rothbart, S.B. (2020). A degenerate peptide library approach to reveal sequence determinants of methyllysine-driven protein interactions. Front. Cell Dev. Biol. *8*, 241.

127. Solovei, I. (2010). Fluorescence in situ hybridization (FISH) on tissue cryosections. Methods Mol. Biol. *659*, 71–82.

128. Ding, F., and Elowitz, M.B. (2019). Constitutive splicing and economies of scale in gene expression. Nat. Struct. Mol. Biol. *26*, 424–432.

129. Gaspar, I., Wippich, F., and Ephrussi, A. (2017). Enzymatic production of single-molecule FISH and RNA capture probes. RNA *23*, 1582–1591.

130. Xiao, W., Yeom, K.-H., Lin, C.-H., and Black, D.L. (2023). Improved enzymatic labeling of fluorescent in situ hybridization probes applied to the visualization of retained introns in cells. RNA *29*, 1274–1287.

131. Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of transcripts in 3D FISH images. Nat. Methods *10*, 277–278.

132. Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., et al. (2017). Epigenetic resetting of human pluripotency. Development *144*, 2748–2763.

133. Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269.

134. Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C.T., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. Science *353*, 598–602.

135. Thévenaz, P., Ruttimann, U.E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. IEEE Trans. Image Process. *7*, 27–41.

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

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

138. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

139. TxDb. Hsapiens.UCSC.hg38.knownGene Bioconductor. http://bioconductor.org/packages/TxDb.Hsapiens.UCSC.hg38.knownGene/.

140. Messmer, T., von Meyenn, F., Savino, A., Santos, F., Mohammed, H., Lun, A.T.L., Marioni, J.C., and Reik, W. (2019). Transcriptional heterogeneity in naive and primed human pluripotent stem cells at single-cell resolution. Cell Rep. *26*, 815–824.e4.

141. Crowell, H.L., Soneson, C., Germain, P.-L., Calini, D., Collin, L., Raposo, C., Malhotra, D., and Robinson, M.D. (2020). muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. Nat. Commun. *11*, 6077.

142. Stirparo, G.G., Boroviak, T., Guo, G., Nichols, J., Smith, A., and Bertone, P. (2018). Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human pre-implantation epiblast. Development *145*, dev158501.

143. Bonora, G., Deng, X., Fang, H., Ramani, V., Qiu, R., Berletch, J.B., Filippova, G.N., Duan, Z., Shendure, J., Noble, W.S., et al. (2018). Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. Nat. Commun. *9*, 1445.

144. Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., et al. (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575–579.

145. Horakova, A.H., Moseley, S.C., McLaughlin, C.R., Tremblay, D.C., and Chadwick, B.P. (2012). The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. Hum. Mol. Genet. *21*, 4367–4377.

146. Chu, X., and Wang, J. (2020). Microscopic chromosomal structural and dynamical origin of cell differentiation and reprogramming. Adv. Sci. (Weinh) *7*, 2001572.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-CIZ1	Novus Biologicals	Novus Cat# NB100-74624; RRID:AB_1048573
anti-H2AK119ub	Cell Signaling	Cell Signaling Technology Cat# 8240T; RRID: AB_10891618
anti-H3K27me3	Cell Signaling	Cell Signaling Technology Cat# 9733T; RRID: AB_2616029
anti-H3K27me3	Millipore	Millipore Cat# 07-449; RRID: AB_310624
anti-H3K27me3	Active Motif	Active Motif Cat# 39155; RRID: AB_2561020
anti-H3K4me3	Abcam	Abcam Cat# ab8580; RRID: AB_306649
anti-H3K4me2	Abcam	Abcam Cat# ab7766; RRID: AB_2560996
anti-H3K4me1	Abcam	Abcam Cat# ab8895; RRID: AB_306847
anti-H3K27ac	Abcam	Abcam Cat# ab4729; RRID: AB_2118291
anti-H3K36me3	Abcam	Abcam Cat# ab9050; RRID: AB_306966
anti-H3K79me2	Active Motif	Active Motif Cat# 39143; RRID: AB_2561018
anti-H3K9me3	Abcam	Abcam Cat# ab8898; RRID: AB_306848
anti-H3K9ac	Abcam	Abcam Cat# ab4441; RRID: AB_2118292
Chemicals, Peptides, and Recombinant Proteins		
Recombinant Human FGF basic Protein	PeproTech	100-18B
Recombinant Activin A	Cell Guidance Systems	GFH6
Recombinant human LIF	Cell Guidance Systems	GFH200
PD0325901	Cell Guidance Systems	SM26
SB431542	Cell Guidance Systems	SM33
WH-4-023	Cell Guidance Systems	SM16
Y27632	Cell Guidance Systems	SM02
IM-12	Cell Guidance Systems	SM04
Fastdigest BbsI	Thermo Scientific	FD1014
T4 PNK	New England Biolabs	M0201S
T4 DNA ligase	New England Biolabs	M0202S
DNAse I	Sigma-Aldrich	4716728001
DNA polymerase I	Thermo Fischer	EP0042
Critical Commercial Assays		
CytoTune-iPS 2.0 Sendai Reprogramming Kit	Thermo Fisher	A16517
Human X chromosome and chromosome 11 Paints	MetaSystems Probes	D-0323-100-OR D-1420-050-OR D-0311-050-OR
TruSeq Stranded mRNA Library Prep	Illumina	20020594
RNAeasy	Qiagen	74104
KAPA HiFi hot start Taq	Kapa biosystems	Kk2502
Quick-DNA Miniprep Plus Kit	Zymo Research	D4069
Vectashield	Vector Labs	Cat #: H-1000
PlasmidSafe exonuclease kit	Fischer Scientific	Cat# NC9046399
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE241444
RNA-seq for UCLA1 hESCs	Sahakyan et al.[34]	GEO: GSE87239

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Expression fold change upon EZH2 inhibition in hESCs	Kumar et al.[75]	Kumar et al.[75]
UCLA1 hESC single-nucleotide polymorphisms	Sahakyan et al.[34]	GEO: GSE87239
Single-cell RNA-seq of human preimplantation embryos	Petropoulos et al.[33]	ArrayExpress: E-MTAB-3929
RNA-seq during capacitation of naïve hPSCs	Rostovskaya et al.[47]	GEO: GSE123055
XCI status	Tukiainen et al.[85]	Tukiainen et al.[85]
Mouse RAP-seq	Markaki et al.[27]	GEO: GSE181236
GM12878 Hi-C	Rao et al.[49]	GEO: GSE63525
H9 hESCs Hi-C	Zhang et al.[51]	GEO: GSE116862
Mouse ESC SPRITE	Quinodoz et al.[86]	GEO: GSE114242
Experimental Models: Cell Lines		
Normal Human Dermal Fibroblasts	Lonza	CC2511
UCLA1 hESCs	Diaz Perez et al.[87]	N/A
H9 hESCs	WiCell Research Institute	N/A
KIST KO H9 hESCs (Clone7)	This paper	N/A
KIST KO H9 hESCs (Clone18)	This paper	N/A
Human iPSC	This paper	N/A
WIN1 hESCs (naïve)	Whitehead Institute; Theunissen et al.[50]	N/A
HNES1 hESCs (naïve)	University of Cambridge; Guo et al.[54]	N/A
HNES3 hESCs (naïve)	University of Cambridge; Guo et al.[54]	N/A
UCLA9	Patel et al.[44]	N/A
129S4/SvJae/ castaneus (F1 2-1) mouse ESCs	Panning et al.[88]	N/A
Female MEFs isolated from 129SV/Jae mice	This paper	N/A
Male MEFs isolated from tetO Xist, R26 M2rTTA MEFs (used without dox induction)	This paper	N/A
Oligonucleotides		
5' Biotin-labeled single-standed DNA probes capturing human or mouse XIST/Xist RNA	Eurofins	Table S1
gRNA for generating XIST KO hESCs	IDT	Table S1
Primers for PCR validation of the XIST KO allele	IDT	Table S1
Primers for human XIST RAP-quantitative PCR	Invitrogen	Table S1
Oligo probes for DNA FISH	Biosciences	Table S1
Oligo probes for RNA smFISH	IDT	Table S1
Human si*SPEN* ID s22831	Thermo Fisher	4427037
Human si*SPEN* ID s22829	Thermo Fisher	4427037
Control siRNA	Thermo Fisher	AM4611
Recombinant DNA		
pSpCas9(BB)-2A-Puro (PX459) plasmid	Ran et al.[89]	Addgene plasmid # 48139
Human XIST BAC	Bacpac resource center	RP11-13M9
Human HUWE1 BAC	Bacpac resource center	RP11-975N19
Human XACT BAC	Bacpac resource center	RP11-35D3
Human GPC3 BAC	Bacpac resource center	RP11-678F20
Human THOC2 BAC	Bacpac resource center	RP11-121P4
Human UTX BAC	Bacpac resource center	RP11-256P2
Human SMS BAC	Bacpac resource center	RP11-147O5
Human SMARCA1 BAC	Bacpac resource center	RP11-137A15
Human SPON1 BAC	Bacpac resource center	RP11-774G22
Human PAGE1 BAC	Bacpac resource center	RP11-315L18

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human CDKL5 BAC	Bacpac resource center	RP11-106N3
Human LAMP2 BAC	Bacpac resource center	RP11-158I12
Software and Algorithms		
R v3.6.3	https://cran.r-project.org/[90]	https://cran.r-project.org/
FastQC v0.11.4	Babraham Institute, Bioinformatics[91]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
TrimGalore v0.4.1	https://doi.org/10.5281/zenodo.5127899[92]	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ https://github.com/FelixKrueger/TrimGalore
HISAT2 v2.2.0	Kim et al.[93]	https://daehwankimlab.github.io/hisat2/
STAR v2.5.3a	Dobin et al.[94]	https://github.com/alexdobin/STAR
deepTools v3.5.1	Ramírez et al.[95]	https://deeptools.readthedocs.io/en/develop/content/about.html
Samtools v1.9	Li et al.[96]	http://samtools.sourceforge.net/
HTSeq v0.6.1p1	Anders et al.[97]	https://htseq.readthedocs.io/en/release_0.11.1/count.html
DESeq2 v1.24.0	Love et al.[98]	http://bioconductor.org/packages/release/bioc/html/DESeq2.html
edgeR v3.26.8	Robinson et al.[99]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Cell Ranger v3.0.2	Zheng et al.[100]	https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome
Seurat v3.1.5	Butler et al.[101] and Stuart et al.[102]	https://satijalab.org/seurat/
scater v1.12.2	McCarthy et al.[103]	https://bioconductor.org/packages/scater/
CrossMap v0.2.9	Zhao et al.[104]	http://crossmap.sourceforge.net/
Bowtie2 v2.2.9	Langmead and Salzberg[105]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Picard v2.1.0	Picard Tools – By Broad Institute[106]	http://broadinstitute.github.io/picard/ https://github.com/broadinstitute/picard
bedtools v2.26.0	Quinlan and Hall[107]	http://bedtools.readthedocs.io/en/latest/
MACS2 v2.2.7.1	Zhang et al.[108]	https://pypi.org/project/MACS2/
DiffBind v2.12.0	Ross-Innes et al.[109]	https://bioconductor.org/packages/release/bioc/html/DiffBind.html
regioneR v1.14.0	Gel et al.[110]	https://bioconductor.org/packages/release/bioc/html/regioneR.html
LOLA v1.14.0	Sheffield and Bock[111]	http://code.databio.org/LOLA
topGO v2.36.0	Alexa et al.[112]	https://bioconductor.org/packages/release/bioc/html/topGO.html
Fiji v2.0.0	Schindelin et al.[113]	https://imagej.net/Fiji
Juicebox v1.5.3	Durand et al.[114]	https://github.com/aidenlab/Juicebox
ggplot2 v3.3.0	https://ggplot2.tidyverse.org/[115]	https://ggplot2.tidyverse.org/
ChromHMM v1.18	Ernst and Kellis[116]	http://compbio.mit.edu/ChromHMM
Genomation v1.16.0	Akalin et al.[117]	https://github.com/BIMSBbioinfo/genomation
ComplexHeatmap v2.11.1	Gu et al.[118]	https://jokergoo.github.io/ComplexHeatmap/

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. Kathrin Plath (kplath@mednet.ucla.edu).

Materials availability

All unique materials generated in this study will be available to researchers from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- All genomics data (RAP-seq, bulk RNA-seq, single cell RNA-seq, CUT&Tag and ChIP-seq) generated in this study have been deposited in the Gene Expression Omnibus (GEO) database GEO: GSE241444 and is publicly available as of the date of publication. Accession numbers of reanalyzed publicly available data are listed in the key resources table as well as Table S2.
- This paper does not report original code. All computational approaches are described in the STAR Methods and software/packages used listed in the key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human cell lines and culturing conditions

hESC/iPSC lines used in this study include UCLA1 (46, XX), iPSC (46, XX), H9 (46, XX), WIN1 (46, XY), HNES1 (46, XY), HNES3 (46, XX) and UCLA9 (46, XX). The primed UCLA1, UCLA9 and H9 hESCs were obtained from the Human Embryonic and Induced Pluripotent Stem Cell Core of the BSCRC at UCLA. The XIST KO was generated in primed H9 hESCs as described in detail below. The human iPSC line was generated by reprogramming of female human fibroblasts (NHDFs-Lonza, lot #472033) and maintained in primed or naïve pluripotent culture conditions (see below). Male naïve WIN1 hESC, derived directly from the human embryo,[50] were kindly provided by the Jaenisch lab. Naïve male HNES1 and female HNES3 hESC lines, derived directly from human blastocysts, were kindly provided by Austin Smith.[54] All hESCs and iPSCs studies received approval from the UCLA Embryonic Stem Cell Research Oversight (ESCRO) Committee (protocols 2008-015 and 2007-009). Mycoplasma test (Lonza, LT07-418) was performed routinely for all cell lines to confirm lack of these pathogens. All cell lines used in this study are not on the list of commonly misidentified cell lines (international cell line authentication committee).

Female naïve and primed hiPSCs were generated by reprogramming of female NHDFs using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher A16517) following manufacturer's instructions for feeder-dependent reprogramming. Briefly, 100,000 cells were plated into a well of a 6-well plate and 24 hours later subjected to an overnight transduction with the Sendai virus. Fibroblasts culture medium was renewed daily for one week. Seven days post transduction, cells were dissociated by TrypLE Express Enzyme (Thermo Fisher) and plated on a monolayer of inactivated mouse embryonic fibroblasts (MEFs) in fibroblast medium. After one day, the medium was changed to either naïve 5iLAF hESC medium or primed hESC medium. We confirmed that the female naïve and primed hiPSC lines used in this study express XIST.

Primed hPSCs (UCLA1, UCLA9, H9 and hiPSCs) were cultured on irradiated MEFs in hESC medium composed of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 0.1mM β-Mercaptoethanol (GIBCO, 21985-023), 10ng/mL recombinant human FGF basic (R&D systems, 233-FB), 1x Penicillin-Streptomycin (GIBCO, 15140-122) in DMEM/F12 media (Sigma, D8437). Primed hESCs and hiPSCs were split every 5-6 days using Collagenase type IV (GIBCO, 17104-019).

To convert primed hPSC lines (UCLA1, UCLA9 and WT and XIST KO H9) to the naïve state,[34,50] primed hPSCs were dissociated to single cells using 0.05% trypsin for 5 minutes at 37°C (Gibco 25300054), trypsin activity was quenched with fibroblasts medium (10% FBS (Life Technologies, 10099141), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 0.1mM β-Mercaptoethanol (GIBCO, 21985-023) in DMEM (Sigma, D6429)) and cells were centrifuged for 5 minutes at 500g. The cell pellet was resuspended in DMEM/F12 medium supplemented with 5% KnockOut Serum Replacement (KSR), 15% heat inactivated FBS, penicillin/streptomycin, nonessential amino acids, GlutaMAX, 0.1mM β-Mercaptoethanol, and 10ng/ml FGF2 and supplemented with 10uM ROCK inhibitor Y-27632. Cells were counted and $2x10^5$ cells were plated on irradiated MEFs. After two days, the culture medium was switched to 5iLAF naïve hESC medium compsed of 1:1 mix of DMEM/F12 and Neurobasal (Life Technologies) supplemented with 1x N2 (Life Technologies), 1x B27 (Life Technologies), 1x Penicillin-Streptomycin, 1x NEAA, 0.5x GlutaMAX, 0.5% KSR, 0.1mM β-Mercaptoethanol, 50μg/ml bovine serum albumin (Sigma), 20ng/ml rhLIF (Cell Guidance Systems), 20ng/ml Activin A (Cell Guidance Systems), 8ng/ml FGF2 (PeproTech), 1μM MEK inhibitor PD0325901 (Cell Guidance Systems), 0.5μM B-Raf inhibitor SB590885 (Cell Guidance Systems), 1μM GSK3β inhibitor IM-12 (Cell Guidance Systems), 1μM Src inhibitor WH-4-023 (Cell Guidance Systems), and 10μM ROCK inhibitor Y-27632 (Cell Guidance Systems).

The naïve hESC lines UCLA1, UCLA9, WT and XIST KO H9, hiPSCs and WIN1 were maintained by passaging every 5-6 days with StemPro Accutase and re-plated after passing through a 40 micron cell strainer in 5iLAF medium. Cells were cultured at 5% CO2 and 37°C, either at 5% O2 (hypoxia) or at atmospheric oxygen levels (normoxia) as indicated (Table S2).

Naïve HNES1 and HNES3 hESC lines were maintained on irradiated MEFs in t2iLGö culture medium consisting of N2B27 medium supplemented with 1 μM PD0325901 (Stemgent or Bio-Techne), 10 ng/ml human LIF (EMD Millipore), 2 μM Gö6983 (Tocris Bio-Techne, 2285), and 2 μM XAV939 (Tocris Bio-Techne, 3748). 10μM ROCK inhibitor (Y-27632, Millipore) was added for 24 hours after passaging cells with StemPro Accutase. Naïve HNES1 and HNES3 hESCs were cultured at 5% CO2 and 5% O2 (hypoxia) at 37°C.

Normal human dermal fibroblasts (NHDFs) (46, XX) (Lonza, lot #472033) were cultured in fibroblast media composed of 10% FBS (Life Technologies, 10099141), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 0.1mM β-Mercaptoethanol (GIBCO, 21985-023) in DMEM (Sigma, D6429).

For ChIP-seq and CUT&Tag experiments, naïve hPSCs were cultured in 5iLAF medium as described above. Prior to the experiments, cells were dissociated with StemPro Accutase and the cell pellet was resuspended in fresh 5iLAF medium. Next, a feeder depletion was performed before the cells were utilized for genomics analyses. To this end, we allowed feeder cells to attach to cell-culture plates for 5 minutes at 37°C; this step was repeated twice and only the non-attached hPSC fraction was collected for the downstream genomics experiments.

Mouse cell lines and culturing conditions

Mouse epiblast-like cells (EpiLCs) were differentiated from female mouse polymorphic *129S4/SvJae/castaneus* F1 2-1 mESCs,[88] which were cultured on 0.5% gelatin-coated plates seeded with irradiated DR4 MEFs, obtained from day 14.5 mouse embryos. mESCs were maintained in mESC medium containing Knockout-DMEM (Life Technologies), 15% FBS (Omega), 2mM L-glutamine (Life Technologies), 1x NEAA (Life Technologies), 0.1mM β-Mercaptoethanol (Sigma), 1x Penicillin/Streptomycin (Life Technologies) and 1000U/mL mouse LIF (homemade) in 5% CO2, 37°C incubators. For EpiLC differentiation,[119] mESCs were adjusted to feeder-free conditions on 0.5% gelatin-coated plates in the presence of 1000U/mL LIF and the two inhibitors CHIR99021 (3 μM) and PD0325901 (0.4 μM) (henceforth referred to as 2i+LIF) in serum-free N2B27 medium containing 1x N2 supplement and 1x B27 supplement (Thermo Fisher), 2mM L-glutamine (Life Technologies), 1x NEAA (Life Technologies), 0.1mM β-Mercaptoethanol (Sigma) and 0.5 x Penicillin/Streptomycin (Life Technologies), for at least 3 passages. To induce differentiation of 2i/LIF-cultured mESCs, cells were dissociated with accutase and seeded at a density of 2×10^5 cells/mL in N2B27 media supplemented with 20 ng/mL Activin A and 12 ng/mL bFGF on geltrex-coated plates. Medium was exchanged daily for mESC and EpiLC cultures. As for all our human cell lines, mycoplasma test (Lonza, LT07-418) was performed routinely for all mouse cell lines to confirm lack of these pathogens. All cell lines used in this study are not on the list of commonly misidentified cell lines (international cell line authentication committee).

Female and male MEFs were cultured on non-gelatin-coated plates under feeder-free conditions in DMEM (Life Technologies), 15% FBS (Omega), 2mM L-glutamine (Life Technologies), 1x NEAA (Life Technologies), 0.1mM β-Mercaptoethanol (Sigma) and 1x Penicillin/Streptomycin (Life Technologies) in 5% CO2, 37°C incubators. Fresh medium was exchanged every three days.

METHOD DETAILS

Generation of XIST KO H9 hESC lines

To define how to best engineer the loss-of-function allele for XIST, we first evaluated the nature of XIST transcripts in naïve hPSCs. We found that the transcriptional start site, 3' end, and exon/intron boundaries are the same in naïve hPSCs as in somatic cells (Figure S5A). CRISPR-Cas9-based genome editing was therefore applied to homozygously excise the first ~2kb of XIST, a region known to be critical for the control of the expression and function of Xist in mouse XCI,[22,57] in the female primed hESC line H9. We have previously shown that the primed H9 hESC line used here carries an eroded Xi without XIST expression.[44] Cells were electroporated with two PX459 plasmids (Addgene #48139) each carrying a different gRNA, one targeting XIST promoter (gRNA1 chrX:73852766-73852785; hg38) and one targeting XIST exon 1 (gRNA2 chrX:73850789-73850808; hg38). Successful editing by both gRNAs generates the ~2kb deletion (Figure S5B). gRNA sequences are given in Table S1. gRNA oligos were ordered from IDT and cloned into PX459.[89] To this end, forward and reverse strand gRNA oligos (100uM) were phosphorylated and annealed using T4 PNK (NEB, cat. no. M0201S) in a thermocycler using following parameters: 37°C for 30 min; 95°C for 5 min; ramp down to 25°C at 5°C min^{-1}. Annealed oligos were diluted 1:200 and cloned into PX459 plasmid in one cut-ligation process, using Fastdigest BbsI (Thermo Scientific, cat. no. FD1014) enzyme and T4 DNA ligase (NEB M0202S). The reaction was done in total of 1 h consisting of 37°C for 5 minutes followed by 21°C for 5 minutes, for 6 cycles total. Residual linearized DNA was digested with PlasmidSafe exonuclease (Fischer Scientific, cat. no. NC9046399) according to manufacturer's directions for 30 minutes at 37°C, followed by 70°C for 30 minutes. 2ul of final product was transformed into Stbl3 chemically competent cells. The sequence of gRNA inserts was confirmed by sanger sequencing and large amounts of the plasmids were prepared using Maxiprep kit (MN 740414.50). Cells were electroporated using Lonza 4D-Nucleofector and P3 Primary Cell kit (Cat. V4XP-3024) according to manufacturer's instructions. Electroporated cells were seeded on feeder-coated plates in primed hESC media with 10uM ROCK inhibitor Y-27632. After 48 hours, cells were selected with puromycin (Puro resistance cassette was expressed from the PX459 plasmid) and surviving colonies were further propagated. To confirm the deletion, genomic DNA was isolated from the bulk population using Zymo Quick-DNA isolation kit (Cat. D4069) for PCR genotyping. PCR primers XP-Frw (CACAAAGATGTCCGG CTTTCA, chrX:73852799-73852819; hg38) and XE-Rev (CCTGCTGAATGCAAATGGGG, chrX:73849335-73849354; hg38) generate a ~1.5kb band upon deletion of the intervening ~2kb genomic sequence. After this step, individual colonies were isolated from the targeted hESC population and screened for homozygous and heterozygous XIST deletions. To detect the WT allele, we used the primers XP-Frw and WT-Rev (CTCTGCCAAAGCGGTAGGTAC, chrX:73851878-73851898; hg38), which amplify a 942bp region from the WT allele and not upon deletion as the WT-Rev complementary sequence is not present then. With this screening strategy, 2 homozygous clones were selected for the follow-up experiments referred to as clone 7 and clone 18 and confirmed by Sanger sequencing.

siRNA-mediated knockdown of *SPEN*

For siRNA-mediated depletion of *SPEN*, we used a mix of two different siRNAs targeting different exons of human *SPEN* (Thermo Fisher #4427037, IDs: s22831, s22829). Equal amounts of these siRNAs were mixed prior to transfection. As a negative control, non-targeting siRNA was used (Thermo Fisher #AM4611). Cells were subjected to *SPEN* depletion and control siRNA treatment as follows: female naïve hESCs (UCLA1) and male naïve hESCs (WIN1) were treated with si*SPEN* and negative control siRNA; and each clone of female naïve XIST KO H9 hESCs (clones 7 and 18) was treated with si*SPEN* and negative control siRNA. Respective naïve hPSCs were transfected using Lipofectamine RNAiMAX reagent (Life Technologies #13778150). After 24 hours of transfection, cells were transfected again using the same siRNA mix. After two rounds of transfection (in total 48 hours), cells were harvested, and RNA was isolated using Qiagen RNAeasy kit (#74104) according to manufacturer's instructions. Knockdown efficiency was quantified by RT-qPCR. First cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen 18064-014). Real time quantitative PCR was performed using TaqMan® Universal PCR Master Mix (Applied Biosystems, 4304437) and the expression levels of *SPEN* were normalized to the expression of the housekeeping gene GAPDH and subsequently relative enrichment was calculated based on non-targeting siRNA control. The Taqman probes are: GAPDH (Applied Biosystems, Hs99999905_m1), SPEN (Applied Biosystems, Hs00209232_m1). Upon confirmation of *SPEN* knockdown, RNA-seq libraries were constructed as described below.

Biochemical fractionation

To fractionate female naïve hESCs (UCLA1) into cytoplasmic, soluble nuclear, and chromatin compartments,[120,121] cells were washed twice with 1X PBS and gently dissociated with Accutase before collection of the cell pellet by centrifugation. 2×10^6 cells from the pellet were retrieved in a 2.0 mL low adhesion microcentrifuge tube (USA Scientific) and washed twice with PBS to remove cells from prematurely-lysed cells. Cell pellets were then resuspended in 200uL ice-cold NP-40 lysis buffer (10 mM Tris-HCl [pH 7.5], 0.05%–0.15% NP40 [Sigma], 150 mM NaCl) and incubated on ice for 5–10 minutes (cell-line dependent). The resulting lysate was immediately layered on top of a chilled 24% sucrose solution (hypotonic buffer in 24 % (w/v) sucrose without detergent) and centrifuged for 10 minutes, 4°C, 6000 x g. 10% of the supernatant (cytoplasmic fraction) was used for immunoblot to check for contamination from nuclear materials, and the rest of the cytoplasmic extract was mixed with an equal volume of 2x proteinase K buffer and proteinase K (20uL of 20mg/mL or equivalent) and incubated at 37°C for 1 hour. Cytoplasmic RNA was then extracted using phenol/chloroform with ethanol precipitation. The nuclear pellet was washed once with 1× PBS/1 mM EDTA buffer before being resuspended in 100 μl of chilled glycerol buffer (20mM Tris-HCl [pH 7.9], 75 mM NaCl, 0.5 mM EDTA [pH 8.0], 50% glycerol (v/v), 0.85 mM DTT, and 0.125 mM PMSF). 100 μl of cold nuclear lysis buffer (10 mM HEPES [pH 7.6], 1 mM DTT, 7.5 mM MgCl2, 0.2 mM EDTA, 0.3 M NaCl, 1 M UREA, 1% NP-40) was then added and the mixture incubated on ice for 2 minutes. After this incubation period, 1/8 of the nuclear lysate was transferred to a separate 1.5 mL microcentrifuge tube (this sample was used to generate the western blot for the chromatin fraction, see below). After centrifugation for 2 minutes, 4°C, 6000 x g, the supernatant (soluble nuclear fraction) from both tubes (the $1/8^{th}$ and $7/8^{th}$ samples) were combined and 10% of the pooled supernatant was used to check its purity by western blot and the rest of the supernatant mixed with an equal volume of 2x proteinase K buffer and proteinase K (20uL of 20mg/mL or equivalent) and incubated at 37°C for 1 hour. Nucleoplasmic RNA was extracted from the proteinase K-treated supernatant using phenol/chloroform with ethanol precipitation. The insoluble nuclear pellet from the $7/8^{th}$ sample was washed once with washing buffer before incubation in Trizol at 50°C, which was performed until the pellet was completely solubilized to extract chromatin-associated RNA. The $1/8^{th}$ insoluble nuclear pellet was resuspended in 5% SDS sample buffer (5% SDS, 95% NuPAGE LDS Sample Buffer (4X) (Thermo Fisher) at 95°C for 5 minutes to extract proteins from the chromatin fraction for Western Blotting. The fractionation experiment was performed twice.

RNA was extracted using the Direct-Zol RNA miniprep kit (Zymo Research) and quantified using a NanoDrop spectrophotometer ND-1000 (Thermo Scientific). RNA sequencing libraries were prepared using TruSeq Stranded mRNA with 4μg starting material. All libraries were amplified for 8-12 cycles and assessed by 4200 TapeStation (Agilent). Libraries were subsequently quantified using the Qubit dsDNA High-Sensitivity Kit (Life Technologies) and sequenced on Illumina NovaSeq 6000 instruments. Read quality was confirmed with FastQC[91] v0.11.4. Reads were aligned to hg38 and gencode annotations v27 using STAR v2.5.3a with default settings. Tracks were created using deepTools bamCoverage function with –normalizeUsing BPM and –effectiveGenomeSize 2913022398.

RAP-seq of human and mouse XIST/Xist

To explore where on DNA the RNA localizes, RAP-seq,[16] one of the biochemical methods that enables high-resolution mapping of RNA localization on chromatin[17,122] was applied. RAP followed by DNA sequencing was adapted from previously described RAP-seq[16] and CHART[17] approaches as described below.

Human XIST RAP-seq was done in two batches with small differences in lysis, sonication, biotinylated oligonucleotide concentration and sequencing approaches, which are indicated below. In addition to the female naïve hPSCs, primed XIST-positive hESCs and female human fibroblasts, we included the XIST-negative primed hESC line H9[44] and female naïve XIST KO hESCs in the RAP-seq experiments to rule out the possibility of non-specific detection of XIST localization arising from hybridization of RAP probes to other RNAs or open chromatin regions.

Typically, 30-50 million cells were harvested from confluent cell cultures, washed with 1x PBS, and incubated with freshly-made 10ml 2mM DSG in 1x PBS at room temperature for 45 minutes. Cells were spun down and subsequently crosslinked further at room temperature with 10ml 3% formaldehyde in 1x PBS for 10 minutes, and the reaction was stopped by the addition of 2ml 2.5M glycine and cells incubated for an additional 5 min at room temperature. Cells were then pelleted at 2,000g for 5 min at 4°C, washed twice in ice-cold 1xPBS, aliquoted into 10 million cells and snap-frozen in liquid nitrogen and frozen at -80 °C.

We lysed crosslinked cells in batches of 10 million cells on ice in Cell Lysis Buffer (10mM HEPES at pH7.5, 20mM KCl, 2mM MgCl2, 1mM EDTA with 0.1% NP-40, 1mM TCEP, 0.5mM PMSF) for 10 minutes, then lysis was completed either with glass dounce homogenizer (batch 1) or without (batch2).[16] The cells were spun and nuclei lysed in Nuclear Lysis Buffer (20mM HEPES at pH7.5, 50mM KCl, 1.5mM $MnCl_2$, 1% NP-40, 0.4% sodium deoxycholate, 0.1% N-lauroylsarcosine, 1mM TCEP, 0.5mM PMSF) for 10 minutes on ice. Chromatin was solubilized by sonication. For batch 1, five cycles of sonication were performed, with each cycle lasting 45 seconds on high power. For batch 2, sonication was done with a Misonix S-400 sonicator with microtip (model number: U1240A0418) at an amplitude of 12, total process time of 2 minutes with pulse on for 1 second and pulse off for 3 seconds. After sonication, chromatin was further segmented by incubating it with TURBO DNase digestion at a concentration of 0.1-0.4U/ul for 12-20 minutes at 37°C (time was optimized for each sample) and upon addition of 1/100th 100x DNAse-cofactor solution (250mM $MnCL_2$, 50mM $CaCl_2$). The digestion was stopped by addition of 1/25th 25x DNase Stop Solution (mixture of 250mM EDTA and 125mM EGTA). Lysate was diluted to hybridization conditions by adding 1.4x RAP hybridization buffer (1x: 20 mM Tris pH 7.5, 7 mM EDTA, 3 mM EGTA, 150 mM LiCl, 1% NP-40, 0.2% N-lauroylsarcosine,0.1% sodium deoxycholate, 3 M guanidine thiocyanate, 2mM TCEP). The lysate is frozen in aliquots represent 5 million cells and snap-frozen.

Lysate in hybridization buffer was pre-cleared by adding streptavidin-coated C1 beads and incubating at 37°C for 30 minutes. 5% of the pre-cleared lysate was taken as input sample and the remaining used for the XIST pulldown. Denatured probes were mixed (85°C for 3 min and snap-cooled) with 37°C-heated lysate and incubated at 37°C for two hours 30 min to capture XIST RNA and associated chromatin. The biotinylated probes were captured by addition of Streptavidin C1 beads for an additional 30 min at 37°C. The samples were transferred to a magnetic rack and washed three times with RAP Wash Buffer (20mM Tris pH 7.5, 1M LiCl, 10 mM EDTA, 1% NP-40, 0.2% N-lauroylsarcosine, 0.1% sodium-deoxycholate, 3 M guanidine thiocyanate, 2.5mM TCEP) at 37°C for 3 minutes and three times with RAP Wash Buffer (20mM Tris pH 7.5, 10 mM EDTA, 1% NP-40, 0.2% N-lauroylsarcosine, 0.1% sodium-deoxycholate, 3 M guanidine thiocyanate, 2.5mM TCEP) at 37°C for 3 minutes.

XIST RNA pulldown from 5 million cells was done using 1ug (batch 1) or 50pmol (batch 2) of non-overlapping 90bp long single stranded DNA probes with the first nucleotide (the 5′) labeled with Biotin-5 (Eurofins) (Table S1). Human XIST probes for RAP-seq were designed and gifted from Guttman Lab at Caltech (https://www.guttmanlab.caltech.edu/). The biotinylated single-stranded 5' biotinylated DNA probes are antisense to the XIST transcript and hybridize to the spliced human XIST transcript, allowing the purification of the RNA and its associated genomic DNA from crosslinked cell lysates. For the probe mix, the individually synthesized oligonucleotides were pooled in equimolar amounts. To test the effect of a larger amount the biotinylated probes in the pulldown, a higher concentration (5ug probes) was added to the second replicate of fibroblasts (R2). 1mg (batch 1) or 0.6mg (batch 2) of C1 Streptavidin beads were used.

DNA was eluted by Rnase H digestion, and crosslinking was reversed via proteinase K digestion of eluted DNA at 60°C for 10-12 hours. The DNA libraries were prepared using NEBNext Ultra End Rpair/dA-Tailing Module (NEB) and TruSeq DNA adapters (Illumina) ligated using Quick Ligase (NEB). Libraries were amplified by KAPA HiFi Polymerase (Roche), pooled, and sequenced on the Illlumina HiSeq platform to generate single-end (batch 1) or paired-end (batch 2) reads.

RAP-seq was also utilized to explore Xist's localization at day 2 (D2) and day 4 (D4) of female mouse EpiLC differentiation (two replicates per time point) and in female and male MEFs. 50 million EpiLCs (at D2 or D4) or MEFs were harvested after accutase-mediated dissociation. Cells were crosslinked and quenched as described above, subsequently pelleted at 2,000g for 5 minutes at 4°C and aliquoted into vials of 10 million cells. Cells from one aliquot were then subjected to lysate preparation as follows: Cell pellets were resuspended in 10ml of nuclear extraction buffer LB1 containing 50mM HEPES-KOH (pH 7.5), 140mM NaCl, 1mM EDTA, 10% (v/v) glycerol, 0.5% (v/v) NP-40/Igepal CA-630 and 0.2% (v/v) Triton X-100, and incubated for 10 minutes with rotation at 4°C, then pelleted at 2,000g for 5 minutes and the same procedure was performed with buffer LB2, which contains 10mM Tris-HCL (pH 8.0), 200mM NaCl, 1mM EDTA and 0.5mM EGTA. For cell lysis, nuclei were resuspended in 500μL of buffer LB3 containing 10mM Tris-HCl (pH 8.0), 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% (w/v) sodium deoxycholate and 0.5% (v/v) N-lauroylsarcosine, then sonicated on ice using the Misonix S-400 sonicator with microtip (model number: U1240A0418) for 2 minutes at an amplitude of 12 with 1 second pulses intermitted by 3 second pauses. Next, chromatin was further digested using TURBO DNase at a concentration of 0.1-0.4U/μl at 37°C for 15 minutes. The digestion was stopped by the addition of DNase Stop Solution (mixture of EDTA and EGTA). The rest of the protocol is as describe above except that Streptavidin beads after pulldown were washed six times with RAP Wash Buffer (20mM Tris pH 7.5, 10 mM EDTA, 1% NP-40, 0.2% N-lauroylsarcosine, 0.1% sodium-deoxycholate, 3 M guanidine thiocyanate, 2.5mM TCEP) at 45°C for 5 minutes. The mouse Xist RNA pulldown was performed using 50pmol of 90nt-long single stranded 5' biotinylated oligonucleotide probes targeting Xist (for every 5 million cells). Mouse Xist probes for RAP-seq were initially designed and gifted by the Guttman Lab at Caltech[16] (https://www.guttmanlab.caltech.edu/) (Table S1). Libraries were sequenced on the Illumina NovaSeq 6000 platform to generate 50bp paired-end reads.

Quantitative polymerase chain reaction (qPCR)

The RAP-seq data revealed that a higher percentage of reads align to the X-chromosome in female human fibroblasts compared to female naïve hPSCs (Figure 2A). This finding could be explained by a higher accumulation of XIST on the Xi compared to the Xd and/or the localization of XIST to autosomes in naïve hPSCs. To begin to address this, we performed qPCR measurements of X-linked regions after XIST pulldown. In addition, we performed qPCR measurements for autosomal regions upon XIST pulldown to confirm autosomal binding of XIST. We designed qPCR primers to regions with diverse XIST enrichment (determined based on RAP-seq data) (Table S1). These regions include a region on the X chromosome (qPCR_1 - ChrX:17608575-17608725; hg38), an autosomal region showing high XIST enrichment specifically in female naïve hPSCs and overlaps with the gene *SPON1* (qPCR_2 - Chr11:14251638-14251822; hg38), and two autosomal regions showing low XIST enrichment in female naïve hPSCs (qPCR_3 - Chr11:11424936-11425088; hg38 (Ctrl1) and qPCR_4 - Chr12:52747252-52747407; hg38 (Ctrl2)). As control, a region with no XIST enrichment in all samples (fibroblasts, female primed and naïve hPSCs) was amplified with qPCR_5 (Chr10:56359977-56360174; hg38 (qPCR control)). qPCR was performed using PowerUp SYBR green PCR master mix (A25742, Applied Biosystems) in 10 ul reactions with the QuantStudio3Real-Time PCR system (Applied Biosystems). 0.5ul of samples from inputs or pulldown from the RAP-seq experiments in female naïve hESCs (H9) and primed iPSC (XIST-positive) were used as templates for each reaction. Mean Quantification Cycle (Cq) of three technical replicates from each sample was used. Relative XIST binding was calculated as $2^{-\Delta\Delta Ct}$: $2^{-((\text{Selected primer in pulldown} - \text{Selected primer in input}) - ((\text{Control primer in pulldown} - \text{Control primer in input}))}$.

Bulk RNA-seq

For bulk RNA-seq, cells were washed with 1xPBS and dissociated with accutase. Harvested cells were then lysed using Trizol reagent (Life Technologies #15596018) and RNA was isolated using Qiagen RNAeasy kit (#74104) according to manufacturer's instructions. RNA-seq libraries were prepared with the TruSeq Stranded mRNA Library Prep Kit (Illumina 20020594) according to manufacturer's instructions. RNA-seq data for female naïve UCLA1 replicates R1, R2, and R5, and of early naïve UCLA1 hESCs (UCLA1$^{\text{pre-XIST}}$) were described before.[34] Library quality was assessed by TapeStation (Agilent) and subsequently quantified using the Qubit dsDNA High-Sensitivity Kit (Life Technologies) prior to sequencing on Illumina NovaSeq 6000.

scRNA-seq for naïve WT and XIST KO hESCs

For scRNA-seq, each clone of female naïve XIST KO H9 hESCs (clone 7 and clone 18) was converted independently to the naïve state, together with female naïve WT H9 hESCs. Cells were dissociated with accutase for 5 minutes, and cell pellets washed with 1xPBS+0.04% BSA and pelleted again at 500rcf for 5 minutes. Afterwards cells were resuspended in 1xPBS+0.04% BSA and passed through a 40 micron strainer to deplete cell clumps. The cell concentration was adjusted to 800-1200cells/ul before loading cells onto the 10X Genomics Chromium instrument. scRNA-seq libraries were generated using the Chromium single cell 3' reagent kit V3 following manufacturer's instructions and library fragment size distribution was determined by BioAnalyzer. Individual libraries were designed to target 10,000 cells. Afterwards libraries were pooled and sequenced on the Illumina Novaseq 6000 platform.

Immunofluorescence staining

For immunofluorescence staining, naïve hPSCs were seeded on feeder-coated coverslips and human fibroblasts were seeded directly on the glass coverslips. After 24-48 hours, cells were fixed with 4% PFA in 1xPBS for 10 minutes and washed with 1xPBS, afterwards permeabilized with 0.5% Triton X-100 in 1xPBS, and then blocked with 1% BSA in 1xPBS with 0.05% Tween-20 for 40 minutes. Primary antibody incubation was conducted in 1% BSA for 1 hour at RT. Samples were again washed three times with 1xPBS-Tween and incubated with fluorescent secondary antibodies for 45 minutes, then washed and counterstained with DAPI for 5 minutes and mounted using Vectashield (Vector labs: H-1000). The secondary antibodies used in this study were all from Life Technologies used at 1:400 dilution. Images were taken using LSM 880 Confocal Instrument (Zeiss) or Zeiss Axio Imager M1. Fiji (ImageJ) was used for image processing. Antibodies used in immunofluorescence staining include anti-H3K27me3 (Cell Signaling C36B11, Cat# 9733 for all images except Figures S7C and S7D which also uses Millipore cat# 07-449 as a comparison), anti-CIZ1 (Novus Biologicals, NB100-74624), and anti-H2AK119ub (Cell Signaling, Cat# 8240T).

H3K27me3 Antibody Comparison

Prior studies yielded different findings with respect to the accumulation of H3K27me3 under the XIST domain in female naïve hPSCs. Specifically, several studies reported the lack of an accumulation of H3K27me3 on the XIST-associated X chromosome in naïve hESCs[35,123] and human pre-implantation embryos,[32,124] even though it was described by our group in female naïve hESCs in a prior study[34] as well as in this study. One possible reason for this discrepancy might be the use of different antibodies in the respective immunostaining experiments since Vallot et al.,[35] An et al.,[123] and Teklenburg et al.,[124] utilized the anti H3K27me3 from Millipore (cat#07-449), whereas the antibody from Active Motif (cat#39155) was used in the Sahakyan et al. study[34] and in this study (for ChIP-seq). Additionally, in this study, we used the anti-H3K27me3 antibody from Cell Signaling (C36B11, cat#9733) for CUT&Tag and immunostaining experiments. To directly investigate whether different antibodies detect the X chromosome enrichment of H3K27me3 differently, immunofluorescence staining was done using the Cell Signaling (cat#9733) and Millipore (cat#07-449) anti H3K27me3 antibodies in female naïve H9 hESCs. The staining was performed in parallel, with the same secondary antibody. 86% of cells showed an enrichment of H3K27me3 across the XIST domain in naïve H9 cells when the Cell Signaling antibody was used and 10% showed a H3K27me3 accumulation with the Millipore antibody (Figures S7C and S7D). Both H3K27me3 antibodies (Millipore and Cell Signaling) detect an enrichment of H3K27me on the Xi in somatic cells in the vast majority of cells (not shown), and both were used to detect H3K27me3 in EZH2 inhibition experiments in naïve and primed H9 hESCs showing global genomic loss of H3K27me3 upon EZH2 inhibition supporting the specificity of both antibodies.[75,125] A higher specificity of the Cell Signaling antibody compared to the Millipore one was revealed by Kupai et al.[126] Overall, we conclude that different H3K27me3 antibodies differ in their ability to detect the accumulation of this histone mark on the Xd in female naïve hPSCs.

RNA and RNA/DNA FISH

For RNA FISH for the lncRNAs XIST and XACT and the nascent transcript signals of X-linked genes and *SPON1*, hPSCs were seeded on feeder-coated coverslips 24-48 hours before fixation, to keep hPSC colony size small, with coverslips coated with 0.5% gelatin overnight prior to culturing. Fibroblasts were seeded directly on coverslips. Specimens were washed with 1xPBS, fixed with 4% formaldehyde in 1xPBS for 10 minutes, permeabilized with cold (4°C) 0.5% Triton X-100 in 1xPBS for 10 minutes, dehydrated in cold (4°C) 70-100% ethanol series for 10 minutes each step, and air dried. Cells were then hybridized with labeled DNA probes in a humidified chamber at 37°C overnight, washed for three 5 min intervals with 50% formamide in 2× SSC, 2× SSC, then 1× SSC at 42°C, counterstained with DAPI and mounted with Vectashield (Vector Labs, H-1000). Double-stranded DNA probes were generated from full length cDNA constructs or BACs using nick translation (NT) reaction.[127] 1 μg of DNA was labeled in a 50 μL NT reaction overnight (12-15h) at 15°C, using 1.3-2.5 μL fluorescently-labeled dUTPs, 1 μL of DNA Polymerase I (Thermo EP0042) and 2 μL of DNase I (Sigma-Aldrich 4716728001) which was always freshly prepared by a 1:200 dilution in ice-cold H_2O. After the overnight incubation, NT reactions were quenched using 0.5 μl 0.5M EDTA and products were purified using magnetic beads. Afterwards, probes were resuspended in Nuclease-free H_2O and ethanol-precipitated together with Salmon sperm and Cot1 DNA at −80°C overnight. After precipitation and washes with ethanol series (70%–100%), probes were resuspended in 50ul deionized formamide with shaking at 37°C for 6-8h. Finally, 50ul of hybridization buffer (4× SSC, 20% dextran sulfate) was added to the probes and stored at −20°C (final hybridization buffer concentrations: 10% Dextran Sulphate, 2xSSC, 50% formamide). The BACs used include *XIST* (RP11-13M9), *XACT* (RP11-35D3), *GPC3* (RP11-678F20), *SMS* (RP11-147O5), *SMARCA1* (RP11-137A15), *THOC2* (RP11-121P4), *UTX* (RP11-256P2), and *SPON1* (RP11-774G22). Every new batch of probes was first tested on fibroblasts. For *XACT* signal quantification, datasets were converted to 8-bit images in Fiji and signal intensities were measured using the intensity Plot Profile tool. Intensity values were exported as csv files for plotting. For signal quantifications[128] for *THOC2*, *SMARCA1*, *GPC3*, *UTX* and *SMS*, images were acquired using the Zeiss Axio Imager M1. Z-stack images acquired from the microscope were projected to 2D images with maximum local intensity threshold for subsequent analysis. The single transcript locations and their intensities were identified and quantified by 2D Gaussian fitting which subtracts the background.[128] If an XIST cloud was presented in the cell, its location was identified similarly by 2D Gaussian fitting method. The minimal distance from each studied RNA single transcripts to the XIST cloud was calculated by their locations within the cell.

For sequential RNA and DNA FISH experiments with chromosome paints or oligonucleotide probes and Xist/XIST probes,[27] cells were grown on high precision. Coverslips were transferred to new multi-well plates, washed three times with PBS and fixed with 3% formaldehyde in 1x PBS for 10min, followed by two washes with 1xPBS. Samples were quenched for 10min with 20mM glycine in 1xPBS, washed with 1xPBS, and permeabilized with 0.5% Triton X-100 in 1xPBS for 15min (PBST), washed twice with PBST, equilibrated in 2xSSC for 10min and incubated for 30min to 2 hrs with 50% formamide in 2xSSC. RNA FISH for XIST was performed first, samples were post-fixed and DNA FISH followed. For DNA FISH, a denaturation step was performed for 2min at 76°C before probe hybridization at 37°C overnight. After demounting coverslips, unbound probes were washed-off with three 20min washes with 2xSSCT (2xSSC, 0.5% Tween 20) under mild shaking, followed by three 5min washes with 4xSSCT, and three additional 5min washes with 0.1xSSC. Chromosome X and 11 paint probes were purchased from Metasystems. Custom 45bp anti-sense oligonucleotide probes designed to target genomic DNA sequences were used to detect the 8Mb region of chromosome 11 (chr11:11730501-19952055; hg38) enriched in XIST binding based on female naïve hPSC RAP-seq (Table S1).

High-resolution imaging was performed on the LSM880 confocal laser scanning microscope with a 63x/1.4 NA plan Apochromat oil objective. For image processing and analysis Fiji (ImageJ) was used.[113] For Figures 2F and 2G, image data from Figure 2E were converted to 8-bit tifs by using the automatic Otsu threshold to obtain binary masks of the corresponding chromosome territories or XIST signals. The surface area or Feret's diameter of each chromosome or XIST territories was calculated using the Analyzer Particle function. The Plot Profiler was used to measure distances of the farthest XIST foci to the Xd territory in hPSCs.

To quantify the frequency of overlap between XIST RNA and the 8Mb chromosome 11 target region, RNA/DNA FISH image data were projected, and the Plot Profiler function was used to measure signal intensities across overlapping regions between the different channels. Overlapping spectra where both channels exhibited an SNR (Signal-to-noise-Ratio) above 2 were marked as 'hits'. The total number of nuclei was counted from each Field of View (FOV) and the number of hits in each FOV were given.

Immuno-RNA FISH

Immunostaining was performed as described above. In all buffers and antibody solutions, RNaseOUT 1:200 (Thermo Fisher Scientific, 10777019) was added to preserve RNA. Before performing RNA FISH, samples were fixed again with 4% PFA in 1xPBS for 10 min at room temperature. RNA FISH was then performed using DNA probes as described above.

Single molecule RNA FISH

Exonic and intronic single molecule RNA FISH probes were enzymatically labeled with distinct spectrally fluorescent dyes using modified FISH probe labeling protocols.[129,130] Specifically, an equimolar mix of single stranded oligos (SMACRA1-exonic, GPC3-exonic, XIST exonic/intronic oligos) complementary to target genes (48 oligos per set; Table S1) were labeled with ddUTP at the 3'end by TdT enzyme. The ddUTP conjugated oligo mixture was then labeled with NHS-coupled fluorescent dyes. Probes were precipitated by ethanol and purified with DNA purification column (Promega, A9282). The concentration of fluorescently-labeled probes was measured by Nanodrop. For RNA FISH, female naïve H9 hESCs were cultured with feeders in a 12-well plate. When cell density reached 70-80%, cells were washed once with 1xPBS, fixed with 4% PFA, and permeabilized with 70% ethanol at 4°C overnight. Cells were rehydrated in 2xSSC buffer with 10% formamide and immersed in hybridization buffer (Biosearch: SMF-HB1-10, 10% formamide added freshly) with a final probe concentration of 1 ng/ul for >16 hours. The samples were then washed twice with wash buffer A (Biosearch: SMF-WA1-60) for 30 minutes at 37°C. DAPI (Invitrogen: D1306) was added in the second wash to a final concentration of 0.5 ug/ml. Samples were washed once with wash buffer B (Biosearch: SMF-WB1-20) for 5 minutes. The coverslips were mounted on microscope slides in ProLong Gold antifade media (Invitrogen: P36930) overnight before microscopy. Cells were imaged on a Leica TCS SP8 light-sheet microscope equipped with a 63x, 1.4 Numerical Aperture (NA) oil-immersion objective (Leica) and a Leica sCMOS-camera. Images were acquired in ~200 nm z-dimension axis steps across ranges of approximately 4 μm. For image analysis, multicolor z-stack images obtained from the Leica confocal SP8 were split to individual channels and imported into FISH-quant.[131] Nuclei and cytoplasm were segmented by drawing outlines manually via DAPI and fluorescent dye signals, respectively.

Super-resolution microscopy

For Figure 7A, 3D-structured illumination microscopy (3D-SIM) was performed on a DeltaVision OMX-SR system (Cytiva, Marlborough, MA, USA) equipped with a 60x/1.42 NA Plan Apo oil immersion objective (Olympus, Tokyo, Japan), sCMOS cameras (PCO, Kelheim, Germany) and 405, 488, 642 nm diode lasers and a 568 nm DPSS laser. Image stacks were acquired on the OMX AcquireSR software package 4.4.9934 with a z-steps of 125nm and with 15 raw images per plane (five phases, three angles). Raw data were computationally reconstructed with the soft-WoRx 7.0.0 software package and subsequently imported into ImageJ/Fiji for conversion into 16-bit tiff files.

Higher-order organization of the X chromosome

To define the higher-order three-dimensional organization of the Xd, Xi and Xa, sequential RNA/DNA FISH was performed on H9 naïve hESCs (at passage 8 after naïve conversion from primed H9 hESCs at passage 65) cultured in t2iLGö media[132,133] at 5% O2, 5% CO2, as well as on female human fibroblasts (Lonza). For sequential FISH,[27,134] cells were plated into μ-Slide 8 Well Glass Bottom chamber slides (Ibidi) after adding 0.1 μm TetraSpeck™ microspheres (beads) and Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (Gibco). For fixation, cells were washed with 1xPBS, beads were added again, and fixation was achieved with 4% PFA in 1xPBS for 10 minutes. XIST RNA FISH was executed using Atto 488 (Sigma-Aldrich)-labeled probes. Imaging took place on a motorized stage-equipped LSM 880 Confocal Instrument (Zeiss) using the Plan-Apochromat 100X/1.40 Oil DIC objective using Zen Blue and Zen Black. The Stitching tool created 20x20 tiled images to map well topography for relocation. Z-stacks of FOVs (Fields of View) at 16-bit resolution with voxel size of 0.07x0.07x0.29 μm3 were used. Before DNA FISH, 0.1mg/mL RNAse A (Invitrogen) and 5U/ml RNAse H (New England Biolabs) were added for 20 minutes at 37°C and slides were washed with 2xSSC for 15 minutes. For each of the three sequential rounds of DNA FISH shown in Figure S6P, probes targeting three different genomic locations were used, labeled with Atto 488, Cy3 (Sigma-Aldrich), Texas Red (TR) (Thermo Fisher Scientific), or Cy5 (VWR), through nick translation, respectively. Specifically, the following BACs were used (brackets indicating the DNA FISH round (2, 3 or 4) and the fluorophore): *XIST* (RP11-13M9) (2,488), *HUWE1* (RP11-975N19) (2,Cy3), *THOC2* (RP11-121P4) (2,Cy5), *XACT* (RP11-35D3) (3,488), *GPC3* (RP11-678F20) (3,TR), *UTX* (RP11-256P2) (3,Cy5), *PAGE1* (RP11-315L18) (4,488), *CDKL5* (RP11-106N3) (4,Cy3), and *LAMP2* (RP11-158I12) (4,Cy5). After imaging each round of DNA FISH, probes were stripped in 2xSSC+70% Formamide for 5 minutes on a 70°C heat block. To image and relocate the FOV, we focused on a corner of a well and tiled images, moving the focus towards the center of each FOV.

To align all images of the same FOV, the first round of DNA FISH hybridization was used as a reference (baseline dataset). Signals of the multi-spectral beads and probes in each channel were extracted using 3D Objects Counter. Using the multi-spectral beads, the 488 channel was used to record the shift in their coordinates (x,y,z) over sequential rounds of hybridization and the MultiStackReg[135] function of Fiji[113] was applied to adjust the coordinates in each round of hybridization to the baseline stack and the transformation file was saved. The coordinates of each channel's bead from the 488 channel were subtracted yielding a chromatic shift that was applied to the probes' coordinates in all channels. The Euclidian distance was calculated between all pairwise regions in each round for the two X chromosomes in naïve and somatic cells.

ChIP-seq

ChIP-seq was performed for the following nine histone modification: H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me3, H3K27me3, H3K27ac, H3K36me3, H3K79me2.[136] Male naïve hESCs (WIN1) and female naïve hiPSCs were grown to a final concentration of 1×10^8 cells. For all antibodies used, cells were chemically cross-linked at room temperature by the addition of formaldehyde to 1% final concentration for 10 minutes and quenched with 0.125 M final concentration glycine. Cross-linked cells were re-suspended in sonication buffer (50mM Hepes-KOH pH 7.5, 140mM NaCl, 1mM EDTA, 1% TritonX-100, 0.1% Na-deoxycholate, 0.1% SDS) and sonicated using a Diagenode Bioruptor for three 10-minute rounds using pulsing settings (30 seconds ON; 1 minute OFF). 10 ug of sonicated chromatin was then incubated overnight at 4°C and under constant stirring with 5 ug of antibodies-conjugated to magnetic beads (Active Motif; 53014). The antibodies used were anti-H3K9ac (Abcam; ab4441), anti-H3K4me3 (Abcam; ab8580), anti-H3K4me2 (Abcam ab7766), anti-H3K4me1 (Abcam; ab8895), anti-H3K27me3 (Active Motif; 39155), anti-H3K27ac (Abcam; ab4729), anti-H3K36me3 (Abcam; ab9050), anti-H3K79me2 (Active Motif, 39143) and anti-H3K9me3 (abcam, ab8898). Following the immunoprecipitation, beads were washed twice with RIPA buffer (50mM Tris-HCl pH8, 150 mM NaCl, 2mM EDTA, 1% NP-40, 0.1% Na-deocycholate, 0.1% SDS), low salt buffer (20mM Tris pH 8.1, 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS), high salt buffer (20mM Tris pH 8.1, 500mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS), LiCl buffer (10mM Tris pH 8.1, 250mM LiCl, 1mM EDTA, 1% Na-deoxycholate, 1% NP-40), and 1xTE. Finally, DNA was extracted by reverse crosslinking at 60°C overnight with proteinase K (20ug/ul) and 1% SDS followed by phenol:chloroform:iso-amylacohol purification. DNA from whole cell extract was extracted and sonicated to 150bp. Input libraries were constructed using 10 ng of staring material and amplified for a similar number of cycles to the other libraries. All protocols for Illumina/Solexa sequencing library preparation, sequencing, and quality control were performed as recommended by Illumina, with the minor modification of limiting the PCR amplification step to 10 cycles. All constructed libraries were sequenced using paired-end 50 bp sequencing reactions.

CUT&Tag

CUT&Tag experiments[137] for H3K27me3 in female naïve WT H9 hESCs, two clones (clone 7 and clone 18) of female naïve XIST KO H9 hESCs and in male naïve WIN1 hESCs were performed each with two replicates. Additionally, we performed a CUT&Tag control experiment with a rabbit IgG antibody as negative control. Briefly, for each sample, 1×10^5 cells were harvested by trypsinization and washed twice with 1ml wash buffer (20mM HEPES, 150mM NaCl, 0.5mM Spermidine), then immobilized on 10ul of concanavalin-A beads by incubating cell suspension and concanavalin-A beads at room temperature for 10 minutes. The primary antibody (anti-H3K27me3, Cell Signaling C36B11, Cat# 9733) was diluted in antibody buffer (20mM HEPES, 150mM NaCl, 0.5mM Spermidine, 0.05% Digitonin, 2mM EDTA and 0.1% BSA) at a 1:100 ratio, and then incubated with cells attached concanavalin-A beads for 2 hours at room temperature. Next, the secondary antibody (guinea pig anti-rabbit IgG ABIN101961) was incubated with cells attached concanavalin-A beads in Dig-wash buffer (20mM HEPES, 150mM NaCl, 0.5mM Spermidine, 0.05% Digitonin) at 1:100 ratio for 1 hours and washed by Dig-wash buffer. pA-Tn5 was then diluted with Dig300 wash buffer (20mM HEPES, 300mM NaCl, 0.5mM Spermidine, 0.01% Digitonin) at 1:250 ratio and incubated with cells attached concanavalin-A beads for 1 hour. After washing the cells attached concanavalin-A beads with Dig300 wash buffer, the tagmentation reaction was activated by adding tagmentation buffer (20mM HEPES, 300mM NaCl, 0.5mM Spermidine, 0.01% Digitonin, 10mM MgCl 2) to the cells attached concanavalin-A beads and incubated at 37°C for 1 hour. Tagmentation was terminated by adding EDTA to 16.7mM and DNA fragments were solubilized by adding SDS to 0.1% and 50ug of protease K. Lastly, genomic DNA was extracted with phenol chloroform isoamyl, precipitated by adding ethanol to 75% and cleaned up by washing with 100% ethanol and resuspended in water. The extracted DNA proceeded to library amplification.

QUANTIFICATION AND STATISTICAL ANALYSIS

RAP-seq alignment

RAP-seq DNA sequencing reads were trimmed using trim_galore (https://github.com/FelixKrueger/TrimGalore[92]) with default parameters to remove the standard Illumina adaptor sequence. Bowtie2[105] was used to align reads to the human (hg38) or mouse (mm9) genome with the default parameters. Reads with mapping quality of less than 30 were removed using Samtools,[96] and Picard MarkDuplicates[106] was used to mark and remove PCR duplicates.

Calculation of RAP-seq enrichment

Bedtools makewindows[107] was used to create different size genomic windows across the chromosomes including: 1) 100kb windows every 25kb for enrichment analysis (Figures S1D, S1E, and S1H) for correlations of RAP-seq data between samples (Figure 1D), correlations of RAP-seq data with histone marks (Figure 3J), and X chromosome clusters analysis (Figures 6H and S7H); 2) 1Mb every 250kb for correlations with genomic features (Figures 3I, 7J, and S2F; Table S5) and chromosome-wide visualization (Figures 1C and S1J); 3) 1kb for visualization of small genomic regions (Figures 2B, 4G, 4N, and 7D) and comparison of XIST levels relative to RAP-qPCR (Figures 2D and S1F); 4) 1Mb for correlations with inter-chromosomal interactions (Figure 7L); and 5) 5kb starting from transcription start sites (TSSs) and extending over the gene body, to capture the XIST level around genes (Figures 4B, 4F, 5N, S4D, and S6L). bedtools intersect[107] was used to count RAP-seq reads in the respective genomic regions defined above, from each sample. To account for differences in sequencing depth, the read counts in each genomic region were normalized to the sum of all reads in that sample.

RAP-seq enrichment scores in each genomic region were calculated using the ratio of normalized read counts in the RAP-seq XIST pulldown to the input of each sample. A region was defined as unmappable using the inputs of all samples. Specifically, the R function isOutlier (scater package[103]) was used to identify outliers (having less or more than expected read counts) based on the minimal number of reads across all input samples (nmads>4 for all genomic windows, nmads>6 for 1Mb or 1Mb every 250kb). Genomic regions identified as outliers were removed from all downstream analysis.

The normalized enrichment ratios of the pulldowns and inputs were used in all computational analysis. We note that while the read counts of genomic intervals were defined in 100kb or 1Mb windows every 25kb and 250kb, respectively, enrichment scores were assigned to the 25kb or 250kb windows in the center of the 100kb and 1Mb windows. This was done to prevent overlapping windows in the downstream analysis.

XIST peak calling and binding comparison

In human cells, the following steps were applied to define peaks of autosomal XIST binding based on RAP-seq data:

1) XIST peak calling was performed using MACS2 callpeak[108] with broad and max-gap=1000 parameters and using the input of the respective XIST pulldown sample as control. Only autosomal peaks were retained. These peaks are referred to as 'pre-filtered MACS2-peaks' in this methods section.
2) Low confidence peaks with either MACS2 q-value $\geq$0.05, fold change $\leq$2 for all samples except for for human naïve and primed iPSCs where we applied a $\leq$3 cutoff, or short length (<500bps) were removed. Peaks overlapping with centromeric and telomeric regions (downloaded from the UCSC Genome Browser, track gap, table gap[138]) were also removed. We note that the right arm of chromosome 8 is likely duplicated in naïve UCLA1 hESCs, based on the increased number of reads in the UCLA1 input sample. Therefore, we removed XIST peaks found in UCLA1 in this region (chr8, peak start >121,000,000). This resulted in 6903, 9845 and 9347 filtered autosomal peaks for H9, UCLA1 and hiPSCs, 79 and 100 peaks in the fibroblast replicates R1 and R2, 319 peaks in primed (XIST-positive) hiPSCs, 109 peaks in primed (XIST-negative) H9 hESCs, and 7 peaks in naïve H9 XIST KO hESCs. These peaks are referred to as 'MACS2 peaks' throughout the manuscript and methods, and are used in Figures 2C and S1I. They were also deposited to GEO.
3) To facilitate the comparison of peaks between samples, bedtools merge[107] was used with pre-filtered MACS2 peaks from all samples (from step 1 above; reasoning is given below). This approach merges overlapping and consecutive (less than 1 bps apart) peaks from all human RAP-seq data sets into a single peak, referred to as 'pre-filtered merged peak' (or pf-merged peak for short) throughout the methods. 55895 pf-merged peaks were constructed.
4) Next, sample identity was assigned to each pf-merged peak. To this end, bedtools intersect[107] was used to intersects the pf-merged peaks (from step 3) and the MACS2 peaks of each sample (from step 2), using -wa -wb -a pf-merged_peaks.bed -b MACS2_peaks_in_each_sample.bed. Each pf-merged peak was assigned sample identity/identities based on this intersect. Since this intersect only considers highly significant peaks (derived from step 2), it is possible that a given pf-merged peak can be assigned to none of the samples, only one sample, all samples, or anything in between. pf-merged peaks overlapping with no MACS2 peak (due to the filtering in step2) were removed, resulting in a final list of 19046 merged peaks. These merged peaks are referred to as 'peaks' throughout the text. The sample assignment of these peaks is as follows: 5456, 8858, and 8506 peaks for H9, UCLA1 and hiPSCs, 79 and 100 peaks for fibroblast replicates R1 and R2, 319 peaks for primed, XIST-positive hiPSCs, 109 peaks for primed, XIST-negative H9 hESCs, and 7 peaks for naïve XIST KO H9 hESCs, and were plotted in Figures 3A, S1K, and S2A. We note that all pre-filtered MACS2-peaks from step 1 are used to define the pf-merged peak coordinates in step 3, but only confident MACS2 peaks from step 2 are used to assign sample identity. Thereby, peak boundaries are defined using the pre-filtered MACS2 peaks, but only pf-merged peaks containing high-confidence MACS2 peaks are retained for downstream analysis. Visual inspection of the pf-merged and the resulting peaks together with the RAP-seq XIST enrichment over autosomal regions showed that this approach was better at capturing the broad autosomal localization of XIST. The -$\log_{10}$(q-value) of the MACS2 peak was used for plotting peak significance in Figure 3D. In cases where a peak intersected with more than one MACS2 peak of a given sample, the max -$\log_{10}$(q-value) was used in this figure.
5) Finally, naïve hPSCs-specific peaks, conserved peaks, and UCLA1 and H9 peaks were defined. *a)* Naïve hPSCs-specific peaks: these peaks represent the 18564 peaks from step 4 assigned to naïve hPSC lines but not to fibroblast, primed hiPSCs, primed H9 hESCs or naïve H9 XIST KO hESCs samples, and were plotted in Figures 3B, 3D, S2C, and S2D. Of them, 5301, 8674, and 8341 were assigned to naïve H9, UCLA1 and hiPSCs, respectively. To test if the overlap between peaks from the three different naïve female hPSC data sets is statistically significant, which would suggest that XIST enrichment is conserved across the different naïve hPSC lines, Z-scores were calculated using overlapPermTest function of regioneR package (using ntimes=1000 and autosomal regions as genome background) for all pairs of the three naïve hPSCs (H9 vs iPSC, H9 vs UCLA1 and iPSC vs UCLA1). Briefly, this function performs permutation tests to assess if sample1 peaks significantly overlap with sample2 peaks. The peaks of all three naïve pairwise comparisons (H9 vs iPSC, H9 vs UCLA1 and iPSC vs UCLA1) significantly overlapped (p-value<0.001), with z-scores = 29, 32, and 29, respectively. *b)* Conserved naïve hPSC peaks (724 peaks), referred to as conserved XIST peaks throughout the text, are defined as naïve hPSCs-specific peaks that are present in all three female naïve hPSC lines (H9, UCLA1 and hiPSCs). These conserved peaks are used in Figures 2B, 3C, 3E–3H, 3K, 3L, 4E, 4G, 4H, 4K, 4M, 5L, 5M, 6L, 6N–6P, S2B, S2E, S2G–S2I, and S4C. *c)* Naïve H9 hESC peaks (5301 peaks; referred to as H9 hESC-derived peaks throughout in the text) and are used in Figures 5L, 5M, and S6K. *d)* Naïve UCLA1 hESC peaks (8674 peaks; referred to as UCLA1 hESC-derived in the main text) and are used in Figure S6K.

Similar approaches were used to identify Xist peaks in mouse EpiLC differentiation and MEF RAP-seq data:

1) MACS2 callpeak[108] was used to call peaks in each sample (pre-filtered MACS2-peaks').
2) Low confidence peaks with either q-value $\geq$ 0.05, fold change $\leq$ 2 or short length (<500bps) were removed. Peaks overlapping with centromeric and telomeric regions (downloaded from the UCSC Genome Browser, track gap, table gap[138]) were also removed, resulting in 8227 and 7544 filtered autosomal peaks in D2 replicates R1 and R2, 1553 and 1148 in D4 replicates R1 and R2, 1020 in female MEFs and 326 in male MEFs ('MACS2 peaks').
3) To facilitate the comparison of peaks between samples, similar to the human XIST peak curation, bedtools merge[107] was used with pre-filtered MACS2-peaks from all samples from step 1. 18401 pf-merged peaks were constructed.
4) bedtools intersect[107] was used to assign sample identity to each of the pf-merged peaks. 11870 peaks were assigned as following: 7625 and 6200 D2 replicates R1 and R2, 1030 and 770 D4 replicates R1 and R2, 839 female MEFs and 320 male MEFs and were used in Figure S7Q. The max -$\log_{10}$(q-value) of the intersecting peak was used as peak significance in Figure 7G.
5) Finally, D2-specific, D4-specific, D2&D4 overlapping, and background peaks were defined as follows: D2-specific peaks were defined as peaks with assigned identity of both D2 replicates and not of any of the D4 replicates, female or male MEFs. This yielded 2529 D2-specific peaks. D4-specific peaks were defined as peaks with identity assigned to both D4 replicates and not of any of the D2 replicates, female or male MEFs. This yielded 24 D4-specific peaks. D2&D4 overlapping peaks were defined as peaks with assigned identity of both replicates of D2, both replicates of D4, and not of female or male MEFs. This yielded 149 D2&D4 peaks. All other peaks were defined as background peaks and include those that could be only assigned to one D2 or D4 replicate of the female and male MEF samples, yielding 9168 background peaks. D2-specific, D4-specific and D2&D4 peaks are used in Figures 7C–7G. The background peak set is additionally included in Figure 7G. In this figure, the peak significance (-$\log_{10}$(q-value)) from D2 RAP-seq data is plotted for different peaks sets (background, D2-specific and D2&D4 peaks). If a peak was only significant in one D2 replicate, the peak significance is given for only the significant peak. If a peak was significant in both D2 replicates, the average peak significance of both D2 replicates is plotted. The D2 peaks in Figure 7K include both the D2-specific peaks and D2&D4 peaks, totaling 2678 peaks and Figure 7H the genes associated with these D2 peaks.

To define if a gene overlaps or not with XIST/Xist peaks, the gene body (transcription start to transcription end site) was used in the intersect, requiring at least 1bps overlap.

Analysis of XIST distribution on the X

To explore the differential localization of XIST on the X chromosome between female naïve hPSCs and fibroblasts (Figure S1G), MACS2[108] was first used to call narrow peaks in each XIST pulldown, using the input of each sample as control, as described above. Next, the R package DiffBind[109] was used to identify differential peaks comparing all three naïve female hPSCs to the two fibroblast replicates using: dba function with minOverlap=1 and AnalysisMethod=DBA_DESEQ2 to construct the dba object; dba.count function with filter=10 and filterFun=sum and keeping only X chromosome peaks, to count RAP-seq reads in peaks; dba.contrast function to establish design and contrast and; dba.analyze function to generate the results of the differential localization of XIST. We note that XIST is more enriched on autosomes in naïve hPSCs compared to fibroblast cells (Figure 2A) and filtering out autosomal reads allowed us to identify differences in the relative enrichment between somatic and naïve hPSCs.

Characterization of XIST/Xist localization

We used different approaches to identify genomic features associated with XIST/Xist localization: 1) For correlations with histone marks (Figure 3J), Pearson correlation was calculated between XIST enrichment in autosomal windows (100kb every 25kb) and histone mark enrichment in each window, for various histone marks from male naïve hESCs (WIN1) (using ChIP-seq data). 2) For correlation with different genomic features (Figures 3I, 7J, and S2F; Table S5), 1Mb genomic intervals every 250kb were used. XIST/Xist enrichment in each window was compared to the interval score of various genomic features, including gene density and repeat sequences. For this, a gene annotation file was obtained from UCSC,[139] and DNA repeat annotations from the UCSC Genome Browser, track RepeatMasker, table rmsk (using hg38 or mm9).[138] The number of annotated genes and repeats in each genomic interval was calculated and Pearson correlation was used to define the correlation between XIST/Xist enrichment and feature counts in each genomic interval. The *XIST/Xist* locus along with windows covering the 10Mb at either side of the *XIST/Xist* locus were removed for these analyses, similar to Engreitz et al.[16] 3) For the correlation between inter-chromosomal interaction scores and Xist localization (Figure 7L), Xist enrichment in 1Mb windows along the autosomes was correlated with contact frequencies of *Xist* locus and the autosomal windows (1Mb). Contact frequencies were extracted from SPRITE data of mESCs[86] as described below. 4) The R package LOLA[111] was used to identify enriched features in XIST/Xist autosomal peaks identified in each of the three female naïve hPSCs, in conserved autosomal XIST peaks of female naïve hPSCs, or in autosomal Xist peaks of day 2 female mouse EpiLCs. Autosomal regions were subset from LOLA's tiles25000.hg38.bed (http://big.databio.org/LOLAweb/universes/hg38/) and used as background. Features include diverse TF binding sites (ENCODE ChIP-seq) extracted from LOLA Core database,[111] Hi-C subcompartments extracted from GM12878[49] (GEO: GSE63525) ect., and are listed in Table S5.

Processing of bulk RNA-seq data

RNA-seq reads were trimmed using trim_galore (https://github.com/FelixKrueger/TrimGalore[92]) with default parameters to remove the standard Illumina adaptor sequences. Reads were then mapped to the human genome (hg38 assembly) using HISAT2[93] with default parameters. Reads with mapping quality less than 30 were removed using samtools.[96] For paired-end libraries, unpaired reads were removed using samtools view -f 0x02. Read counts for each gene were calculated using HTSeq with the following parameters: –format=bam –order=pos –stranded=reverse –minaqual=0 –type=exon –mode=union –idattr=gene_name.[97] Gene expression analysis was done separately for different combinations of samples (as marked by the "Normalization group" column in Table S2), including: 1) Female and male naïve hPSCs (5 female UCLA1 hESC, 2 female hiPSC, 2 female H9 hESC and 2 male WIN1 hESC datasets, all in normoxia culture conditions); 2) Female (XIST and pre-XIST) and male naïve hPSCs (2 female UCLA1$^{pre\text{-}XIST}$ hESC, 5 female UCLA1 hESC, 2 female iPSC, 2 female H9 hESC and 2 male WIN1 hESC datasets, all in normoxia culture conditions); 3) female naïve XIST KO and WT H9 hESCs (22 XIST KO H9 hESC and 10 WT XIST H9 hESC datasets, all in hypoxia culture conditions); and 4) si*SPEN* experiments in female naïve XIST KO and WT hESCs and male naïve WT hESCs (2 si*SPEN* and 1 siCTRL in female UCLA1 hESC, 2 si*SPEN* and 1 siCTRL in male WIN1 hESC, 2 si*SPEN* and 2 siCTRL in XIST KO H9 hESC (clone 18), and 2 si*SPEN* and 2 siCTRL in XIST KO H9 hESC (clone 7) datasets).

Counts-per-million (cpm) were calculated using cpm from the edgeR package, and genes with low read counts were removed (keeping genes with counts-per-million > 0.5 in at least two samples). RPKMs (Reads Per Kilobase per Million mapped reads) values were calculated for each gene. Regularized log transformation of each gene in each sample were calculated using rlog from DESeq2 package[98] and used as normalized gene expression values. Differential gene expression analysis was done using DESeq from the DESeq2 package[98] using fitType = "local". Genes were categorized as upregulated (adjusted p-value < 0.05 & $\log_2$(Fold change) > 0.5), downregulated (adjusted p-value < 0.05 & $\log_2$(Fold change) < -0.5), or with no significant change, except for the comparison of expression changes upon si*SPEN* in the different samples which were categorized as upregulated (adjusted p-value < 0.05 & $\log_2$ (Fold change) > 0), downregulated (adjusted p-value < 0.05 & $\log_2$(Fold change) < 0), or with no significant change.

Gene expression (read counts) obtained from Rostovskaya et al.[47] were used for differential gene expression analyses between primed and naïve hPSCs (Figure 3G), between female and male naïve hESCs (Figures 4A, S4C, and S4D) and during a time course of capacitation of naïve hPSCs (Figure 3F). Regularized log transformation of each gene in each sample was calculated using rlog from DESeq2 package[98] and used as normalized gene expression values. DESeq2 was used to identify genes that are: 1) Differentially expressed between primed and naïve hPSCs. For this comparison, naïve hPSCs (day 0) were compared to primed hESCs in E8 medium. This was done separately for male HNES1 hESCs (Figure 3G, replicate R1) and female H9 hESCs (Figure 3G, replicate R2) naïve and primed hESCs. 2) Differentially expressed between female and male naïve hESCs by comparing female naïve H9 and male naïve HNES1 hESCs (day 0; Figures 4A, S4C, and S4D). Genes were categorized as upregulated (adjusted p-value < 0.05 & $\log_2$(Fold change) > 0.5), downregulated (adjusted p-value < 0.05 & $\log_2$(Fold change) < -0.5), or with no significant change. 3) Upregulated during capacitation (Figure 3F). For this, gene expression during a time course of capacitation of male naïve hESCs (HNES1) was used, comparing day 10 to days 0, 1, 2, 3 or 7. A subset of expressed genes was used, keeping 28689 genes with baseMean>1 in at least one of the comparisons. Upregulated genes were defined as having adjusted p-value < 0.05 and $\log_2$(Fold change) > 0.5 in any of the comparisons of day 10 to days 0, 1, 2, 3 or 7 of male hESCs. Hypergeometric test comparing autosomal genes significantly upregulated and genes overlapping conserved autosomal XIST peaks in naïve hPSCs was used to test for significant overlap (expressed autosomal genes as described above were used as background).

Multiple datasets were used to explore differential gene expression between primed and naïve hPSCs: 1) male primed and naïve HNES1 hESCs extracted from Rostovskaya et al.[47] as described above (Figure 3G, replicate R1); 2) female primed and naïve H9 hESCs extracted from Rostovskaya et al.[47] as described above (Figure 3G, replicate R2); and 3) female primed and naïve H9 hESCs extracted from Collier et al.[48] (Figure 3G, replicate R3). For the latter, gene expression (read counts) in naïve and primed hPSCs were obtained from Collier et al.[48] and analyzed using DESeq2 as described above. Gene expression fold changes upon EZH2 inhibition were extracted from Kumar et al.[75]

Expression ratios between chromosome X and autosomes (X/A ratios) were calculated using the sum of raw read counts or the median and mean of RPKMs of X-linked and autosomal genes.

Processing of single cell expression data

Reads were aligned using the cellranger[100] (https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome) count function against a custom reference genome of human (hg38, GRCh38-3) and mouse (mm10), created using cellranger[100] mkref function. This was done to account for mouse feeder cells. Aligned reads were processed further through the Seurat pipeline[101,102] and filtered based on multiple criteria: 1) An initial filtering was done keeping genes detected in at least three cells and cells with at least 800 detected genes. 2) Because hESCs were cultured with mouse feeder cells, filtering out mouse cells was done by keeping cells classified as human by Cell Ranger's GEM classification. 3) Mitochondrial RNA was quantified per cell. A quality control metric was calculated using R function calculateQCMetrics (scater package)[103] and the function isOutlier (scater package)[103] was applied to identify and remove outliers based on the library size (nmads>3), number of genes (nmads>3), and the percentage of mitochondrial genes in each cell (nmads>1).

Upon filtering, the data were then normalized and scaled, and highly variable genes were detected, using the SCTransform function (Seurat package)[101,102] with default parameters. Principal Component Analysis (PCA) was done using the Seurat function RunPCA on the most highly variable genes. The top 20 principal components were used to find the 20 nearest neighbors of each cell (using FindNeighbors function from Seurat package) following by application of the FindClusters function to identify cell clusters. Clusters were visualized by uniform manifold approximation and projection (UMAP) dimensional reduction technique using the Seurat RunUMAP function. The vast majority of female naïve WT and KO H9 hESCs reproducibly cluster together (>92% of WT and KO cells are in cluster 0 in each scRNA-seq experiment) (Figures S5G and S5H). A small proportion of WT and KO cells was assigned to one or two other clusters that express naïve pluripotency markers at lower and differentiation signatures at higher levels (Figures S5G–S5J). To account for the possibility that the lower X/A ratio in WT cells in bulk RNA-seq data (Figures 5F, S5E, and S5F) is due to XCI induced in the subset of non- or less-naïve cells, only cells with the highest expression of naïve markers were used to explore X/A ratios (i.e. cluster 0 cells). The focus on naïve cluster 0 cells also excludes the possibility that the up- or down-regulation of autosomal XIST target genes, which contain many differentiation-associated genes, is due to departure from the naïve pluripotent state. Module scores were calculated using AddModuleScore from the Seurat package for cell makers defined in Messmer et al.[140] (Figure S5I) or for genes overlapping XIST autosomal peaks (Figure 5M), and represent the normalized average expression of cell markers or genes overlapping XIST autosomal peaks in individual cells respectively. The sum reads extracted from the counts slot in the RNA assay of the Seurat object was extracted for X-linked and autosomal genes and was used to explore X/A ratios in each single cell (Figure 5G).

A pseudobulk approach was used to perform differential expression analysis between naïve WT and XIST KO H9 hESCs, as this approach was shown in recent benchmarking studies to perform best in single cell differential expression analysis.[141] To this end, pseudo-replicates were created by splitting each sample and cluster into three, equal size, random groups. The sum of the raw counts for each gene was calculated for the ‘count’ slot in the ‘RNA’ assay of the Seurat object, grouping by sample, Seurat cluster and pseudo-replicate. This approach was used for creating a new raw count matrix for each gene, with three pseudoreplicates for each sample and Seurat cluster. DESeq2[98] was then used to calculate differential gene expression between the KO and WT in each Seurat cluster, using the default Wald test. Shrunken $\log_2$ fold-change values were generated using the lfcShrink function with type=“ashr”. Genes with adjusted p-value<0.05 and $\log_2$(fold change)>0.5 were marked as significantly upregulated upon XIST KO.

To explore gene expression in pre-implantation embryos, single cell RNA-seq data for male and female human pre-implantation embryos were obtained from Petropoulos et al.[33] RPKMs and expression fold changes between female and male cells at each developmental stage were obtained from Supplemental Table 7 in Petropoulos et al.[33] In addition, to explore if genes overlapping with conserved autosomal XIST peaks have lower expression in female compared to male pre-implantation embryos (Figure 4M), read counts for each gene and single cell were downloaded (counts.txt file from ArrayExpress accession ArrayExpress: E-MTAB-3929).[33] To obtain annotation of each single cell in the read count matrix, Supplemental Table 4 describing lineage specification of cells was downloaded from Stirparo et al.[142] and the “Stage”, ”sex” and “Assigned.lineage..original.study” columns were used to assign the stage, sex and lineage for each single cell in the original data. Autosomal genes were defined as overlapping or not with conserved autosomal XIST peaks from naïve hPSCs, and the sum of read counts in each single cell was calculated for autosomal gene overlapping with conserved autosomal XIST peaks and normalized by the sum of read counts of all autosomal genes of that single cell, providing normalized sum of read counts for genes associated with conserved autosomal XIST peaks in each single cell. To obtain normalized expression levels of individual genes (Figure 4L), a pseudobulk approach was used. For this, read counts for each gene and single cell were used and cpm scores calculated using cpm from the edgeR package. Genes with low read counts were removed (keeping genes with cpm > 0.5 in at least two cells). Single cells were grouped by their stage, lineage and sex, and the sum of read counts in each group was calculated. Regularized log transformation of each gene in each group was calculated using rlog from DESeq2 package.[98]

Assessment of pluripotency state

Published naïve pluripotency markers[33,48] were used to confirm the successful conversion of primed hESCs to the naïve pluripotency state. Specifically, the DotPlot function (using dot.scale=8) from the Seurat package[101,102] was used to explore expression of the pluripotent markers in the single cell data, where the size of the circle indicates the percentage of cells in a cluster expressing the marker and the color indicates average expression levels across all cells in the cluster.

Definition X-dosage compensation classes

Gene expression fold change between female and male cells was obtained for the following comparisons: 1) Female and male epiblast cells from day 6 and 7 pre-implantation from Petropoulos et al.[33]; 2) Female naïve H9 hESCs vs male naïve HNES1 hESCs from Rostovskaya et al.[47] (the day 0 time point of the capacitation time course, which represents the naïve state); 3) Female naïve hPSCs (H9, UCLA1, and iPSC) vs male naïve hESCs (WIN1); and 4) Female naïve HNES3 hESC line vs the male naïve HNES1 hESC line. Fold changes in each comparison were calculated as described above. X-linked genes detected in all comparisons (377 genes) were used to cluster genes into four X-dosage compensation classes using kmeans, resulting in 79, 95, 127 and 76 genes in classes 1 through 4 (Figure 4A). Published classifications[85] were used to assign genes as escaping XCI or subject to XCI, with 20, 11, 3 and 8 genes escaping XCI within the X-dosage compensation classes 1 through 4, which represents 25, 12, 2 and 11% of genes in each class. The *XIST* gene was excluded from this analysis.

Haplotype phasing

For UCLA1 hESCs, a list of known X-linked Single-Nucleotide Polymorphisms (SNPs) was obtained from Sahakyan et al.[34] The genotype coordinates (in hg19) in the original VCFs were transferred to the hg38 genome built through the liftOver function in CrossMap.[104] Since the SNP data in UCLA1 hESCs were taken form unphased genotypes, a read overlapping a known SNP could not be directly assigned to one of the two parental X chromosomes (haplotypes). To directly explore gene expression from each X chromosome, phasing haplotypes analysis, which identify the variant alleles that are co-located on the same chromosome, was done. For this, RNA-seq data from primed UCLA1 hESCs which display non-random XCI[34] were used together with the list of known SNPs. Due to the non-random silencing of one of the two X chromosomes, all mono-allelically expressed alleles were assigned to the Xa, while the not expressed alleles were assigned to the Xi in primed hESCs (note that the primed UCLA1 line carries a slightly eroded Xi,[44] allowing us to phase only SNPs within silenced genes that have not become reactivated on the eroded Xi). To prevent confusion, we refer to the Xa as X1 and the Xi as X2. Through this process, SNPs were assigned to either the X1 or X2 in primed UCLA1 hESCs. Specifically, for haplotype phasing we followed the following steps. Step 1: For each SNP the allelic coverage was determined in primed UCLA1 hESC RNA-seq data. To this end the SAMtools mpileup tool[96] was used to generate SNP coverage pileups, using min-BQ = 20, and max-depth = 1000000. For each SNP, the read coverage of the reference and alternative allele were calculated. 2) Each allele was assigned to the Xa (X1) and Xi (X2), by focusing on X-linked SNPs with mono allelic expression (for which the total number of reads > 5 and the absolute ratio of reference to alternative allele < 0.2 (80% of reads aligned to either the reference or alternative allele). In cases where the reference allele overlapped with more than 80% of reads, the reference allele was assigned to the Xa while the alternative allele was assigned to the Xi, and vice versa. This approach provided a list of alleles detected on the Xa (X1) and Xi (X2) for UCLA1 (39 X-linked phased SNPs).

Determination of allelic X-linked expression

To begin to understand whether X-chromosome dosage compensation observed in naïve hPSCs occurs on the XIST-expressing X chromosome, X-linked gene expression was examined at allelic resolution. SAMtools mpileup tool[96] was used to generate SNP coverage pileups for bulk RNA-seq datasets of naïve hESCs (UCLA1) (5 RNA-seq replicates), using min-BQ = 20, and max-depth = 1000000. The ratios of the number of reads covering the X1 and X2 extracted from the coverage pileups were calculated for each SNP.

Gene ontology analysis

Enrichment for gene ontologies was performed using the R package topGO[112] with ontology=''BP'', mapping=''org.Hs.eg.db'' and statistic=''fisher''. For human gene analyses, a subset of autosomal genes with detectable gene expression in female (UCLA1, H9 and iPSC) and male (WIN1) hPSCs (using cpm > 0.5 in at least two samples as describe above) was used as a background (17,906 genes). Ontologies enrichment was tested for autosomal genes overlapping with conserved autosomal XIST peaks of female naïve hPSCs and detected as expressed. For the ontology of autosomal XIST target genes in human and mouse, autosomal genes that overlapped both conserved autosomal XIST peaks of female naïve hPSCs and mouse day 2 EpiLC peaks (D2-specific and D2 & D4 peaks) were used against a background of all human and mouse autosomal orthologs.

3D contact frequencies data analysis

To explore the 3D contact frequencies of the X chromosome in human (Figure S1G), Xa and Xi Hi-C contacts (.hic files) in GM12878 somatic cells were downloaded from Rao et al.[49] (GEO: GSE63525) (https://data.4dnucleome.org/files-processed/4DNFIYECESRC/#file-overview). For intra-chromosomal Xa and Xi interactions, Knight-Ruiz (KR) normalized observed-over-expected contact matrixes were generated using Juicebox dump command of the Juicebox tool.[114] Distance normalized interaction signals (observed/expected) were calculated at 1Mb resolution for the X chromosomes.

To explore the inter-chromosomal interactions of the human *XIST* locus with autosomal windows (Figures 3K and 3L), Hi-C contacts (.hic files) in undifferentiated H9 hESCs were downloaded from Zhang et al.[51] (GEO: GSE116862) (https://data.4dnucleome.org/files-processed/4DNFIRGUR82F/#file-overview). Observed contact matrix was generated using Juicebox dump command of the Juicebox tool[114] in 1Mb resolution. Inter-chromosomal interactions of the *XIST* locus with autosomal windows were extracted and split into windows overlapping conserved XIST peaks, and windows that do not. Similar results obtained using GM12878 somatic cells[49] (data not shown).

Similarly, to explore the inter-chromosomal interactions of the mouse *Xist* locus with autosomal windows (Figure 7K), SPRITE data of mouse embryonic stem (mES) cells was used. Specifically, inter-chromosomal interactions maps from mES were downloaded from Quinodoz et al.[86] at 1Mb resolution (using none_1000_iced normalization). Inter-chromosomal interactions of the *Xist* locus with autosomal windows were extracted and split into windows overlapping D2 Xist peaks, and windows that do not.

ChIP-seq analysis

Reads from ChIP-seq experiments were mapped to the human genome (hg38) using Bowtie2 software[105] and only those reads that aligned to a unique position with no more than two sequence mismatches were retained for further analysis. Multiple reads mapping to the exact same location and strand in the genome were collapsed to a single read to account for clonal amplification effects.

Similar to the RAP-seq analysis, read counts in 100kb windows every 25kb were extracted for each modification and sample. To account for differences in sequencing depth, read counts in each genomic region were normalized to the sum of all reads in that modification and sample. The ratios of the normalized read count in the ChIP-seq pulldown and the input of each sample were calculated and used as enrichment scores. Similar to the RAP-seq analysis, to prevent overlapping windows in downstream analysis, the enrichment scores were assigned to the centered 25kb of each window.

To explore histone mark enrichment around genes in each four X-dosage compensation classes, read per million (RPM) within 1000 bps up and downstream of gene TSSs were calculated using ScoreMatrixList function of the genomation package.[117] Enrichment of histone marks in each X-dosage compensation class was calculated as the Log_2(fold change) comparing the median score (RPM) of genes in each class to the median of genes in the other three classes, for each modification. Wilcoxon p-values comparing the RPM of genes assigned to each class versus the RPM of genes assigned to the other three classes was used. For this analysis, histone marks extracted from ChIP-seq of male naïve hESCs (WIN1) were used.

ChromHMM modeling parameters

To derive chromatin state segmentations for male naïve WIN1 hESCs and female naïve iPSC, ChromHMM was used (version v1.18)[116] with default parameters and the ChIP-seq data as described above. First, reads were binarized into 200 base pair windows using the BinarizeBam command for all chromatin marks with a p-value cutoff of 0.0001 and using the hg38 genomic as a reference. To reduce effects of artifacts, redundancy in the input data was removed by keeping only one sequencing read in cases where multiple reads mapped to the same genomic position and strand orientation. Multiple models with different numbers of states ranging from 10 to 30 were examined and a model with 22 chromatin states that is both interpretable and able to capture the combinatorial complexity of chromatin marks was selected. Candidate annotations were assigned based on the combination of enriched histone marks and known genomic features, including genes bodies, transcription start site (TSS), transcription end site (TES), exons and CpG islands. Genomic locations exhibiting aberrant enrichment for all chromatin marks that reside within known backlisted genomic regions (ENCODE blacklist regions) were marked as repetitive state. The chromatin state annotations include: Prom A: Active promoter; Prom P: Poised promoter; Enh A: Active enhancer; Tx EnhA: transcribed active enhancer; Tx 5': transcribed 5'; Tx 3': transcribed 3'; Poly R: PolyComb Repressed; Het: Heterochromatin; Het Transc: Heterochromatin transcribed; Low: Low signal; Repetitive: Repetitive region.

CUT&Tag analysis

Alignment of the CUT&Tag reads was done similar to the alignment of the XIST/Xist RAP-seq reads described above. Read counts in 100kb windows every 25kb were extracted for each sample. To account for differences in sequencing depth, read counts in each genomic region were normalized to the sum of all reads in that sample. The ratios of the normalized read count in the CUT&Tag and the IgG control of each sample were calculated and used as enrichment scores. For gene enrichment analysis, RPMs of H3K27me3 were calculated in 1000 bps up and downstream of gene TSSs as described for ChIP-seq enrichment analysis. Three approaches were used to explore differential H3K27me3 deposition between female naïve WT and XIST KO H9 hESCs: 1) To investigate whether loss of H3K27me3 signal upon XIST KO occurs at specific chromatin states (Figure 6P), read counts within each ChromHMM chromatin state window (200bps) were calculated. Differential H3K27me3 analysis was done using DESeq2[98] between CUT&Tag reads of female naïve WT and XIST KO hESCs (both clones), and the $\log_2$ fold change in each window was used for the downstream analysis. 200bps windows with less than 10 reads in all samples were removed. Windows were divided into those that overlap with conserved autosomal XIST peaks and those that do not. Windows were further divided by their overlap with different ChromHMM chromatin states extracted from male naïve hESCs (WIN1) described above. For each state, Wilcoxon test was used to compare H3K27me3 $\log_2$ fold changes for regions overlapping with conserved XIST peaks or not. 2) To explore whether the loss of H3K27me3 signal upon XIST KO occurs at regions marked by XIST-independent H3K27me3 (Figure S7N), H3K27me3 peak calling was first performed using CUT&Tag experiment in female naïve XIST KO hESCs. The peak calling was performed using MACS2 callpeak[108] using broad and the IgG as a control sample. Next, the differential H3K27me3 scores between female naïve XIST WT and KO hESCs from 1) were compared between regions (200bp windows) that overlapped or not with H3K27me3 peaks in female naïve XIST KO hESCs. 3) To identify genes that lost H3K27me3 upon XIST KO (Figure S7M), read per million (RPM) within 1000 bps up and 1000 bps down-stream of gene TSSs were calculated using ScoreMatrixList function of the genomation package.[117] RPMs from multiple replicates were averaged, and the RPM fold change between female naïve WT and XIST KO hESC was calculated for each gene.

Histone marks enrichment on the X chromosome

To explore if different histone marks are enriched or depleted on the X chromosome (Figures 6A and 6F), the percentage of reads aligned to each chromosome was calculated for each sample and modification in the ChIP-seq and CUT&Tag. To account for differences in chromosome sizes, normalized counts were extracted by dividing the percentage of reads aligned to each chromosome, by the percentage of reads aligned to that chromosome in the corresponding input. Z-scores were calculated by comparing the normalized counts for the X chromosome to those of all autosomes. P-values of the Z-scores, representing the probability of obtaining specific Z-scores considering one-tail distribution, were also calculated.

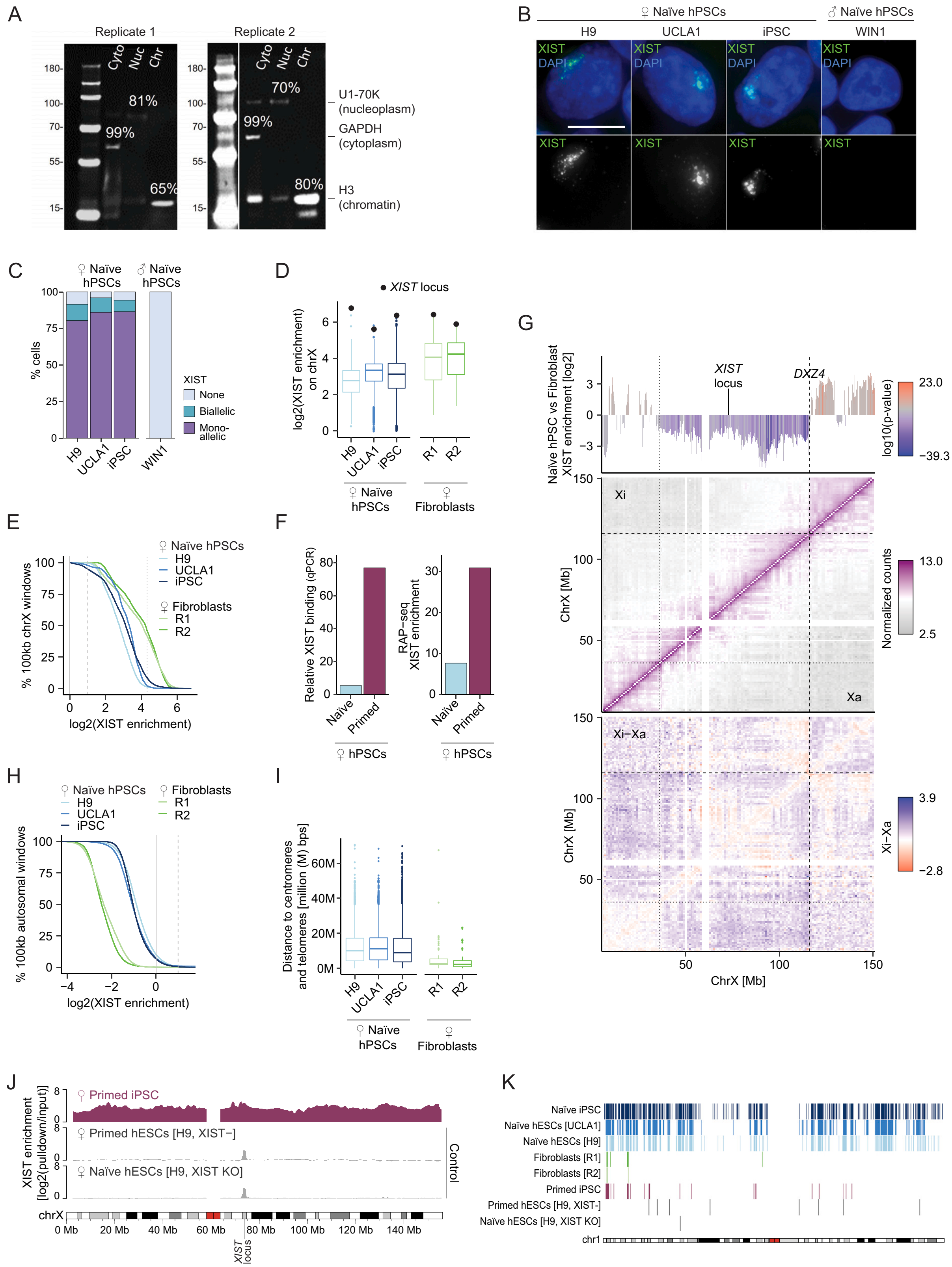
H3K37me3 and XIST enrichment clustering

Enrichment scores along the X chromosome (100kb windows, every 25kb) extracted from CUT&Tag H3K27me3 data in female naïve WT H9 hESCs, the corresponding XIST KO cells (both clones) and male naïve WIN1 hESCs, and XIST RAP-seq data in the female naïve hPSC lines H9, UCLA1 and iPSC, were scaled and clustered into 5 groups using pheatmap function of the ComplexHeatmap package[118] with kmeans_k=5. 24, 18, 24, 13 and 21% of windows on the X chromosome were assigned to clusters 1 through 5. To explore if the X-chromosome clusters are marked by specific chromatin states on the active X chromosome (i.e. without XIST-dependent regulation), we paired the XIST/H3K27me3 clusters with chromatin annotations from naïve male hESCs (Figures S7F and S7G). We also paired the clusters with dosage compensation state annotations from Figure 4A and gene expression data (Figure S7L).

Similarly, autosomal windows overlapping conserved XIST peaks were clustered into four groups based on CUT&Tag H3K27me3 data in female naïve WT H9 hESCs, the corresponding XIST KO cells (both clones) and male naïve WIN1 hESCs, with kmeans_k=4. 27, 22, 32 and 19% of XIST-targeted conserved autosomal windows are in clusters I through IV. A gene was assigned to a cluster for downstream gene analyses based on the location of its TSS.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. XIST localization in female naive hPSCs and human fibroblasts, related to Figures 1 and 2

(A) Western blots confirming successful fractionation of female naive hESCs (UCLA1) into cytoplasmic (Cyto), nucleoplasmic (Nuc), and chromatin (Chr) compartments marked by GAPDH, U1-70K (SNRP70), and histone H3, respectively. Intensity percentages for each are given. Results for the two replicate experiments used in this manuscript are shown.

(B) Representative RNA FISH images for XIST (green) in female naive hESC lines H9 and UCLA1, a female naive hiPSC line, and the male naive hESC line WIN1. DAPI staining (blue) marks nuclei. Scale bars, 10 μm.

(C) Quantification of the percentage of cells from (B) with biallelic, monoallelic, or no XIST signal.

(D) Boxplot of XIST enrichment for all 100 kb windows (every 25 kb) along the X chromosome in the indicated cell lines. R1 and R2 represent the two independent replicate XIST RAP-seq experiments for fibroblasts. The genomic window containing the *XIST* locus is highlighted.

(E) Cumulative distribution plot of XIST enrichment across all 100 kb windows (every 25 kb) on the X chromosome from XIST RAP-seq data of indicated cell lines. The solid vertical line marks no enrichment (no difference between pull-down and input); the dashed line a 2-fold enrichment; and the dotted line a 20-fold enrichment of XIST pull-down over input.

(F) Relative XIST binding at a specific region on the X chromosome (chrX: 17,608,575–17,608,725; hg38) in naive female H9 hESCs (light blue) and primed XIST-positive hiPSCs (fuchsia) determined by RAP-qPCR (left) and RAP-seq-based XIST enrichment for the 1 kb window overlapping the region amplified by the qPCR primers (right). The lower XIST level in naive compared to primed hESCs suggests a lower accumulation of XIST on the Xd compared with the Xi. However, additional quantitative experiments are needed to explore the difference in overall XIST enrichment between the Xd and Xi.

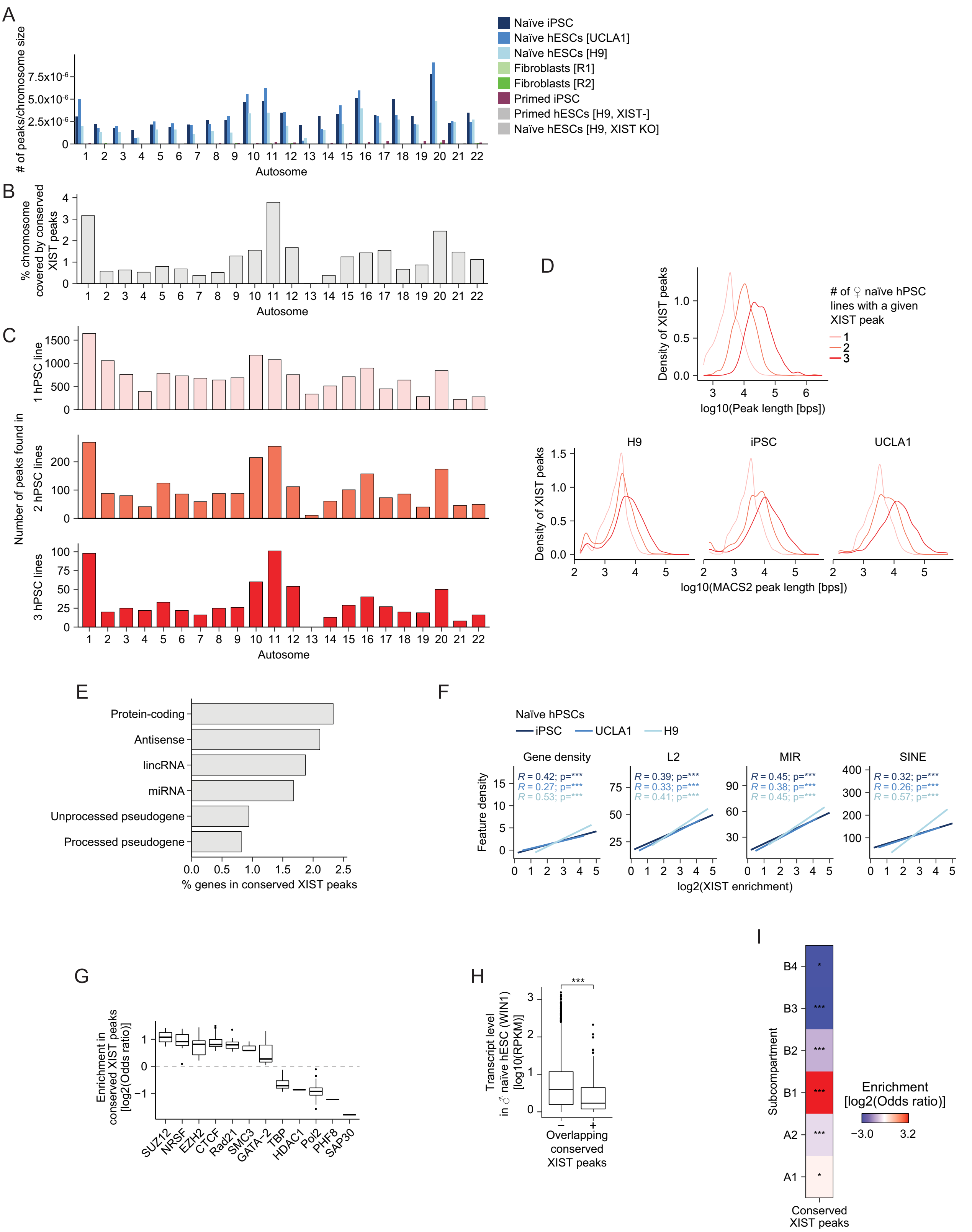
(G) Top: comparison of the relative level of XIST and its distribution along the X chromosome between female naive hPSCs (H9, UCLA1, and hiPSCs) and female human fibroblast replicates (using DiffBind on X-linked RAP-seq reads; see STAR Methods). Color represents $\log_{10}(p)$ indicating the confidence of the differences in XIST binding in each region and is positive (red) in genomic regions with higher XIST enrichment in naive hPSCs and negative (blue) when it is higher in fibroblasts. Only regions with significant differences in XIST enrichment ($p < 0.01$) are presented. The *XIST* locus is also indicated. This analysis identified an 80 Mb region across the center of the X chromosome that is preferentially associated with XIST in fibroblasts. Conversely, the $\sim$34 Mb region at the 3′ end of the X is preferentially bound by XIST in naive hPSCs. The switch in XIST localization between these two regions occurs at the macro-satellite repeat locus *DXZ4*, which is known to partition the Xi in somatic cells into two spatial super domains[42,43,49,143–145] and marked by the dashed vertical line. The distribution across the first $\sim$36 Mb of the X chromosome (before the dotted vertical line) is more similar between naive hPSCs and fibroblasts, which may be related to the fact that many escapers of XCI are concentrated in this region. Middle: normalized Hi-C contacts on the Xi (top half) and Xa (bottom half) in the human female somatic cell line GM12878. Bottom: normalized Hi-C contact differences between the Xi and Xa (top triangle mirroring the bottom one). The comparison of the Hi-C maps and XIST enrichment indicates that differences in 3D chromatin organization between the Xi and Xa take place at the genomic regions on the X chromosome that display the switch in XIST distribution between naive hPSCs and fibroblasts. The intra-chromosomal interactions in the 5′ 36 Mb region are more similar between the Xa and Xi of somatic cells than elsewhere on the X chromosome, consistent with the similarities in XIST distribution. Together, this analysis suggests that chromatin structure plays a role in the differential localization of XIST along the X chromosome in different cell states.

(H) As in (E), except for autosomal XIST enrichment.

(I) Boxplot for the closest distance of autosomal XIST MACS2 peaks (see STAR Methods) to centromeric or telomeric regions in *cis* in indicated RAP-seq datasets.

(J) XIST enrichment along the X chromosome in female primed XIST-expressing hiPSCs (fuchsia) and two control samples without XIST, including female primed H9 hESCs lacking XIST expression due to Xi erosion (gray; middle) and female naive XIST KO H9 hESCs (gray; bottom). Unmappable regions are masked in white and the centromeric region is highlighted in red. The single peak in the two control samples represents the genomic *XIST* locus.

(K) Location of autosomal XIST peaks (colored ticks) along chromosome 1 based on RAP-seq data from indicated female cell lines.



(legend on next page)

Figure S2. Characterization of autosomal XIST localization, related to Figure 3

(A) Bar graph of the number of autosomal XIST peaks on each autosome in indicated female cell lines, normalized by chromosome size.

(B) Percentage of base pairs of each autosome covered by conserved XIST peaks.

(C) Number of XIST peaks per autosome for naive hPSC-specific peaks present in only one (top), two (middle), or all three (bottom; representing conserved autosomal XIST peaks) naive hPSC lines.

(D) Density plot showing the length (in base pairs) of naive hPSC-specific autosomal XIST peaks detected in one, two, or three female naive hPSC line(s) from (C) (top). The density plots below show the length of the respective MACS2 peaks in each cell line (see STAR Methods for the definition of MACS2 peaks).

(E) Percentage of autosomal genes in indicated categories overlapping with conserved autosomal XIST peaks.

(F) Linear regression line between XIST enrichment along the X chromosome (1 Mb windows every 250 kb) for indicated female naive hPSC lines and the density of genes, L2, MIR, and SINE elements within the same genomic interval. Pearson correlation (R) and p value (***$p < 0.001$) are given.

(G) Enrichment ($\log_2$(odds ratio)) of binding sites of indicated regulators in conserved autosomal XIST peaks vs. the rest of the autosomal genome, based on ENCODE data from diverse cell types and tissues.

(H) Boxplots of transcript levels in male naive hESC (WIN1) for genes overlapping (+) with conserved autosomal XIST peaks or not (−). Wilcoxon p value: ***$p < 0.001$.

(I) Enrichment ($\log_2$(odds ratio)) of conserved autosomal XIST peaks in different Hi-C sub-compartments extracted from the GM12787 cell line.[49] Fisher's exact test p values: *$p < 0.05$, ***$p < 0.001$.

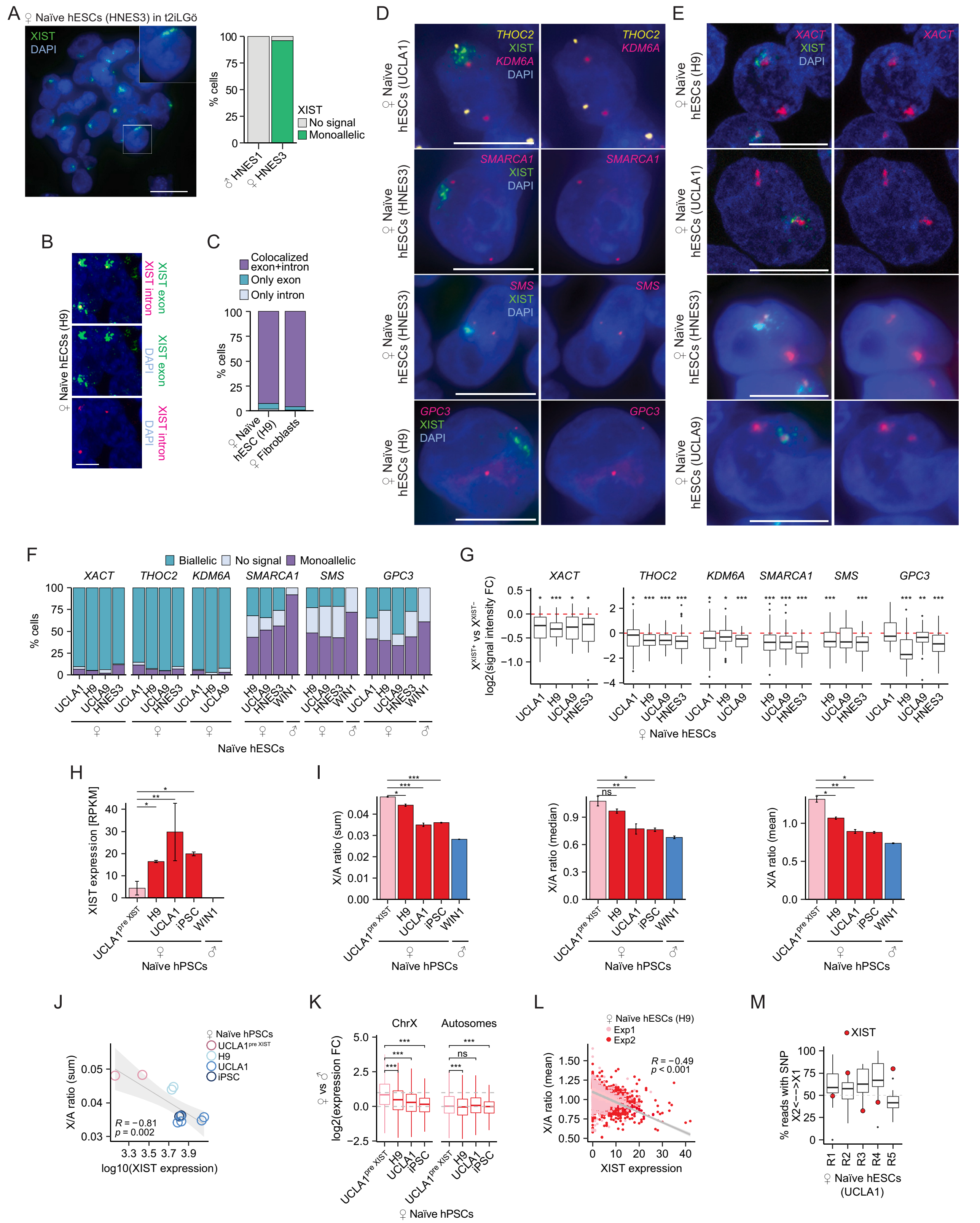


Figure S3. Characterization of gene expression regulation on the XIST-coated X chromosome in female naive hPSCs, related to Figure 4
(A) Left: representative image for XIST RNA FISH (green) in the female naive hESC line HNES3, with DAPI marking nuclei. Scale bars, 20 μm. Right: quantification of the percentage of cells with monoallelic or no XIST signal in naive male (HNES1) and female (HNES3) hESC lines. The HNES1 and HNES3 hESC lines were derived directly from blastocysts in the t2iLGö culture medium.[54]

(legend continued on next page)

(B) Representative single-molecule RNA FISH image detecting exonic (green) and intronic (pink) XIST sequences in female naive H9 hESCs, with DAPI marking the nuclei, shown merged or separated into channels. Scale bars, 5 μm.

(C) Quantification of the percentage of cells with colocalization of the XIST intronic and exonic signals from (B), indicating that the intronic signal is associated with the exonic signal accumulation in the vast majority of cells, consistent with the localization of XIST on the X chromosome in *cis* in naive hESCs, i.e., on the X chromosome that the RNA is transcribed from.

(D) Representative RNA FISH images for XIST (green) and nascent transcription spots of the X-linked genes *THOC2* and *KDM6A* (first row), *SMARCA1* (second row), *SMS* (third row), and *GPC3* (fourth row) in indicated female naive hPSC lines (marked on the left), with DAPI demarcating the nuclei. Scale bars, 10 μm. The right column shows only the nascent transcription spot channel, for clearer visualization.

(E) Representative RNA FISH images for XIST (green) and the X-linked lncRNA XACT (pink) in the female naive hPSC lines H9, UCLA1, HNES3, and UCLA9, as indicated on the left. Scale bars, 10 μm. DAPI demarcates the nuclei.

(F) Quantification of the expression pattern of XACT and the nascent transcription patterns for *THOC2*, *KDM6A, SMARCA1*, *SMS*, and *GPC3* in indicated female hPSC lines and the male naive hESC line WIN1, based on RNA FISH experiments for the respective X-linked transcript together with XIST. Examples of images on which these counts were based, are shown in (D) and (E). The proportion of cells with no signal, monoallelic, and biallelic signal is given. These quantifications confirm that both X chromosomes are expressed in female naive hPSCs and, therefore, that female naive hPSCs do not display XCI. *XACT*, *THOC2*, and *KDM6A* displayed biallelic signals in most cells. *GPC3*, *SMARCA1*, and *SMS* all displayed biallelic expression, but in a smaller fraction of female naive cells, and otherwise were monoallelically expressed or not detected. It is noteworthy that these results correlate with the X-dosage compensation class assignment (defined in Figure 4A) for these genes, with *GPC3* and *SMARCA1* belonging to class 1 (most strongly dosage-compensated X-linked genes), *SMS* to class 2, and *THOC2* and *KDM6A* to class 4 (least dosage-compensated X-linked genes) (Figure S4A). Moreover, RNA FISH for *SMARCA1*, *SMS*, and *GCP3* in the male naive hESC line WIN1 showed that the nascent transcription signal of these genes, particularly of *SMS* and *GPC3*, was not detected in a fraction of male cells, suggesting that their transcript levels are below the threshold of detection for our RNA FISH approach or that transcription occurs in bursts. Given this result, one would not expect biallelic expression of *SMS*, *SMARCA1*, and *GPC3* in all female cells, which is exactly the result we obtained. Thus, the monoallelic expression of these X-linked genes in a subset of female cells likely is due to detection limitations or transcriptional bursting.

(G) Boxplots of the $\log_2$ fold change (FC) in intensity of the XACT signal and the nascent transcript signals of the X-linked genes *THOC2*, *KDM6A*, *SMARCA1*, *SMS*, and *GPC3* between the XIST-covered (XIST+) and XIST-negative (XIST−) X chromosome in indicated female naive hPSC lines based on RNA FISH experiments discussed in (D)–(F). The dashed line (y = 0) represents no difference in signal intensity between the two X chromosomes. Negative y values represent lower signal on the XIST+ than the XIST− X chromosome. Wilcoxon p values indicate that the signals on the XIST+ are significantly lower than on the XIST− X chromosome: *$p < 0.05$, **$p < 0.001$, ***$p < 0.001$.

(H) XIST expression level in indicated female naive hPSC lines (UCLA1$^{pre\text{-}XIST}$, H9, UCLA1, and iPSC), and the male naive hESC line WIN1, based on bulk RNA-seq data. For this, five replicates for naive UCLA1 hESCs and two replicates each for naive H9 hESCs, naive hiPSCs, naive UCLA1$^{pre\text{-}XIST}$ hESCs, and naive WIN1 hESCs were exploited (Table S2). Naive UCLA1$^{pre\text{-}XIST}$ hESCs have very low levels of XIST as they represent an early naive state in the primed-to-naive conversion where both X chromosomes are active and XIST expression is not yet strongly induced.[34] Mean expression and error bars corresponding to the SD between replicates are shown. t test p values: *$p < 0.05$, **$p < 0.01$. These data confirm that male naive hESCs do not express XIST and that naive UCLA1$^{pre\text{-}XIST}$ hESCs have a lower XIST level than fully established female naive hPSCs (UCLA1, H9, and iPSC).

(I) Ratios of reads aligned to X-linked and autosomal genes (X/A ratios) for the bulk RNA-seq datasets described in (H), based on the sum (left), median (middle), or mean (right) of read counts. For each bar, the mean ratio of replicates and the associated SD are shown. t test p values: ns $\geq$ 0.05, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. These data show that the male naive hESC line WIN1 not expressing XIST has the lowest and naive UCLA1$^{pre\text{-}XIST}$ with two active Xs the highest X/A ratio. XIST-expressing female naive hPSC lines have a reduced X/A ratios approaching that of the male naive hESC line. The female naive hPSC lines do not reach the X/A ratio of male naive hESCs, consistent with dampening of X-linked gene expression occurring on only one of the two X chromosomes due to monoallelic XIST expression.

(J) Scatter plot comparing the X/A ratio, based on the sum of read counts of X-linked and autosomal genes, with XIST expression for bulk RNA-seq datasets from female naive hPSC lines described in (H). Pearson correlation (R) and p value, and linear regression line with 95% confidence interval (gray background) are given. This plot shows that the X/A ratio negatively correlates with XIST levels.

(K) Boxplots of the $\log_2$ fold change (FC) of X-linked (chrX) and autosomal gene expression between the female naive hPSC lines indicated on the x axis and the male naive hESC line WIN1, based on bulk RNA-seq data described in (H). Lower dashed line (y = 0) represents no differences in gene expression between the respective female line and male cells. Upper dashed line (y = 1) represents double dosage in female cells compared with the male naive hESC line. Wilcoxon p values: ns $\geq$ 0.05, ***$p < 0.001$. These data show that the lowered X/A ratio in XIST-expressing female naive hPSCs is due to a decrease of X-linked gene expression relative to autosomal gene expression.

(L) Scatter plot of the X/A ratio (based on the mean expression of X-linked vs. autosomal genes) and XIST expression in individual cells of the female naive WT hESC line H9, based on scRNA-seq data from two independent primed-to-naive conversion experiments. Pearson correlation (R) and p value as well as the linear regression line are given. This plot shows that individual female naive hPSCs show a decrease of the X/A ratio with increasing XIST levels, validating and extending the bulk RNA-seq data above. Overall, the data in (H)–(L) support a role for XIST in dampening X-linked gene expression in female naive hPSCs and suggest that the degree of XCD is sensitive to the level of XIST.

(M) To assess in bulk RNA-seq data whether X chromosome dosage compensation observed in naive hPSCs occurs on the XIST-expressing X chromosome, we examined X-linked gene expression at allelic resolution in female naive UCLA hESC. Reads from the X chromosomes were split according to allele-specific single-nucleotide polymorphisms (SNPs), phased, and assigned to a given X chromosome (X1 or X2), where X1 is the active X chromosome (Xa) and X2 the Xi in primed UCLA1 hESCs, which display non-random XCI. We then assigned X1 or X2 status to XIST and X-linked gene expression in each of the five bulk RNA-seq replicates of naive UCLA1 hESCs. The boxplots show the percentage of reads aligned to X1 and X2. Red dots represent reads with SNPs overlapping the XIST transcript. These data show that, when XIST expression is skewed toward one of the two X chromosomes, X-linked gene expression is biased toward the other X. Given the *cis*-localization of XIST (B) and (C), these data suggest that XIST expression is correlated with lower expression of the same chromosome. These data are consistent with the RNA FISH data in (G) that demonstrate that the XIST-associated X chromosome is more lowly expressed than the other X chromosome and support the conclusion that XIST mediates XCD in female naive hPSCs.

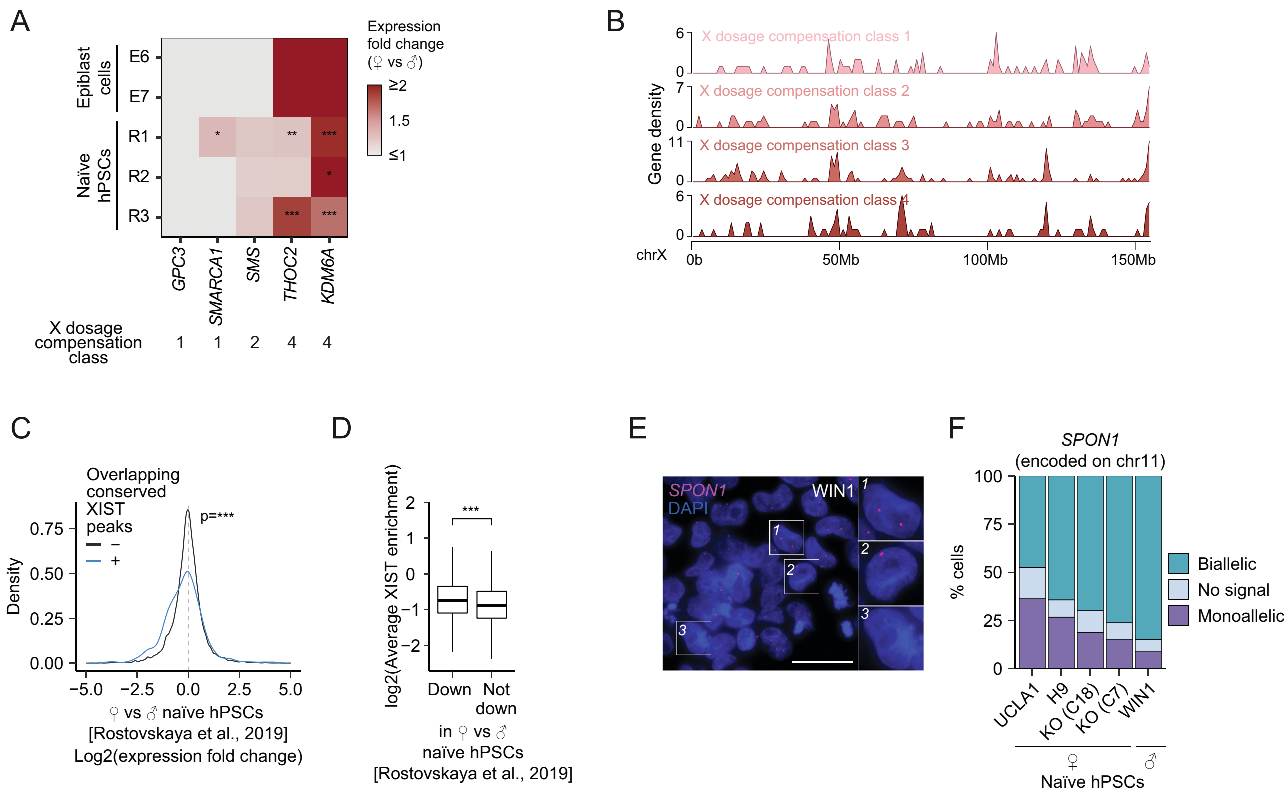


Figure S4. Dampening of X-linked and autosomal XIST target genes in female naive hPSCs, related to Figure 4

(A) Heatmap showing the differences in gene expression between female and male epiblast cells of E6 or E7 pre-implantation embryos or female and male naive hPSCs (R1–R3 as defined in Figure 4A) for the indicated X-linked genes, for which we analyzed the nascent transcript signal by RNA FISH in this study. The X-dosage compensation class (defined in Figure 4A) for each of these genes is given below. Wilcoxon p values: *$p < 0.05$, **$p < 0.01$ ***$p < 0.001$.

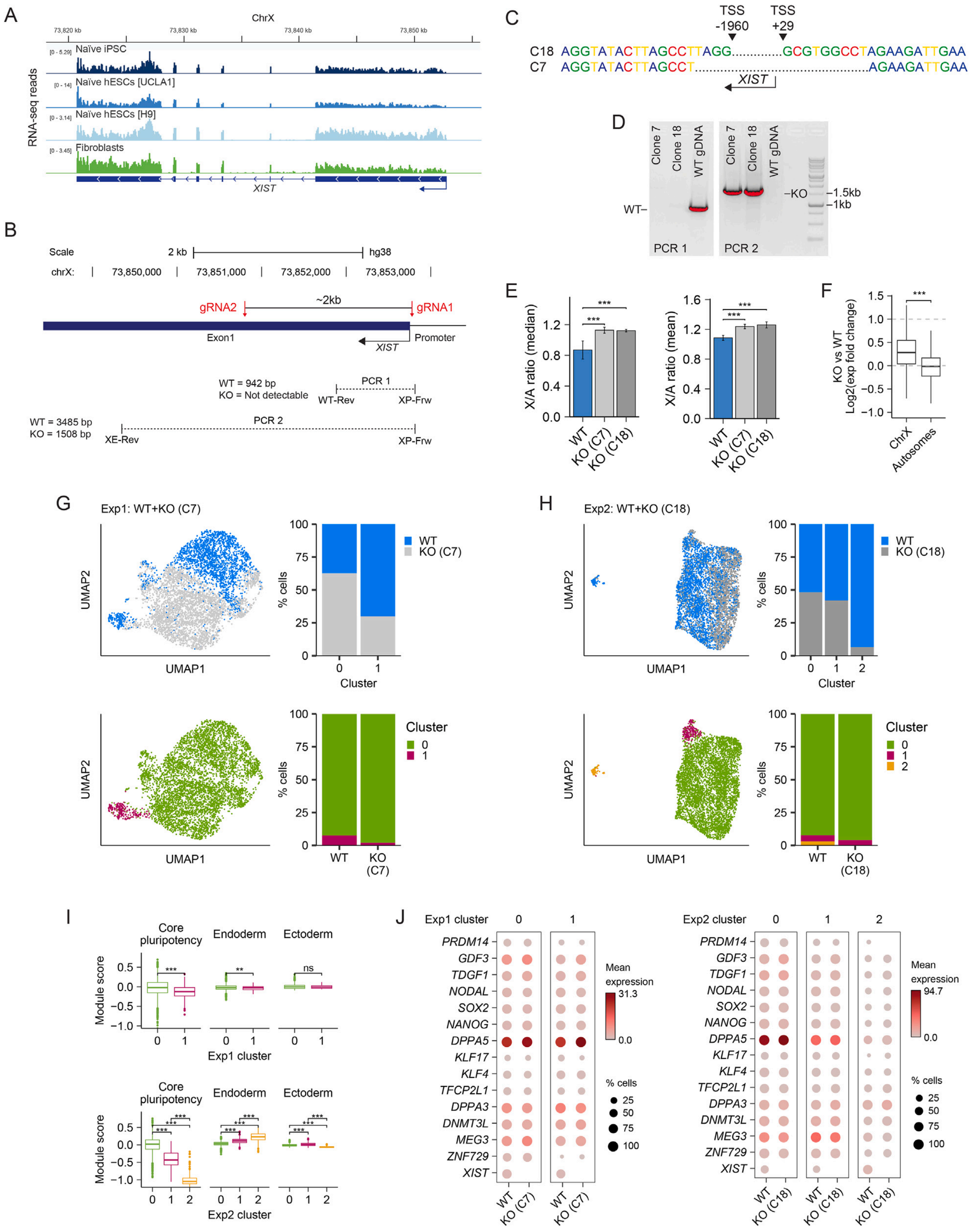
(B) Density of genes assigned to each of the four X-dosage compensation classes (from Figure 4A) across the X chromosome.

(C) Density plot of differences in autosomal gene expression between female (H9) and male (HNES1) naive hESCs from published data[47] for genes overlapping (+) or not overlapping (−) with conserved autosomal XIST peaks. Wilcoxon test p value indicates that autosomal genes with conserved XIST peaks are more lowly expressed in female than male hESCs: ***$p < 0.001$. The dashed line marks no differences in gene expression.

(D) Boxplots of the average XIST enrichment in female naive hPSCs at autosomal genes that have significantly lower expression in female H9 compared with male HNES1 naive hESCs (referred to as "down") and those genes that are not significantly downregulated ("not down"), using the same published data as in (C).[47] Wilcoxon p value: ***$p < 0.001$. Outliers are omitted for clearer visualization.

(E) Representative RNA FISH image for nascent transcription foci of the autosomal XIST target gene *SPON1* (encoded on chr11) in the male naive hESC line WIN1, with DAPI demarcating the nuclei. Marked nuclei are enlarged on the right and represent different nascent expression patterns observed for *SPON1*. Scale bars, 20 μm.

(F) Quantification of the proportion of cells with no, monoallelic, or biallelic *SPON1* nascent transcript signals for the female naive WT hESC lines UCLA1 and H9, the two female naive XIST KO hESC clones (C18 and C7), and the male naive hESC line WIN1. Representative images for each pattern are shown in (E).



(legend on next page)

Figure S5. Characterization of naive XIST KO hESCs, related to Figure 5

(A) Genome tracks showing bulk RNA-seq reads assigned to the genomic *XIST* locus from female naive hPSCs (iPSC, UCLA1, and H9) and human female fibroblast data. In the *XIST* gene model shown below exons are indicated with thicker lines and the direction of transcription is given with triangles. These data show that all expected exons of XIST are transcribed in naive hPSCs, similar to female fibroblasts with an Xi, and suggest that the major transcriptional start site in naive hPSCs is the same as in fibroblasts.

(B) Schematic representation of the *XIST* knockout strategy. CRISPR-Cas9 was used to homozygously excise a ~2 kb region encompassing a small part of the promoter region and the 5′ end of exon1 of the *XIST* transcript. The *XIST* transcription direction and cutting sites of the gRNAs are given. The location of primers used for PCR genotyping (PCR1 and PCR2) and the length of the expected PCR fragments for WT and KO alleles are given below.

(C) Sequence of the PCR product of the XE-Rev and XP-Frw primers (displayed in B as PCR2) around the deletion site for both XIST KO clones (C7 and C18) in primed H9 hESCs is given. The deleted region is marked by dots and the position relative to XIST's transcriptional start site (TSS) given.

(D) PCR genotyping of the primed H9 XIST KO clones 7 and 18. PCR1: the XP-Frw and WT-Rev primer set amplifies a 942 bp region in in WT cells. As the region complementary to WT-Rev is not present upon XIST deletion, no PCR band is obtained for XIST KO clones (see primer location in B). PCR2: the XP-Frw and XE-Rev primers displayed in (B) amplify a ~1.5 kb region from genomic DNA of KO clones. For this PCR, we chose a short extension time that does not enable the amplification of the WT allele in WT cells (~3.5 kb).

(E) Bar graph of the X/A ratio of naive WT and XIST KO H9 hESCs based on median (left) or mean (right) gene expression values from bulk RNA-seq data. Each bar represents the mean ratio of replicate bulk RNA-seq data sets with error bars giving the SD. Wilcoxon p value: ***$p < 0.001$.

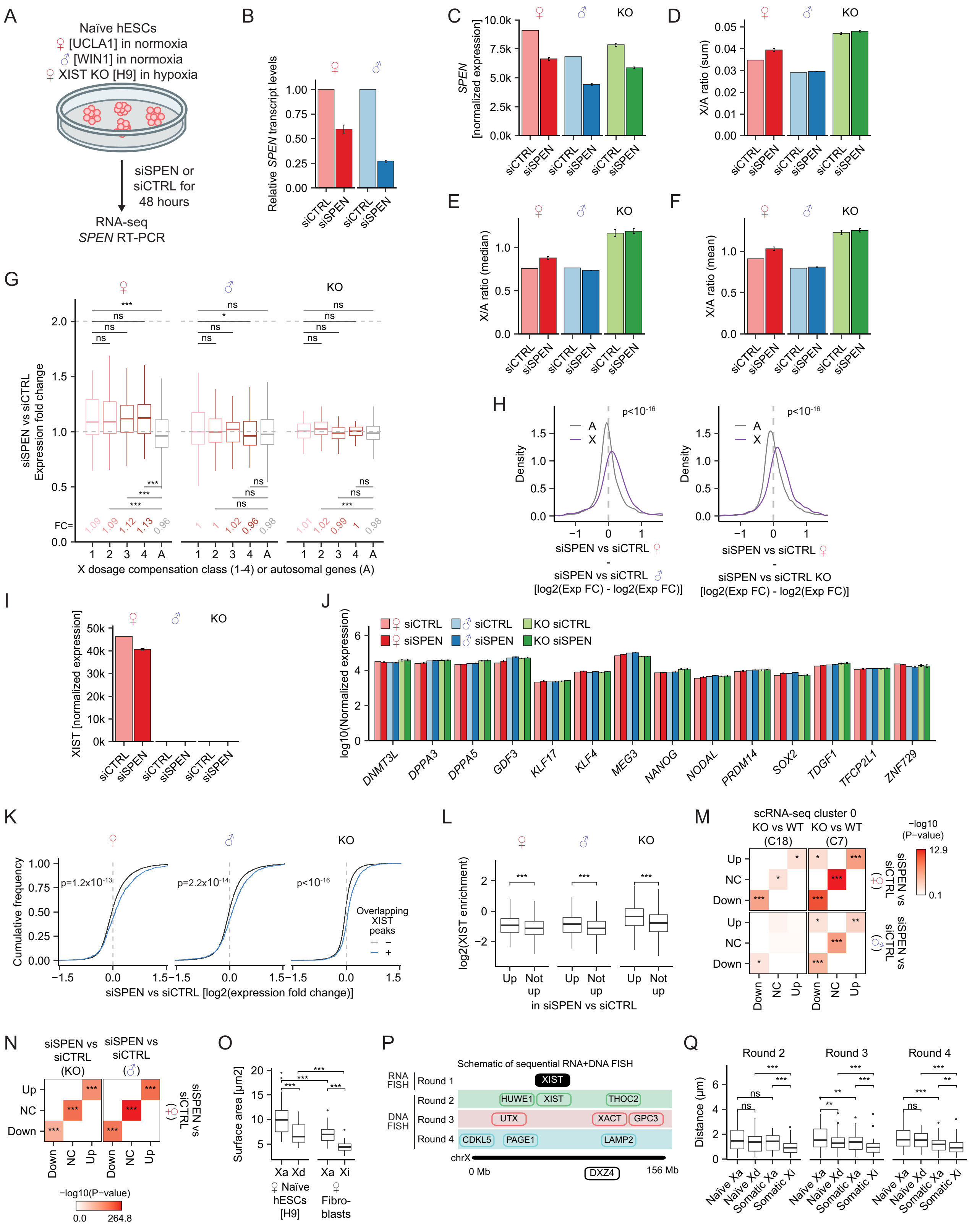
(F) Boxplots of the $\log_2$ gene expression fold change between female naive XIST KO and WT H9 hESCs for X-linked and autosomal genes, based on bulk RNA-seq data. Lower dashed line ($y = 0$) represents no differences in gene expression, upper dashed line ($y = 1$) represents double dosage in XIST KO compared with WT cells. Wilcoxon p values: ***$p < 0.001$.

(G) Uniform manifold approximation and projection (UMAP) of scRNA-seq data from female naive WT and XIST KO (clone 7) H9 hESCs. Primed WT and XIST KO clone 7 H9 hESCs were converted in parallel to the naive pluripotent state and analyzed by scRNA-seq (experiment 1; Exp1). In the top UMAP, cells are colored by origin (WT or XIST KO) and in the bottom UMAP by cluster assignment. The proportion of WT and XIST KO cells in each cluster (top right) and the proportion of each cluster for WT or XIST KO cells (bottom right) are quantified. The vast majority of naive WT and XIST KO cells reproducibly cluster together (>92% of WT and XIST KO cells are in cluster 0).

(H) Similar to (G) except that results for experiment 2 (Exp2) are shown, where primed WT and XIST KO clone 18 H9 hESCs were converted to the naive pluripotent state and then analyzed by scRNA-seq. In this case, we obtained 3 clusters, with cluster 0 capturing the majority of cells.

(I) Boxplots of core pluripotency, endoderm, and ectoderm module scores (using cell makers defined previously[140]) for individual naive WT and XIST KO hESCs, divided by clusters defined in (G) and (H). The small proportion of WT and XIST KO cells assigned to clusters 1 and 2 express naive pluripotency markers at lower and differentiation signatures at higher levels. Wilcoxon p values: ns ≥ 0.05, **$p < 0.01$, ***$p < 0.001$.

(J) Dotplot representation of naive pluripotency marker and *XIST* expression based on scRNA-seq data from (G) on the left and (H) on the right, split by cluster. Color indicates the average expression of the indicated gene in that cluster, and the size of the dot the proportion of cells expressing that gene.



(legend on next page)

Figure S6. Characterization of mechanisms that contribute to gene dampening on the Xd and at XIST-targeted autosomal regions in naive hPSCs, related to Figure 5

(A) Schematic representation of the *SPEN* depletion experiments.

(B) Relative transcript levels of *SPEN* in female UCLA1 and male WIN1 naive hESC lines upon treatment with siCTRL or si*SPEN*, based on qPCR. Mean transcript levels with error bars representing standard deviation (SD) of replicate data are shown.

(C) Normalized *SPEN* expression in naive female UCLA1, male WIN1, and female XIST KO H9 hESCs upon siCTRL or si*SPEN* treatment, based on bulk RNA-seq data. Each bar represents the mean expression from replicate data sets and error bars give the SD. k = ×1,000.

(D) X/A ratios based on the sum of read counts for the same conditions as in (C). Each bar represents the mean ratio of replicates and error bars give the SD.

(E) As in (D), except that the X/A ratio was formed based on median expression values.

(F) As in (D), except that the X/A ratio was formed based on mean expression values.

(G) Boxplots of the fold change of gene expression between si*SPEN* and siCTRL treatments in female, male and in female KO hESCs described in (C), for X-linked genes, split into X-dosage compensation class (1–4, defined in Figure 4A) and autosomal genes (A), based on bulk RNA-seq data. The lower dashed line (y = 1) represents no difference in gene expression. The median fold change (FC) for each gene group is given below the boxes. Wilcoxon p values: ns $\geq$ 0.05, *p < 0.05, ***p < 0.001. These data show that *SPEN* knockdown in male hESCs and XIST KO cells does not dramatically change of the expression of X-linked genes, whereas X-linked genes across all X-dosage compensation classes are significantly upregulated upon *SPEN* depletion in female WT cells. These data suggest that SPEN is required for gene dampening on the X only in cells with an XIST-coated Xd. Notably, *SPEN* depletion in naive female WT hESCs affects gene expression in all X-dosage compensation classes similarly. This result is different from the findings with XIST KO cells (Figure 5H), where X-dosage compensation classes have different susceptibility to XIST loss. Specifically, genes in the X-dosage compensation class 4 are similarly upregulated in XIST KO and *SPEN* depletion, whereas genes with strong dosage compensation (class 1) are more strongly upregulated upon XIST KO compared with SPEN depletion.

(H) Left: density plots of the change in expression between si*SPEN* and siCTRL treatment in naive female WT hESCs compared with naive male WT hESCs, for X-linked (purple) and autosomal (gray) genes. Right: as on the left, except for the comparison between naive female WT hESCs and XIST KO hESCs. The dashed line marks no changes in gene expression upon si*SPEN* treatment. Wilcoxon p values comparing X and autosomal expression are given. This figure shows that the upregulation of X-linked genes upon si*SPEN* treatment is higher in naive female WT hESCs than in naive male or female XIST KO hESCs.

(I) As in (C), except for XIST transcript levels.

(J) As in (C), except for indicated naive pluripotency genes.

(K) Cumulative distribution plots of gene expression changes between si*SPEN* and siCTRL treatments in naive female WT (UCLA1), male WT (WIN1), and female XIST KO (H9) hESCs for autosomal genes that overlap with autosomal XIST peaks (+; blue) or not (−; gray). For the female and male WT comparisons, we used autosomal UCLA1 hESC-derived XIST peaks, and for the XIST KO hESC analysis autosomal XIST peaks from the naive hESC line H9. Wilcoxon test p values are given and indicate that expression of autosomal genes associated with XIST peaks is upregulated upon si*SPEN* treatment in all cell lines.

(L) Boxplots of XIST enrichment around autosomal genes that are significantly upregulated (up) or not significantly upregulated (not up) upon si*SPEN* treatment (compared with siCTRL) in naive female WT (UCLA1), male WT (WIN1), and female XIST KO (H9) hESCs. For the naive female (UCLA1) and male (WIN1) WT hESC analyses, XIST RAP-seq enrichment scores from the naive hESC line UCLA1 were used and for the XIST KO H9 analysis, XIST RAP-seq enrichment scores from naive H9 hESCs. Wilcoxon p value: ***p < 0.001.

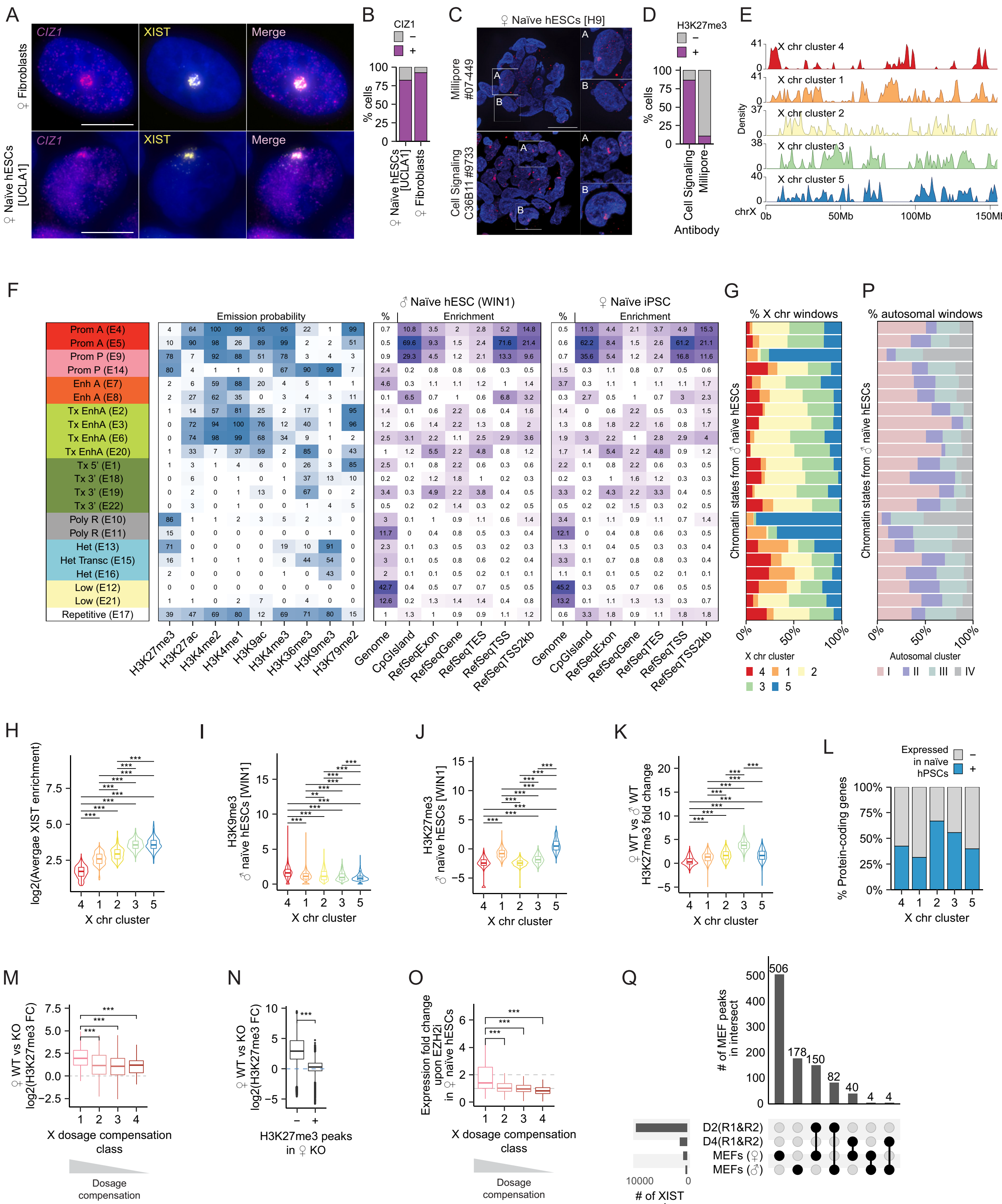
(M) Heatmap of hypergeometric tests for overlap between autosomal genes that are significantly upregulated (up), do not change (NC), or are significantly downregulated (down) upon *SPEN* siRNA treatment (relative to siCTRL) in naive female (UCLA1) or male (WIN1) WT hESCs (based on bulk RNA-seq data, indicated on the right) and upon deletion of XIST in naive H9 hESCs (two clones, XIST KO vs. WT, based on scRNA-seq, marked on the top). Asterisks represent significant p values: *p < 0.05, **p < 0.005, ***p < 0.0005, and the color indicates p values.

(N) As in (M), except for genes regulated upon *SPEN* depletion in naive female WT (UCLA1) hESCs (right) vs. genes regulated upon *SPEN* depletion in naive male (WIN1) hESCs or naive female XIST KO hESCs (top).

(O) Boxplots of the surface area of the two X chromosomes (annotated as Xd, Xa, and Xi) in individual female naive H9 hESCs and fibroblasts based on X paints. The Xd and Xi were identified by XIST RNA FISH as the chromosome associated with the RNA. Wilcoxon p values: ***p < 0.001.

(P) Schematic representation of our sequential RNA/DNA FISH approach on the female naive hESC line H9 and female human fibroblasts. XIST RNA FISH was applied first to identify the Xd and Xi, respectively, in each cell. Then, three sequential rounds of DNA FISH were performed, each targeting three different genomic loci along the X chromosome as indicated.

(Q) Analysis of the RNA/DNA FISH data described in (P). Boxplots of the pairwise distances between the DNA signals in each DNA FISH hybridization round, which were measured in 20 naive H9 hESCs and 16 fibroblast cells. Wilcoxon p values: ns $\geq$ 0.05, **p < 0.01, ***p < 0.001. These data show that the Xi is more compact than the Xa in somatic cells or the Xd and Xa in naive hESCs. It also shows that the Xa in naive hESCs is less compacted than the Xa in fibroblasts consistent with differentiation-induced chromatin compaction.[146]



(legend on next page)

Figure S7. Characterization of XIST-mediated chromatin changes on the X chromosome and autosomes in naive hPSCs, related to Figures 6 and 7

(A) Representative images for XIST RNA FISH combined with immunostaining for CIZ1 in female human fibroblasts and female naive hESCs (UCLA1), demonstrating that CIZ1 accumulates on the XIST-associated Xd in female naive hPSCs, mirroring results on the Xi in fibroblasts.

(B) Quantification of the proportion of cells with CIZ1 accumulation under the strongest section of the XIST signal, i.e., on the Xd in naive hESCs and the Xi in fibroblasts, from data shown in (A).

(C) Representative immunostaining images with indicated antibodies against H3K27me3, in naive female hESCs (H9).

(D) Quantification of the proportion of cells from (C) showing an H3K27me3 accumulation. Prior studies yielded different findings with respect to the accumulation of H3K27me3 under the XIST domain in female naive hPSCs. Specifically, several studies reported the lack of an accumulation of H3K27me3 on the XIST-associated X chromosome in naive hESCs[35,123] and human pre-implantation embryos.[32,124] In contrast, H3K27me3 accumulation on the XIST-coated chromosome was described by our group in female naive hESCs in a prior study.[34] We reasoned that a possible reason for this discrepancy might be the use of different antibodies in the respective immunostaining experiments since Vallot et al.,[35] An et al.,[123] and Teklenburg et al.,[124] utilized the H3K27me3 antibody from Millipore (cat#07-449), whereas we used the antibody from Active Motif (cat#39155) in the Sahakyan et al. study.[34] To directly investigate whether different antibodies detect the X chromosome enrichment of H3K27me3 differently, we compared the Cell Signaling (cat#9733) and Millipore (cat#07-449) H3K27me3 antibodies in this female naive H9 hESCs immunostaining experiment. The staining was performed in parallel, with the same secondary antibody. 86% of cells showed an enrichment of H3K27me3 on the Xd in naive H9 hESCs when the Cell Signaling antibody was used, whereas only 10% showed a H3K27me3 accumulation with the Millipore antibody. Importantly, in female naive XIST KO hESCs, the accumulation of H3K27me3 observed with the Cell Signaling antibody is completely lacking (Figures 6B and 6C), emphasizing that the observed H3K27me3 enrichment on the Xd in naive hPSCs is XIST-dependent. Both H3K27me3 antibodies (Millipore and Cell Signaling) detect the Xi-enrichment of H3K27me in somatic cells in the vast majority of cells (not shown), and showed global loss of signal upon EZH2 inhibition in naive and primed H9 hESCs supporting their specificity.[75,125] Interestingly, a recent paper revealed the higher specificity of the Cell Signaling antibody compared with the Millipore one.[126] We also found that the Active Motif (cat#39155) and Cell Signaling antibodies both detect the XIST-dependent H3K27me3 accumulation on the Xd in ChIP-seq and CUT&Tag experiments (Figure 6). Taken together, we conclude that the H3K27me3 accumulation on the Xd is consistently detected with different H3K27me3 antibodies and methods (ChIP-seq, CUT&Tag, and immunostaining). Yet, certain antibodies are not well suited for the analysis of the H3K27me3 enrichment on the Xd in female naive hPSCs.

(E) Density of genes assigned to each of the five X chromosome clusters described in Figure 6H across the X chromosome.

(F) ChromHMM analysis for histone marks from naive female and male WT hPSCs. Left: heatmap showing 22 chromatin states defined by ChromHMM, color-coded and grouped based on their putative annotation (the label in brackets represent the ChromHMM assigned state number). Annotations include: Prom A, active promoter; Prom P, poised promoter; Enh A, active enhancer; Tx EnhA, transcribed active enhancer; Tx 5′, transcribed 5′; Tx 3′, transcribed 3′; poly R, polycomb repressed; Het, heterochromatin; Het Transc, heterochromatin transcribed; Low, low signal; Repetitive, repetitive region. Middle: frequency of each histone mark for each chromatin state (ChromHMM emission probabilities), colored from white to blue for increasing proportions. Right: percentage of genome occupancy (most left "genome" column) and enrichment of indicated features (TES, transcription end site; TSS, transcription start site) in each chromatin state, for male naive hESC (WIN1) and female naive iPSCs.

(G) Bar graph of the percentage of the respective chromatin state in male naive WIN1 hESCs (same order as in F, in 200 bp windows) overlapping with each of the five X chromosome clusters from Figure 6H. Genomic regions belonging to X chromosome clusters 4 and 1 are most strongly pre-marked with heterochromatin markers, whereas genomic regions of cluster 5 are polycomb repressed in male naive hESCs. Genomic regions belonging to cluster 2 capture the largest fraction of the active chromatin states in male naive hESCs.

(H) Violin plots for XIST enrichment (average of all three females naive hPSC lines) for the five X chromosome clusters from Figure 6H. Wilcoxon p values: ***$p < 0.001$. Outliers were omitted for clearer visualization.

(I) Violin plots for H3K9me3 enrichment in male naive hESCs (WIN1) for the five X chromosome clusters from Figure 6H. Wilcoxon p values: **$p < 0.01$, ***$p < 0.001$. Outliers were omitted for clearer visualization. Genomic regions belonging to clusters 4 and 1 are pre-marked with the repressive histone mark H3K9me3 in male naive hESCs.

(J) As in (I), except for H3K27me3.

(K) As in (I), except for the difference in H3K27me3 enrichment between female (iPSCs) and male (WIN1) naive hPSCs. Genomic regions belonging to cluster 3 show the highest H3K27me3 deposition in female WT cells compared with male.

(L) Bar graph of the percentage of protein-coding genes that are expressed in naive hPSCs or not for each of the five X chromosome clusters from Figure 6H (genes were assigned to a cluster based on the location of the gene's TSS). The few genes assigned to cluster 4 and 1 (Figure 6I) are relatively lowly expressed, supporting their constitutive heterochromatic state. Similarly, more than half of the genes associated with cluster 5 are not detected as expressed, consistent with their polycomb-repressive state in (G). However, those that are expressed tend to be strongly dosage-compensated (Figure 6J).

(M) Boxplots of the $\log_2$ fold change (FC) of H3K27me3 between female naive WT and XIST KO H9 hESCs for genes in each of the four X-dosage compensation classes from Figure 4A. H3K27me3 levels were obtained for the 2 kb window surrounding a gene's TSS (1 kb upstream and downstream of a gene's TSS). Wilcoxon p value: ***$p < 0.001$. Dashed vertical line represents no change in H3K27me3 between naive female WT and XIST KO cells. The dosage compensation extent between female and male cells (from Figure 4A) is given with the triangle below. These data show that promoter regions of the most dosage-compensated X-linked genes (X-dosage compensation class 1) have the highest gain of H3K27me3 in female WT compared with KO cells. This finding is consistent with the fact that genes in this class often belong to X chromosome cluster 3 (Figure 6J) with the largest female WT hESC-specific gain of H3K27me3 as seen in (K).

(N) Boxplots of $\log_2$ fold change (FC) of H3K27me3 between female naive WT and XIST KO H9 hESCs within 200 bp windows along the X chromosome, divided into windows that do or do not overlap with H3K7me3 peaks identified in naive female XIST KO hESCs. Wilcoxon p value: ***$p < 0.001$. Dashed vertical line represents no change in H3K27me3 between WT and XIST KO cells. These data show that unmarked H3K27me3 regions preferentially increase H3K27me3 levels in an XIST-dependent manner.

(O) Boxplots of the fold change in gene expression upon EZH2 inhibition (EZH2i) in naive female WT hESCs from published data[75] for genes in each of the four X-dosage compensation classes from Figure 4A. Lower dashed line ($y = 1$) represents no differences in gene expression between control and EZH2i treatment, upper dashed line ($y = 2$) represents double dosage upon EZH2i. Wilcoxon p values: ***$p < 0.001$.

(P) As in (G), except for the overlapping with the four autosomal clusters in Figure 6L.

(Q) Analysis of MEF Xist peaks. Graph showing the number of autosomal Xist peaks detected in female or male MEF RAP-seq data and their overlap with each other or with day 2 (D2) and day 4 (D4) Xist RAP-seq peaks. The total number of peaks detected in day 2 (union of the two replicates), day 4 (union of the two replicates), female and male MEFs is shown on the left. Please note that most day 2 and day 4 peaks are not detected in MEFs.