Cell

Site-Specific Silencing of Regulatory Elements as a Mechanism of X Inactivation

J. Mauro Calabrese,[1] Wei Sun,[1,2] Lingyun Song,[3] Joshua W. Mugford,[1] Lucy Williams,[1,4] Della Yee,[1] Joshua Starmer,[1] Piotr Mieczkowski,[1] Gregory E. Crawford,[3] and Terry Magnuson[1,*]

[1]Department of Genetics, the Carolina Center for Genome Sciences, and the Lineberger Comprehensive Cancer Center

[2]Department of Biostatistics

University of North Carolina, Chapel Hill, NC 27599, USA

[3]Institute for Genome Sciences and Policy, and Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, North Carolina 27708, USA

[4]Present address: Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

*Correspondence: trm4@med.unc.edu

http://dx.doi.org/10.1016/j.cell.2012.10.037

SUMMARY

The inactive X chromosome's (Xi) physical territory is microscopically devoid of transcriptional hallmarks and enriched in silencing-associated modifications. How these microscopic signatures relate to specific Xi sequences is unknown. Therefore, we profiled Xi gene expression and chromatin states at high resolution via allele-specific sequencing in mouse trophoblast stem cells. Most notably, X-inactivated transcription start sites harbored distinct epigenetic signatures relative to surrounding Xi DNA. These sites displayed H3-lysine27-trimethylation enrichment and DNaseI hypersensitivity, similar to autosomal Polycomb targets, yet excluded Pol II and other transcriptional hallmarks, similar to nontranscribed genes. CTCF bound X-inactivated and escaping genes, irrespective of measured chromatin boundaries. Escape from X inactivation occurred within, and X inactivation was maintained exterior to, the area encompassed by *Xist* in cells subject to imprinted and random X inactivation. The data support a model whereby inactivation of specific regulatory elements, rather than a simple chromosome-wide separation from transcription machinery, governs gene silencing over the Xi.

INTRODUCTION

X chromosome inactivation (XCI) equalizes X-linked gene dosage between mammalian sexes, resulting in transcriptional silencing of one of two female X chromosomes during early development. XCI is critical for mammalian development, and epigenetic processes required for XCI, most notably gene silencing mediated by Polycomb group proteins and noncoding RNA, play important roles in many biological phenomena (Surface et al., 2010). As such, XCI is a paradigm for epigenetic silencing mediated by noncoding RNA.

Two waves of XCI occur in the mouse. The first, imprinted XCI, initiates at the eight-cell stage of development and results in inactivation of the paternally inherited X chromosome (Kalantry et al., 2009; Patrat et al., 2009). Imprinted XCI is maintained in extraembryonic cells, whereas cells from the inner mass reactivate their paternal X during blastocyst maturation (Williams et al., 2011). XCI then reoccurs within the inner cell mass, randomly selecting the paternal or maternal X for silencing (Escamilla-Del-Arenal et al., 2011).

The inactive X chromosome (Xi) is distinguished from autosomes by several salient features. An ~17 kb noncoding RNA, *Xist*, is expressed from and coats the Xi in *cis*, and is required for the early maintenance of XCI (Kalantry et al., 2009; Namekawa et al., 2010). *Xist* coating results in widespread Xi deposition of H3-lysine27-trimethylation (H3K27me3), catalyzed by the Polycomb Repressive Complex 2 (PRC2). After coating, the Xi can be visualized microscopically with antibodies recognizing H3K27me3 or PRC2 components (Mak et al., 2002; Plath et al., 2003; Silva et al., 2003). PRC2 is required for maintenance of XCI during differentiation of extraembryonic lineages (Kalantry et al., 2006; Wang et al., 2001) and acts redundantly with PRC1 to maintain XCI in the embryo (Schoeftner et al., 2006).

The Xi's physical territory is microscopically devoid of transcription-associated hallmarks, including RNA polymerase II (Pol II), histone H3-acetylation, and histone H3-lysine4-methylation (H3K4me) (Escamilla-Del-Arenal et al., 2011). Exclusion of these marks from the Xi's territory is *Xist*-dependent and occurs during initiation of XCI. Movement of genes into the territory is coincident with silencing (Chaumeil et al., 2006). Whether relocation itself causes gene silencing or is simply correlated with XCI is unclear (Escamilla-Del-Arenal et al., 2011).

The content of X-linked DNA is additionally noteworthy. Approximately 35% of human and mouse X-linked DNA is derived from LINE repeats, compared to ~20% of autosomal

DNA (Fujita et al., 2011). Due to this enrichment, LINEs were thought to be conduits for *Xist* and associated silencing factors as they coat the Xi, although recent work argues a more indirect role for LINEs in this process (Tattermusch and Brockdorff, 2011). In this regard, LINEs have been proposed to nucleate formation of a transcriptionally silent spatial core within the Xi, into which X-linked genes are recruited as they are silenced (Chaumeil et al., 2006; Chow et al., 2010; Namekawa et al., 2010).

Finally, although the majority of X-linked genes are silenced by XCI, a minority escapes X inactivation. The proportion and identity of escaping genes differs between cell types and ranges from 3% to 15% of X-linked genes (Carrel and Willard, 2005; Patrat et al., 2009; Yang et al., 2010). Mechanistic models suggest escaping genes are positioned exterior to the Xi's silent domain, in contrast to X-inactivated genes, allowing escapers to efficiently access transcription machinery (Chaumeil et al., 2006). The CTCF insulator protein may also play a critical role in licensing escape (Filippova et al., 2005).

Although well studied on a microscopic level, there is little quantitative information regarding Xi chromatin at submicroscopic resolution. Understanding the epigenetic states of individual regulatory elements over the Xi is critical to a complete mechanistic understanding of XCI. Therefore, via a combination of allele-specific RNA-, ChIP-, FAIRE-, and DNase-Seq, we profiled X-linked chromatin patterns at the submicroscopic scale in mouse trophoblast stem cells (TSCs), which are subject to imprinted XCI. The resulting gene expression and chromatin maps solidify TSCs as a platform for understanding the maintenance of XCI in a stem cell population. Our analysis revealed unexpected properties of X-inactivated and escaping genes, both in terms of their epigenetic signatures, as well as their subnuclear localization patterns in TSCs and also in cells subject to random XCI. Together, our results suggest a model whereby the major mechanism of transcriptional silencing associated with maintenance of XCI is not a simple chromosome-wide separation from transcription machinery but rather localized occlusion of Pol II from specific sites along the Xi.

RESULTS

Quantitative Allele-Specific Expression Map of the TSC Xi

In order to study XCI in a natural context and still differentiate between the active X (Xa) and Xi when analyzing gene expression and chromatin patterns, we chose to study XCI in female TSCs, where inactivation patterns are nonrandom. Because TSCs maintain imprinted XCI, their Xa and Xi are maternally and paternally inherited, respectively, and SNP-overlapping sequence reads in these cells reliably trace chromosome of origin (Quinn et al., 2006).

As a prerequisite for understanding the relationship between Xi chromatin patterns and gene silencing, we began our study by measuring allelic gene expression via strand-specific RNA-Seq in female TSC lines derived from crosses between CAST/EiJ (Cast) and C57BL/6J (B6) mice. Considering our goal of relating expression to Xi chromatin patterns, and that the most accurate build of the genome is B6-derived, downstream chromatin analyses focused on a TSC line with a Cast Xa and B6 Xi (C/B). RNA-Seq was performed on an additional TSC line carrying a B6 Xa and Cast Xi (B/C), in order to help differentiate between strain- and parent-of-origin-specific expression biases.

The two female TSC lines were selected for RNA-Seq based on normal karyotypes and expression of TSC-specific markers (data not shown). TSC poly-adenylated RNA was collected, and replicate cDNA libraries were prepared and sequenced (Ingolia et al., 2009; Table S1 available online). SNP-overlapping reads were identified, and allelic expression was calculated by dividing the number of B6- or Cast-overlapping reads by the total number of allelic reads per gene. The resulting expression maps served as a necessary baseline for the subsequent interpretation of X-linked chromatin patterns.

Internal consistency and experimental validation indicated RNA-Seq data to be of high quality. R^2 values comparing gene expression between C/B and B/C replicates were 0.970 and 0.985, respectively; this same comparison between the two TSC lines gave an R^2 of 0.917. Moreover, nine genes were selected for validation via a quantitative allele-specific RT-PCR assay and in all cases measured allelic ratios and sequencing data were concordant (Figure S1).

We next determined what genes were subject to and escaped TSC XCI, examining genes with $\geq$20 SNP-overlapping reads in both TSC lines. Previous analyses used a cutoff of 10% Xi expression to differentiate between X-inactivated and escaping genes (Carrel and Willard, 2005; Yang et al., 2010). We sought to empirically define TSC escape, reasoning that a 10% cutoff might exclude true escapers. Allelic expression data were fit to a mixture of beta binomial distributions, allowing us to determine the probability of escape per gene (Extended Experimental Procedures). For each gene, the proportion of Xi expression, the total number of escaping reads, and the sequence quality scores of SNP-overlapping bases within these reads were accounted for in escape determination.

Thirteen percent of genes (35 from 262 eligible) escaped XCI to varying degrees (Figure 1; Tables S2 and S3). Escape did not occur in predictable patterns along the Xi; genes escaped in isolation or in groups. With *Xist* as the major exception, escapers remained predominantly Xa expressed, displaying a median of ~15% expression from the Xi. Twenty-seven escapers exhibited at least 10% Xi expression. In contrast, X-inactivated genes had a median of ~0.5% Xi expression, and there was no correlation between inactivation strength and gene distance from *Xist* (R^2 0.07 and 0.04 for C/B and B/C, respectively; Figure 1).

Genes that escaped XCI in TSCs also escaped in other tissues. We analyzed allelic expression via PCR in day 6.5 embryonic tissue and found the four genes tested escaped in the ectoplacental cone and extraembryonic ectoderm (data not shown). Additionally, about 1/2 of escapers identified in mouse tissues subject to random XCI also escaped TSC XCI; 7 of 13 escapers from embryonic kidney cells (Yang et al., 2010), and 10 of 20 escapers from neural precursor cells (Splinter et al., 2011), escaped in TSCs. Finally, we noted that Xi genetic background significantly affected TSC escape profiles. About

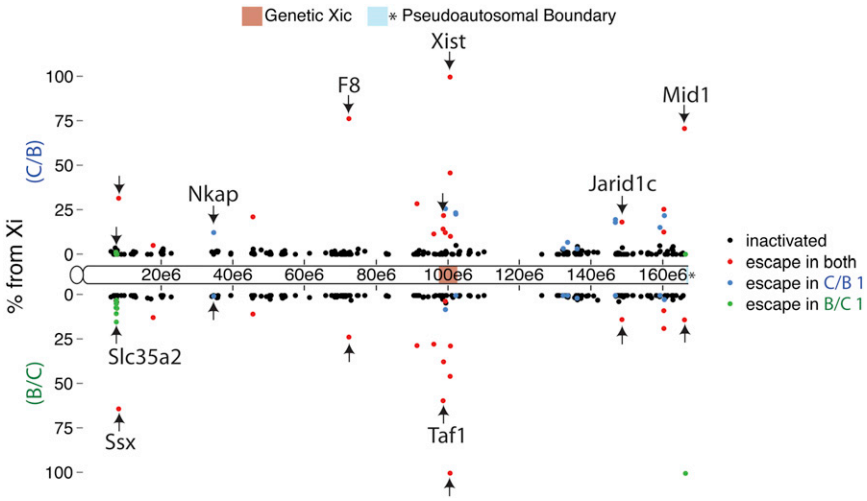


Figure 1. Allele-Specific Expression Map of the TSC Xi

Dots denote X location and XCI status of evaluated TSC genes. y axis is percentage Xi expression in TSC lines C/B (positive scale) and B/C (negative scale). Arrows mark representative escapers. See also Figure S1 and Tables S1, S2, and S3.

1/2 of escapers did so in both TSC lines; however, 11 genes escaped only in C/B TSCs and 8 only in B/C TSCs (blue and green dots, Figure 1). Allelic PCR in independently derived TSC lines verified that the majority of differential escape was heritable and not TSC line specific (Figure S1).

High-Resolution Analysis of TSC Xi Chromatin Environments

Having established an allelic expression map of the TSC X chromosome, we initiated our analysis of X-linked epigenetic patterns at high resolution via ChIP-Seq in C/B TSCs. The overarching goal of these experiments was to characterize the submicroscopic chromatin signatures over the TSC Xi and quantitatively examine how these signatures related to gene expression, DNA features, and known microscopic properties of the chromosome (Escamilla-Del-Arenal et al., 2011). The survey initially centered on H3K27me3, which is required for XCI maintenance in extraembryonic cells (Kalantry et al., 2006; Wang et al., 2001). Additional ChIP-Seq data sets were generated for repression-associated histone H4-lysine20-monomethylation (H4K20me1), which also microscopically coats the TSC Xi, as well as marks associated with active transcription, such as Pol II, H3K4me2, and histone H3-lysine36-trimethylation (H3K36me3), and total histone H3 (H3).

Both autosomal and X-linked ChIP-Seq data showed expected allelic distributions. Although the Cast-to-B6 allelic ratios were approximately 1:1 in all data sets, they showed a slight bias toward the reference genome (Figure 2A). In contrast, allelic biases were skewed over the X chromosomes in a manner consistent with data from classical immunofluorescence (IF) microscopy experiments (Escamilla-Del-Arenal et al., 2011); active marks were biased toward the Cast genome, or Xa, and repressive marks were biased toward the B6 genome, or Xi (Figure 2B).

The tiling densities of H3K27me3 and H4K20me1 were next examined in order to quantitatively address how their microscopic Xi enrichments related to specific chromosomal regions (Figure 2C). Both marks exhibited large-scale density fluctuation over the X with megabase-sized regions of enrichment and depletion (Figure 2C, right). These X-linked profiles had a Spearman coefficient of 0.88, higher than their autosomal correlation of 0.46. X-linked H3K27me3 and H4K20me1 patterns also contrasted with chromosome 1 patterns, for example, where enriched regions were more punctate (Figure 2C, left). Importantly, allelic data showed the major patterns of X-linked H3K27me3 and H4K20me1 enrichment reflected those of the Xi (Figure 2C, right versus 2D).

LINE-Dense Regions of the Xi Are Exterior to H3K27me3 and *Xist* Domains

Previous studies have shown both H3K27me3 and *Xist* colocalize with gene- and not LINE-dense regions of the Xi (Chadwick, 2007; Duthie et al., 1999; Mak et al., 2002; Marks et al., 2009). We addressed the relationship between X-linked H3K27me3 density, genes, and LINEs, continuing to fit our data in context with previous observations before examining more detailed epigenetic aspects of the TSC Xi. Tiling density plots over the TSC X showed reciprocal relationships between H3K27me3 levels and genes and LINEs (Figure 3A). H3K27me3 and gene density positively correlated over the X (Spearman coefficient 0.29; Figure 3A, middle versus bottom). In contrast, X-linked LINE and H3K27me3 density were inversely correlated (Spearman coefficient −0.57; Figure 3A, middle versus top). Almost invariably, peaks of LINE density colocalized with valleys of H3K27me3 density and vice versa (Figures 3A, S2A, and S2B). The continuity of H3 density within LINE-dense regions validated our ability to detect and normalize for H3K27me3 within these regions.

LINEs and other repeats have been proposed to make up the spatial core of the Xi from the initiation of XCI onward, with X-inactivated genes moving into the proposed core as they are silenced (Chaumeil et al., 2006; Chow et al., 2010; Clemson et al., 2006; Namekawa et al., 2010). Considering the different H3K27me3 levels between LINE- and gene-dense regions of the TSC Xi, it was difficult to reconcile how these sequence classes would occupy the same nuclear space. We therefore directly examined the spatial relationship between the gene- and LINE-dense regions of the TSC Xi. Using H3K27me3 IF combined with DNA FISH, we examined the locations of five X-linked FISH probe pairs relative to the H3K27me3 coat. Pairs consisted of probes in neighboring gene- and LINE-dense regions (Figure 3A, labeled tick marks; Figure S2C). H3K27me3 IF and DNA FISH were performed, Z stack images were collected and deconvolved, and the areas surrounding individual Xi's were

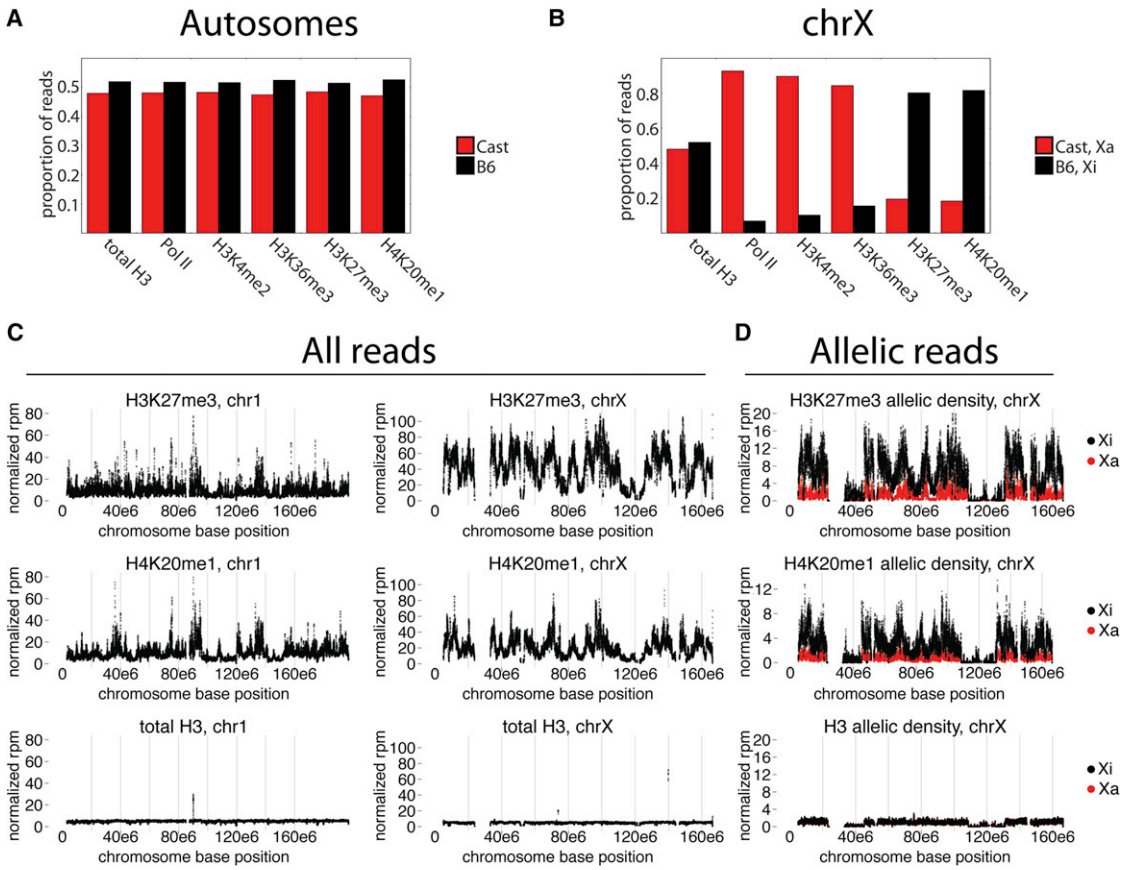


Figure 2. ChIP-Seq Shows X-Linked Epigenetic Biases

(A and B) Proportional allelic distributions over autosomes (A) and chrX (B).

(C) chr1 and chrX nonallelic tiling density plots, in normalized reads per million per 40 kb bin (rpm). Only bins with ≥50% alignability are shown.

(D) chrX allelic tiling density plots, scales as in (C).

See also Table S1.

subjected to 3D reconstruction. After imaging, probes were scored as either inside, on the edge of, or exterior to the Xi's H3K27me3 coat (Figure 3B).

Surprisingly, LINE-dense regions were more frequently exterior to the H3K27me3 domain than gene-dense regions, for four of five probe pairs (Figure 3C). Similar results were obtained with two probe pairs relative to the *Xist* coat (Figure 3D). One of five LINE-dense probes, 15L, was most frequently interior to the H3K27me3 domain, potentially due to higher H3K27me3 levels or proximity to the centromere (Figure S2C). We also examined the spatial relationship between LINE- and gene-dense regions in cell types subject to random XCI: mouse embryonic fibroblasts (MEFs) and embryoid bodies (EBs). The LINE-dense regions examined in these cells were external to *Xist* at a frequency similar to that found in TSCs (Figures 3E and 3F). We conclude that in TSCs, MEFs, and EBs, a majority of the most LINE-dense Xi regions are spatially separated from gene-dense regions of the chromosome, which are encompassed by *Xist* and H3K27me3. Given this spatial separation, our results suggest maintenance of gene silencing during XCI is not associated with translocation into a LINE-dense spatial core.

X-Inactivated Promoters Exclude Active Chromatin Marks and Are Enriched in H3K27me3 and H4K20me1

Having established the large-scale relationship between histone marks and DNA features of the TSC Xi, we began a more high-resolution study of X-linked chromatin. We initially compared two X-linked gene categories that differ in their transcriptional status: (1) X-inactivated genes, on average expressed at 99.5% from the Xa, and (2) nontranscribed genes, defined by their complete lack of expression as assessed by RNA-Seq. Comparison of metagene data between these two gene categories required normalization for gene-set size, as well as SNP density (Figure S3).

Figure 4 shows allelic metagene plots in the 10 kb surrounding X-inactivated and nontranscribed transcription start sites (TSS). As expected from RNA-Seq data, nearly all Pol II, H3K4me2, and H3K36me3 signal over X-inactivated genes derived from the Xa, indicating Pol II binding to X-inactivated TSS is blocked (Figure 4A, i–iii). H3K27me3 density at X-inactivated genes was uniformly higher on the Xi relative to the Xa; however, there was a modest peak of H3K27me3 density in the 4 kb surrounding both X-inactivated and

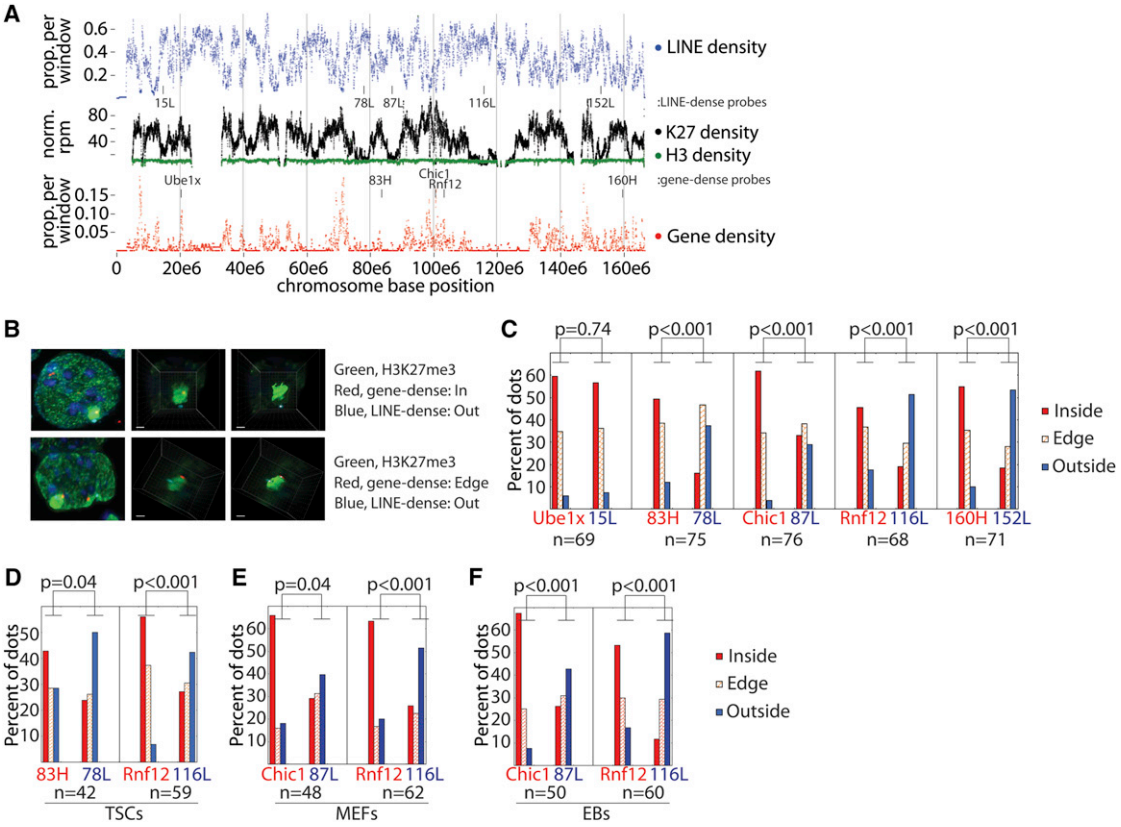


Figure 3. H3K27me3 and LINE Density Inversely Correlate over the Xi

(A) Tiling LINE and gene density over the X. H3K27me3/H3 density are as in Figure 2C. Only bins with ≥50% alignability are shown. FISH probe locations from (B–F) are marked.

(B) DNA FISH quantification process. H3K27me3 IF is green, with gene-dense/LINE-dense probes in red and blue. Left to right shows the progression: a whole nucleus, an extracted Xi, a 3D-rendered Xi image.

(C–F) FISH probe location relative to Xi H3K27me3 (C) or *Xist* (D–F) domains, by percentage. n, Xi's counted per probe pair. p values are from χ^2 tests comparing dots inside/on the edge of versus outside the Xi domain.

See also Figure S2.

nontranscribed TSS (Figure 4B, i). A similar H3K27me3 peak was present in two independently derived TSC lines, confirming TSS enrichment as a general feature of H3K27me3 accumulation over the TSC Xi (Figure 4B, ii and iii). H4K20me1 enrichment over Xi TSS mirrored H3K27me3 patterns (Figure 4B, iv), consistent with the high positive correlation between these two modifications in tiling density plots (Figure 2C).

Xi H3K27me3 and H4K20me1 TSS enrichments were similar between X-inactivated and nontranscribed genes (Figure 4B, i–iv). This contrasts with previous work in EBs, which found H3K27me3 density to be higher on X-inactivated compared to nontranscribed genes, perhaps highlighting a difference between cell types or stages of XCI analyzed (Marks et al., 2009). Our results indicate that TSS-proximal nucleosomes, regardless of Xa transcriptional status, are more likely to be modified with H3K27me3 and H4K20me1 than surrounding Xi sequence, suggesting these sites either have increased capacity to recruit the modifications, or they more stably retain them after deposition.

Regulatory Elements Display DNaseI Hypersensitivity across the TSC Xi

The absence of transcription-associated chromatin signatures at X-inactivated TSS suggested these regions exist in a locally closed state, lacking the nucleosome depletion typically found at utilized TSS. To address this hypothesis, we profiled allelic nucleosome density with two antibody-independent methods, DNase- and FAIRE-Seq (Giresi and Lieb, 2009; Song et al., 2011). DNase-Seq detects genomic regions that are hypersensitive to DNaseI digestion, whereas FAIRE-Seq uses formaldehyde treatment to enrich for genomic regions not crosslinked to proteins. Both techniques identify nucleosome-depleted sites typically found at active promoters and regulatory elements but are also known to detect nonoverlapping sites (Song et al., 2011).

DNase- and FAIRE-Seq patterns were first examined over autosomes to verify both methods performed as expected in TSCs. Indeed, for both techniques, signal was present at highly expressed genes and absent from nontranscribed genes (Figures S4A and S4B). Autosomal Polycomb targets, though

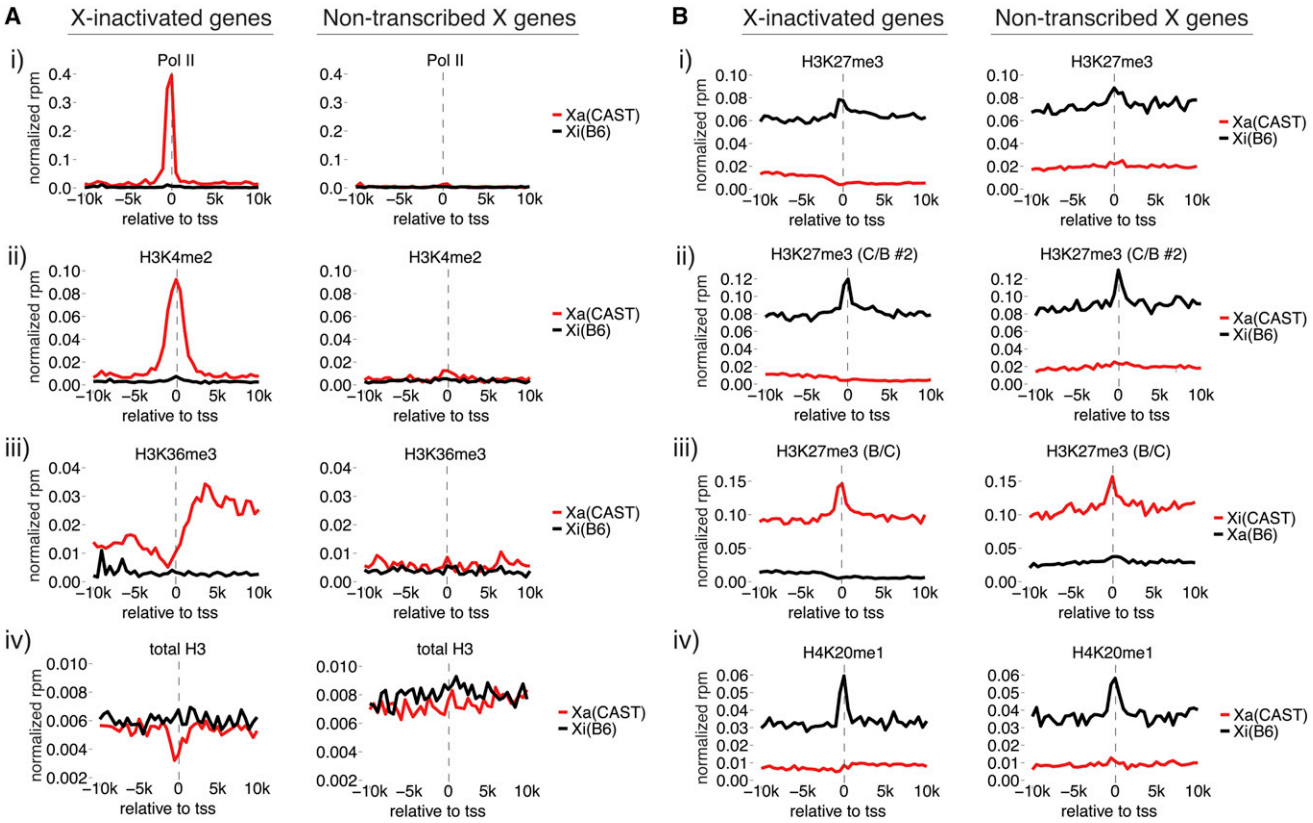


Figure 4. Xi Promoters Exclude Active Chromatin Marks and Are Enriched in H3K27me3 and H4K20me1

(A and B) Allelic metagene profiles of X-inactivated and nontranscribed X-linked genes. Cast (red) and B6 (black) data are shown as rpm per 500 bp bin. (C/B #2), a C/B TSC line not used for RNA-Seq. (B/C), the TSC line from Figure 1 RNA-Seq. See also Figure S3.

expressed at low levels in TSCs, accumulated signal from both methods, expected given the H3 depletion at their TSS (Figure S4C).

We next examined the X-linked DNase-Seq metagene profile. A significant enrichment of DNase-Seq signal was observed at X-inactivated TSS on the Xi (Figure 5A). This result was surprising considering the near-total exclusion of Pol II and other active chromatin marks from these same sites. Although the DNaseI signal surrounding X-inactivated TSS was 6.7-fold lower on the Xi compared to the Xa, Xi signal was well above background, and 2-fold greater than Xi signal at nontranscribed genes (Figure 5A). Importantly, allelic profiles were due to small signal contributions from many genes and likely represent the average DNaseI hypersensitivity (DHS) found at X-inactivated promoters (Figure S4D). In total, 198 genes contributed TSS-associated signal to the Xi profile; median and maximum signal contributions were at 0.3% and 1.9%, respectively. These numbers were similar to those on the Xa, where 201 genes contributed signal, and the median and maximum contributions were 0.2% and 3.4%, respectively. Therefore, X-inactivated TSS exhibited a level of Xi DHS above surrounding sequence and what would have been expected for a nontranscribed gene.

In contrast, FAIRE-Seq did not detect Xi-associated signal at TSS, although robust signal was seen over X-inactivated TSS on the Xa (Figure 5B) and over autosomal Polycomb targets (Figure S4C). The lack of TSS-associated FAIRE signal was consistent with our total H3 profiling, which lacked Xi, TSS-localized H3 depletion (Figure 4A, iv). TSS signal-to-noise estimates from metagene profiles suggested that DNase-Seq had a higher dynamic range than either FAIRE- or total H3-Seq, with a signal-to-noise ratio of 20, as compared to 4 and 2, respectively. Therefore, the lack of FAIRE enrichment and H3 depletion at Xi TSS, despite the presence of DHS, suggests the existence of a nucleosome-depleted site on the Xi that is smaller or less persistent than the equivalent Xa site.

RNA-Seq analysis indicated that X-inactivated genes were expressed from the TSC Xi at low levels, with a median expression value of 0.037 reads-per-kilobase-million (rpkm), similar to the median of 0.089 rpkm for autosomal Polycomb targets. We therefore examined whether the amounts of DNaseI, FAIRE, and Pol II observed at X-inactivated TSS were the same as, or different than, what would be expected for an equivalently expressed autosomal gene. If different than expected, these features might suggest distinct epigenetic properties of X-inactivated TSS that may yield insight into XCI's mechanism. To make the comparison, we selected three autosomal genes classes: those expressed at levels similar to X-inactivated genes on the Xa (A^{Xa}) and Xi (A^{Xi}), and nontranscribed genes (A^{NT}). Metagene

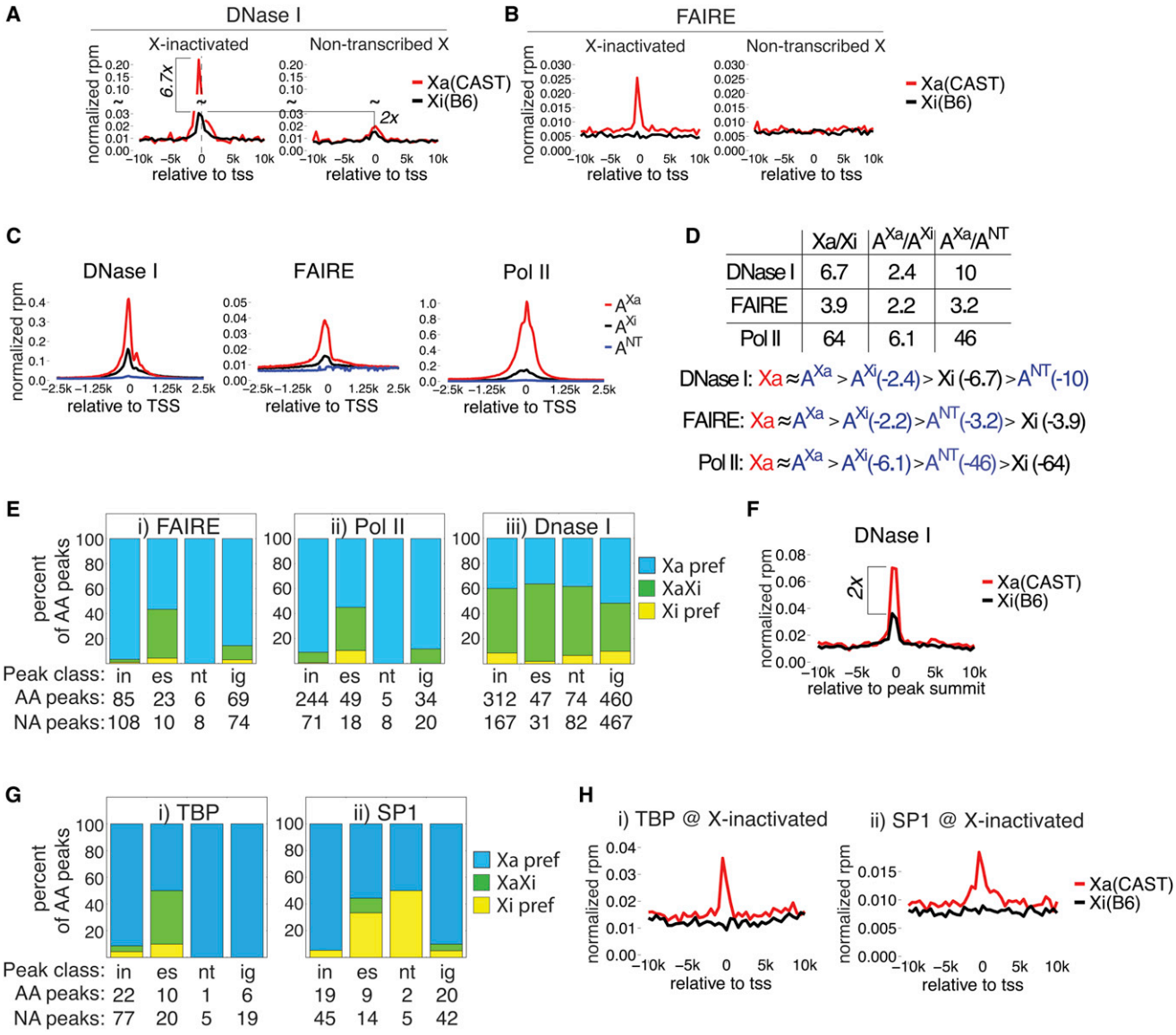


Figure 5. DNaseI Hypersensitivity without FAIRE Enrichment at X-Inactivated TSS

(A and B) DNase-Seq and FAIRE allelic metagene profiles as in Figure 4.

(C) Metagene profiles of autosomal genes, expression-matched to X-inactivated Xa genes (A^{Xa}), X-inactivated Xi genes (A^{Xi}), and nontranscribed genes (A^{NT}).

(D) DNaseI, FAIRE, and Pol II ratios between Xa and Xi genes, A^{Xa} and A^{Xi} genes, and A^{Xa} and A^{NT} genes. Classes were compared assuming that signal from Xa and A^{Xa} genes would be equivalent, given their similar expression levels.

(E) Allelic, X-linked distribution of FAIRE (i), Pol II (ii), and DNaseI (iii) peaks. Categorization is relative to genes falling within 5 kb of a peak start or end; "in," "es," "nt," and "ig" peaks associate with X-inactivated, escaping, nontranscribed, or intergenic regions, respectively. "Xa/Xi pref," peaks detected on one X but not the other; "XaXi," peaks detected on both X's; "AA peaks," peaks that were *A*llelically *A*ssigned to at least one X via SNP-overlapping reads. "NA peaks," peaks that were *N*ot *A*ssignable to either X due to lack of allelic data.

(F) Allelic metagene profile of X-linked, intergenic sites of DHS.

(G) Distribution of TBP (i) and SP1 (ii) peaks over the Xa and Xi as in (E).

(H) Allelic metagene profiles of TBP (i) and SP1 (ii) surrounding X-inactivated genes as in (A).

See also Figures S4 and S5 and Tables S1 and S4.

profiles of DNaseI, FAIRE, and Pol II were created for each (Figure 5C), and the ratio of TSS-associated signal between X-linked and autosomal classes was compared (Figure 5D). Allelic and total data were not directly compared because of their different enrichment scales; instead, ratios between X-linked and expression-matched autosomal genes were calculated separately, and data were compared assuming that signal from Xa and A^{Xa} genes would be equivalent, given their similarly robust expression. The relative Xi DHS of X-inactivated TSS was higher than that observed at nontranscribed autosomal genes and lower

than that observed at expression-matched autosomal genes (Figure 5D; $A^{Xi} > Xi > A^{NT}$). In contrast, Xi FAIRE and Pol II signal at X-inactivated TSS was comparable to autosomal nontranscribed genes, showing levels approaching background (Figure 5D; $A^{Xi} > A^{NT} > Xi$). Therefore, X-inactivated TSS harbor an Xi epigenetic state distinct from lowly expressed and nontranscribed autosomal genes, showing DHS but lacking FAIRE or Pol II enrichment. The transcriptional difference between X-inactivated and nontranscribed genes could be explained if a low level of Xi Pol II binding resulted in efficient target gene elongation.

We next examined whether regulatory elements across the Xi had epigenetic properties similar to those of X-inactivated TSS. MACS-defined peaks of DNaseI, FAIRE, and Pol II were separated into four classes based on relation to their most proximal gene. Individual peaks associated with X-inactivated, escaping, nontranscribed, or intergenic regions, defined as genomic space greater than 5 kb away from a known gene (Figure 5E; "in," "es," "nt," and "ig" peaks, respectively). Allelic binding events were then determined via an empirical background model (Table S4). Consistent with the exclusion observed at X-inactivated TSS (Figures 5B and 4A, i), FAIRE and Pol II peaks were rarely detected surrounding X-inactivated, nontranscribed, and intergenic Xi regions, whereas Xi detection increased around escaping genes (Figure 5E, I and ii). In contrast, but in line with our TSS analysis (Figure 5A), DNaseI peaks were detected across the Xi, mostly at sites shared with the Xa, regardless of associating region (Figure 5E, iii).

Intergenic DHS frequently marks regulatory elements that participate in transcriptional control of nearby and distal genes (Song et al., 2011). To date, such sites have not been investigated in relation to XCI. Of the 460 X-linked, allelically assignable intergenic DNaseI peaks, ~50% maintained Xi DHS despite the near transcriptional silence of most X-linked genes (Figure 5E, iii).

To understand how Xi regulatory elements are processed by XCI, the average chromatin state at intergenic DHS sites was examined via metagene analysis with MACS-defined peak summits as reference points. The analysis focused on the 176 intergenic DHS sites detected on both Xs to avoid confounding contributions from peaks specific to the Xa or Xi. These sites showed significant Xi DHS and had an average Xa:Xi ratio of 2 (Figure 5F). Excluding DHS sites that bound CTCF from this analysis did not change the Xa:Xi ratio (not shown). In contrast, other active marks were excluded from Xi intergenic DHS sites, similar to that observed at X-inactivated TSS (Figure S5A). Consistent with intergenic DHS marking active regulatory elements on the Xa but not Xi, we observed Xa-only enrichment of H3K4me1 and H3K27-acetylation, two marks associated with utilized transcription factor binding (Figure S5B). Intergenic DHS sites also showed small but detectable levels of Xi H3K27me3 and H4K20me1 enrichment (Figure S5C).

Considering that DHS is often indicative of transcription factor binding, we examined whether such binding occurred on the Xi. ChIP-Seq was performed for the general transcription factor TBP, and the gene-specific transcription factor SP1, whose motif was significantly enriched over DHS peaks on the X (not shown). In neither case was binding detected over the Xi, although robust signal was present on the Xa, as expected (Figures 5G and 5H).

In summary, Xi regulatory elements retained DHS, similar to Xa counterparts and autosomal Polycomb targets, yet excluded Pol II and transcription factors, and were not detected as open chromatin via FAIRE, similar to nontranscribed genes (Figures 5 and S4). These data indicate X-inactivated regulatory elements maintain chromatin states distinct from both autosomal Polycomb targets and nontranscribed genes. Moreover, the observed TBP and SP1 exclusion suggests that XCI operates at least in part by preventing efficient transcription factor binding to the Xi.

Variable H3K27me3 Microenvironments Surrounding Genes Escaping XCI

We next examined chromatin modifications surrounding genes that escaped XCI in an effort to understand how localized transcription occurs within the repressive Xi environment. As expected, escapers associated with marks of active transcription (Figure S6A). Additionally, many escapers existed in H3K27me3-depleted microenvironments, such as those in Figure 6A. Generally, escape level inversely correlated with local H3K27me3 density (Figure 6D; r, −0.46, Pearson's coefficient), although not all escapers were locally insulated from the mark. Several escapers, such as *Syap1*, had H3K27me3 levels that closely resembled those found at X-inactivated genes (Figures 6B and 6C). Previous works suggest that the DNA binding protein CTCF, and nuclear position relative to the *Xist* domain, may play important roles in escape (Chaumeil et al., 2006; Filippova et al., 2005). We therefore examined whether these two features might better correlate with Xi expression than local H3K27me3 levels.

CTCF Binds Sites Present on Both X's

We addressed CTCF's role in facilitating escape by localizing Xi binding sites via ChIP-Seq. CTCF binding correlates with insulation between chromatin states genome-wide (Song et al., 2011). Furthermore, CTCF binds the *Jarid1c* gene only in species where it escapes XCI, suggesting a role for CTCF in escape licensing (Filippova et al., 2005). We therefore hypothesized that local CTCF binding would be associated with increased levels of escape. Indeed, there was a moderate positive correlation between CTCF binding and Xi expression, supporting a role for CTCF in escape (Figure 6E; r = 0.25; Pearson's correlation). However, we observed CTCF binding across the Xi, regardless of X-inactivation status or gene presence (Figures 6A–6C, green "CTCF" track). 59, 81, 68, and 71 percent of allelically assignable CTCF peaks within X-inactivated, escaping, nontranscribed, and intergenic regions, respectively, were present on the Xi or both X's (Figure 6F). These results indicate CTCF binding alone is not predictive of escape or local insulation from H3K27me3.

Also unexpectedly, the majority of Xi CTCF peaks were located at sites shared with the Xa and not Xi specific (394 of 428 Xi peaks; Figures 6A–6C, purple stars; Figure 6F, green bars). Xi-specific CTCF binding did exist, for example at the H3K27me3 boundaries flanking the X-inactivation center (Figure 6A, black asterisks), but made up a minority of peaks (34 of 428 Xi peaks; Figure 6F, yellow bars). The presence of shared

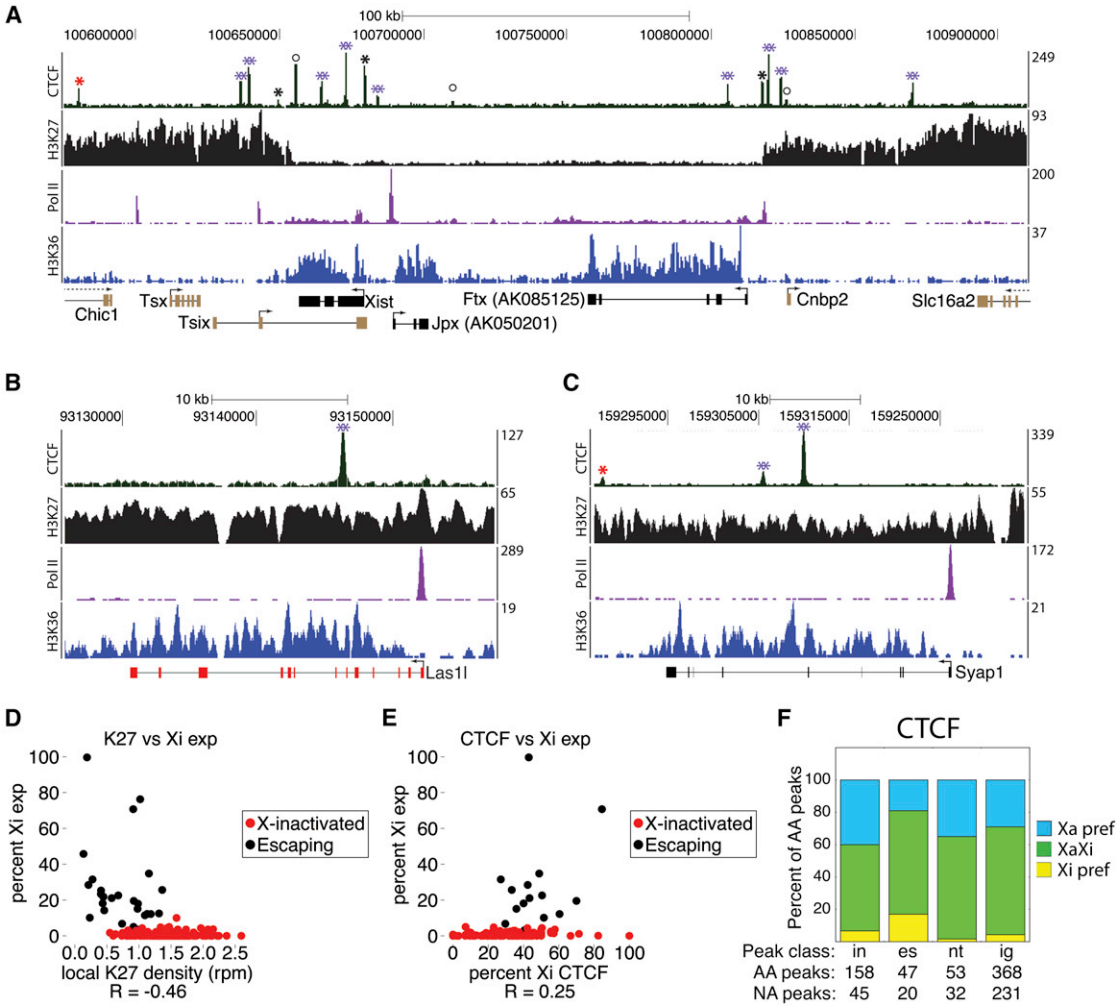


Figure 6. Variable H3K27me3 Levels Associated with Escape, and Widespread CTCF Binding to the Xi

(A–C) Genomic windows surrounding the X-inactivation center (A), *Las1l* (B), and *Syap1* (C). Nonallelic ChIP-Seq data are shown above genome annotations. Black genes escaped, red genes were X-inactivated, gold genes were not allelically assigned. Red/black asterisks (*) mark CTCF peaks on the Xa/Xi; purple double asterisks (**), those on both X's; circles, nonassignable peaks.

(D) Xi expression versus genic H3K27me3 density, for X-linked genes with ≥20 allelic RNA-Seq reads.

(E) Xi expression versus Xi CTCF binding levels, for genes in (D) that contain a CTCF peak within 5 kb of their start or end.

(F) Allelic distribution of CTCF peaks, as in Figure 5E.

See also Figure S6 and Tables S1 and S4.

CTCF peaks across varied Xi environments, regardless of gene or chromatin boundary location, suggests complex utilization of the protein in XCI.

Transcriptional Competence and Position Relative to the *Xist* Domain Are Uncoupled

Lastly, we examined the nuclear position of escaping and X-inactivated genes relative to the Xi's H3K27me3 and *Xist* domains. Previous work has shown the area encompassed by *Xist* to be microscopically devoid of transcription-associated features, such as Pol II, and that escaping genes are preferentially located outside of the *Xist* domain (Escamilla-Del-Arenal et al., 2011). Considering these data, it has been hypothesized that externalization relative to the *Xist* domain may place genes in an environment permissive to transcription, potentially playing a causal role in escape.

To examine the relationship between escape and *Xist* externalization in TSCs, we selected four escaping loci expressed at varying levels from the Xi: *Taf1-Ogt* (18% average Xi exp.), *Jarid1c* (18% Xi exp.), *Nkap* (12% Xi exp.), and *Utx* (5% Xi exp.). The nuclear position of these genes relative to the Xi's *Xist* and H3K27me3 coat was examined together with an X-inactivated gene (*Rnf12* or *Abcb7*) via the DNA FISH assay described in Figure 3. This assay allowed escaping and X-inactivated gene location to be quantified in tandem, providing an internal control per experiment.

As expected, escaping genes were more frequently exterior to the *Xist* domain than X-inactivated genes, in TSCs, MEFs, and

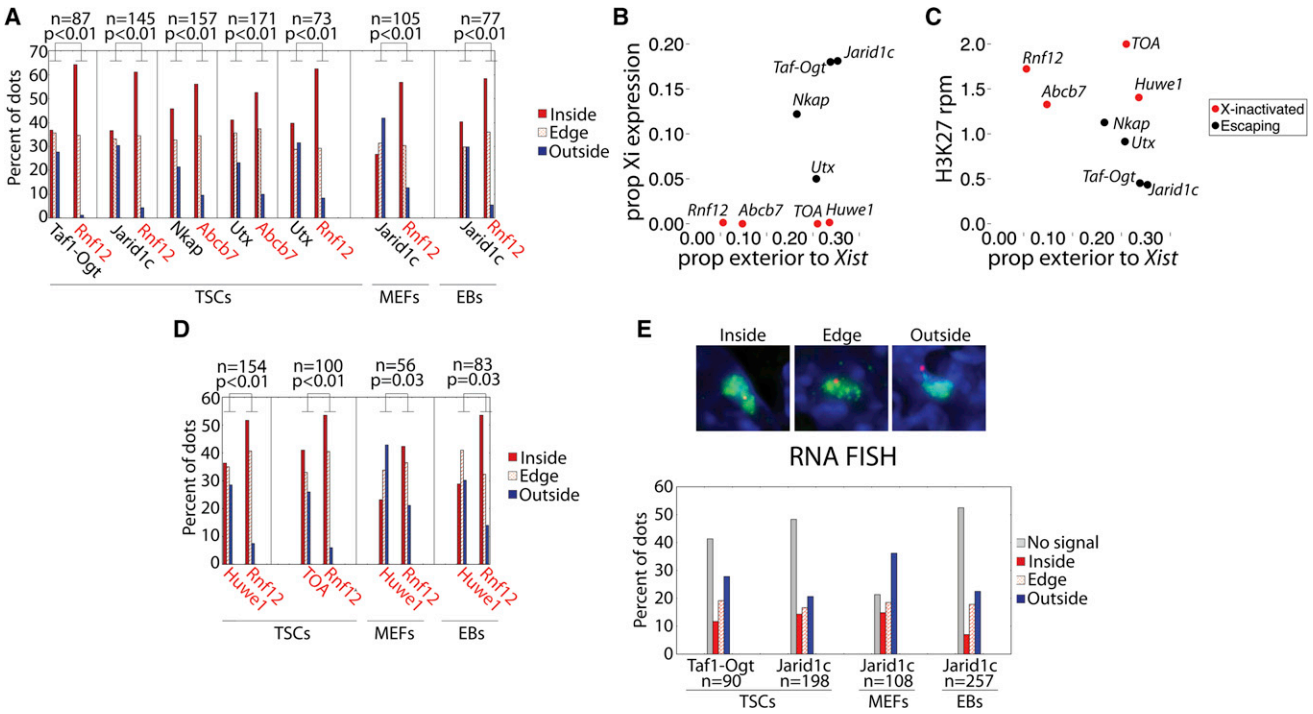


Figure 7. Externalization Relative to the Microscopic *Xist* Domain Does Not Induce Escape

(A) Gene location relative to *Xist* domain, by percentage. Vertical lines separate probe pairs of escaping (black type) and X-inactivated (red type) genes. n, Xi's counted per probe pair. p values are from χ^2 tests comparing dots inside/on the edge of versus outside the Xi domain.

(B and C) *Xist* domain externalization frequency versus (B) Xi expression, or (C) local H3K27me3 density.

(D) Gene location relative to *Xist* domain, as in (A).

(E) RNA FISH signal relative to *Xist* domain. "No Signal," monoallelic expression from the Xa. Representative images are above the bar graph.

See also Figure S6.

EBs (Figure 7A). Localization differences were not found when examining position relative to the Xi's H3K27me3 coat (Figure S6B). For these six loci, externalization frequency correlated well with TSC Xi expression and H3K27me3 levels (r = 0.87 and −0.95, respectively, Pearson's correlation; Figure 7B) with some exceptions: *Utx* and *Nkap* had similar externalization frequencies but different Xi expression levels, and *Rnf12* and *Abcb7* had different externalization frequencies but similar Xi expression levels (p = 0.03 for externalization, χ^2 test; Figure 7B).

Considering these exceptions, we further explored the relationship between *Xist* domain externalization and Xi expression, attempting to parse the difference between causality and correlation. The positions of two X-inactivated loci adjacent to escaping genes were examined relative to the *Xist* domain: *Huwe1*, adjacent to the escaper *Jarid1c* and displaying 0.1% Xi expression, and a ~200 kb locus, referred to as *TOA*, for *Taf1* and *Ogt* Associated, which contained four X-inactivated genes adjacent to the escapers *Taf1* and *Ogt* and displayed 0% aggregate Xi expression. In TSCs, these two loci had externalization frequencies similar to neighboring escaping loci, despite lacking Xi expression and maintaining high levels of H3K27me3 (Figures 7B–7D). The same localization pattern was observed in MEFs and EBs (Figure 7D). RNA FISH confirmed the monoallelic expression of these loci (Figure S6C). Together, these results indicate that gene externalization relative to the *Xist* domain is positively correlated with, but insufficient to induce, escape from XCI.

In light of this result, we examined where escaper transcription occurred on the single-cell level, via RNA FISH and 3D reconstruction of the *Xist* domain. If the microscopic region encompassed by *Xist* is truly impermeable to transcription, as previous analyses would predict (Escamilla-Del-Arenal et al., 2011), then escaping loci would only express when located on the boundary of, or exterior to, the *Xist* domain, where Pol II and associated factors could be accessed. We chose two of the most robustly escaping TSC loci for this analysis, *Jarid1c* and *Taf1-Ogt*, given their high levels of biallelism and clarity of RNA FISH signal. In contrast to what would be predicted from existing models, 28% and 20% of the total observed RNA FISH dots for *Jarid1c* and *Taf1-Ogt* localized within the TSC *Xist* domain (representing 14% and 12% of the total Xi's counted; Figure 7E). Similar colocalization frequencies were observed for *Jarid1c* and *Xist* in MEFs (21%) and EBs (15%) (Figure 7E). The detection of RNA FISH signal within the region encompassed by *Xist* suggests that the microscopic domain is not an obligate silent compartment in TSCs, MEFs, or EBs.

Together, two lines of evidence indicated that transcriptional competence is uncoupled from gene position relative to the microscopic *Xist* domain, in TSCs, MEFs, and EBs. First, X-inactivated loci adjacent to escapers remained inactive despite their

frequent localization outside of the microscopic *Xist* domain (Figure 7D). Second, escape frequently colocalized with the *Xist* domain, as assessed via RNA FISH (Figure 7E). We conclude that XCI is unlikely to be maintained solely by a chromosome-wide separation from transcription machinery. Rather, our results suggest a role for the locus-specific silencing of regulatory elements in the inactivation process.

DISCUSSION

Our allelic gene expression and chromatin analyses revealed a complex set of epigenetic environments over the TSC Xi. Most notably, X-inactivated regulatory elements harbored an epigenetic signature distinct from autosomal Polycomb targets and nontranscribed genes, displaying DHS without enrichment of other transcriptional hallmarks. Together with microscopic analyses examining gene location relative to the *Xist* domain, our data suggest a central role for the site-specific silencing of regulatory elements in maintenance of both random and imprinted XCI.

We present a global analysis of allelic gene expression over an imprinted Xi, finding that 13% of TSC genes escaped XCI. This percentage was similar to that estimated for human cells (Carrel and Willard, 2005) but different from the 3% observed in mouse kidney (Yang et al., 2010), indicating different cell types have varying escape frequency. Degree of TSC escape could be partly, though not perfectly, predicted by local H3K27me3 levels. Considering this imperfect correlation, we hypothesize that many escaping genes fluctuate between H3K27me3 insulated and uninsulated states, with expression occurring most robustly during times of insulation. Alternatively, some genes may escape even in the presence of high local H3K27me3 levels.

Considering its classification as an insulator, we addressed CTCF's role in establishing escape from XCI. Perhaps not surprisingly given its widespread binding patterns (Song et al., 2011), the X-linked distribution of CTCF was complex. Xi CTCF binding positively correlated with escape, supporting a role in the process, as has been proposed (Filippova et al., 2005). However, 92% of the 428 CTCF peaks detected over the Xi were at sites shared with the Xa. Accordingly, Xi CTCF binding did not predict the location of escapers or H3K27me3 boundaries, consistent with results from a previous study examining transgene insulation during XCI (Ciavatta et al., 2006). The function of mirrored CTCF binding over diverse X-linked environments is unknown. We favor the possibility that CTCF binds certain sites without discrimination between X's, but these sites are differentially utilized in the creation of allele-specific chromatin structures.

Regarding the epigenetic properties of Xi regulatory elements, we describe evidence indicating these sites are recognized from surrounding sequence, yet rendered nonfunctional by the XCI machinery. On average, Xi regulatory elements showed enrichment of H3K27me3 and H4K20me1 relative to neighboring Xi DNA, supporting the notion that Xi regulatory elements are recognized by cellular machinery. Remarkably, regulatory elements across the TSC Xi showed DHS, despite excluding Pol II and other transcriptional hallmarks. These data suggest that nucleosomes surrounding Xi regulatory elements are recognized and displaced, but an unidentified property of these loci—a certain conformation, noncanonical nucleosome, or associated protein or modification—precludes efficient binding of Pol II to the Xi.

Binding of two transcription factors, TBP and SP1, was not detected over Xi regulatory elements despite these elements harboring the DHS typical of transcription factor binding. This lack of detectable binding indicates that XCI operates at least in part by preventing the stable association of Pol II recruitment factors to targets. In the complete absence of transcription factor binding, the DHS observed at Xi regulatory elements could be the byproduct of a mechanism excluding such factors from the Xi. This mechanism would be PRC2-independent, considering Xi DHS peaks do not invariably coincide with H3K27me3-enriched regions (data not shown), and TSC autosomal PRC2 targets bind TBP, SP1, and Pol II (Figure S4C). It is also possible that transcription factors bind the TSC Xi at low levels, resulting in DHS, but our allelic ChIP-Seq lacked the sensitivity needed to detect such binding. In this scenario, an estimated upper bound of Xi transcription factor binding might be on par with DHS levels at Xi TSS, about 7-fold less on the Xi as compared to the Xa. Were this the case, XCI could operate combinatorially through inhibition of transcription factor binding and function, given that X-inactivated genes are repressed ~200-fold on the TSC Xi and essentially lack Pol II binding. Continued study of Xi chromatin states will likely yield additional insight into XCI's mechanism.

Informed by our genomic analyses, spatial properties of the Xi were examined relative to pre-existing mechanistic models of XCI. These models suggest that the Xi's spatial core, marked by *Xist*, is a transcriptionally silent, repeat-dense region, into which genes are recruited as they are inactivated (Chaumeil et al., 2006; Chow et al., 2010; Clemson et al., 2006; Namekawa et al., 2010). Entry into this domain may induce, or at a minimum, help maintain silencing by preventing access to transcription machinery. Accordingly, escapers are thought to be actively maintained at the domain's exterior, allowing them access to Pol II (Escamilla-Del-Arenal et al., 2011).

Our work suggests significant revisions to the above model. Site-specific DNA FISH demonstrated that LINE-dense regions did not invariably make up the spatial core of the Xi, as has been proposed (Chaumeil et al., 2006; Chow et al., 2010; Clemson et al., 2006; Namekawa et al., 2010). Rather, gene- and LINE-dense regions occupied separate nuclear territories in TSCs, MEFs, and EBs, with LINE-dense DNA most frequently adjacent to the *Xist* domain. This spatial separation suggests that maintenance of gene silencing during XCI does not require colocalization with a LINE-dense core and supports an indirect role for the Xi's most LINE-dense regions in *Xist*-mediated silencing (Tattermusch and Brockdorff, 2011). Previous works defining the Xi's core as repeat dense have relied on site-nonspecific FISH probes such as Cot-1, which cannot differentiate between repetitive sequence in genic and intergenic regions, perhaps explaining the observed discrepancies.

Our data also suggest that gene externalization relative to the measured *Xist* domain is a consequence rather than cause of escape. At two separate loci, escapers and adjacent X-inactivated genes were found outside of the *Xist* domain at similar

frequencies. If externalization were a primary factor in inducing escape, then these X-inactivated loci would have exhibited increased Xi expression upon externalization. Instead, they maintained silencing regardless of location relative to the measured *Xist* domain, indicating that licensing of escape occurs on a gene-specific level and is not strictly determined by chromosome topology. In support of this, escape was detected within the microscopic *Xist* domain, as assessed via RNA FISH. This notion is further supported by the observation that differential gene regulation can occur within topologically associated genomic regions (Dixon et al., 2012; Nora et al., 2012). We hypothesize escapers and topologically associated X-inactivated genes are externalized due to escaper interactions with transcription factories, which are abundant outside of the measured *Xist* domain.

Considered together, our results support a model for XCI whereby individual regulatory elements are maintained in a silent state by a mechanism that persists regardless of their location relative to a larger X-linked domain. Although a chromosome-wide exclusion of transcription machinery from the Xi's physical territory may play a role in XCI, it ultimately appears secondary to site-specific silencing during XCI maintenance; genes escaping XCI were expressed within the Xi's interior, and X-inactivated genes remained silent when separated from the microscopic *Xist* domain. Separation from the microscopic domain may occur dynamically, resulting in temporary loss of local *Xist* coating at externalized X-inactivated genes. Alternatively, *Xist* binding may persist over externalized regions but at levels that are not detectable by conventional RNA FISH. In either case, silencing is stable throughout externalization, as gene expression levels of external and internal X-inactivated genes were indistinguishable.

The distinct submicroscopic epigenetic signatures of the TSC Xi lend additional credence to a model of XCI where gene silencing is governed by inactivation of individual regulatory elements rather than a chromosome-scale, spatial segregation away from transcription machinery. Our surprising observation that X-inactivated regulatory elements display DHS and proximal H3K27me3 and H4K20me1 enrichment indicates that these sites on the Xi are recognized as such from surrounding DNA. The absence of transcription-associated signals from genes along the Xi, despite their apparent exposure to a nuclear environment permissive to transcription, indicates that XCI-induced epigenetic signatures can be stably maintained independent of a chromosome-scale nuclear compartment dedicated to transcriptional silencing.

EXPERIMENTAL PROCEDURES

TSC Derivation and Culture

TSC lines were derived and cultured as previously described (Quinn et al., 2006). To remove feeder cells for genomic analyses, TSCs were trypsinized, preplated for 40 min, and split 2× or 4× prior to harvesting chromatin and RNA, respectively.

RNA-Seq and Validation

Strand-specific cDNA libraries were prepared from polyA-purified TSC RNA as described in (Ingolia et al., 2009). Quantitative allele-specific RT-PCR (Figure S1) was performed as described in Kalantry et al. (2009); differential sensitivity to restriction enzyme digestion was used as a means to discriminate between amplified alleles.

ChIP-, DNase-, and FAIRE-Seq

For ChIP- and FAIRE-Seq, TSCs were crosslinked for 10 min at room temperature in DMEM and 10% serum with 0.6% formaldehyde, followed by a 5 min quench with 125mM glycine. ChIP conditions varied per antibody, and were performed largely as described in Rahl et al. (2010); 10–40 million feeder-free TSCs and 10 μg of antibody per IP were used. Antibodies used were histone H3 (Abcam ab1791), Pol II (Santa Cruz sc-899), H3K4me2 (Millipore 07-030), H3K36me3 (Abcam ab9050), H3K27me3 (Abcam ab6002), H4K20me1 (Active Motif 39175), CTCF (Pugacheva et al., 2005), TBP (Abcam ab818), and SP1 (Santa Cruz sc-17824). ChIP-Seq libraries were prepared from 10–200 ng of ChIP'd DNA. FAIRE-Seq and DNase-Seq were performed as described in (Giresi and Lieb, 2009) and (Song and Crawford, 2010).

DNA and RNA FISH

For DNA and RNA FISH experiments, cells were fixed for 10 min in 4% paraformaldehyde in PBS and permeabilized for 10 min on ice in 0.5% Triton X-100 in PBS and Ribonucleoside Vanadyl complex. Cells were serially dehydrated in ethanol before heat denaturation at 80°C (DNA FISH only) and probe hybridization overnight at 37°C. Posthybridization cells were washed 3× for 5 min in 50% formamide/2X SSC at 42°C, then 3× for 5 min in 1X SSC at 55°C. Z stack images were acquired with a 100× objective on a Zeiss Axio Imager 2 and deconvolved with an iterative-constrained algorithm. Nuclear regions surrounding 10–15 Xi's per 100× image were selected and imported into Bitplane's Imaris analysis software for 3D reconstruction and analysis. Final counts represent summed data from a minimum of biological replicates. BACs and fosmids were ordered from the BACPAC resource center and fingerprinted with Hind III and BamH I for insert verification. Fluorescent labeling of probes was performed with Invitrogen's BioPrime kit.

Allelic Analysis

SNP data was obtained from (http://www.sanger.ac.uk/resources/mouse/genomes/) on January 15, 2010, and used to create an in silico Cast genome build. Reads that uniquely aligned to B6/Cast genomes with the Bowtie algorithm (Langmead et al., 2009) were retained for downstream analyses. A nonredundant list of mouse genes (Table S3) was annotated from the set of UCSC Known Genes (mm9, downloaded on January 20, 2010) and used for downstream analyses (Fujita et al., 2011).

ACCESSION NUMBERS

The NCBI accession number for the raw data are reported in this article is GSE39406.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.10.037.

ACKNOWLEDGMENTS

Thanks to Magnuson lab members for critical comments, to Terry Furey and Darin London for DNase-Seq help, and to Victor Lobanenkov for kindly providing the CTCF antibody mixture. This work was supported by the NIH grant R01-GM101974 and UNC Cancer Research Fund. JMC was supported by the American Cancer Society postdoctoral fellowship 117571-PF-09-124-01-DDC.

Received: November 15, 2011
Revised: June 5, 2012
Accepted: October 9, 2012
Published: November 20, 2012

REFERENCES

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.

Chadwick, B.P. (2007). Variation in Xi chromatin organization and correlation of the H3K27me3 chromatin territories to transcribed sequences by microarray analysis. Chromosoma *116*, 147–157.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. *20*, 2223–2237.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956–969.

Ciavatta, D., Kalantry, S., Magnuson, T., and Smithies, O. (2006). A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation. Proc. Natl. Acad. Sci. USA *103*, 9958–9963.

Clemson, C.M., Hall, L.L., Byron, M., McNeil, J., and Lawrence, J.B. (2006). The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. Proc. Natl. Acad. Sci. USA *103*, 7688–7693.

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

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.

Escamilla-Del-Arenal, M., da Rocha, S.T., and Heard, E. (2011). Evolutionary diversity and developmental regulation of X-chromosome inactivation. Hum. Genet. *130*, 307–327.

Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.P., Hu, Y.J., Nguyen, D.K., Tsuchiya, K.D., and Disteche, C.M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. Dev. Cell *8*, 31–42.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., et al. (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res. *39*(Database issue), D876–D882.

Giresi, P.G., and Lieb, J.D. (2009). Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Methods *48*, 233–239.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Kalantry, S., Mills, K.C., Yee, D., Otte, A.P., Panning, B., and Magnuson, T. (2006). The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. Nat. Cell Biol. *8*, 195–202.

Kalantry, S., Purushothaman, S., Bowen, R.B., Starmer, J., and Magnuson, T. (2009). Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. Nature *460*, 647–651.

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

Mak, W., Baxter, J., Silva, J., Newall, A.E., Otte, A.P., and Brockdorff, N. (2002). Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. Curr. Biol. *12*, 1016–1020.

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

Namekawa, S.H., Payer, B., Huynh, K.D., Jaenisch, R., and Lee, J.T. (2010). Two-step imprinted X inactivation: repeat versus genic silencing in the mouse. Mol. Cell. Biol. *30*, 3187–3205.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Patrat, C., Okamoto, I., Diabangouaya, P., Vialon, V., Le Baccon, P., Chow, J., and Heard, E. (2009). Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. Proc. Natl. Acad. Sci. USA *106*, 5198–5203.

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.

Pugacheva, E.M., Tiwari, V.K., Abdullaev, Z., Vostrov, A.A., Flanagan, P.T., Quitschke, W.W., Loukinov, D.I., Ohlsson, R., and Lobanenkov, V.V. (2005). Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. Hum. Mol. Genet. *14*, 953–965.

Quinn, J., Kunath, T., and Rossant, J. (2006). Mouse trophoblast stem cells. Methods Mol. Med. *121*, 125–148.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J. *25*, 3110–3122.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., 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.

Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb. Protoc. *2*, pdb prot5384.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., Gräf, S., Huss, M., Keefe, D., et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. *21*, 1757–1767.

Splinter, E., de Wit, E., Nora, E.P., Klous, P., van de Werken, H.J., Zhu, Y., Kaaij, L.J., van Ijcken, W., Gribnau, J., Heard, E., and de Laat, W. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. *25*, 1371–1383.

Surface, L.E., Thornton, S.R., and Boyer, L.A. (2010). Polycomb group proteins set the stage for early lineage commitment. Cell Stem Cell *7*, 288–298.

Tattermusch, A., and Brockdorff, N. (2011). A scaffold for X chromosome inactivation. Hum. Genet. *130*, 247–253.

Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. Nat. Genet. *28*, 371–375.

Williams, L.H., Kalantry, S., Starmer, J., and Magnuson, T. (2011). Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass. Development *138*, 2049–2057.

Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. *20*, 614–622.

EXTENDED EXPERIMENTAL PROCEDURES

Experimental Analysis

TSC Derivation and Culture

TSC lines were derived and cultured as previously described (Quinn et al., 2006). C57BL/6J and CAST/EiJ mice, originally obtained from Jackson Labs, were kind gifts from F. Pardo-Manuel de Villena. Of the four TSC lines used in this work, three had a 40N chromosome content. The remaining line, B/C #2, had a 41N chromosome content, although DNA and RNA FISH experiments definitively indicated that the additional chromosome was not an X. To eliminate feeder cell contamination of TSC RNA and chromatin preparations, TSC cultures grown on sublethally irradiated feeders were trypsinized and preplated for 40 min before transferring to new plates. TSC cultures were then split an additional 2× or 4× before harvesting cells for chromatin and RNA, respectively. Cell density counts consistently showed feeder contamination to be <1:1,000 or <1:10,000 at the time of chromatin and RNA harvesting, respectively.

RNA-Seq and Validation

Strand-specific cDNA libraries were prepared from TSC RNA as described in (Ingolia et al., 2009), with additional protocol modifications suggested by NT Ingolia. Biological replicate RNA preparations were obtained from feeder-free C/B and B/C TSC cultures with Trizol (Invitrogen). RNA from each cell line was then pooled, and cDNA libraries were constructed in duplicate from polyA RNA purified from 100μg of pooled total RNA (Dynal/Invitrogen). cDNA libraries were sequenced on Illumina's Genome Analyzer IIx instrument. For the C/B cell line, three lanes of sequence were obtained for replicate 1 and two lanes of sequence for replicate 2; an additional cDNA library was prepared from an entirely separate biological RNA preparation and sequenced to a depth of one lane. R-squared values comparing resulting gene expression values between all three C/B replicates were approximately 0.97. For the B/C cell line, three lanes of sequence data were obtained for replicates 1 and 2. The R-squared value for the comparison of gene expression levels between B/C replicates was 0.985.

Allele-specific, heteroduplex-minimizing RT-PCR assays were performed essentially as in (Kalantry et al., 2009). Per TSC line, 0.5 μg of RNA was reverse-transcribed with Superscript III (Invitrogen) and 1/100th of each cDNA reaction was used for PCR quantification. Linear ranges of amplification per primer pair were established prior to performing heteroduplex assays. Primers derive from either (Huynh and Lee, 2003; Stavropoulos et al., 2001), or this work. Allelic assays are described below in the format [*gene name*: (forward primer, reverse primer, restriction enzyme used, allele cut, source of assay)]: *Ogt*: (agtttgagcccaaatcatgc, ttctgcttcagcaacactgc, Cac8I, Cast, This work); *Suv39h1*: (GCATCACATCGCCTTCTAGTC, CTTTGAAAGCCCCACAGAAA, Cac8I, B6, This work); *Syap1*: (GCACAACGTTGAATTTGTGGT, CACCCAGTTTGACATGCTGA, ScrFI, B6, This work); *Ak157237*: (TTGGGTCTTCAAGGAGCATC, AGGCGAAAACAGATTGTGCT, HpyCH4IV, B6, This work); *Pgk1*: (CGTGATGAGGGTGGACTTC AAC, TAGTTTGGACAGTGAGGCTCGG, Mse1, Cast, S&L); *Jarid1c*: (acctcacccaagaccttcct, gtcagccatctctccaaagc, Hpy188I, Cast, S&L)Utx: (tcagtaccaggcctcctcat, aactctcacgaaggcaggaa, HpyCH4IV, B6, This work); *Slc35a2*: (ACCTACAGC TCGGCCTCTTT, GCAACAGTGGACAGCACAAT, ScrFI, Cast, This work); *2610029G23Rik*: (CCTGCCTTGTTCATCCTCTC, CTCCCTCCCATTTCAAATCA, NlaIII, Cast, This work); *Nkap*: (catcctcgtccgcctattac, ttagaggtgtgtgccactgc, Sfc I, Cast, This work); *Xist*: (CCCGCTGCTGAGTGTTTGATATG, CAGAGTAGCGAGGACTTGAAGAG, ScrF, B6, H&L).

To test whether the observed escape variation between C/B and B/C cell lines was due to heritable genetic differences between Cast and B6 mice, or random variation between TSC lines, allelic expression of 6 differential escapers was measured in a panel of TSC lines with a PCR-based method (Figure S1; Kalantry et al., 2009). Four of 6 differentially escaping genes tested recapitulated strain-specific escape in additional TSC lines not used for sequencing, suggesting that the majority of observed differential escape (~67%) was due to heritable genetic differences between B6 and Cast mice and not TSC line specific.

ChIP-, DNase-, and FAIRE-Seq

ChIP was performed largely as described in (Rahl et al., 2010), with 10–40 million feeder-free TSCs and 10 μg of antibody per IP. Antibodies used were: histone H3 (Abcam ab1791), Pol II (Santa Cruz sc-899), H3K4me2 (Millipore 07-030), H3K36me3 (Abcam ab9050), H3K27me3 (Abcam ab6002), H4K20me1 (Active Motif 39175), H3K27ac (Abcam 4729), H3K4me1 (Abcam ab8895), CTCF (Kim et al., 2007; Pugacheva et al., 2005), TBP (Abcam ab818), and SP1 (Santa Cruz sc-17824).

TSCs were crosslinked for 10 min at room temperature in DMEM and 10% serum with 0.6% formaldehyde, followed by a 5 min quench with 125mM glycine.

For histone ChIPs, TSCs were incubated with buffer 1 (50 mM HEPES pH 7.3, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, and 0.25% Triton X-100) for 10 min at 4C, then incubated with buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA pH 8.0 and 0.5 mM EGTA pH 8.0) for 10 min at RT, before re-suspension in buffer 3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.1% Na-deoxycholate, 0.5% N- lauroylsarcosine). Extracts were then sonicated to generate 200-500 bp DNA fragments, cleared via centrifugation, and diluted to 20 million cells equivalents per ml of buffer 3 containing 1% Triton X-100. Antibodies were preconjugated to Dynal Protein A or G beads (Invitrogen) overnight in 0.5% BSA/PBS and added to cell extracts for overnight ChIP. Post-ChIP, beads were washed 4× in (50 mM HEPES pH 7.3, 500 mM LiCl, 1 mM EDTA, 1% NP-40 and 0.7% Na-Deoxycholate) for 4 min each then once in TE containing 50mM NaCl. Antibody complexes were eluted for 15 min at 65 C in (50mM Tris pH 8.0, 10mM EDTA, and 1% SDS), crosslinks were reversed overnight at 65 C, eluates were incubated with Proteinase K and RNase A, and DNA was extracted with phenol/chloroform and precipitated.

For Pol II and CTCF ChIPs, cells were re-suspended in buffer 4 (50 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.1% Na-deoxycholate, 0.1% SDS), sonicated to generate 200-500 bp DNA fragments, cleared via centrifugation, and diluted to 20 million cells equivalents per ml of buffer 4 containing 1% Triton X-100. Post-ChIP, beads were washed 3x with buffer 4 containing 1% Triton X-100, once with buffer 4 containing 1% Triton X-100 and 500mM NaCl, once with buffer 5 (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), and once with TE before eluting as above.

For TBP and SP1 ChIPs, cells were re-suspended in buffer 6 (40 mM Tris-HCl pH 7.6, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, and 0.5% NP-40), sonicated to generate 200-500 bp DNA fragments, cleared via centrifugation, and diluted to 20 million cells equivalents per ml of buffer 6 containing 1% Triton X-100. Post-ChIP, beads were washed 3x with buffer 6 containing 1% Triton X-100, 2x with buffer 6 containing 1% Triton X-100 and 500mM NaCl, and once with TE before eluting as above.

ChIP-Seq libraries were then prepared according to Illumina instructions from 10–200 ng of ChIP'd DNA. All libraries except for Pol II, H3K36me3, H3K27me3 C/B#2, and H3K27me3 B/C were prepared from a minimum of biological duplicates and were sequenced on Illumina's Genome Analyzer IIx or HiSeq2000 instrument. Pol II, H3K36me3, H3K27me3 C/B#2, and H3K27me3 B/C data are from a single library preparation and one lane of sequencing.

FAIRE-Seq was performed essentially as described in (Giresi and Lieb, 2009), except TSCs were crosslinked with 0.6% instead of 1% formaldehyde. FAIRE-Seq libraries were prepared from biological duplicates.

DNase-Seq was performed as described in (Song and Crawford, 2010). Libraries were prepared from biological duplicates.

DNA and RNA FISH

DNA and RNA FISH/IF experiments were performed essentially as described, (Chaumeil et al., 2006; Kalantry and Magnuson, 2006), (http://www.epigenome-noe.net/WWW/researchtools/protocol.php?protid=3). Cells were fixed for 10 min in 4% paraformaldehyde/PBS, followed by a 10 min permeabilization on ice in 0.5% Triton X-100 in PBS and 1:200 NEB Ribonucleoside Vanadyl complex. Cells were then dehydrated by serial 3 min incubations with 75%, 85%, 95%, and 100% ethanol, and air-dried for 5 min before heat denaturation. For DNA FISH/IF, H3K27me3 IF was performed prior to dehydration and denaturation, with an antibody from Millipore (07-449) at 1:200 dilution. After incubation with a biotinylated rabbit secondary antibody (Invitrogen), αH3K27me3-treated cells were fixed with 4% paraformaldehyde for 5 min, washed, and placed into the ethanol dehydration series. For DNA FISH/IF experiments, cells were heat denatured at 80 C for 20 min in 70% formamide/2X SSC. For DNA/*Xist* RNA FISH, cells were denatured for 9 min at this temperature in the same solution. Following denaturation, cells were washed in cold 2X SSC and probes hybridized overnight at 37 C. Posthybridization cells were washed 3× for 5 min in 50% formamide/2X SSC at 42 C, then 3× for 5 min in 1X SSC at 55 C. Postwashing, DNA FISH/IF cells were incubated with streptavidin Alexa Fluor 488 (Invitrogen).

Using a 100× objective, Z stack images were acquired on a Zeiss Axio Imager 2 and deconvolved with an iterative-constrained algorithm (AxioVision). Nuclear regions surrounding 10-15 inactive X chromosomes per 100× image were selected by visualizing the H3K27me3 or *Xist* channel in isolation, without signal from DAPI or region-specific FISH probes, to prevent any selection bias. Selected regions were then imported into Bitplane's Imaris analysis software for 3D reconstruction. Reconstruction was performed sequentially, first by defining the 3D boundary of the H3K27me3/*Xist* domain in the absence of region-specific FISH signal, then by marking the center of each region-specific FISH signal with a 0.25μm sphere in the absence of the H3K27me3/*Xist* signal. All image annotations were then visualized contemporaneously and the location of region-specific FISH signals were determined relative to the defined 3D boundary of H3K27me3/*Xist*. The signal was called "Inside" if the sphere was completely covered by the reconstructed H3K27me3/*Xist* boundary, "Edge" if the sphere intersected at all with the boundary, and "Outside" if the sphere did not intersect at all with the boundary. Final counts represent the summed data from a minimum of biological replicates.

BACs and fosmids were ordered from the BACPAC resource center and fingerprinted with Hind III and BamH I for insert verification. *Xist* was detected with the fosmid G135P63425C4, *Abcb7* with BAC RP24-274B9, *Rnf12* with fosmid G135P605237C7, *Jarid1c* with fosmid G135P603627A9, *Taf1-Ogt* with BAC RP24-325M16, *Nkap* with fosmid G135P601281C2, *Utx* with BAC RP23-174N2, *Huwe1* with BAC RP23-224F24, and *TOA* with BAC RP24-286J22. Other BACs and fosmids are described in Figure S2C. Fluorescent labeling was performed with Invitrogen's BioPrime kit.

Bioinformatic Analysis

Sequence Alignment and Genome Annotations

Sequence reads were aligned to genomic sequence with Bowtie (Langmead et al., 2009). All mm9 genome annotations were obtained from the UCSC genome browser (Rhead et al., 2010). Variant sequence data were obtained from the Sanger Institute (http://www.sanger.ac.uk/resources/mouse/genomes/). Only reads that uniquely aligned to B6 or Cast genomes were used for all downstream analyses. A slight bias toward the B6 genome was observed (Figure 2A) due to unannotated SNPs and insertions and deletions present in the Cast genome that were absent from Sanger variant data.

Gene Annotation

The annotation of splice and transcript variants has created considerable redundancy among the set of UCSC Known Genes. To eliminate redundancy but retain the complexity of known annotations for gene expression and chromatin analyses, a set of spatially distinct genes was annotated from the set of UCSC Known Genes (build mm9, downloaded 1/10/2010). Any group of Known Genes that overlapped by at least 90% of their length and matched in strand were collapsed into a spatially distinct gene. Gene sets were named by creating a comma-delimited list of the distinct Gene Symbol names associated with the spatially distinct gene. By this definition, mm9 contains 26,396 spatially distinct genes from a total of 49,409 Known Gene annotations (Table S3).

SNP-overlapping reads and allelic expression ratios per gene

To find SNP-overlapping reads, Sanger-annotated Cast SNPs were downloaded on 1/15/10, and Cast alleles were substituted into their corresponding mm9 positions. Sequence reads were then aligned to both mm9 and the Cast version of mm9, selecting for unique perfect matches. Sequences that overlapped at least one SNP from one genome and did not match anywhere in the other were counted as SNP overlapping. SNP-overlapping reads falling between the start and end of each spatially distinct gene annotation were counted to determine allelic expression ratios. Early in the analysis it became apparent that a small proportion of Cast SNPs were either located in unannotated nonunique portions of the genome (at 35 bp resolution) or were incorrectly annotated as SNPs. Such SNPs unidirectionally skewed allelic expression ratios of overlapping genes toward a single genotype, regardless of parent-of-origin. For example, the X-linked genes *Rps4x* and *Msn* showed initial allelic expression ratios skewed toward the B6 and Cast genomes, respectively. In C/B TSCs, *Rps4x* was skewed toward the B6 genome, or Xi, with a B6-to-Cast ratio of 177:30. In B/C TSCs, *Rps4x* again showed heavy skewing toward the B6 genome, now the Xa, with a B6-to-Cast ratio of 2,141:3. *Msn* showed an opposite pattern of skewing, biased toward the Cast genome. In C/B TSCs, *Msn* had a B6-to-Cast ratio of 9:739 and in B/C the B6-to-Cast ratio was 643:150. For each gene, the bulk of the skewed allelic data derived from reads overlapping a single SNP. PCR-based validation of these genes showed very different allelic ratios, strongly suggesting that the single outlier SNP for each gene was incorrectly annotated; either the genomic position surrounding the SNP was nonunique at 35 bp resolution, or the annotated SNP did not exist in our Cast and B6 mice from Jackson Labs. We noted that such anomalous SNPs frequently colocalized with large clusters of nonuniquely mapping RNA-Seq reads. We therefore decided to exclude SNPs that exhibited such colocalization, reasoning that true allelic ratios should remain relatively constant even if a small number of correctly annotated SNPs were removed from the analysis. Ultimately, only 7,121 SNPs were excluded genome-wide from the total set of 17,744,681. Reanalyzing allelic expression of *Rps4x* and *Msn* with the filtered set of SNPs gave allelic ratios that were much more in line with the values obtained by our PCR-based assay (not shown). Therefore, all subsequent allelic analysis was performed with the filtered set of SNPs.

Gene Expression

Sequence data were transformed into gene expression levels similar to (Mortazavi et al., 2008), with some alterations. For each spatially distinct gene, the total set of distinct exonic 35mers was extracted from associated UCSC Known Gene mRNAs, including 35mers spanning annotated splice junctions. The proportion of these 35mers that uniquely matched either mm9 or UCSC Known Gene exon junctions was then determined by sequential alignment to mm9 and the set UCSC Known Gene mRNAs. Final gene expression values were displayed in Reads Per uniquely mapping Kilobase of 35mers per Million total reads, or rpkm. Highly expressed genes are those expressed at greater than 10 rpkm. Expression values are reported in Table S3.

Determination of XCI Status, Overview

The beta-binomial model has been used to evaluate allelic expression differences from human HapMap cell lines (Pickrell et al., 2010). We extended this approach in order to quantitatively assess XCI status in TSCs. Posterior probabilities of X inactivation were determined by fitting allelic expression data to a mixture of two beta-binomial distributions, one accounting for X-inactivated genes and the other escaping genes. Both the total number of allele-specific reads per gene and the sequence quality score of the SNP-overlapping base of each of these reads were used to model uncertainty. This method had the added advantage of accounting for the slight bias of data toward the reference/B6 genome (51:49 [B6:Cast], over the entire genome). This bias was expected considering our build of the Cast genome lacked all Cast-specific insertions and deletions.

Ultimately, genes with posterior probabilities of inactivation of less than 5% were considered to escape XCI. These genes could also be thought of as having a >95% chance of escaping XCI, as determined by the optimal fit to the mixture of beta-binomial distributions. Because posterior probabilities can be interpreted as False Discovery Rates (FDRs), our method implicitly makes adjustments for multiple testing and limits the number of false positives. To increase confidence in annotation of escaping genes, only genes with 20 or more allele-specific reads were considered eligible to be evaluated for escape. Using this cutoff, 660 X-linked genes were excluded from the analysis; 370 were expressed but did not have 20 or more SNP-overlapping reads in both TSC lines, and 290 were not transcribed in TSCs. The prior probability that a gene did not escape XCI was modeled by a logistic regression with two predictors: the total number of sequence reads escaping XCI and the summation of quality scores of the escaping reads. Based on the prior knowledge that most X-linked genes did not escape XCI, prior probabilities were restricted to be no smaller than 0.2 to avoid excessive influence on posterior probabilities. The final results were insensitive to minimum prior probability cutoffs between 0.05 and 0.3. The parameters in our model were estimated by maximizing the overall mixture model likelihood, and the posterior probability that each gene did not escape XCI was then calculated based on the parameter estimates.

The resulting output of these calculations ultimately allowed for a more complete annotation of TSC XCI status than would have been achieved with a 10% cutoff to define escape. For example, the gene *Rnf12* was represented by 2,200 SNP-overlapping reads in the C/B TSC line, only three of which, or 0.1% of the total, mapped to the Xi. By all accounts *Rnf12* would be considered X-inactivated. In contrast, the gene *Rnf128* was represented by 1,293 SNP-overlapping reads in C/B TSCs, 40 of which, or 3.1% of the total, mapped to the Xi. Using a cutoff of 10% Xi expression to define escape, both *Rnf12* and *Rnf128* would be grouped together as inactivated genes. However, it stands to reason that the mechanisms governing *Rnf12* and *Rnf128* expression from the Xi are different, given that *Rnf128* exhibited ~30 times more Xi expression than *Rnf12* in C/B TSCs. Under this logic, *Rnf128* could be considered an escaper, and our statistical framework classified it as such. Similarly, the known escaping gene *Utx* exhibited 5% expression from the Xi in TSCs and would have been classified as inactivated with a 10% cutoff to define escape. Lastly, for the handful of genes at the low end of our allele-specific read cutoff, modeling escape with the beta binomial distribution allowed both the total number, and

sequence quality scores, of Xi-overlapping reads to impact the classification of inactivation status. For example, the gene *Magee1* had the same percentage of Xi-overlapping reads as *Utx*, but had 10-fold fewer allelic reads, with only 20. *Magee1* was not classified as an escaper and Utx was. This stands to reason, considering there was a higher probability that *Magee1*'s single B6-associated read derived from sequencing error, whereas with 10 times the amount of B6-associated reads, the probability that *Utx*'s Xi data derived from sequencing error was exceeding low.

Determination of XCI Status, Statistical Methods

For gene i, let the number of allele-specific RNA-Seq reads mapped to the inactivated/activated chromosomes be $n_{i,0}$ and $n_{i,1}$, respectively, and let $n_{i,0} = n_{i,0} + n_{i,1}$. We first model $n_{i,0}$ by a binomial distribution:

$$p(n_{i,0}|n_i, p_i) = \binom{n_i}{n_{i,0}} p_i^{n_{i,0}} (1 - p_i)^{n_i - n_{i,0}},$$

where p_i indicates the expected proportion of reads from the inactivated chromosome. We further assume that p_i follows a mixture of two beta distributions:

$$f(p_i) = \pi_{i0} f_0(p_i; \alpha_0, \beta_0) + (1 - \pi_{i0}) f_1(p_i; \alpha_1, \beta_1), \quad (1)$$

where $f_0(p_i; \alpha_0, \beta_0)$ and $f_1(p_i; \alpha_1, \beta_1)$ are two beta distributions for inactivated genes and genes that escape inactivation, respectively, and α_0, β_0, α_1, and β_1 are the unknown parameters to be estimated. Known inactivated genes, such as Rnf12, have p_i approaching 0. Therefore, in general, p_i's from $f_0(p_i; \alpha_0, \beta_0)$ are small (e..g, <0.01), reflecting possible sequencing errors. π_{i0} is the prior probability that gene i is inactivated. We integrate out p_i to obtain the posterior distribution of $n_{i,0}$ in terms of α_0, β_0, α_1, and β_1:

$$p(n_{i,0}|n_i, \alpha_0, \beta_0, \alpha_1, \beta_1) = \int p(n_{i,0}|n_i, p_i) f(p_i) dp_i = \pi_{i0} h_{i0} + (1 - \pi_{i0}) h_{i1}$$

where h_{i0} and h_{i1} are two beta-binomial distributions:

$$h_{i0} = \binom{n_i}{n_{i,0}} \frac{B(n_{i,0} + \alpha_0, n_i - n_{i,0} + \beta_0)}{B(\alpha_0, \beta_0)}$$

$$h_{i1} = \binom{n_i}{n_{i,0}} \frac{B(n_{i,0} + \alpha_1, n_i - n_{i,0} + \beta_1)}{B(\alpha_1, \beta_1)}$$

and $B(\alpha, \beta)$ is beta function with parameters α and β. The beta-binomial distribution is a generalization of binomial distribution to allow extra variance, which has been used to model RNA-Seq data before (Pickrell et al., 2010). In this study, the extra variability comes from the fact that each gene has its own proportion of reads escaping inactivation.

For each read, we can obtain a base-calling quality score at the SNP location. We model the prior probability that one gene escapes inactivation by a logistic regression with two predictors: the total number of escaping reads and the summation of quality scores of these reads (denoted by q_i):

$$\log\left(\frac{\pi_{i0}}{1 - \pi_{i0}}\right) = b_0 + b_1 n_{i,0} + b_2 q_i, \quad (2)$$

where b_0, b_1, and b_2 are regression coefficients to be estimated.

Now we have finished the model setup and there are altogether seven parameters to be estimated: α_0, α_1, β_0, β_1, b_0, b_1, and b_2. We estimated these parameters by a Maximum Likelihood approach by using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). For the robustness of the algorithm, and based on the prior belief that most of genes are inactivated, we impose an extra restriction that $\pi_{i0} \geq 0.2$. This is equivalent to adding a large penalty $\lambda I(\pi_{i0} < 0.2)$ to the likelihood, where λ is an arbitrary large positive number and $I(\pi_{i0} < 0.2)$ is an indicator function that equals 1 if $\pi_{i0} < 0.2$ and 0 otherwise. To maximize this alternative likelihood, we simply maximize the original likelihood and set π_{i0} to be 0.2 if its estimate is smaller than 0.2. Our final results remain similar for any π_{i0} cutoff from 0.05 to 0.3. Given the parameter estimates from the EM algorithm, we can estimate the posterior probability that one gene is inactivated by:

$$\hat{\tau}_{i0} = \frac{\hat{\pi}_{i0} \hat{h}_0}{\hat{\pi}_{i0} \hat{h}_{i0} + \hat{\pi}_{i1} \hat{h}_{i1}},$$

where the hat sign ^ indicates the estimate of the corresponding parameter. We then assign one gene as activated or inactivated based on $\widehat{\tau}_{i0}$, which can also be interpreted as local False Discovery Rate (FDR) (Efron et al., 2001). If we claim one gene is activated when $\widehat{\tau}_{i0} \leq \tau_C$, then the overall FDR is $\sum_i \widehat{\tau}_{i0} I(\widehat{\tau}_{i0} \leq \tau_C) / \sum_i I(\widehat{\tau}_{i0} \leq \tau_C)$, where $I(\widehat{\tau}_{i0} \leq \tau_C)$ is an indicator function.

Determination of Absolute Xi Expression Levels

RPKM expression levels for X-inactivated genes were estimated by multiplying the total RPKM expression level per X-inactivated gene by the proportion of allelic reads detected from the Xi. Considering the median Xi-to-Xa ratio was about 1:200, genes with allelic read counts around or below 200 might not be expected to display Xi signal, even if these genes were expressed from the Xi. Therefore, only those genes with nonzero allelic expression values from the Xi were included in this calculation, to avoid bias resulting from nondetection of Xi expression.

Genome Alignability

The proportion of mm9 that could be uniquely mapped with 20, 35, 45, or 90 bp sequence tags, depending on the data set in question, was defined here as genome alignability.

ChIP-Seq Peak Calling

Regions of ChIP-Seq enrichment were annotated with the MACS algorithm, with the H3 data set as a control (Zhang et al., 2008). For analysis of autosomal H3K27me3, peaks above the top 20th percentile of enrichment were used.

Chromosome Tiling and Box Plots

ChIP-Seq tiling plots in Figures 2 and 3 were created by summing data in 40 kb bins across each chromosome, moving in 4 kb increments. To determine normalized reads per million per bin, total reads per bin were divided by the proportion of alignable bases per bin, then multiplied by 1 million and divided by the total number of reads per data set. Bins with alignability of less than 0.5 (i.e., less than 50% alignable at 35bp resolution) were excluded from tiling density plots to avoid potential uncertainty that would be introduced by normalizing highly nonunique regions. LINE and gene density tiling plots in Figure 3 were constructed with 200 kb bins tiled in 20 kb increments. Exon density served as a surrogate for gene density in these analyses. The proportion of DNA per bin that derived from LINEs/exons was determined with data available from the UCSC genome browser (Fujita et al., 2011). Spearman coefficients comparing different tiling density plots were determined with 40 kb tiled bins, selecting every 10th bin in order to avoid autocorrelation effects. Tiling and box plots were constructed with R and the ggplot2 package (http://www.R-project.org).

Metagene Analyses

Metagene plots are representative images of the average histone modification density over a particular class of gene. Those from Figures 4, 5, S4, and S5 were constructed by recording counts of sequence read starts per 25bp or 500bp bin surrounding annotated TSS for the gene class in question. In addition to normalization for gene number, allelic metagene analysis required normalization for the number of uniquely alignable SNPs present in each bin for the specific gene set being analyzed. SNPs were assigned a numerical value between 0 and 1, which represented the fraction of uniquely aligning bases that overlapped the SNP relative to the total number of bases overlapping the SNP. As expected, the total number of alignable SNPs per bin increased with read length but did not vary substantially between B6 and Cast genome builds, as visualized in Figure S3 for X-inactivated and nontranscribed X-linked genes.

Comparison of X-Linked and Autosomal Gene Chromatin Patterns

Expression-matched autosomal gene sets for comparison with X-inactivated genes were selected by taking those autosomal genes with expression levels between the 1st and 3rd quartile of RPKM values for X-inactivated genes from the Xa and Xi. Ultimately three classes of autosomal genes were examined: those expressed at similar levels to X-inactivated genes on the Xa (A^{Xa}), those expressed at similar levels to X-inactivated genes on the Xi (A^{Xi}), and nontranscribed genes (A^{NT}). Because the set of X-linked genes with large amounts of allelic data suggested that X-inactivated genes were expressed at low levels from the Xi, the 1st quartile RPKM expression cutoff for the A^{Xi} gene class was pushed above "0", to avoid contaminating the A^{Xi} gene class with the ~11,000 autosomal nontranscribed genes.

Xa/Xi, A^{Xa}/A^{Xi} and A^{Xa}/A^{NT} ratios were calculated by dividing the metagene signals of the 500 base pairs surrounding the TSS-associated peak for the compared classes. Because allelic data from X-linked metagene profiles could not be directly compared to total metagene data from expression-matched autosomal gene sets, given their different metagene profile bin sizes, ratios between X-linked and autosomal gene classes were calculated separately, and data were compared under that the assumption that signal from Xa and A^{Xa} genes would be equivalent, given their similarly robust expression levels.

Allelic Peak Assignment

Allelic assignment of peaks was a function of read and SNP density within each peak, as well as background levels for each data set. Therefore, prior to allelic peak assignment, estimates of antibody-associated background for ChIP-, DNase, and FAIRE-Seq data sets were calculated. SNPs that were not within 10 kb of a MACS-annotated peak were considered to be located within data set background regions, given that they overlapped genomic regions that were not detected by MACS as statistically enriched within the data set in question. The number of reads overlapping each background SNP was then stored in a list and subsequently used to calculate significance of allelic events.

To calculate p-values for significant allelic binding events, the number of allelic reads falling within each peak (the actual signal) was compared to what would be expected were those reads due to background overlap, based on the total number of SNPs per peak. The expected amount of background per peak was calculated by performing 100,000 simulations with the set of background SNPs

described above, and counting the number of times the background signal was greater than or equal to the actual signal. For example, a particular CTCF peak had 42 reads overlapping B6 alleles, from 7 total SNPs. Using the CTCF B6 background set, 7 randomly selected SNPs never accounted for 42 reads in 100,000 permutations. This peak was clearly enriched on the B6 allele relative to background. Another peak had 2 reads overlapping B6 alleles, from 4 total SNPs. Using the CTCF B6 background set, 4 randomly selected SNPs accounted for at least two reads in 48% of the 100,000 permutations. This peak was not enriched on the B6 allele relative to background.

The threshold of significance to call a peak allelically assignable was set at a p value that gave a false discovery rate per data set of less than or equal to 5%. Peaks could either be detected with significance only on the Xa or Xi (referred to as ''Xa pref'' or ''Xi pref'' in the manuscript), both chromosomes (''XaXi''), or on neither chromosome (''NA peaks''). The locations and properties of allelically assigned peaks are shown in Table S4.

SUPPLEMENTAL REFERENCES

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. *20*, 2223–2237.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Series B Stat. Methodol. *39*, 1–38.

Efron, B., Tibshirani, R., Storey, J.D., and Tusher, V.G. (2001). Empirical Bayes analysis of a microarray experiment. J. Am. Stat. Assoc. *96*, 1151–1160.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., et al. (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res. *39*(Database issue), D876–D882.

Giresi, P.G., and Lieb, J.D. (2009). Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Methods *48*, 233–239.

Huynh, K.D., and Lee, J.T. (2003). Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature *426*, 857–862.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Kalantry, S., and Magnuson, T. (2006). The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. PLoS Genet. *2*, e66.

Kalantry, S., Purushothaman, S., Bowen, R.B., Starmer, J., and Magnuson, T. (2009). Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. Nature *460*, 647–651.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

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

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature *464*, 768–772.

Pugacheva, E.M., Tiwari, V.K., Abdullaev, Z., Vostrov, A.A., Flanagan, P.T., Quitschke, W.W., Loukinov, D.I., Ohlsson, R., and Lobanenkov, V.V. (2005). Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. Hum. Mol. Genet. *14*, 953–965.

Quinn, J., Kunath, T., and Rossant, J. (2006). Mouse trophoblast stem cells. Methods Mol. Med. *121*, 125–148.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2010). The UCSC Genome Browser database: update 2010. Nucleic Acids Res. *38*(Database issue), D613–D619.

Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb. Protoc. *2*, pdb prot5384.

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

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

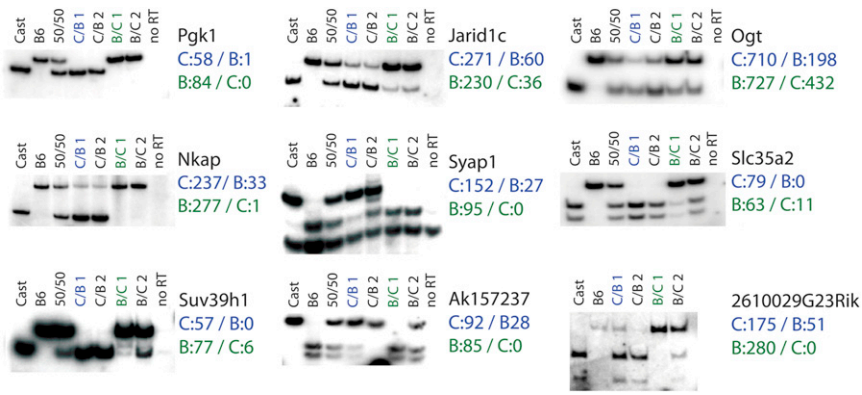
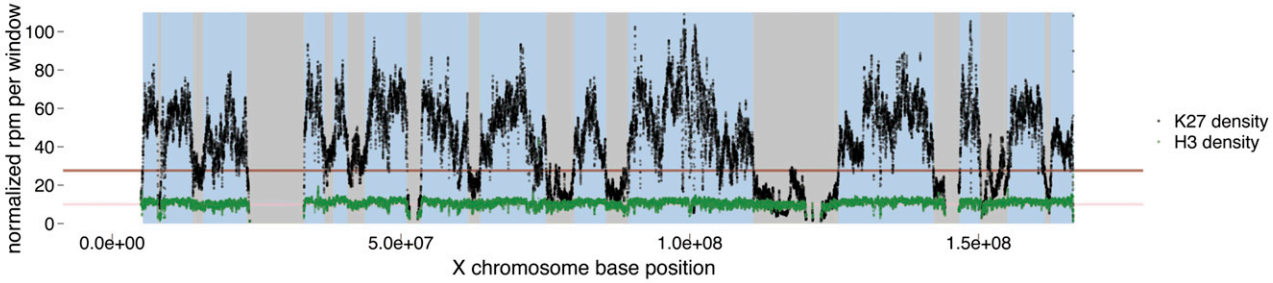


Figure S1. RNA-Seq Validation and Examination of Escape in a Panel of TSC Lines, Related to Figure 1

Quantitative allele-specific RT-PCR assays in four independently derived TSC lines. Allelic expression was distinguished by differential mobility of PCR amplicons after restriction digestion of SNP-containing regions. cDNA template source is indicated above each gel lane. "Cast" and "B6" represent cDNA from inbred TSC lines. "50/50" represents an equal mixture of Cast and B6 cDNA. C/B 1 in blue and B/C 1 in green were used for RNA-Seq. Allelic counts from RNA-Seq data are shown for each line. *Pgk1* was X-inactivated; *Jarid1c* and *Ogt* escaped in all tested TSCs; *Nkap* and *Syap1* escaped in C/B not B/C; *Slc35a2* and *Suv39h1* escaped in B/C and not C/B; *Ak157237* and *2610029G23Rik* showed nonpredictable TSC escape patterns.

A



B

P/V #	Start	End	Length	Alignability	Prop. LINE
peak1	5300000	7900000	2600000	0.8491	0.2391
valley1	7900000	8450000	550000	0.4197	0.4959
peak2	8450000	14000000	5550000	0.8389	0.2051
valley2	14000000	15750000	1750000	0.7439	0.4986
peak3	15750000	23250000	7500000	0.7706	0.3712
valley3	23250000	33250000	10000000	0.0996	0.516
peak4	33250000	36750000	3500000	0.7468	0.2521
valley4	36750000	38250000	1500000	0.774	0.49
peak5	38250000	40750000	2500000	0.8476	0.2747
valley5	40750000	43750000	3000000	0.7601	0.5153
peak6	43750000	51000000	7250000	0.8333	0.2871
valley6	51000000	53500000	2500000	0.2989	0.3444
peak7	53500000	61600000	8100000	0.8128	0.3363
valley7	61600000	63750000	2150000	0.7431	0.5095
peak8	63750000	75100000	11350000	0.7509	0.3333
valley8	75100000	80000000	4900000	0.716	0.4883
peak9	80000000	85500000	5500000	0.8472	0.3033
valley9	85500000	89350000	3850000	0.6877	0.4864
peak10	89350000	111000000	21650000	0.7952	0.363
valley10	111000000	125750000	14750000	0.6394	0.4429
peak11	125750000	142250000	16500000	0.8274	0.2912
valley11	142250000	146750000	4500000	0.4037	0.3341
peak12	146750000	150250000	3500000	0.8092	0.3051
valley12	150250000	155000000	4750000	0.758	0.3905
peak13	155000000	161450000	6450000	0.8595	0.2256
valley13	161450000	162500000	1050000	0.7712	0.4185
peak14	162500000	166450000	3950000	0.8792	0.2065

C

Figure 3 probe ID:	Ube1x	15L	83H	78L	Chic1	87L	Rnf12	116L	160H	152L
start	20224883	14720724	83580255	78014372	100537038	86998950	101145921	115875046	159658600	152749993
stop	20264898	14919689	83773430	78205725	100577585	87167836	101187238	116079262	159867586	152931686
length	40015	198965	193175	191353	40547	168886	41317	204216	208986	181693
% align	92	80	85	74	92	78	94	74	82	75
% LINE	5	55	30	54	15	51	7	56	27	56
K27me3 (rpk)	513	185	410	95	638	101	574	45	368	113
% cells w/2 dots	91	94	96	95	90	94	98	100	100	100
BACPAC ID	G135-P65743A11	RP23-204P18	RP23-351J22	RP23-267N5	G135-P66518D5	RP23-368B24	G135-P605237C7	RP23-468D22	RP23-133E13	RP23-259M13

Figure S2. Approximate Divisions of Major H3K27me3 Peaks and Valleys along the Xi, Related to Figure 3

(A) Tiling density plots of H3K27me3 and total H3 across the X, with approximate boundaries of H3K27me3 major peaks (blue boxes) and valleys (gray boxes). The red line represents the 1st quartile of H3K27me3 density along the X, and was used to roughly divide the X into its major peaks and valleys. The pink line represents the average genomic H3K27me3 density per 40 kb bin. Considering the inverse correlation between LINEs and H3K27me3, and the high LINE density within regions of low alignability, continuous regions of tiled density with less than 50% alignability were annotated as major valleys.

(B) Genomic coordinates of major peaks and valleys, shown with the proportion of each region that is uniquely alignable at 35bp resolution, and the proportion of sequence within each region that is LINE derived.

(C) Characteristics of BACs and Fosmids used in Figure 3B–F. *% align*, the percent of each region that is uniquely alignable at 35bp resolution. *% LINE*, the percent of each region that is LINE derived. *K27me3 (rpk)*, the number of H3K27me3 reads per kilobase, normalized for alignability, within each region. *% cells w/2 dots*, the percent of visualized cells for each probe that had 2 probe dots per Xi per nucleus, included to provide the reader with an estimate for the specificity of LINE-dense probes; LINE-dense probes showed at least as much specificity as gene-dense probes, perhaps not surprising, considering that despite their repetitive origin, the examined LINE-dense regions were at least 74% alignable at 35bp resolution. *BACPAC ID*, the reference ID for each BAC/Fosmid.

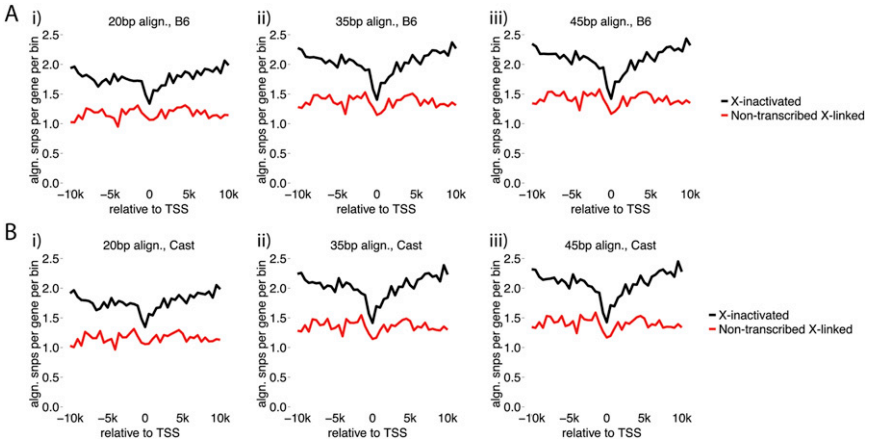


Figure S3. SNP Density Is Variable within Metagene Profiles and between Gene Sets, Related to Figure 4

Number of alignable SNPs per 500bp bin surrounding X-inactivated and nontranscribed X-linked genes. Data are shown for both B6 (A) and Cast genomes (B), as well as for 20, 35, and 45bp reads (i, ii, and iii, respectively).

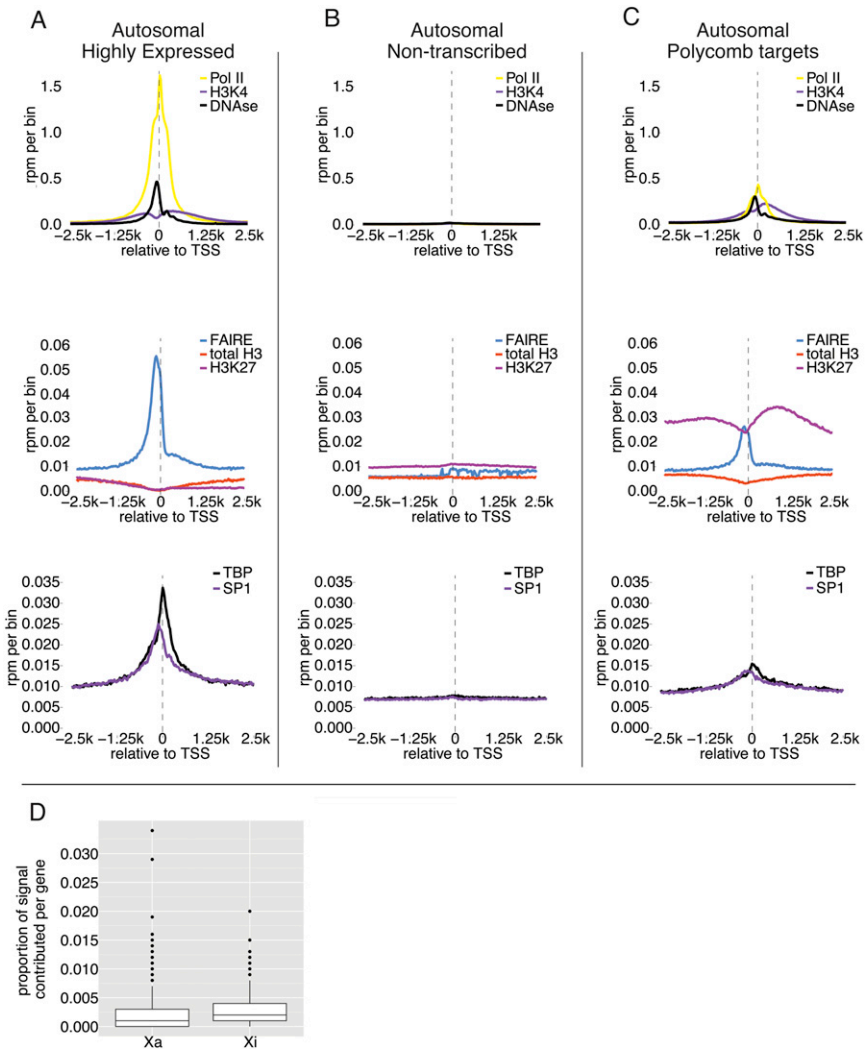


Figure S4. Additional Metagene Chromatin Data for Autosomal and X-Linked Genes, Related to Figure 5

(A–C) Metagene profiles of –Seq data sets at autosomal (A) highly expressed, (B) nontranscribed, and (C) Polycomb-target genes. ChIP-Seq data are represented by the proportion of total data per 25 bp bin.

(D) Boxplot showing proportion of total DNase-Seq data contributed per gene for Xa and Xi metagene profiles, respectively.

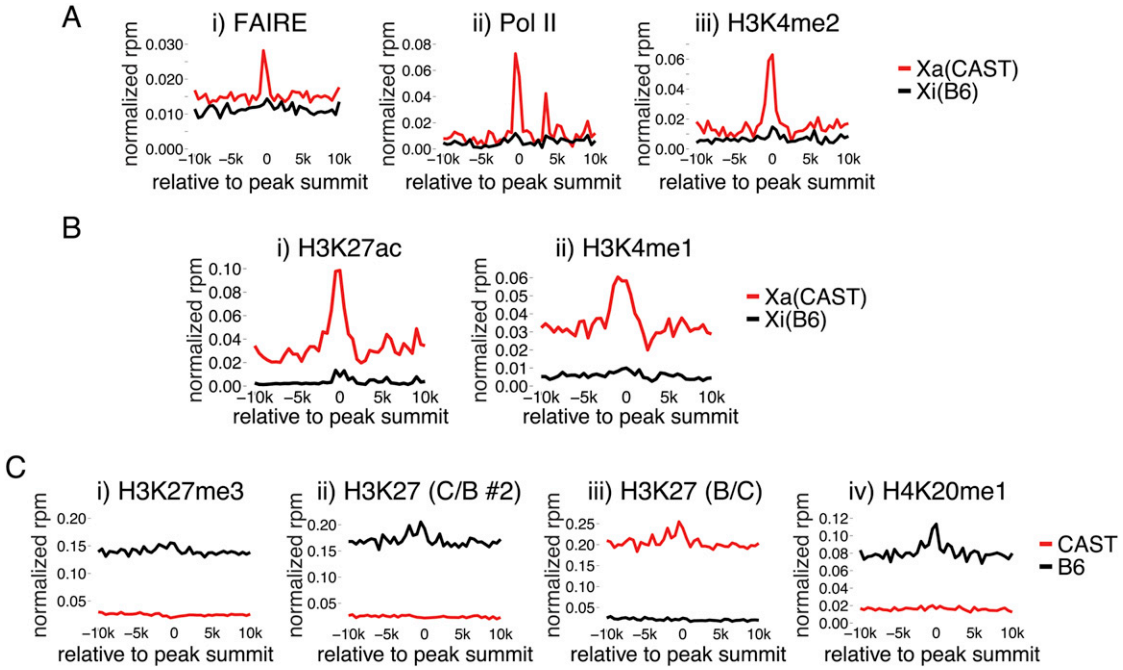


Figure S5. Additional Properties of X-Linked Regulatory Elements, Related to Figure 5

(A) Allelic metagene profiles of FAIRE, Pol II, and H3K4me2 at intergenic peaks of DNase I hypersensitivity.

(B) Allelic metagene profiles of H3K27-acetylation and H3K4me1 at intergenic peaks of DNase I hypersensitivity.

(C) Allelic metagene profiles of H3K27me3 and H4K20me1 at intergenic peaks of DNase I hypersensitivity.

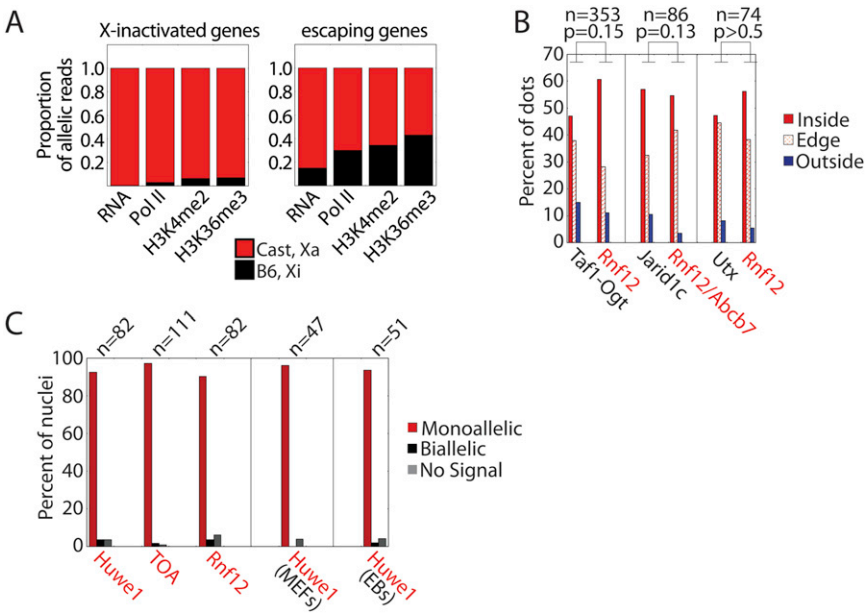


Figure S6. Additional Escaping Gene Properties, Related to Figures 6 and 7

(A) Allelic distribution of sequence data over X-inactivated and escaping genes. ChIP-Seq data represents SNP-overlapping reads within MACS-defined regions of factor enrichment.

(B) Location of analyzed DNA FISH probes relative to the Xi's H3K27me3 domain, broken down by percentage. The total number of Xi's counted per probe pair is shown. P-values derive from Chi-square tests comparing the total number of dots inside/on the edge versus outside the H3K27me3 domain per probe pair. Black and red gene names denote escapers and X-inactivated genes, respectively. Black lines separate analyzed probe pairs.

(C) RNA-FISH confirming that the *Huwe1* and *TOA* loci are subject to XCI in TSCs, MEFs, and EBs.