

RESEARCH

Open Access



Bipartite structure of the inactive mouse X chromosome

Xinxian Deng^{1†}, Wenyu Ma^{2†}, Vijay Ramani², Andrew Hill², Fan Yang¹, Ferhat Ay², Joel B. Berletch¹, Carl Anthony Blau^{3,4}, Jay Shendure², Zhijun Duan^{3,4*}, William S. Noble^{2,5*} and Christine M. Disteche^{1,6*}

Abstract

Background: In mammals, one of the female X chromosomes and all imprinted genes are expressed exclusively from a single allele in somatic cells. To evaluate structural changes associated with allelic silencing, we have applied a recently developed Hi-C assay that uses DNase I for chromatin fragmentation to mouse F1 hybrid systems.

Results: We find radically different conformations for the two female mouse X chromosomes. The inactive X has two superdomains of frequent intrachromosomal contacts separated by a boundary region. Comparison with the recently reported two-superdomain structure of the human inactive X shows that the genomic content of the superdomains differs between species, but part of the boundary region is conserved and located near the *Dzx4/DXZ4* locus. In mouse, the boundary region also contains a minisatellite, *Ds-TR*, and both *Dzx4* and *Ds-TR* appear to be anchored to the nucleolus. Genes that escape X inactivation do not cluster but are located near the periphery of the 3D structure, as are regions enriched in CTCF or RNA polymerase. Fewer short-range intrachromosomal contacts are detected for the inactive alleles of genes subject to X inactivation compared with the active alleles and with genes that escape X inactivation. This pattern is also evident for imprinted genes, in which more chromatin contacts are detected for the expressed allele.

Conclusions: By applying a novel Hi-C method to map allelic chromatin contacts, we discover a specific bipartite organization of the mouse inactive X chromosome that probably plays an important role in maintenance of gene silencing.

Background

Chromosomes occupy specific territories within the nucleus [1]. In diploid cells, homologous chromosomes occupy separate territories, but expression from genes located on either the paternal or maternal homolog is usually similar. In contrast, X-linked genes are subject to silencing by X chromosome inactivation (XCI) on one of the two homologs in female somatic cells [2], and a subset of autosomal genes are subject to imprinting and expressed from either the paternal or maternal allele [3]. These exceptional genomic regions thus exhibit radically

different expression levels of each allele. At the same time, allele-specific chromatin conformation/contact differences are observed at these regions [4–7]. The inactive X chromosome (Xi) becomes highly condensed compared with the active X (Xa) and forms the Barr body, often visible as a dense region within the nucleus [8]. The heterochromatization of one of the X chromosomes in female somatic cells is initiated by the long non-coding RNA (lncRNA) *Xist* that coats the Xi in early embryogenesis and silences transcription by recruiting specific proteins that put in place repressive histone modifications such as tri-methylation of histone H3K27, ubiquitination of histone H2AK119, and de-acetylation [9–12]. Additional layers of silencing involve DNA methylation at the CpG islands of X-linked genes and late replication [13]. Some genes, representing about 10–15 % of X-linked genes in human and 3–5 % in mouse, escape XCI and thus remain expressed from both alleles [14–16]. In addition to becoming condensed, the Xi occupies a particular nuclear

* Correspondence: zjduan@uw.edu; william-noble@uw.edu; cdistech@u.washington.edu

†Equal contributors

³Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, Washington, USA

²Department of Genome Sciences, University of Washington, Seattle, Washington, USA

¹Department of Pathology, University of Washington, Seattle, Washington, USA

Full list of author information is available at the end of the article

compartment near the nuclear membrane or the nucleolus [17, 18]. Imprinted genomic regions also undergo epigenetic and conformational changes associated with silencing of one allele [3]. Such conformation changes are considered to involve the formation of loops that bring enhancers and promoters together at the expressed allele [6, 7].

Little is known about the three-dimensional (3D) structure of the X chromosomes and of alleles at the imprinted regions. Previous studies proposed that the Xi condenses around a core of LINE1 (L1) repetitive elements [19], with genes located in the outer layer and escape genes in the most outer layer [20]. Limited analyses of chromatin contacts at specific regions of the mouse X chromosome have been done using chromatin conformation capture approaches such as 4C and 5C [4, 5]. Approaches to visualize the 3D configuration of the whole nucleus and of entire chromosomes include Hi-C, a method to identify chromatin contacts within chromosomes (intrachromosomal) or between chromosomes (interchromosomal). Topologically associated domains (TADs) representing domains (median size 800 kb) of enhanced intrachromosomal contacts have been defined along the human and mouse genomes by Hi-C [6, 21, 22]. TADs are separated by boundary regions often enriched in CTCF [22], and disruption of a boundary can affect adjacent TADs, supporting a functional role in 3D organization [4].

We have previously reported a novel type of Hi-C method, DNase Hi-C, which uses DNase I rather than restriction enzymes for chromatin fragmentation, leading to improved efficiency and mappability compared with Hi-C methods based on restriction enzyme digestion [23]. By combining our original DNase Hi-C protocol with nuclear ligation [24], we have now implemented an “in situ” version of DNase Hi-C, which is dramatically simplified and much easier to use. As observed with in situ Hi-C [6], this updated protocol also reduces the frequency of spurious contacts due to random ligation. Here, we used both DNase Hi-C and its in situ extension to obtain an allelic 3D view of the mouse genome, including the X chromosomes and imprinted regions *in vitro* and *in vivo*.

To discriminate between alleles, we employed F1 mouse hybrid systems that we previously developed, based on single nucleotide polymorphisms (SNPs) in conjunction with skewed XCI [15, 25, 26]. To maximize the number of SNPs that could be used for allele discrimination, C57BL/6J (BL6) female mice were bred to *Mus spretus* males. These two mouse species differ by SNPs with a frequency of 1/70–96 bp, depending on the chromosome. The parent BL6 mice had either an *Hprt* mutation in order to skew XCI (to the BL6 X) in a cell line (Patski) [27], or an *Xist* mutation to skew XCI (to

the *spretus* X) in mouse tissues [25]. Imprinted regions can also be examined using these F1 hybrid systems in which the paternal *spretus* and maternal BL6 alleles can be identified. Allelic gene expression measurements by RNA-seq and quantitative RT-PCR previously verified complete XCI skewing and mono-allelic expression of imprinted genes in the F1 hybrid systems [15, 25, 26].

By applying our DNase Hi-C approach [23] as well as the new adapted *in situ* DNase Hi-C protocol to the mouse F1 hybrid systems described above, we demonstrate that the mouse Xi condenses in a bipartite structure both in the Patski cell line and in mouse brain, representing the first such analysis *in vivo*. The genomic content of the two superdomains differs between mouse and human where a similar structure of the Xi was also recently reported by Hi-C in cultured human cells [6]. However, the boundary region between the two superdomains is partially conserved and contains elements that bind CTCF and nucleolar proteins. We also show that chromatin contacts differ between loci on the paternal and maternal alleles at imprinted genes as well as between genes that escape XCI and genes subject to XCI, suggesting a functional link between chromatin contacts and transcription.

Results

The inactive mouse X chromosome forms a bipartite structure in cultured cells and tissue

We used DNase Hi-C and a modified *in situ* DNase Hi-C (see details in “Materials and methods”) to obtain allele-specific maps of intrachromosomal contacts on the mouse X chromosomes and autosomes. While our published DNase Hi-C [23], like *in situ* Hi-C [6], markedly reduces cellular input requirements, the protocol also requires a time-consuming agarose gel proximity ligation. Inspired by single cell Hi-C [28] and *in situ* Hi-C [6], we simplified our protocol by carrying out proximity ligation within intact nuclei instead of in solid agarose gel. Our updated protocol, termed *in situ* DNase Hi-C, requires only 2–3 days instead of 6–7 days to generate a library, with considerably less hands-on time and lower costs than the original DNase Hi-C.

Data were obtained from two biological replicates of Patski fibroblast cells using *in situ* DNase Hi-C. The Patski cell line in which the Xi is from BL6 was originally derived from the kidney of an 18 days postcoitum F1 female embryo obtained by mating a BL6 female with an *Hprt* mutation to a *spretus* male and growing cells in hypoxanthine-aminopterin-thymidine (HAT) medium [27]. Chromosome analyses confirmed a near-diploid karyotype with two X chromosomes. Importantly, we also obtained Hi-C data *in vivo* by applying DNase Hi-C and *in situ* DNase Hi-C to a whole brain specimen from an F1 adult female mouse derived from a cross between

a BL6 female with an *Xist* mutation and a *spretus* male, in which the Xi is from *spretus* [25]. To identify reads that map to each parental genome in F1 mice, a “pseudo-*spretus*” genome was assembled by substituting known SNPs between BL6 and *spretus* into the BL6 mm9 reference genome as previously described [15]. SNPs were obtained from the Sanger Institute (SNP database Nov/2011 version) and from in-house analysis [26]. After aligning reads separately to the BL6 and to the pseudo-*spretus* genomes we segregated all high-quality uniquely mapped reads (MAPQ ≥30) into three categories: (1) BL6-SNP reads containing only BL6-specific SNP(s); (2) *spretus*-SNP reads containing only *spretus*-specific SNP(s); (3) reads that do not contain valid SNPs. Additional file 1 summarizes the total number of mapped reads and allele-specific reads for each library, and Additional file 2 lists the number of intra-chromosomal and interchromosomal contacts, as well as the number of contacts in relation to distance (from close range to ultra-long range) obtained for each experiment. About one-third of intrachromosomal contacts are >100 kb apart, which accounts for TADs and higher scale chromatin organization (Additional file 2).

The contact maps obtained for each homologous chromosome at 1 Mb resolution by DNase Hi-C or in situ DNase Hi-C in F1 brain and Patski cells show a striking bipartite structure for the Xi, which is very different from that of the Xa (Fig. 1a). In contrast, homologous autosomes appear to have similar structures (Additional file 3). While the Xa forms small contact domains similar to those found on autosomes and representing TADs, the Xi is less topologically organized and shows a high frequency of distant chromosomal contacts within each of two large superdomains (Fig. 1a). Finer-scale analyses at 100 kb resolution for a region around the *Xist* domain (chrX:98.5–103.5 Mb) also showed less defined TADs on the Xi versus the Xa (Additional file 4). Allelic TAD calling at 40 kb was not possible due to insufficient sequence depth. However, combining data from both X chromosomes yielded 102 TADs in F1 brain and 61 TADs in Patski cells, similar to a previous study in cultured fibroblasts [12]. Many of these TADs would be contributed by the Xa.

The two superdomains on the Xi are separated by a boundary region centering at position 72.8–72.9 Mb (mm9) in F1 brain and in Patski cells (Fig. 1a; Additional file 5). These coordinates represent a region with no contact (or the least number of contacts) between superdomains in contact maps at 100 kb and 40 kb resolution (Additional file 5). Hereafter, we refer to superdomain 1 (approximate size 73 Mb) for the domain adjacent to the centromere and superdomain 2 (approximate size 94 Mb) for the distal domain. Within each superdomain we observed a high frequency of contacts (intra-superdomain)

compared with the frequency of contacts between superdomains (inter-superdomain) (Table 1). A bipartite index score was calculated as the ratio between the frequency of intra-superdomain to inter-superdomain contacts. In both F1 brain and Patski cells DNase Hi-C data, the observed bipartite index at the boundary region is significant ($p = 1.25E-3$ and $5.67E-3$, respectively) for the Xi, but not significant for the Xa (Table 1; Fig. 1).

The contact maps obtained by in situ DNase Hi-C for two biological replicates of Patski cells or by regular or in situ DNase Hi-C for F1 brain at 1 Mb resolution show remarkably similar features between replicates, between methods, and between in vitro and in vivo mouse systems (Additional file 6). These results indicate that condensation of the Xi follows a similar pattern in two very different cell types at different developmental stages (embryonic kidney fibroblasts and whole adult brain). A similar structure was reported in a mouse fibroblast cell line [12]. To increase coverage with reads spanning SNPs, data from biological replicates of Patski cells or F1 brain were combined for further analyses. Despite similarities between systems, the index ratio for the bipartite structure was higher for the Xi in F1 brain than Patski cells, suggesting a more condensed structure of the superdomains in brain (Table 1; Fig. 1a; Additional file 6).

Next, we generated 3D models of the mouse X chromosomes, which allowed us to visualize the bipartite structure of the Xi (Fig. 1b). These models are consistent with a turning point or hinge at the boundary region (hereafter named hinge region) between the superdomains. The 3D coordinates of the telomeric ends of the chromosomes are uncertain due to the presence of unmappable regions at the telomeres, especially at the centromeric end that represents a large genomic region enriched in highly repeated, alpha-satellite DNA [29].

The Xi superdomains differ between human and mouse but the hinge region is near DXZ4/Dxz4 in both species

To determine whether contact maps were conserved in mammals, we compared topological domains within the mouse Xi with those previously reported in a human lymphoblastoid cell line [6] (Fig. 2a). In human, the two superdomains on the Xi are of unequal size: superdomain 1 (115 Mb) contains the short arm, centromere and proximal long arm, while superdomain 2 (40 Mb) contains the distal long arm. In contrast, the two mouse superdomains are closer in size (72 Mb for superdomain 1 and 94 Mb for superdomain 2). Maps of synteny between human and mouse X chromosomes show that the gene content of the superdomains differs between species. There are several inversions of genomic material between species, indicating that the Xi 3D structure is not conserved. For example, loci included in mouse superdomain 1 are found in separate superdomains in human (Fig. 2a).

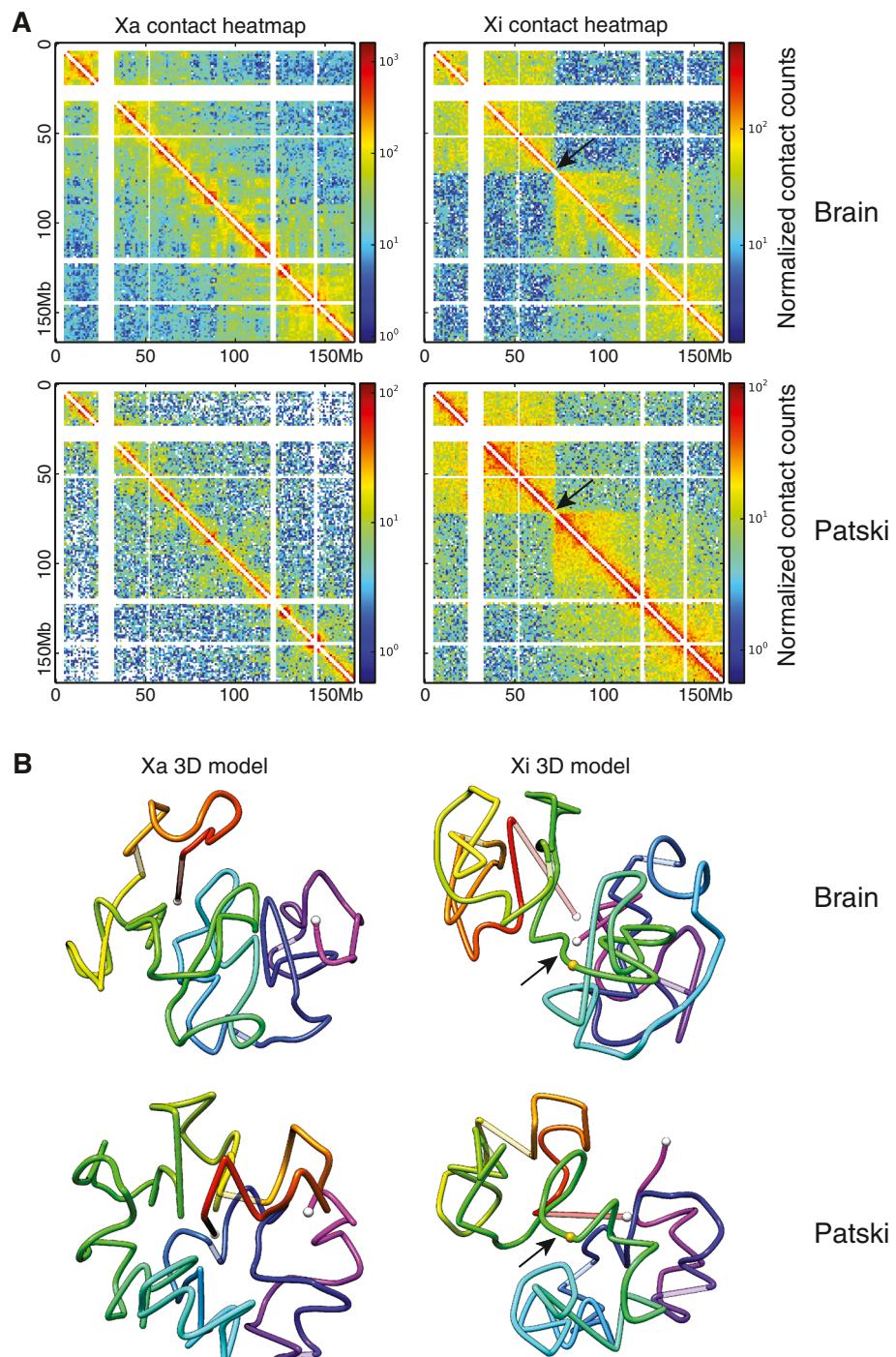


Fig. 1 Bipartite structure of the inactive X chromosome in mouse F1 brain and Patski cells. **a** Allelic intrachromosomal chromatin contact heatmaps of the Xa and Xi based on SNP reads at 1 Mb resolution obtained by DNase Hi-C and in situ DNase Hi-C in female F1 brain (*spretus* Xi) and in Patski cells (BL6 Xi). **b** 3D models of the Xa and Xi built on contact frequency at 1 Mb resolution. White dots represent chromosome ends; lines are colored from red to purple in the direction from centromere to telomere; unmappable regions (corresponding to the white strips in the heatmaps) are set at 75 % transparency; the arrow indicates the hinge region of transition between the two condensed superdomains; the orange dot indicates the position of *Dzx4*

Table 1 Frequency of contacts within superdomains and between superdomains

Library	Chromosome	Frequency of intra-superdomain 1 contacts ^a	Frequency of intra-superdomain 2 contacts ^a	Frequency of inter-superdomain contacts ^a	Bipartite index ^b	P value ^c
Combined-brain	Xa	46.72	55.22	13.80	3.69	0.534
Combined-brain	Xi	32.10	33.42	7.81	4.19	1.25E-03
Combined-Patski	Xa	5.26	6.15	2.32	2.46	0.832
Combined-Patski	Xi	12.03	12.97	3.76	3.32	5.67E-03

^aThe frequency of contacts is calculated as the average number of contacts between two 1 Mb-size windows

^bThe bipartite index represents the ratio between the frequency of intra-superdomain contacts to inter-superdomain contacts

^cThe p value is calculated by one-sided Z-test to measure the significance of the observed superdomain boundary compared with randomly shifted boundaries

Despite differences in the genomic content of the superdomains, the hinge region is partially conserved. In mouse, the end of the hinge region distal to the centromere is near the *Dxz4* locus (nucleotides 72,970,797–73,010,038) that transcribes the lncRNA *4933407K13Rik* (Fig. 2b; Additional file 4). Similarly, one end of the hinge region is near the *DXZ4* locus on the human Xi [6]. The conserved *Dxz4/DXZ4* loci represent macrosatellite repeats previously shown to bind CTCF on the Xi in both human and mouse by comparison of male and female samples [30, 31]. Interestingly, the mouse hinge region also contains a 29 kb gamma satellite repeat, *Ds-TR* (downstream inverted tandem repeat; nucleotides 72,888,859–72,917,881) flanked by a potential promoter region [31]. This minisatellite and flanking promoter region are apparently not present in human.

To determine whether the Xi superdomains and the hinge region could be visualized by microscopy, we performed RNA-fluorescence *in situ* hybridization (RNA-FISH) for *Xist*, which showed coating of two separate regions on the Xi in a subset of nuclei (<10 %) in female primary neurons, mouse embryonic fibroblasts (MEFs), and Patski cells (Fig. 3a–c). The low frequency of nuclei in which we observed a bipartite structure using *Xist* RNA-FISH may be due to the limitation of two-dimensional FISH and/or to the loss of the 3D structure under the denaturation and hybridization conditions of the FISH procedure. DNA-FISH for *Dxz4* following *Xist* RNA-FISH showed a single signal located between the two *Xist*-coated regions on the Xi, suggesting that these regions represent the superdomains detected by Hi-C (Fig. 3b). DNA-FISH using a whole mouse X chromosome paint together with a probe for *Dxz4* confirmed that *Dxz4* is preferentially located on the outside of the condensed Xi, even though the bipartite structure was not clearly visible using the X paint (Fig. 3d).

The hinge region on the mouse Xi binds CTCF and associates with the nucleolus

We previously reported allele-specific profiles of CTCF binding and of RNA polymerase II phosphorylated at serine 5 (hereafter PolII) occupancy obtained by ChIP-seq

in F1 brain and Patski cells [15]. Examination of the hinge region in these datasets shows strong CTCF and PolII peaks at the *Ds-TR* promoter (nucleotides 72,919,240–72,919,749) on both the Xi and Xa, especially in F1 brain, suggesting that *Ds-TR* is expressed on both alleles (Fig. 4a). Note that enrichment was always lower on the Xi than the Xa, similar to what is observed for other genes that escape XCI [15]. No CTCF binding was apparent at *Dxz4* due to the low mappability of repeated sequences. However, by ChIP-chip analysis CTCF binding at *Dxz4* was much higher in female than male mouse liver (Fig. 4b), consistent with Xi binding and with previous studies [31, 32]. Except for strong binding at the promoter, there was no evidence of CTCF binding along the minisatellite *Ds-TR* by either ChIP-seq or ChIP-chip (Fig. 4a, b). CTCF motif analysis using FIMO (Find Individual Motif Occurrences) [33] identified three adjacent CTCF binding motifs at the *Ds-TR* promoter location.

Our previous studies have shown that the lncRNA loci *Dxz4* and *Firre* associate with the nucleolus surface when located on the Xi [32], and the Xi is known to visit the nucleolus [18]. To determine whether the hinge region between the two superdomains on the mouse Xi represents a nucleolus-associated domain (NAD), we isolated nucleoli from Patski cells after fixation to capture genomic regions that associate with the nucleolus [34, 35]. Quantitative PCR (qPCR) showed that *Dxz4*, *Ds-TR*, and its promoter were all enriched in the nucleolar fraction, with *Ds-TR* showing the highest enrichment representing a 6.5-fold increase (Fig. 4c). A 12–25-fold increase was seen for a positive control represented by the 18S ribosomal RNA gene known to associate with the nucleolus (Fig. 4c). The lncRNA loci *Firre* and *Xist* showed a 3.1- and 1.3-fold enrichment, respectively. Three control autosomal genes (*Msl2*, *ActB*, *Gapdh*) and five control X-linked genes (*Mid1*, *Kdm5c*, *Rlim*, *Ddx3x*, *Magee1*) showed low enrichment (0.4–2.6-fold).

We next performed ChIP-chip analysis for nucleophosmin, a protein located at the periphery of the nucleolus. Enrichment was observed at *Dxz4*, *Ds-TR*, and the *Ds-TR* promoter in female liver while in male liver

A

Human Xi

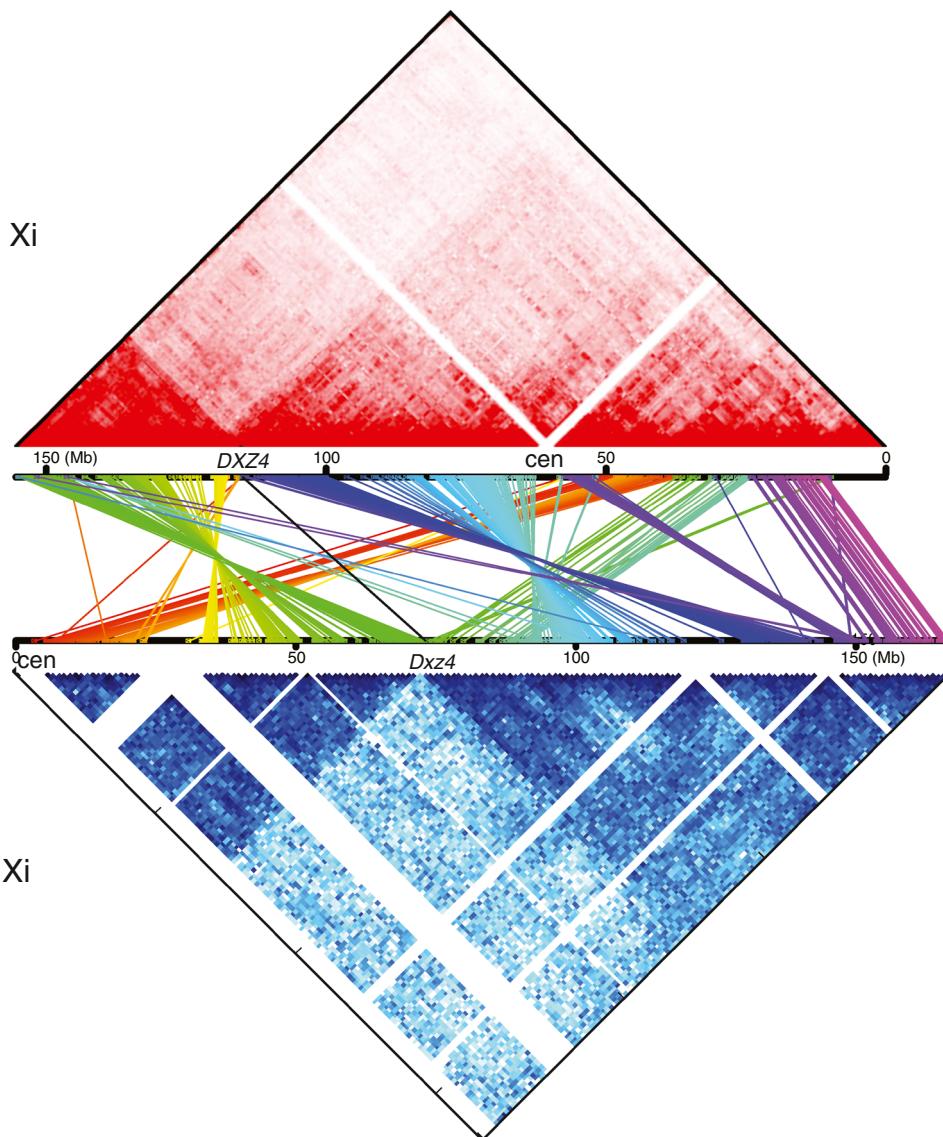
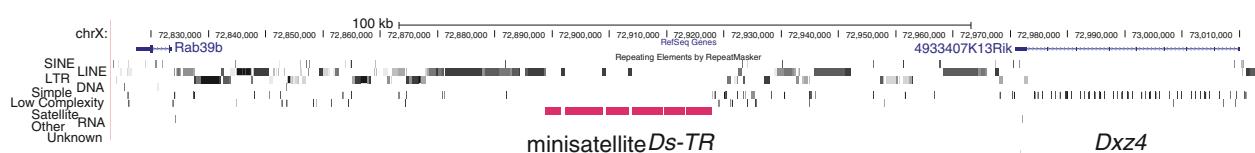
**B**

Fig. 2 Comparison between superdomains on the Xi in human and mouse. **a** Topological domains on the Xi compared between human and mouse. An allelic contact map of the human Xi generated based on published Hi-C data obtained in human lymphoblastoid cell line GM12878 [6] is shown on top (red) and aligned to a contact map of the mouse Xi based on our data obtained in mouse F1 brain by DNase Hi-C and in situ DNase Hi-C at bottom (blue). The Xi maps are compared based on the position of homologous genes between human (hg19) and mouse (mm9) (see "Materials and methods"). The human and mouse X chromosomes were oriented such that the position of *DXZ4*/*Dxz4* and the adjacent *PLS3*/*Pls3* genes are in the same orientation at the right extremity of the hinge region. Each pair of homologous genes is connected by a colored line between the contact maps and genes within blocks of conserved regions are indicated in a similar color. Several inversions and transpositions are apparent and the content of the superdomains is only partially conserved. **b** Gene content of the hinge region in mouse (~72.8–72.9 Mb). *Dzx4* is located at one extremity of the hinge region, which also contains the minisatellite *Ds-TR*. The size and location of the hinge region were estimated as described in the text (see also Additional file 5)

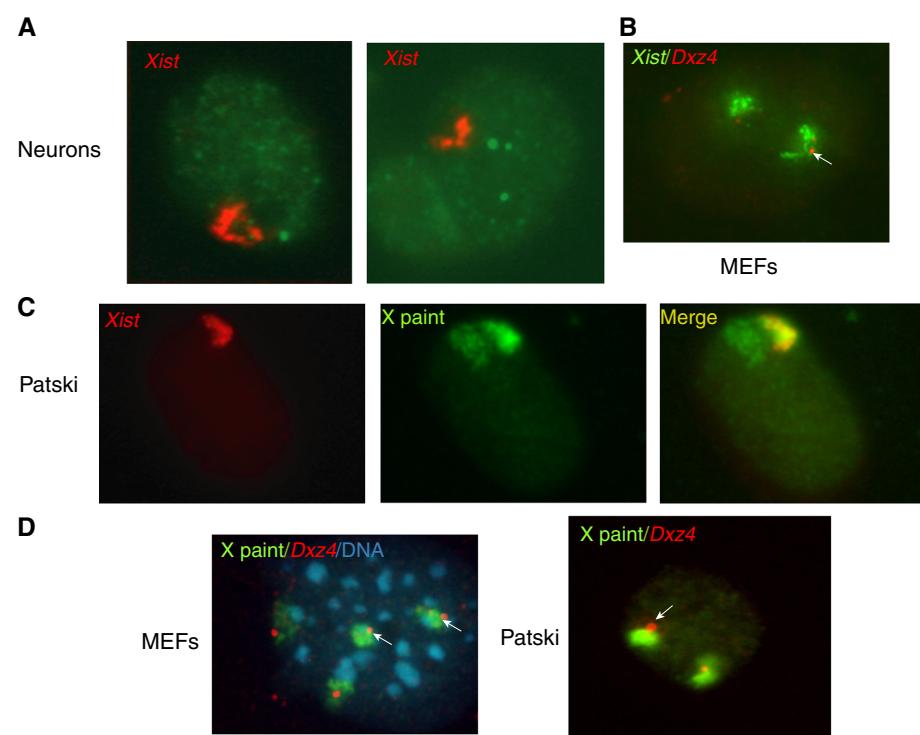


Fig. 3 FISH analysis of the mouse Xi. **a** Examples of nuclei after RNA-FISH for *Xist* (red) in neuronal cells show a bipartite structure for the Xi, consistent with the 3D structure detected by DNase Hi-C. **b** DNA-FISH for *Dzx4* (red) following RNA-FISH for *Xist* (green) in MEFs shows that *Dzx4* is located between the two regions coated by *Xist* RNA, consistent with *Dzx4* location at the hinge region. Note that there are two inactive Xs marked by *Xist* RNA clouds in this MEF line. **c** While RNA-FISH for *Xist* (red) in Patski cells shows the bipartite structure, DNA-FISH using a mouse X chromosome paint probe (green) shows only a condensed structure for the Xi. **d** DNA-FISH using both a mouse X paint (green) and *Dzx4* (red) shows that *Dzx4* is located at the edge of the condensed Xi in MEFs and Patski cells. MEF nucleus is stained with DAPI (blue). The arrows mark *Dzx4* on the Xi

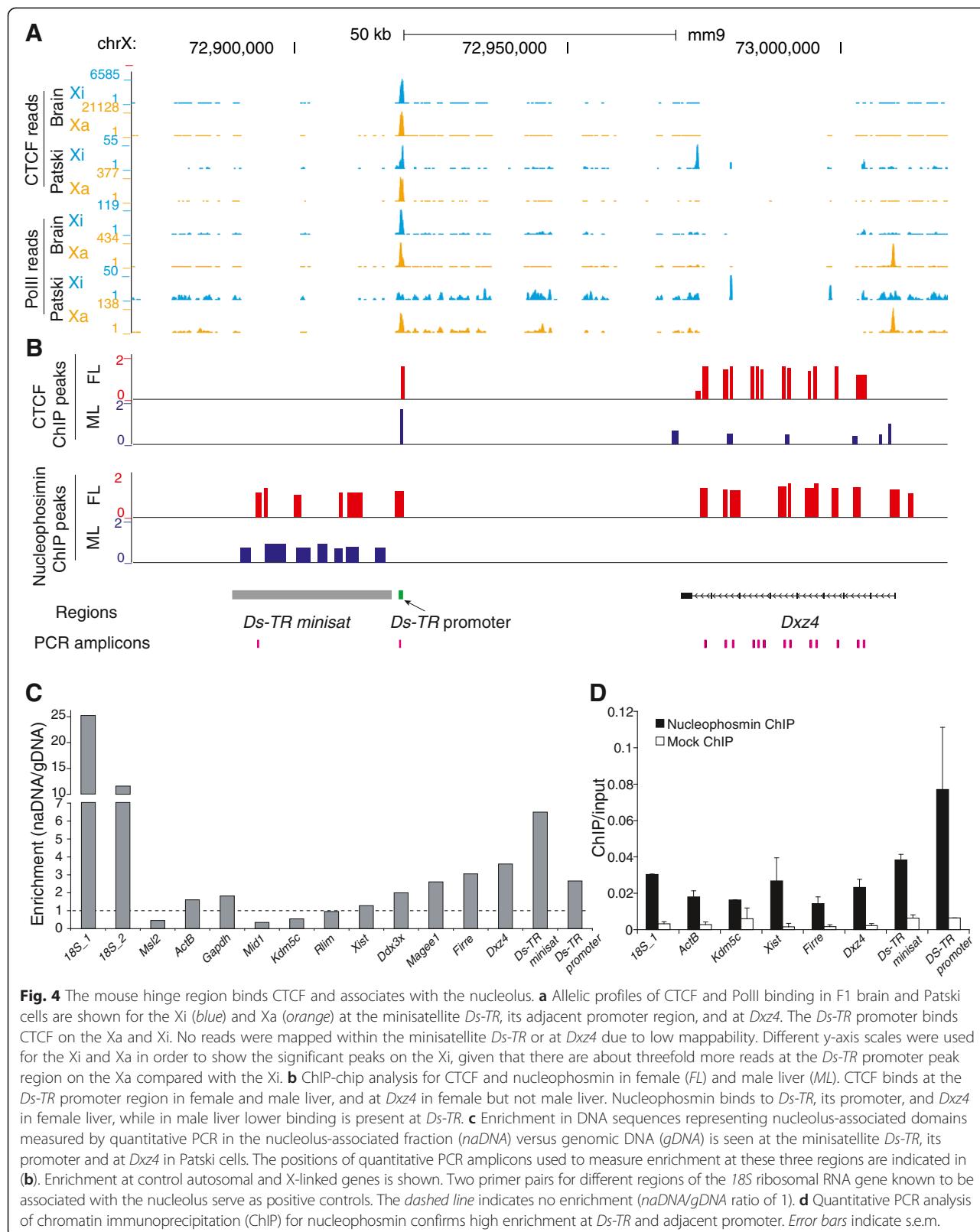
enrichment was seen only at *Ds-TR*, suggesting Xi-specific nucleophosmin binding at specific loci (Fig. 4b). ChIP-qPCR confirmed enrichment in nucleophosmin at these loci especially at the *Ds-TR* promoter (Fig. 4d). Taken together, our results indicate that the hinge region between the two superdomains on the Xi represents an NAD in mouse. Interestingly, the human *DXZ4* locus has also been reported to represent a NAD in HeLa cells [34].

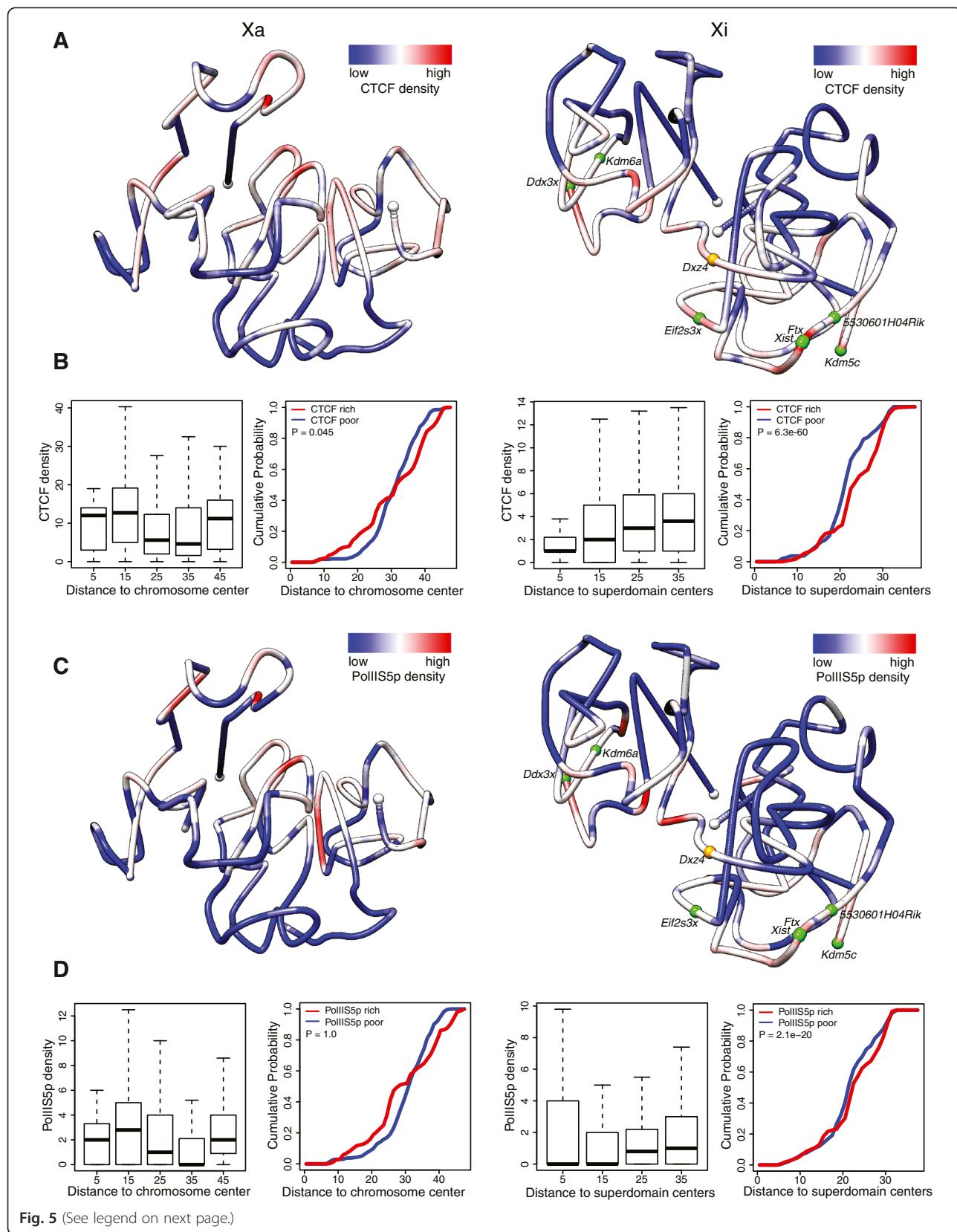
Distribution of genes, PolII occupancy, CTCF binding and L1 density in relation to the 3D structure of the Xi

Allelic distributions of CTCF and PolII datasets [15] were combined with the 3D models of the X chromosomes to visualize the position of regions enriched in CTCF and in active transcription (Fig. 5). Visual inspection of the 3D models of the X chromosomes indicates that CTCF and PolII tend to bind to regions on the outside of the 3D structure of the Xi but not the Xa (Fig. 5a, c). The density of CTCF binding or PolII occupancy in 1 Mb bins along the Xi was positively correlated with the bin distance to the center of each superdomain, confirming significant enrichment in CTCF binding and in

active transcription at the periphery of the Xi, but not the Xa (Fig. 5b, d). In contrast, the reverse pattern was observed for regions enriched in L1 elements, which were preferentially located on the inside of the 3D structure (Fig. 6a, b). Note that the density of CTCF binding appears greater on one side of the surface of the Xi 3D structure, possibly representing attachment to the nucleolus surface or to the nuclear membrane (Fig. 5a); however, further studies will be needed to confirm this arrangement. Interestingly, the non-random distribution pattern of regions enriched with CTCF, PolII or L1 elements on the Xi 3D model is more evident in F1 brain compared with Patski cells (data not shown), supporting a more constrained organization of the Xi in brain.

We next determined the position within the 3D structure of the Xi of a subset of seven genes that were previously shown to consistently escape XCI in F1 brain and other tissues as well as Patski cells [15, 26] (Additional file 7). These escape genes were found to be located in the outer layer of the 3D structure of the Xi in F1 brain ($p = 0.004$, Z-test; Fig. 5a, c). We then compared intrachromosomal contacts at autosomal genes (22,874), X-linked genes (975), and the seven escape genes in F1





(See figure on previous page.)

Fig. 5 Distribution of CTCF and PolIII binding on 3D models of the Xa and Xi. **a** 3D models of the Xa (left) and Xi (right) at 1 Mb resolution in mouse F1 brain colored to display the density of allelic CTCF binding (red indicates more binding). CTCF binding tends to be denser at the periphery of the Xi 3D structure, possibly on one face of the model. White dots indicate chromosome ends, orange dot *Dxz4*, green dots escape genes. **b** Box plots for the Xa (left) and the Xi (right) showing allele-specific CTCF-peak density at 1 Mb resolution grouped by the corresponding distances of the 1 Mb regions to the Xa chromosomal center or to the Xi superdomain centers, and empirical cumulative curves of 1 Mb regions binned based on their distance to the Xa chromosomal center or to the Xi superdomain centers for the CTCF-rich (red line, top 25 % CTCF-binding regions) and CTCF-poor regions (blue line, bottom 25 %). The empirical cumulative density as a function of the distance to the chromosome or superdomain centers for the feature-rich and feature-poor regions were compared using one-side Wilcoxon rank-sum test. **c, d** Same analysis for allelic PolIII occupancy. Like CTCF, PolIII occupancy tends to be higher at the periphery of the Xi 3D structure

brain in terms of the ratio of contacts on the maternal chromosome (BL6) to those on the paternal chromosome (*spretus*) (Fig. 7a). Most autosomal genes had a similar number of intrachromosomal contacts on each allele, although there was a slight shift towards maternal contacts, probably due to biased mapping of reads to the reference genome (BL6). In contrast to autosomal genes, X-linked genes, the majority of which are subject to XCI, had more contacts on the Xa than the Xi allele, indicating fewer

contacts on the silent copy (Fig. 7a). Genes that escape XCI and are expressed from both alleles showed contact ratios intermediate between autosomal and X-linked genes ($p = 0.0046$, Kolmogorov-Smirnov test; Fig. 7b). Our results show that escape genes have a greater number of specific contacts than inactivated genes on the Xi. Based on a previous 4C study in which contacts among a subset of escape genes were reported in mouse neuronal cells, we also considered potential interactions between specific

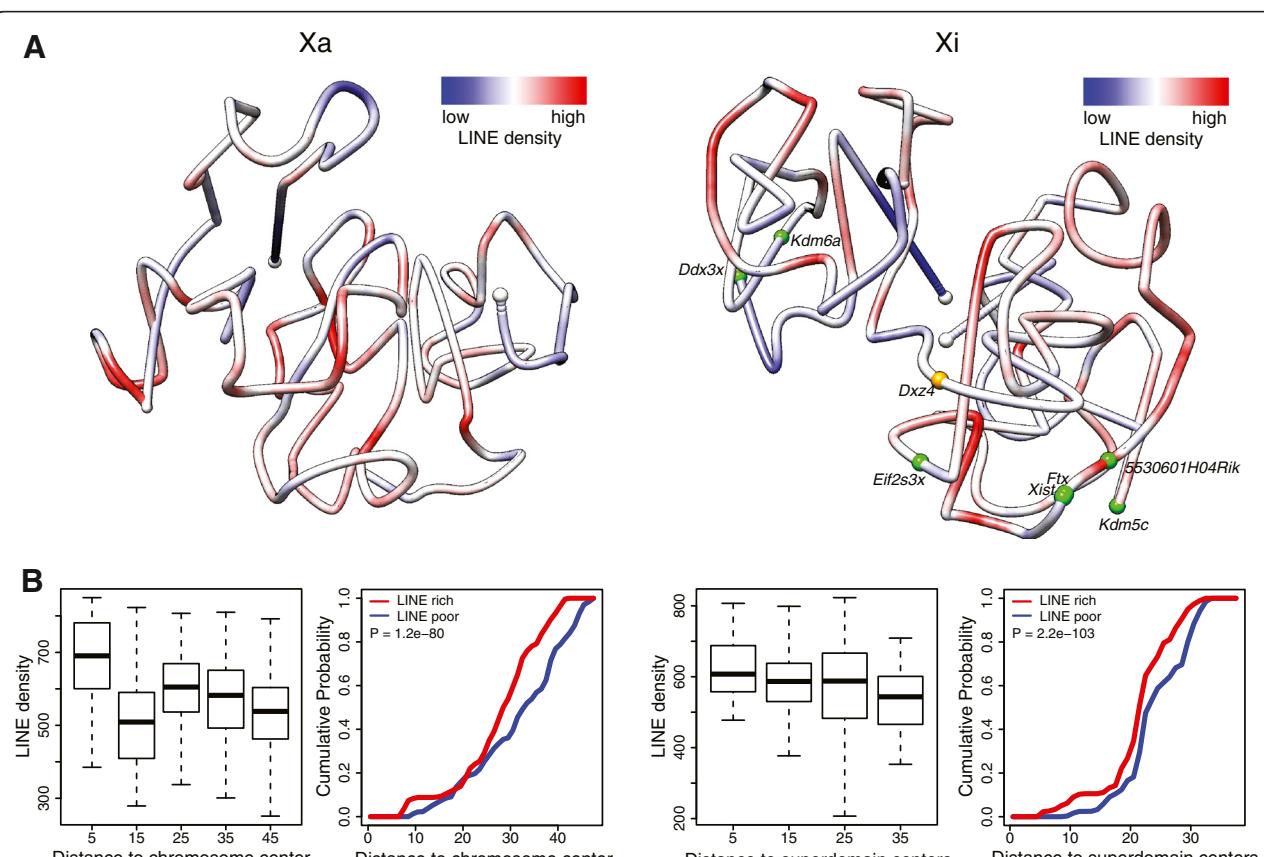


Fig. 6 Distribution of L1 elements on 3D models of the Xa and Xi. **a** 3D models of the Xa and Xi at 1 Mb resolution in mouse F1 brain colored to display the density of L1 elements (red indicates more L1 elements). L1 elements tend to be located at the inside of the Xi 3D structure. **b** Box plots for the Xa (left) and the Xi (right) showing L1 density at 1 Mb resolution grouped by the corresponding distances of the 1 Mb regions to the Xa chromosomal center or to the Xi superdomain centers for the L1-rich (red line, top 25 %) and L1-poor regions (blue line, bottom 25 %). The empirical cumulative density as a function of the distance to the chromosome or superdomain centers for the feature-rich and feature-poor regions were compared using one-side Wilcoxon rank-sum test

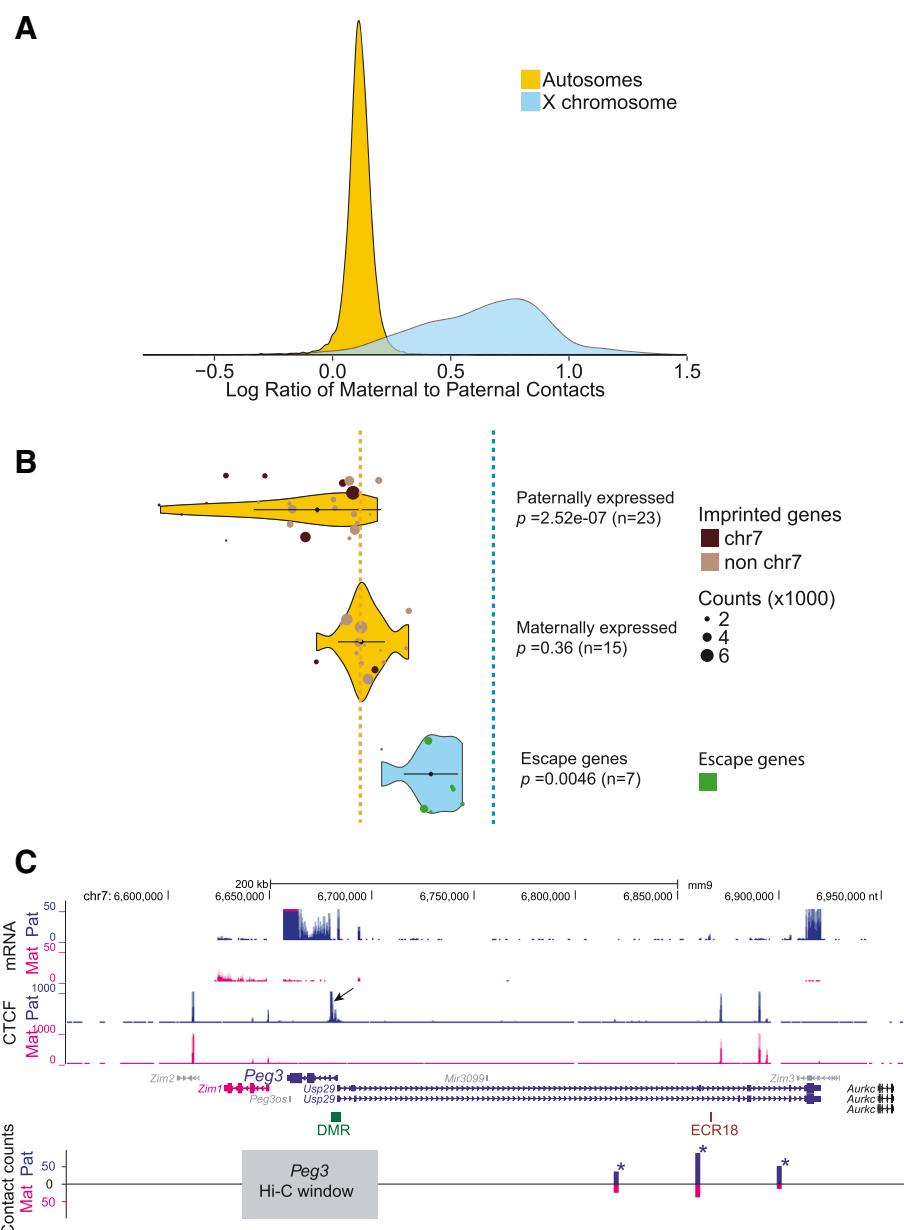


Fig. 7 Intrachromosomal contacts at X-linked genes and at imprinted genes. **a** Distribution of maternal-to-paternal allelic contacts at autosomal genes and X-linked genes determined by DNase Hi-C at 40 kb resolution in mouse F1 brain in which the paternal autosomes and the Xi are from *spretus*. Compared with autosomal genes, X-linked genes show high maternal-to-paternal ratios, indicating less frequent contacts at silent genes on the Xi. **b** Violin plots show the distribution of maternal-to-paternal allelic contacts at maternally and paternally imprinted genes and at genes that escape XCI at 40 kb resolution in F1 brain. Compared with other autosomal genes paternally expressed imprinted genes have a lower maternal-to-paternal contact ratio as shown by a long tail. These genes are preferentially located on chromosome 7 and when they are removed from the analysis, the shape of the distribution changes (as shown by a shorter tail in Additional file 8). The chromosomal location of imprinted genes is indicated by dots color-coded to indicate the chromosome of origin. The distribution of maternal-to-paternal allelic contacts for genes that escape XCI differs from the rest of X-linked genes, reflecting a higher number of contacts at expressed alleles. Dotted lines indicate median ratios of maternal-to-paternal contacts at autosomal and X-linked genes. **c** Significant contacts are detected between the imprinted paternally expressed gene *Peg3* and neighboring regions on the paternal allele. Allelic RNA-seq confirms *Peg3* expression on the paternal allele. Allelic CTCF profiles show binding to the differentially methylated region (DMR) adjacent to *Peg3* promoter region only on the paternal allele (arrow), presumably facilitating the formation of contacts between the *Peg3* promoter region and the distant enhancer ECR18 (evolutionarily conserved region 18) [63]. The needle plot of contacts counts between a 40 kb window that overlaps *Peg3* (grey bar) and nearby regions shows more interactions on the paternal (blue, Pat) than the maternal allele (pink, Mat). Genes with maternal or paternal expression are colored in pink or blue, non-imprinted genes in black, and non-expressed genes in grey. Contact regions showing significant allelic biases are marked by asterisks.

escape genes [5]. However, we did not detect such specific interactions, probably due to the limited allelic read coverage and resolution in our Hi-C analysis.

Differential interactions at paternally or maternally imprinted regions

Maternal and paternal alleles of imprinted genes are differentially expressed and thus expected to have differential structure in terms of intrachromosomal contacts [3]. We examined a total of 38 genes imprinted in mouse brain, representing 15 genes expressed on the maternal allele and 23 genes expressed on the paternal allele (Additional file 7). The list of genes considered here was based on a previous study [36] and confirmed by examining our own allelic RNA-seq data in F1 brain (data not shown). Measurements of intrachromosomal contacts in F1 brain in which the maternal allele is from BL6 and the paternal allele from *spretus* showed a higher contact frequency on the expressed allele (Fig. 7b). Examples of significant cis contacts on the expressed allele are shown using needle plots for a 40 kb Hi-C window at the paternally expressed gene *Peg3* and at the maternally expressed gene *Kcnk9* (Fig. 7c; Additional file 8: Figure S5a). When considering all imprinted genes, we confirmed that intrachromosomal interactions were more frequent at the active allele based on violin plots (paternally expressed $p = 2.5e-7$, maternally expressed $p = 0.36$, Kolmogorov-Smirnov test; Fig. 7b). Genes expressed from the paternal allele have an especially high number of contacts, and hence show a long tail toward low maternal-to-paternal contact ratios. A tail toward high maternal-to-paternal contact ratios is not observed for maternally expressed genes, which show a tighter distribution of ratios. Interestingly, almost all genes with high contact number (1000 contacts or more) on the active paternal allele are located on mouse chromosome 7 (Fig. 7b). This is not surprising because chromosome 7 has 9 out of 23 paternally expressed and 2 out of 15 maternally expressed imprinted genes. The distribution of the maternal-to-paternal ratios for all genes on each autosome was similar (data not shown), indicating that the effect seen for imprinted genes on chromosome 7 is unique to paternally expressed genes on this chromosome (Fig. 7b). Accordingly, when removing genes located on chromosome 7 the tail of the distribution for paternally expressed genes is shorter, confirming that paternally expressed genes on chromosome 7 contribute to a high contact frequency (Additional file 8: Figure S5b).

Discussion

Using DNase Hi-C [23] and a novel in situ DNase Hi-C we discovered that the mouse Xi condenses in two three-dimensionally defined superdomains. We applied our new Hi-C methods to an *in vivo* system (mouse

brain) demonstrating the feasibility and reproducibility of this method for the determination of the structure of individual chromosomes in tissues. Our allele-specific approach provides a comprehensive contact map of paternal and maternal homologous chromosomes and will help better understand tissue-specific and homolog-specific differences in nuclear organization.

Comparisons between published human data [6] and our mouse data revealed surprising differences between superdomains identified in these species. Large differences in the sequence content and organization of the superdomains between species implies that the folding of the Xi may only be partially determined by its sequence. However, the hinge region is partially conserved and located near the *Dxz4/DXZ4* macrosatellite locus in both species, which suggests that this locus has a conserved role in terms of organization of the 3D Xi structure. The *Dxz4/DXZ4* loci transcribe lncRNAs and bind CTCF on the Xi [30–32], which may facilitate the formation of the two superdomains. lncRNAs have been proposed as key elements of nuclear organization [37]. Expression of *Dxz4* (4933407K13Rik) in brain and Patski cells was very low (<1 RPKM (reads per kb of exon per million mapped reads)), possibly due to failure to detect small transcripts and/or low mappability of the repeat. Thus, whether the *Dxz4* lncRNA plays a role in the formation of the hinge region is still unclear. Substantial structural differences have been found in mouse versus the primate *Dxz4/DXZ4* [31, 38]. The most significant difference is that in human and other primates *DXZ4* is composed of as many as 100 copies of a 3 kb GC-rich repeat, while in mouse *Dxz4* contains about seven repeats ranging in size from 3.8 kb to 5.7 kb that are not particularly GC-rich. Interestingly, we found that the mouse hinge region also contains a novel gamma minisatellite *Ds-TR* not found elsewhere in the mouse genome. *Ds-TR* consists of a palindrome repeat spanning ~30 kb, located ~50 kb downstream of *Dxz4* and absent on the human or rat X chromosomes. Although no RNA-seq reads were observed at *Ds-TR*, probably due to low mappability and/or failure to detect small transcripts, *Ds-TR* is apparently expressed from the Xi as evident by our findings of allelic PolII occupancy at least in F1 brain, in agreement with a previous study that compared transcription in female and male cells [31]. The role of gamma satellites is poorly understood but one such repeat has been shown to prevent the spread of heterochromatin at pericentromeric regions, suggesting that *Ds-TR* could help form a boundary between the two superdomains on the Xi [39]. We speculate that *Dxz4* and *Ds-TR* may function together as chromatin boundaries between the Xi superdomains. However, the molecular mechanisms of formation of a bipartite structure remain to be analyzed to determine whether

expression of *Dxz4* and *Ds-TR*, and/or CTCF binding, and/or nucleolus association are independent or related factors in the formation of superdomains on the Xi. Recent studies have reported that the lncRNA *Xist* recruits structural proteins [10–12]. Interestingly, we found that *Xist* RNA-FISH could reveal the Xi bipartite structure, which was not clearly seen by DNA-FISH using an X-paint, suggesting that *Xist* may facilitate contacts within the Xi superdomains.

One important chromosomal organizer protein is the zinc finger protein CTCF often found at the transition between TADs [6, 21, 22]. CTCF-binding at domain boundaries helps anchor chromatin loops and frequently link promoters and enhancers [6]. However, CTCF binding at the hinge region of the Xi is clearly not sufficient to explain the formation of the two superdomains because CTCF binding can be found concentrated elsewhere on the Xi where no superdomains are detected. For example, we and others reported a cluster of CTCF binding specifically on the Xi at the *Firre* locus [32, 40] and CTCF also binds near *Xist* on the mouse Xi [41]. The human homologs of these loci (*XIST*, *FIRRE*) together with *DXZ4* and another lncRNA locus, *LOC550643*, were previously shown to contact each other and to function as anchor regions for superloops on the human Xi [6, 42]. Our datasets did not have sufficient resolution to detect long-range contacts between the corresponding loci in mouse. However, our previous DNA-FISH studies in mouse fibroblasts failed to show association between *Dxz4* and *Firre*, suggesting differences between species [32].

We found that regions enriched in CTCF binding on the Xi tended to be located at the periphery of the 3D structure, suggesting that these loci may serve as attachment sites. Indeed, the mouse and human Xi often occupy specific locations in the nucleus near the lamina or the nucleolus [17, 18]. Based on recent studies of lamina-associated domains (LADs) and NADs, it may be that the nucleolus and nuclear periphery act as “velcro” for heterochromatin including the Xi [43]. An early study proposed that the Barr body represents a looped structure formed by telomeric association with the nuclear membrane [44]. Due to low mappability, we unfortunately could not determine the structure of the telomeric ends of the Xi. However, we found that the *Dxz4/Ds-TR* region represents a NAD that binds the nucleolar protein nucleophosin. Thus, the hinge region between superdomains represents a large NAD whose tethering to the nucleolus may govern the formation of a bipartite structure. Whether the hinge region provides flexibility to the Xi 3D structure is unknown. Positioning of the Xi within the nucleus is important for maintenance of its heterochromatic structure. We recently reported that both *Dxz4* and *Firre* associate with the surface of the

nucleolus and that *Firre* helps maintain H3K27me3, a repressive histone modification that marks the Xi [32].

Our 3D analyses of the Xi show that genes that escape XCI are located at the periphery of the 3D structure, as previously reported [19]. However, we did not detect specific contacts between these genes, in contrast to a previous 4C study [5], which may be due to the lower resolution of our Hi-C data. Our observation of more short-range contacts on the Xi at genes that escape XCI versus genes subject to XCI may reflect random contacts between inactivated genes on the Xi, while specific interactions would take place at expressed genes. Similarly, a larger number of specific intrachromosomal contacts were also found on the Xa versus the Xi, resulting in more defined topological domains on the Xa, consistent with previous studies [4–6, 12]. These observations do not exclude inactivated genes having many intrachromosomal contacts on the Xi, if those contacts were variable from cell-to-cell and thus not detected by Hi-C done on bulk tissue. Indeed, high-resolution FISH combined with 3C analyses and single cell Hi-C analyses have shown cell-to-cell variability in TAD conformation [28, 45]. The superdomains on the mouse Xi appear less condensed in Patski cells than in F1 brain. Interestingly, our previous studies show a lower density of CTCF sites and more genes that escape XCI in Patski cells, suggesting a less compact Xi structure in these cells [15]. A recent study has also shown that deletion of CTCF sites causes disruption of TADs and spreading of euchromatin into heterochromatin [46]. Similarly, deletion of CTCF sites at the boundary of a Polycomb repressed domain results in transcriptional activation of genes in that domain [47].

Similar to our findings at escape genes, imprinted loci also show more contacts on the expressed allele, which probably reflects interactions between promoters and enhancers facilitated by allelic CTCF binding [3, 6, 7]. We discovered that the contacts are more frequent at paternally expressed genes (especially those on chromosome 7) than at maternally expressed genes. The cause of this bias is unclear and whether reported differences in the repeat (SINE) content and/or DNA sequence (GC content) of paternally and maternally expressed imprinted genes play a role is unknown [48, 49].

Conclusions

Our 3D structure analysis of the mouse Xi reveals a bipartite structure. Two superdomains of frequent long-range contacts are separated by a hinge region that is partially conserved in human. The hinge region, which contains *Dxz4* and the minisatellite *Ds-TR*, represents a nucleolus-associated domain that may help target the Xi to the nucleolus. Together with CTCF and PolII binding, expressed escape genes tend to be located at the periphery of the Xi. In addition, analyses of genes that escape

X inactivation and of imprinted genes indicate that expressed genes/alleles have more specific contacts compared with silenced genes/alleles.

Materials and methods

Tissues and cell lines

The Patski fibroblast line in which the Xi is from BL6 and the Xa from *M. spretus* was originally derived from embryonic kidney [27]. The presence of normal X chromosomes was verified by karyotyping. Whole brain was collected from female F1 adult mice obtained by mating *spretus* males (Jackson Labs) with females that carry an *Xist* mutation (B6.Cg-Xist<tm5Sado>) [50], in which there is complete skewing of inactivation of the *spretus* X. Liver specimens were collected from male and female BL6 adult mice [32]. Female MEFs [26] were cultured in standard complete medium. Primary neuron cultures were established on poly-lysine-coated coverslips from the hippocampus dissected from 0–2-day-old BL6 mouse pups.

RNA-FISH, DNA-FISH, and immunostaining

RNA-FISH using a 10 kb *Xist* cDNA plasmid (pXho, which contains most of exon 1 of *Xist*) [51], and DNA-FISH for *Dzx4* (BAC clone RP23-299L1 from BACPAC) were done as described [32]. A whole mouse X chromosome painting probe (XMP X green from MetaSystems) was used for DNA-FISH together with *Dzx4* followed by *Xist* RNA-FISH using a standard protocol.

ChIP-chip

ChIP-chip using an antibody for nucleophosmin (Abcam) was done as described [32]. NimbleScan software (Nimblegen Roche) was used to search for significant enrichment regions using a 500 bp sliding window. Enriched regions with a false discovery rate score less than 0.05 were considered as significant binding peaks.

DNase Hi-C and *in situ* DNase Hi-C

DNase Hi-C was done on mouse F1 brain and Patski cells using a previously published method [23]. *In situ* DNase Hi-C is described below.

Preparation of crosslinked cells

The whole brain from one F1 hybrid mouse was isolated and homogenized in 1× phosphate-buffered saline (PBS) with protease inhibitors followed by crosslinking with 1.5 % formaldehyde as described previously [25]. For Patski cells, one million cells were crosslinked in T-75 flasks with 1 % formaldehyde for 10 min followed by quenching with 125 mM glycine. Cells were scraped, washed in 1× PBS (Gibco), pelleted, and snap frozen in liquid nitrogen.

Chromatin digestion

Cell pellets containing approximately one million cross-linked cells were resuspended in cold lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2 % NP-40) and incubated on ice for 10 min. Nuclei were pelleted at 2500g for 60 s, resuspended in 100 μL of 0.5× DNase I digestion buffer [0.5× DNase I digestion buffer (Thermo), 0.5 mM MnCl₂] containing 0.2 % SDS, and incubated at 37 °C for 30 min. An equal volume of 0.5× DNase I digestion buffer containing 2 % Triton X-100 and 4 U RNase A (Thermo) was added and incubation at 37 °C was continued for 10 min. Then, 1.5 U DNase I (Thermo) was added and digestion carried out at room temperature for 4 min. DNase I digestion was stopped by adding 40 μL of 6× Stop Solution (125 mM EDTA, 2.5 % SDS), followed by centrifugation at 2500g for 60 s. Nuclei were resuspended in 150 μL nuclease-free H₂O (Ambion), and purified with two volumes (300 μL) of AMPure XP SPRI magnetic beads (Beckman Coulter). The resulting mixture was well mixed, incubated at room temperature for 5 min, collected via DynaMag-Spin magnet (Invitrogen), washed twice with 80 % ethanol, and air dried for 2 min.

Chromatin end-repair and dA-tailing

The purified bead-nuclei pellet was resuspended in 200 μL 1× T4 DNA Ligase Buffer (New England Biolabs) containing 0.25 mM dNTPs, 0.075 U/μL T4 DNA polymerase (Thermo) and 0.15 U/μL Klenow Fragment (Thermo), and incubated at room temperature for 1 h. The end-repair reaction was stopped by adding 5 μL of 10 % SDS. The bead-nuclei mixture was pelleted at 2500 g for 60 s, resuspended in 200 μL 1× NEB buffer 2 (New England Biolabs) containing 0.5mM dATP, 1 % Triton X-100 and 0.375 U/μL Klenow (exo-) (Thermo), and incubated at 37 °C for 1 h. dA-tailing reaction was stopped by adding 5 μL of 10 % SDS.

Bridge adaptor ligation

The bead-nuclei mixture was again pelleted at 2500g for 60 s, and resuspended in 30 μL H₂O, 20 μL biotinylated bridge-adaptor (see Ma et al [23] for sequences and adaptor preparation), 20 μL blunt bridge-adaptor, 10 μL 10× T4 DNA Ligase Buffer with ATP, 10 μL polyethylene glycol (PEG)-4000 (Thermo), 5 μL 10 % Triton-X100, and 5 μL T4 DNA ligase (5 U/μL; Thermo). This mixture was incubated at 16 °C overnight to ligate T-tailed biotinylated bridge adaptors to the termini of A-tailed, digested chromatin. Following incubation, the reaction was stopped by adding 5 μL 10 % SDS. The bead-nuclei mixture was then pelleted at 2500g for 60 s and resuspended in 300 μL H₂O. To remove excess unligated adapter, 250 μL 20 % PEG in 2.5 M NaCl was added to the mixture, which was incubated at room temperature for 5 min, collected via DynaMag, and washed once with

80 % ethanol. Beads were then resuspended in 200 μ L H₂O and purified further using 0.8 volumes of 20 % PEG in 2.5 M NaCl as above, to further remove unligated adaptors.

Adaptor phosphorylation and proximity ligation

The air-dried bead-nuclei mixture was resuspended in 100 μ L 1× T4 DNA Ligase Buffer with ATP containing 1 U/ μ L T4 Polynucleotide Kinase (PNK) (Thermo), and incubated at 37 °C for 1 h to phosphorylate ligated bridge adaptors. Following incubation, 90 μ L 10× T4 DNA Ligase Buffer with ATP, 6 μ L T4 DNA ligase (5 U/ μ L; Thermo), and 804 μ L of H₂O were added to the reaction mix. In situ proximity ligation was then carried out at room temperature for 4 h.

Reversal of crosslinking and purification of DNA

Following proximity ligation, bead-nuclei complexes were pelleted at 2500g for 60 s. Pellets were resuspended in 400 μ L 1× NEBuffer #2, 40 μ L 10 % SDS, and 40 μ L Proteinase K (20 mg/ml; Thermo). This mixture was incubated overnight at 60 °C to reverse crosslinks and liberate ligated DNA. After incubation, DNA was precipitated by adding 3 μ L GlycoBlue (Ambion), 50 μ L 3 M sodium acetate pH 5.2, and 550 μ L isopropanol and incubating the mixture at -80 °C for 2 h prior to centrifugation for 30 min at 15,000 rpm at 4 °C. The resulting bead-DNA pellet was resuspended in 100 μ L H₂O, then purified further using 100 μ L AMPure XP beads, which were collected and washed as above. DNA was eluted using 100 μ L H₂O. Typical yields for experiments were 3–5 μ g DNA per one million cells.

Sequencing library preparation

DNA (1.5–2.5 μ g) was used for sequencing library preparation. End-repair was carried out by mixing 1.5–2.5 μ g DNA in 170 μ L H₂O with 20 μ L 10× End-repair reaction buffer (Thermo) and 10 μ L Fast DNA End Repair Enzyme Mix (Thermo), and incubating the resulting mixture at 18 °C for 10 min. DNA was then purified using one volume (200 μ L) AMPure XP beads, which were incubated, washed, and air-dried as above, then resuspended (including beads) in 50 μ L 1× NEBuffer #2 containing 0.6 mM dATP and 12.5 U Klenow (exo-). This bead-enzyme mixture was then incubated at 37 °C for 30 min, after which 5 μ L 10 % SDS was added to stop the dA-tailing reaction. The dA-tailed DNA-beads mixture was purified further by adding 1.6 volumes 20 % PEG in 2.5 M NaCl to the reaction. This mixture was incubated for 5 min, precipitated via DynaMag, washed twice with 80 % ethanol, air-dried, and resuspended in 50 μ L 1× Rapid Ligation Buffer (Thermo) containing 5 μ L 10× TruSeq Adapter (Illumina) and 20 U T4 DNA ligase. This mixture was incubated at room temperature for 1 h

or at 16 °C overnight to ligate sequencing adaptors, followed by quenching with 5 μ L 10 % SDS. The ligation mixture was then brought to 200 μ L with H₂O and purified by adding 1 volume (200 μ L) 20 % PEG in 2.5 M NaCl, immobilizing, washing, and air-drying beads as above. After air-drying, beads were resuspended in 200 μ L H₂O and purified further using 0.8 volumes of 20 % PEG in 2.5 M NaCl as above, to further remove unligated sequencing adaptors. DNA was eluted off of air-dried beads in 100 μ L H₂O, then pulled down with 30 μ L MyOne C1 beads (Life Technologies) that had been washed and resuspended in 100 μ L 2× Bind and Wash buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 2 M NaCl). Streptavidin pull-down was carried out for 20 min at room temperature with rotation. Immobilized DNA was precipitated via DynaMag, washed once with 600 μ L 0.5× Bind and Wash buffer mixed with 0.5× TE lysis buffer (25 mM Tris-HCl, 0.5 mM EDTA, 0.5 % SDS), once with 600 μ L 1× Bind and Wash buffer, once with 600 μ L 1× NEBuffer #2, once with 600 μ L Buffer EB (10 mM Tris-HCl pH 8.5), and resuspended in 20 μ L Buffer EB. Libraries were then amplified for sequencing using 2× Robust Master Mix (KAPA), 10× PCR Primer Cocktail (Illumina) and half the volume of resuspended streptavidin beads, for 12 cycles, purified using 0.8× volumes of AMPure XP beads, then sequenced. Sequencing was carried out using Illumina HiSeq 2000 and NextSeq 500 instruments to generate paired-end 80 bp or paired-end 101 bp reads.

NAD analysis

Nucleoli were isolated from fixed Patski cells using 1 % formaldehyde using a modified method [34, 35]. In brief, two to three million cells were fixed for 10 min at room temperature and quenched using 0.125 M glycine. The fixed cells were resuspended in 1 ml of high magnesium buffer (10 mM HEPES, 0.35 M sucrose, 12 mM MgCl₂ plus protease inhibitors) and sonicated for six rounds of 10-s bursts (full power) using a Misonix Sonicator3000. The dirty nucleoli preparation was centrifuged for 30 s at 15,000g and resuspended in 0.5 ml low magnesium buffer (10 mM HEPES, 0.88 M sucrose, 1 mM MgCl₂ plus protease inhibitors), which was sonicated one more time with a 10-s burst (full power) and centrifuged again. The nucleoli pellet was used for DNA extraction and qPCR. Release of nucleoli was monitored by microscopy after immunostaining of the preparation with nucleophosim antibody (Abcam).

Quantitative PCR

qPCR was performed using a SYBR green system as described before [32]. The primers used are listed in Additional file 9.

Computational analyses

Mapping and filtering of sequence reads

We sequenced the DNase Hi-C libraries using paired-end reads 150 bp in length and the *in situ* DNase Hi-C libraries using paired-end reads 80 bp in length. We performed an exhaustive search and cleaning of the Illumina primer and adaptor sequences in the full-length reads and extracted the remaining read fragments of various lengths from 25 to 80 bp using an in-house script, as described in [23]. We then mapped each end of these cleaned paired-end reads separately to the BL6 genome using the NCBI build v37/mm9 reference genome assembly obtained from the UCSC Genome Browser [52] and the *pseudo-spretus* genome using BWA/v.0.5.9 [53]. The *pseudo-spretus* genome was assembled by substituting available SNPs (from Sanger Institute, SNP database Nov/2011 version) into the BL6 reference genome, as described in [15]. We retained only the reads that mapped uniquely, allowing at most three mismatches and requiring a mapping score MAPQ ≥ 30 to either the BL6 genome or the *pseudo-spretus* genome, for further analyses.

Allele-specific contact maps

Using heterozygous SNPs between the BL6 genome and the *pseudo-spretus* genome, we segregated all high-quality uniquely mapped reads (MAPQ ≥ 30) into three categories: (1) BL6-SNP reads containing only BL6-specific SNP(s); (2) *spretus*-SNP reads containing only *spretus*-specific SNP(s); (3) reads that do not contain valid SNPs. We refer to both BL6-SNP reads and *spretus*-SNP reads as “allele-specific reads”, and reads that do not contain valid SNPs as “allele-uncertain reads”. Furthermore, to eliminate the bias due to the PCR duplication step, we removed redundant paired-end reads. We define two reads as redundant if both ends of the reads are mapped to identical locations in the same genome assembly.

After PCR duplicate removal, we generated allele-specific whole-genome contact maps at 1 Mb, 100 kb and 40 kb resolutions. To do so, we partitioned the genome into non-overlapping bins and counted the number of allele-specific contacts (i.e., uniquely mapped paired-end reads) observed between each pair of bins. The dimension of the resulting contact map is the total number of bins in the genome, and entry (i, j) is the contact count between bins i and j . Specifically, in the allele-specific contact map of the X_a , C_{i_1, j_1} denotes the contact counts between bins i and j on the X_a . Whereas in the allele-specific contact map of the X_i , C_{i_0, j_0} denotes the contact counts between bins i and j on the X_i .

Inference of allele-uncertain reads

Using a similar approach to previous methods [54, 55], we model the contact frequencies between genomic loci pair

as a binomial distribution $X_{i,j} \sim \text{Binomial}(M, p_{i,j})$, where M is the total number of observed contacts (high-quality uniquely mapped and non-redundant paired-end reads) in a given (*in situ*) DNase Hi-C experiment. Since M is large and $p_{i,j}$ is very small, we approximate the binomial distribution by a Poisson distribution $X_{i,j} \sim \text{Poisson}(\lambda_{i,j})$, where $\lambda_{i,j} = Mp_{i,j}$. Adapting to the diploid genome, we assume the observed allele-specific chromatin contact counts follow the Poisson model:

$$X_{i_\odot, j_\otimes} \sim \text{Poisson}(\lambda_{i_\odot, j_\otimes}),$$

where $i_\odot \in \{i_0, i_1\}$ and $j_\otimes \in \{j_0, j_1\}$, $\lambda_{i_\odot, j_\otimes}$ is the expected allele-specific contact counts between loci pair i_\odot and j_\otimes . Furthermore, we assume the Poisson parameter $\lambda_{i_\odot, j_\otimes}$ follows a gamma prior distribution:

$$\lambda_{i_\odot, j_\otimes} \sim \text{Gamma}(\alpha_{G(i_\odot, j_\otimes)}, \beta_{G(i_\odot, j_\otimes)}).$$

The hyper-parameters α and β depend on $G(i_\odot, j_\otimes)$, which is the genomic group assignment of loci pair i_\odot and j_\otimes for accommodating the systematic differences of expected contacting frequencies between intrachromosomal contacts and interchromosomal contacts.

Based on the observations that the intrachromosomal contact frequency decreases as the genomic distance increases and interchromosomal contacts are rare, we model that the hyper-parameter α and β are shared across intrachromosomal contacts between similar genomic distance as well as interchromosomal contacts between two separate chromosomes. Thus, we have:

$$G(i_\odot, j_\otimes) = \begin{cases} g_{k_\odot, d(i_\odot, j_\otimes)} & \text{if } \text{chr}(i_\odot) = k_\odot = k_\otimes = \text{chr}(j_\otimes) \\ g_{k_\odot, k_\otimes} & \text{if } \text{chr}(i_\odot) = k_\odot \neq k_\otimes = \text{chr}(j_\otimes). \end{cases}$$

That is, for intrachromosomal contacts, all loci pairs i_\odot and j_\otimes on the same chromosome k_\odot and with the same genomic distance $d(i_\odot, j_\otimes)$ (binned at given resolution) share the same gamma prior hyper-parameters. On the other hand, for interchromosomal contacts, all loci pairs from the same pair of chromosomes k_\odot and k_\otimes share the same gamma prior hyper-parameters.

Posterior mean estimates of allele-specific contact frequencies $\lambda_{i_\odot, j_\otimes}$ and the hyper-parameters α and β are obtained via the expectation-maximization (EM) algorithm [56]: (1) we assign allele-uncertain reads to the allele-specific contact maps based on the estimates of allele-specific reads; (2) we estimate the hyper-parameters α

and β using the empirical Bayes approach and calculate the posterior mean estimates of $\lambda_{i,j}$; (3) we re-assign allele-uncertain reads based on the current estimation $\lambda_{i,j}$ and update the inferred allele-specific contact maps. We repeat steps 2–3 until convergence. For 1 Mb resolution analysis, we use contact maps containing only allele-specific reads, while for finer-resolution at 100 kb or 40 kb, we use the inferred allele-specific contact maps in our analyses.

Normalization

We normalized the allele-specific contact maps obtained from DNase Hi-C and in situ DNase Hi-C data using an iterative correction method [57]. Here we only used intrachromosomal contacts to normalize the allele-specific contacts. This is based on the observation that interchromosomal (including interhomologous) contacts are rare. We first preprocessed the allele-specific intrachromosomal contact maps at 1 Mb, 100 kb or 40 kb resolution by setting the entries that may be dominated by self-ligation products to 0. These entries are the diagonal, super-diagonal (+1 off-diagonal) and sub-diagonal (-1 off-diagonal) contact counts. In addition, we excluded bins with the lowest 2 % read coverage. Lastly, we applied the iterative correction procedure on each preprocessed intrachromosomal contact map separately to obtain a normalized contact map with near-equal row and column sums.

Topological domain calling

We identified topological domains using a previously described hidden Markov model-based software tool [21]. We applied the topological domain calling on normalized diploid contact maps at 40 kb resolution. As in previous work [21], we classified the regions between the topological domains either as “domain boundaries” (≤ 400 kb) or “unorganized chromatin” (> 400 kb).

Assigning statistical significance to normalized contact maps

To obtain a set of high-confidence contacts, we subjected the diploid contact maps at 40 kb resolution to a statistical confidence estimation procedure, fit-hi-c [54]. The procedure accounts for the effect of genomic distance on the intrachromosomal contact probability by fitting a smoothing spline. We then accounted for biases using the normalization procedure described above. Finally, we applied multiple hypothesis testing to compute q values, which are used to filter contacts at a desired false discovery rate at 0.05.

Assessing superdomain contact density on X_i

To measure density of the two superdomains on the X_i , we calculated the ratio of intra- versus inter-superdomain contact frequencies, called the bipartite index (BI), as follows:

$$\frac{\sum_{i=1}^h \sum_{j=1}^h C_{ij}}{h^2} + \frac{\sum_{i=h+1}^n \sum_{j=h+1}^n C_{ij}}{(n-h)^2}, \\ 2 \frac{\sum_{i=1}^h \sum_{j=h+1}^n C_{ij}}{h(n-h)},$$

where C_{ij} is the allele-specific contact counts for the X chromosome of interest, n is the total number of bins in the chromosomal contact map, and h is the index of superdomain boundary (that is, the hinge region). We calculated the BIs for X_i and X_a in both F1 brain and Patski datasets. A higher BI value represents more condensed packaging of the chromatins within the two superdomains. To measure the significance of the bipartite structure of the X_i , we randomly shifted the superdomain boundary to estimate the null distribution of the bipartite index. We then used the one-sided Z-test to calculate the p value for the observed BI at the hinge region (Table 1).

Comparison between human and mouse X chromosome contact profiles

To construct the synteny map between human and mouse X chromosomes, we used the UCSC liftOver utility [58] to convert the mouse/mm9 coordinates of all refSeq genes on chromosome X [59] to the human/hg19 coordinates. We only used mouse X-linked genes that have a homologous human X-linked counterpart to plot the synteny map in Fig. 2. We used the same tool to convert the human/hg19 coordinates of the 27 superloops reported on the human X chromosome [6] to mouse/mm9 coordinates.

Inference of the 3D structure of X chromosomes

We inferred the 3D structure of the X_a and X_i chromosomes, separately, using the Pastis software [60]. Each X chromosome is modeled as a series of beads on a string, spaced 1 Mb apart. We denote by $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}^3$ the coordinate matrix of the structure, where n denotes the total number of beads on the chromosome ($n = 167$ for the mouse X chromosome), and $x_i \in \mathbb{R}^3$ represents the 3D coordinates of the i -th bead.

The Pastis model assumes that the observed contact counts C_{ij} between beads i and j follows a Poisson distribution, where the Poisson parameter of C_{ij} is a decreasing function of $d_{ij}(X)$ of the form $\beta d_{ij}(X)^\alpha$, and $d_{ij}(X) = ||x_i - x_j||$ is the Euclidean distance between the beads i and j . Therefore, the problem of 3D structure

inference is formulated as the following optimization problem:

$$\max_{\beta, X} \mathcal{L}(X) = \sum_{1 \leq i < j \leq n} C_{i,j} \alpha \log d_{ij}(X) + C_{i,j} \log \beta - \beta d_{ij}(X)^{\alpha}.$$

Here we set $\alpha = -3$ and optimize the structure and β using IPOPT, an interior point filter algorithm [61].

Enrichment of escape genes at X chromosome periphery

To measure the 3D positional preference of escape genes with regards to the X chromosome periphery, we calculated the radial distances of escape genes to the chromosome center and superdomain centers as described below.

The center of the X chromosome is located at the origin, that is, $\sum_{i=1}^n x_i = (0, 0, 0)$. For each escape gene g , the distance of gene g to the chromosome center is $d_g = ||x_k||$, where k is the index of the bin that is closest to the middle point of the gene. In addition, given the observation that X_i forms a bipartite structure and the hinge region is located at locus h , we computed the centers of the two superdomains as $c_1 = \frac{1}{h} \sum_{i=1}^{h-1} x_i$ and $c_2 = \frac{1}{n-h} \sum_{i=h+1}^n x_i$, respectively. Then the distances of escape gene g to the superdomain centers are calculated as $||x_k - c_1||$ and $||x_k - c_2||$.

To test the enrichment of escape genes at the chromosome periphery or at the superdomain periphery, we randomly sampled 100 X-linked genes to estimate the expected distance to the chromosome or superdomain center and then evaluated the significance of observed distances of escape genes using a Z-test.

Correlation between one-dimensional genomic features and 3D structure

To investigate the spatial distribution of genetic and epigenetic features of the X chromosomes in the 3D nucleus space, we performed the following analyses for three different genetic and epigenetic features on both X_a and X_i : (1) allele-specific CTCF binding peaks in brain and Patski cells [15]; (2) allele-specific PolII peaks in brain and Patski cells [15]; (3) L1 elements (downloaded from UCSC Genome Browser).

First, we asked whether feature-rich regions are enriched at the chromosome or superdomain periphery. For each non-overlapping 1 Mb bin i along the X chromosome, we computed the density of the given feature at bin i . Then we visualized the feature density as a function of the radial distance to the chromosome or superdomain center using a boxplot. In addition, we asked whether feature-rich regions tend to locate near the periphery or interior of the

chromosome or superdomains. We define feature-rich regions as bins that fall within the top 25 % in terms of feature density, and we define feature-poor regions as bins within the bottom 25 % of feature density. Then we investigated the empirical cumulative density function of the distance to the chromosome or superdomain center for the feature-rich and feature-poor regions, and then evaluated the difference between the two distributions using one-side Wilcoxon rank-sum test.

Analysis of contacts at escape genes and imprinted genes

For this analysis we used allele-specific Hi-C contact maps at 40 kb resolution. Starting with a list of imprinted genes from a published study [36] we filtered the list to include only genes where our RNA-seq data indicated biased expression towards the putatively expressed allele based on a binomial test. The binomial parameter was derived from the set of all autosomal genes by taking the average ratio of maternal to paternal read counts. Multiple testing correction was performed with the Benjamini-Hochberg procedure, and a q value of less than 0.05 was considered significant. For the list of genes that escape XCI we used a previously established list [15] from which we selected those with an average PolII SNP read count of ≥ 5 in 100 bp intervals at 0.5 kb upstream and downstream of the transcription start site. For the list of autosomal and X-linked genes we downloaded the UCSC knownGenes table, retaining entries that were also listed in Ensembl. We excluded genes that overlap within the same Hi-C window (40 kb resolution) and any of the genes in our lists of imprinted genes or X escape genes. In cases where multiple genes in the remaining set fell within exactly the same Hi-C window, we included only one gene in the background distribution. Finally, we also eliminated any genes for which no contacts were observed in the bulk Hi-C contact map. Background distributions were separated for autosomal and X-linked genes.

For each gene across our imprinted, X escape, and background sets, we performed a virtual 4C analysis, where we extracted one or more columns from the paternal and maternal allele-specific contact maps. The contacts in these columns were summed for each allele prior to calculating $\log_{10}\left(\frac{\text{maternal contact count}+1}{\text{paternal contact count}+1}\right)$. Note that the +1 in the numerator and denominator acts as a pseudocount. For each imprinted gene set, we used a two-sided Kolmogorov-Smirnov test to test for significant deviation from the autosomal gene background distribution. The same test was performed for escape genes in comparison to all X-linked genes used to determine the background distribution. Similar significance values were obtained from alternative significance metrics such as a Wilcoxon rank sum test.

Accession numbers

The RNA-seq, ChIP-chip, ChIP-seq, and DNase Hi-C data are available in the Gene Expression Omnibus (GEO) database, under the accession numbers GSE30761 and GSE59779 (subseries GSE68992).

Ethics statement

For mice sacrificed, euthanasia was accomplished using two methods (carbon dioxide asphyxiation followed by cervical dislocation) as required by the University of Washington's Office of Animal Welfare. Husbandry and all other procedures were approved by the University of Washington's Office of Animal Welfare (Protocol 2254).

Additional files

Additional file 1: Table S1. Summary of allele-specific contacts. (XLSX 10 kb)

Additional file 2: Table S2. Summary of the number of contacts obtained by DNase Hi-C and in situ DNase Hi-C in relation to distance. (XLSX 9 kb)

Additional file 3: Figure S1. Allelic differences between contact maps for the X chromosomes are not seen for homologous autosomes. Allelic intrachromosomal chromatin contact heatmaps of chromosome 1 homologs based on SNP reads at 1 Mb resolution obtained by DNase Hi-C and in situ DNase Hi-C in F1 brain and in Patski cells. Contact maps for chromosomes 1 appear remarkably similar between maternal (BL6) and paternal (*spretus*) chromosomes. See Fig. 1a for comparison with contacts maps obtained for the Xa and Xi, which demonstrate striking differences. (PDF 1178 kb)

Additional file 4: Figure S2. TADs are more prominent on Xa versus Xi. Contact maps for the Xa and Xi at 100 kb resolution for chrX:98,500,00–103,499,999 in F1 brain (top) and Patski cells (bottom). (PDF 717 kb)

Additional file 5: Figure S3. Mapping of the hinge region. Contact maps at the hinge region in F1 brain and Patski cells. Data are shown at a 100 kb resolution for a region of chrX:70,000,000–75,099,999 (left) and at a 40 kb resolution for chrX:70,000,000–75,039,999 (right). The estimated boundaries of the superdomains 1 and 2 (at 100 kb or 40 kb) are marked by dotted red lines and the hinge region located in between the superdomains contains the least number of contacts. (PDF 825 kb)

Additional file 6: Figure S4. Reproducibility of Xi contact maps obtained by DNase Hi-C and in situ DNase Hi-C approaches in F1 brain and Patski cells. Allelic intrachromosomal chromatin contact heatmaps are shown for the mouse Xa and Xi based on SNP reads at 1 Mb resolution using DNase Hi-C and in situ DNase Hi-C to the same whole brain specimen, and for two independent biological replicates using in situ DNase Hi-C on Patski cells. These contact maps show remarkably similar features between replicates, between methods, and between in vitro and in vivo mouse hybrid systems. Note that XCI is reciprocal between F1 brain (*spretus* Xi) and Patski cells (BL6 Xi). (PDF 1598 kb)

Additional file 7: Table S3. List of imprinted genes and genes that escape XCI in mouse F1 brain. (XLSX 12 kb)

Additional file 8: Figure S5. Intrachromosomal contacts at *Kcnk9* and at imprinted genes, excluding those located on chromosome 7. **a** Significant contacts are detected at the maternally expressed gene *Kcnk9* on the maternal allele (Mat). Allelic mRNA-seq profiles show expression on the maternal allele at imprinted genes *Kcnk9*, *Trappc9*, *Chrac1* and *Ago2*, and on the paternal allele (Pat) at *Peg13*, in agreement with a previous study [62]. Allelic CTCF profiles show absence of binding to the differentially methylated region (DMR) on the maternal allele (arrow) presumably facilitating the formation of contacts between the *Kcnk9* promoter region and an unidentified distant enhancer, similar to the

situation in human [62]. The needle plot of contact counts between a 40 kb window that overlaps *Kcnk9* (grey bar) and other regions shows more interactions on the maternal (pink) than the paternal allele (blue). Genes with maternal or paternal expression are colored pink or blue, non-imprinted genes black, and non-expressed genes grey, respectively. Contact regions showing significant allelic biases are marked by asterisks. **b** Distribution of maternal-to-paternal allelic contacts at autosomal genes and X-linked genes determined by DNase Hi-C at 40 kb resolution in F1 brain in which the paternal chromosomes are from *spretus*. **c** Violin plots show the distribution of maternal-to-paternal allelic contacts at maternally and paternally imprinted genes at 40 kb resolution in F1 brain, after removing genes located on chromosome 7, which changes the shape of the distribution, due to fewer genes showing a low maternal-to-paternal contact ratio (see also Fig. 7). Dotted line indicates median ratios of maternal-to-paternal contacts at autosomal genes. (PDF 1263 kb)

Additional file 9: Table S4. List of primers. (XLS 20 kb)

Abbreviations

3D: three-dimensional; Bl: bipartite index; bp: base pair; ChIP: chromatin immunoprecipitation; FISH: fluorescence in situ hybridization; HAT: hypoxanthine-aminopterin-thymidine; L1: LINE1; LAD: lamina-associated domain; lncRNA: long non-coding RNA; MEF: mouse embryonic fibroblast; NAD: nucleolus-associated domain; PBS: phosphate-buffered saline; PEG: polyethylene glycol; PolII: RNA polymerase II phosphorylated at serine 5; qPCR: quantitative PCR; SNP: single nucleotide polymorphism; TAD: topologically associated domain; Xa: active X chromosome; XCI: X chromosome inactivation; Xi: inactive X chromosome.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XD, VR, and ZD performed the Hi-C experiments. XD and FY performed the ChIP-chip, ChIP-seq, and RNA-seq experiments. JB provided the mouse resources. VR and JS did the sequence analyses. WM, AH, FA, and WN performed the bioinformatics analyses. XD, WM, CB, JS, WN and CD discussed the study and wrote the paper. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by grants GM046883 (CMD, XD), and GM098039 (CMD and WSN) from the National Institute of Health (NIH.gov). XD is also supported by a Junior Faculty Pilot award from the Department of Pathology at the University of Washington. We thank Tomas Babak for providing us with a list of imprinted genes ahead of the publication of his manuscript. We thank Jun Xu and Gengze Wei (Washington State University) for the mouse primary neurons, Galina Filippova (Fred Hutchinson Cancer Research Center, Seattle) for female MEFs, and Barbara Panning for the *Xist* plasmid. We thank Xiaoqin Chen (University of Washington) for her help with imaging and cytogenetics analyses. We thank Nelle Varoquaux and Jean-Philippe Vert (Mines ParisTech) for their help with the 3D structure modeling.

Author details

¹Department of Pathology, University of Washington, Seattle, Washington, USA. ²Department of Genome Sciences, University of Washington, Seattle, Washington, USA. ³Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, Washington, USA. ⁴Division of Hematology, University of Washington, Seattle, Washington, USA. ⁵Department of Computer Science and Engineering, University of Washington, Seattle, Washington, USA. ⁶Department of Medicine, University of Washington, Seattle, Washington, USA.

Received: 12 May 2015 Accepted: 20 July 2015

Published online: 07 August 2015

References

1. Bickmore WA. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet*. 2013;14:67–84.
2. Lyon M. Gene action in the X-chromosome of the mouse (*Mus musculus* L). *Nature*. 1961;190:372–3.

3. Bartolomei MS, Ferguson-Smith AC. Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol.* 2011;3:a002592.
4. Nora EP, Lajoie BR, Schulz EG, Giorgotti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 2012;485:381–5.
5. Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, Zhu Y, et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* 2011;25:1371–83.
6. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014;159:1665–80.
7. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, et al. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A.* 2006;103:10684–9.
8. Barr ML, Bertram EG. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature.* 1949;163:676.
9. Gendrel AV, Heard E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu Rev Cell Dev Biol.* 2014;30:561–80.
10. McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature.* 2015;521:232–6.
11. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, et al. Systematic discovery of Xist RNA binding proteins. *Cell.* 2015;161:404–16.
12. Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, et al. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science.* 2015;349:aab2276.
13. Heard E, Disteche CM. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* 2006;20:1848–67.
14. Cotton AM, Ge B, Light N, Adoue V, Pastinen T, Brown CJ. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol.* 2013;14:R122.
15. Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Deng X, et al. Escape from X-inactivation varies in mouse tissues. *PLoS Genet.* 2015;11:e1005079.
16. Deng X, Berletch JB, Nguyen DK, Disteche CM. X chromosome regulation: diverse patterns in development, tissues and disease. *Nat Rev Genet.* 2014;15:367–78.
17. Bourgeois CA, Laquerriere F, Hemon D, Hubert J, Bouteille M. New data on the in-situ position of the inactive X chromosome in the interphase nucleus of human fibroblasts. *Hum Genet.* 1985;69:122–9.
18. Zhang LF, Huynh KD, Lee JT. Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell.* 2007;129:693–706.
19. Chow JC, Ciando C, Fazzari MJ, Mise N, Servant N, Glass JL, et al. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell.* 2010;141:956–69.
20. Heard E, Bickmore W. The ins and outs of gene regulation and chromosome territory organisation. *Curr Opin Cell Biol.* 2007;19:311–6.
21. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;485:376–80.
22. Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell.* 2013;49:773–82.
23. Ma W, Ay F, Lee C, Gulsoy G, Deng X, Cook S, et al. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods.* 2015;12:71–8.
24. Cullen KE, Kladde MP, Seyfred MA. Interaction between transcription regulatory regions of prolactin chromatin. *Science.* 1993;261:203–6.
25. Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, et al. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell.* 2013;25:55–68.
26. Yang F, Babak T, Shendure J, Disteche CM. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* 2010;20:614–22.
27. Lingenfelter PA, Adler DA, Poslinski D, Thomas S, Elliott RW, Chapman VM, et al. Escape from X inactivation of Smcx is preceded by silencing during mouse development. *Nat Genet.* 1998;18:212–3.
28. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature.* 2013;502:59–64.
29. Kuznetsova I, Podgornaya O, Ferguson-Smith MA. High-resolution organization of mouse centromeric and pericentromeric DNA. *Cytogenet Genome Res.* 2006;112:248–55.
30. Chadwick BP. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res.* 2008;18:1259–69.
31. Horakova AH, Calabrese JM, McLaughlin CR, Tremblay DC, Magnuson T, Chadwick BP. The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. *Genome Biol.* 2012;13:R70.
32. Yang F, Deng X, Ma W, Berletch JB, Rabaia N, Wei G, et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* 2015;1–17.
33. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics.* 2011;27:1017–8.
34. Nemeth A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, et al. Initial genomics of the human nucleolus. *PLoS Genet.* 2010;6:e1000889.
35. Sullivan GJ, Bridger JM, Cuthbert AP, Newbold RF, Bickmore WA, McStay B. Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *EMBO J.* 2001;20:2867–74.
36. Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, et al. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat Genet.* 2015;47:544–9.
37. Rinn J, Guttmann M. RNA Function. RNA and dynamic nuclear organization. *Science.* 2014;345:1240–1.
38. McLaughlin CR, Chadwick BP. Characterization of DXZ4 conservation in primates implies important functional roles for CTCF binding, array expression and tandem repeat organization on the X chromosome. *Genome Biol.* 2011;12:R37.
39. Kim JH, Ebersole T, Kouprina N, Noskov VN, Ohzeki J, Masumoto H, et al. Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res.* 2009;19:533–44.
40. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol.* 2014;21:198–206.
41. Kung JT, Kesner B, An JY, Ahn JY, Cifuentes-Rojas C, Colognori D, et al. Locus-specific targeting to the X Chromosome revealed by the RNA interactome of CTCF. *Mol Cell.* 2015;57:361–75.
42. Horakova AH, Moseley SC, McLaughlin CR, Tremblay DC, Chadwick BP. The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum Mol Genet.* 2012;21:4367–77.
43. Padeen J, Heun P. Nucleolus and nuclear periphery: velcro for heterochromatin. *Curr Opin Cell Biol.* 2014;28:54–60.
44. Walker CL, Cargile CB, Floy KM, Delannoy M, Migeon BR. The Barr body is a looped X chromosome formed by telomere association. *Proc Natl Acad Sci U S A.* 1991;88:6191–5.
45. Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell.* 2014;157:950–63.
46. Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, et al. Transcription: CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science.* 2015;347:1017–21.
47. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell.* 2014;159:374–87.
48. Ke X, Thomas NS, Robinson DO, Collins A. The distinguishing sequence characteristics of mouse imprinted genes. *Mamm Genome.* 2002;13:639–45.
49. Greally JM. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci U S A.* 2002;99:327–32.
50. Hoki Y, Kimura N, Kanbayashi M, Amakawa Y, Ohhata T, Sasaki H, et al. A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development.* 2009;136:139–46.
51. Marahrens Y, Loring J, Jaenisch R. Role of the Xist gene in X chromosome choosing. *Cell.* 1998;92:657–64.
52. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12:996–1006.

53. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589–95.
54. Ay F, Bailey TL, Noble WS. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* 2014;24:999–1011.
55. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, et al. A three-dimensional model of the yeast genome. *Nature*. 2010;465:363–7.
56. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B Methodol.* 1977;39:1–38.
57. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods*. 2012;9:999–1003.
58. UCSC genome LiftOver tool. <https://genome.ucsc.edu/cgi-bin/hgLiftOver>.
59. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* 2014;42:D756–63.
60. Varoquaux N, Ay F, Noble WS, Vert JP. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics*. 2014;30:i26–33.
61. Wachter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program.* 2006;106:25–57.
62. Court F, Camprubí C, Garcia CV, Guillaumet-Adkins A, Sparago A, Seruggia D, et al. The PEG13-DMR and brain-specific enhancers dictate imprinted expression within the 8q24 intellectual disability risk locus. *Epigenetics Chromatin*. 2014;7:5.
63. Thiaville MM, Kim H, Frey WD, Kim J. Identification of an evolutionarily conserved cis-regulatory element controlling the Peg3 imprinted domain. *PLoS One*. 2013;8:e75417.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

