

Structural organization of the inactive X chromosome in the mouse

Luca Giorgetti^{1*†}, Bryan R. Lajoie^{2*}, Ava C. Carter^{3*}, Mikael Attia^{1*}, Ye Zhan², Jin Xu³, Chong Jian Chen¹, Noam Kaplan², Howard Y. Chang³, Edith Heard^{1,4} & Job Dekker^{2,5}

X-chromosome inactivation (XCI) involves major reorganization of the X chromosome as it becomes silent and heterochromatic. During female mammalian development, XCI is triggered by upregulation of the non-coding *Xist* RNA from one of the two X chromosomes. *Xist* coats the chromosome in *cis* and induces silencing of almost all genes via its A-repeat region^{1,2}, although some genes (constitutive escapees) avoid silencing in most cell types, and others (facultative escapees) escape XCI only in specific contexts³. A role for *Xist* in organizing the inactive X (Xi) chromosome has been proposed^{4–6}. Recent chromosome conformation capture approaches have revealed global loss of local structure on the Xi chromosome and formation of large mega-domains, separated by a region containing the *DXZ4* macrosatellite^{7–10}. However, the molecular architecture of the Xi chromosome, in both the silent and expressed regions, remains unclear. Here we investigate the structure, chromatin accessibility and expression status of the mouse Xi chromosome in highly polymorphic clonal neural progenitors (NPCs) and embryonic stem cells. We demonstrate a crucial role for *Xist* and the *DXZ4*-containing boundary in shaping Xi chromosome structure using allele-specific genome-wide chromosome conformation capture (Hi-C) analysis, an assay for transposase-accessible chromatin with high throughput sequencing (ATAC-seq) and RNA sequencing. Deletion of the boundary disrupts mega-domain formation, and induction of *Xist* RNA initiates formation of the boundary and the loss of DNA accessibility. We also show that in NPCs, the Xi chromosome lacks active/inactive compartments and topologically associating domains (TADs), except around genes that escape XCI. Escapee gene clusters display TAD-like structures and retain DNA accessibility at promoter-proximal and CTCF-binding sites. Furthermore, altered patterns of facultative escape genes in different neural progenitor clones are associated with the presence of different TAD-like structures after XCI. These findings suggest a key role for transcription and CTCF in the formation of TADs in the context of the Xi chromosome in neural progenitors.

To investigate the structure of the Xi chromosome, we performed allele-specific Hi-C in a clonal neural progenitor cell (NPC) line that was derived from highly polymorphic F₁ mouse embryonic stem (ES) cells (129 × Cast, Extended Data Fig. 1a–f). We first performed Hi-C in ES cells, in which XCI has not yet occurred, and found that autosomes and both active X chromosomes displayed prominent active/inactive (A/B) compartmentalization and TAD structures (Extended Data Figs 2a–c and 3). In NPCs, compartments and TADs were similarly detected on autosomes and the active X (Xa) chromosome (Fig. 1a, Extended Data Fig. 3a). Notably, however, the Xi chromosome displayed no A/B compartments (Extended Data Fig. 2c), but was instead partitioned into two massive interaction domains separated by a hinge region of ~200 kb including the *DXZ4* macrosatellite^{11,12} (Fig. 1a),

as recently reported for the human and mouse Xi chromosomes^{7–9}. Furthermore, TADs were found to be largely absent on the Xi chromosome (Fig. 1a), as previously suggested^{9,10}.

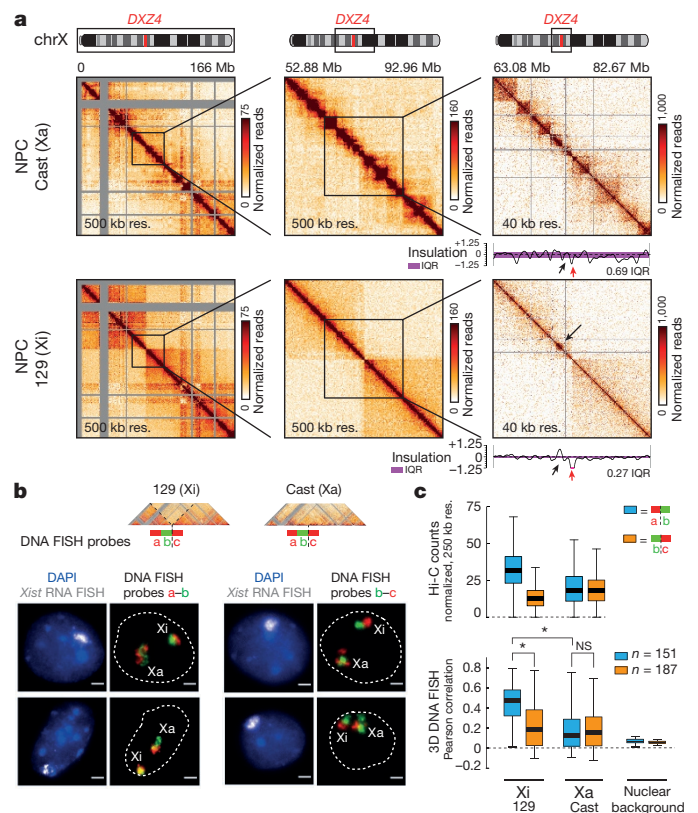


Figure 1 | The distinct conformation of the Xi and Xa chromosomes. **a**, Allele-specific Hi-C maps of the Xa and Xi chromosomes in NPCs (left), and two increasingly smaller regions centred around *DXZ4* (centre and right). Purple areas in the insulation score plots indicate the interquartile range (IQR) of insulation scores over the entire X chromosome, to illustrate reduced insulation scores along the Xi chromosome (indicating loss of TADs). Black arrow: position of a residual TAD. Red arrow: position of *DXZ4*. **b**, Top, scheme of DNA FISH probes. Bottom, probes a–b (within the same mega-domain) are more overlapping and spherical on the Xi than on the Xa chromosome, whereas signals from b–c (across the mega-domain boundary) show partitioning on the Xi chromosome into two separate domains. **c**, Loci detected by probes a–b are more interacting than b–c on the Xi chromosome, in both Hi-C (top) and 3D-DNA FISH (bottom). * $P < 8 \times 10^{-17}$ (Wilcoxon's rank sum test corrected with Bonferroni). NS, not significant. n denotes number of cells analysed in DNA FISH. Centre lines: medians, all experiments were performed in biological duplicates.

¹Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, 26 Rue d'Ulm, 75248 Paris Cedex 05, France. ²Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, Massachusetts 01605, USA. ³Center for Personal Dynamic Regulomes and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ⁴Collège de France, 11 place Marcelin-Berthelot, Paris 75005, France. ⁵Howard Hughes Medical Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, Massachusetts 01605, USA. †Present address: Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland.

*These authors contributed equally to this work.

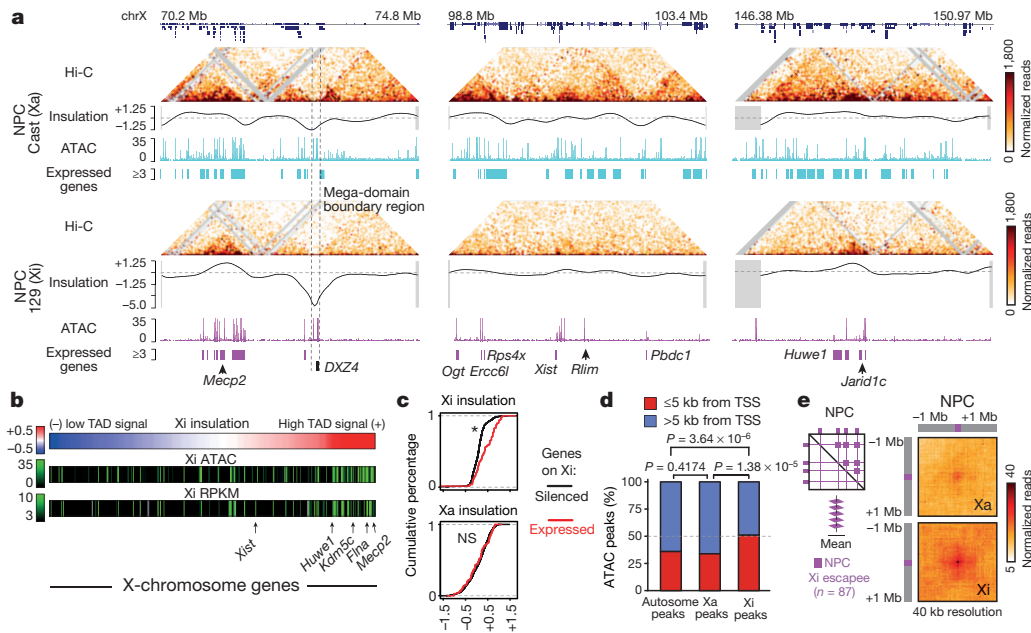


Figure 2 | Expression, chromatin accessibility and chromatin conformation along the Xi chromosome. **a**, Hi-C interactions, insulation score, ATAC-seq signal and location of expressed genes (via RNA-seq) in three regions of the Xa and Xi chromosome. Left to right: the cluster of facultative escape genes containing *Mecp2*, the *Xist* locus and the region encompassing *Jarid1c*. Hi-C data are shown at 40-kb resolution. **b**, Regions with increased TAD structure harbour promoters that are expressed and accessible on the Xi chromosome, as shown by analysis of Hi-C insulation (TAD structure), ATAC-seq read counts, and RNA-seq reads per kilobase of transcript per million mapped reads (RPKM). Each column represents a promoter, sorted by insulation score in the 40-kb interval containing the promoter, lowest-to-highest (weakest-to-strongest TAD signal).

To investigate the spatial organization of mega-domains at the single-cell level, we performed DNA fluorescence *in situ* hybridization (FISH) with 18-Mb probe sets located within one mega-domain (probes a–b), or spanning the mega-domain boundary (probes b–c) (Fig. 1b, Extended Data Fig. 4a). Despite extensive cell-to-cell variation, regions within the same mega-domain showed greater overlap on the Xi than on the Xa chromosome (Fig. 1b, c, Extended Data Fig. 4b), whereas regions spanning the boundary showed lower overlap on the Xi chromosome, in agreement with Hi-C data (Fig. 1c). Similar results were obtained in an NPC clone with a Cast Xi chromosome and in astrocytes (Extended Data Fig. 4c). Using two independent quantification methods (Extended Data Fig. 4d–f), the volume of each 18-Mb region was found to be modestly (approximately 20%) but significantly smaller on the Xi than the Xa chromosome, consistent with observations on the human Xi chromosome^{13–15}. Thus, the mouse Xi chromosome is moderately compacted and partitioned into two large, spatially distinct domains that show varying degrees of overlap within the cell population (Fig. 1b, c, Extended Data Fig. 4).

Although the Xi chromosome globally presented no TADs in NPCs, we could detect a few residual TAD-like structures (Fig. 1a, black arrow). Integration of Hi-C, RNA sequencing (RNA-seq)¹⁶ and ATAC-seq data produced in the same lines revealed that these Xi chromosome TAD-like structures correspond to hotspots of residual transcription and open chromatin. The RNA-seq and ATAC-seq profiles are consistent with global inactivity of the Xi chromosome (Extended Data Fig. 5a), with a 75% reduction in both the number of expressed genes and accessible elements compared to the Xa chromosome. Most ATAC-seq peaks on the Xi chromosome fall in the pseudoautosomal region, the *Xist* locus and at the promoters of genes that escape XCI either facultatively or constitutively³ (Extended Data Fig. 5a).

ATAC counts are extracted from promoters (TSS \pm 500 base pairs (bp)). **c**, Xi-chromosome-expressed genes (escapees) fall within regions with higher insulation scores on the Xi chromosome as compared to Xi-chromosome-silenced genes ($*P = 4.44 \times 10^{-16}$, Fisher exact test), despite having similar insulation scores on the Xa chromosome ($P = 0.43114$, Kolmogorov–Smirnov test). **d**, ATAC-seq peaks on the Xi chromosome tend to be closer to TSSs (within 5 kb) than peaks on autosomes and the Xa chromosome (statistical significance was assessed by Fisher exact test). **e**, Interaction pile-up maps showing mean Hi-C signal for all pairwise combinations of the 87 wild-type NPC Xi chromosome escapees on the both the Xa and the Xi chromosome.

The amount of local structure on the Xi chromosome correlates with the density of transcribed loci and accessible elements (Extended Data Fig. 5d), as shown in Fig. 2a by three examples: a cluster of 19 facultative escapees including *Mecp2*, overlapping a strong \sim 800-kb TAD-like structure; the *Xist* locus with moderate interactions across a \sim 250-kb region; and 5 escapees including the constitutive *Jarid1c* (also known as *Kdm5c*) and the facultative *Huwe1* genes, embedded in a \sim 500-kb TAD. Importantly, residual TAD-like structures on the Xi chromosome occasionally coincide with sub-TAD structures on the Xa chromosome (Fig. 2a). Generally, escapees are located in Xi chromosome regions with higher TAD strength and chromatin accessibility as compared to silenced genes (Fig. 2b, c, Extended Data Fig. 5b). Thus, in NPCs, the appearance of TAD structures is intimately linked to gene expression on the Xi chromosome, unlike on the Xa chromosome and autosomes, where TADs are present even in the absence of transcription.

Notably, 51% of Xi chromosome accessible sites are $<$ 5 kb from a promoter compared to \sim 35% on the Xa chromosome (Fig. 2d), suggesting that escape is often regulated through promoter-proximal sites. Most ATAC-seq peaks on the Xi chromosome were found at CTCF-binding sites (Extended Data Fig. 5c), implicating CTCF in escape. Recent findings suggest that cohesin (which co-localizes with CTCF) is globally lost on the Xi chromosome, which may lead to the chromosome-wide loss of TADs⁹. Our discovery that only escapees show TAD-like structures on the Xi chromosome, and that they are associated with putative CTCF sites, is consistent with a role for CTCF in TAD formation and/or maintenance^{17–19}.

Escapees on the Xi chromosome also tend to interact with each other even across the mega-domain boundary, consistent with previous circularized chromosome conformation capture (4C) analysis results⁵ (Fig. 2e and Extended Data Fig. 2d, e).

To investigate the importance of the unusual bipartite organization of the Xi chromosome we deleted the ~200 kb boundary region encompassing the *DXZ4* locus, specifically on the 129 allele in ES cells (Δ FT) (Extended Data Figs 1b and 6a). After differentiation, many NPC clones were isolated with a 129 (Δ FT) Xi chromosome. Deletion of the boundary did not affect XCI establishment, as NPC clones with either a wild-type or Δ FT Xi chromosome were obtained. Hi-C performed on one such clone (D9B2) revealed massive reorganization of the Δ FT Xi chromosome resulting in fusion of the two mega-domains (Fig. 3a). No effect was visible on the Cast Xa chromosome (Extended Data Fig. 7a). DNA FISH confirmed that sequences on either side of the deleted boundary overlap significantly more on the Δ FT Xi chromosome compared to wild type, consistent with Hi-C data (Extended Data Fig. 8a).

Notably, in the D9B2 clone, facultative escapees (*Mecp2*, *Huwe1*) no longer escaped from XCI on the Δ FT Xi chromosome (Extended Data Fig. 6b-c). ATAC-seq and RNA-seq confirmed substantially reduced escape on the Δ FT Xi chromosome, with only 29 expressed genes (Extended Data Fig. 8b). Transcription and chromatin accessibility were lost at 66 of the 87 facultative escapee genes (~76%) (Fig. 3b), but maintained at all 6 constitutive escapees (Fig. 3b, Supplementary Table 1 and Extended Data Fig. 8c). However, analysis of multiple NPC clones, both wild type and Δ FT, revealed only a slight tendency for Δ FT NPC clones to show less escape than wild type, with varying degrees of facultative escape between clones, with or without the mega-domain boundary (Extended Data Fig. 6d). Nevertheless, clone D9B2 provided a unique opportunity to study the relationship between transcriptional activity and chromosome conformation on the Xi chromosome.

TAD-like structures were absent on the Xi chromosome when expression was lost at facultative escapees in the D9B2 clone (Fig. 3b and Extended Data Fig. 7b). Xi chromosome-wide comparisons between the D9B2 (Δ FT) and wild-type NPCs revealed a strong correlation between loss of escape, loss of chromatin accessibility and reduction in TAD signal (Fig. 3c). Specific long-range interactions between escapees were also lost on the Δ FT Xi chromosome (Fig. 3d). Eight genes showed *de novo* escape in the D9B2 clone (Supplementary Table 1). These *de novo* escapees were not clustered or highly accessible by ATAC-seq, and their expression did not lead to strong TAD formation (data not shown).

Of the 224 Xi chromosome ATAC-seq peaks that we detected in the wild-type clone, 139 were lost in the D9B2 clone (Fig. 3e, Extended Data Fig. 8b). These lost sites were enriched for promoter-proximal location (64%) (Fig. 3f, g). In total, 93% of these promoter-proximal sites contain CTCF-binding sites, an enrichment compared to the 64% of promoter-distal sites (Fig. 3f). These CTCF sites are closer to escape gene transcription start sites (TSSs) than unchanging sites, again pointing to a role for promoter-proximal CTCF in escape gene regulation (Fig. 3h).

We next investigated the role of *Xist* in establishing the unusual organization of the Xi chromosome. We induced *Xist* expression in undifferentiated male (XY) ES cells carrying a tetracycline-inducible promoter at the endogenous *Xist* locus¹. Hi-C revealed that 48 h of *Xist* induction resulted in notable structural changes along the X chromosome, and these changes were not observed after induction of a *Xist* mutant lacking the A-repeat region (Fig. 4a, b)—which cannot silence genes (Extended Data Fig. 9a), but is competent for *Xist* coating and exclusion of RNA polymerase II (refs 1, 4). Wild-type *Xist* induction did not lead to detectable changes in TAD structure (data not shown), but resulted in increased interaction frequencies along the chromosome (Fig. 4b, d). The contact map of the wild-type *Xist*-coated X chromosome was found to be more similar to that of the NPC Xi chromosome than either the non-induced or the A-repeat mutant *Xist*-coated X chromosome (Fig. 4a). Notably, physical separation across the mega-domain boundary occurred (Fig. 4b), confirmed by RNA/DNA FISH (Fig. 4c, Extended Data Fig. 9b). Induction of wild-type

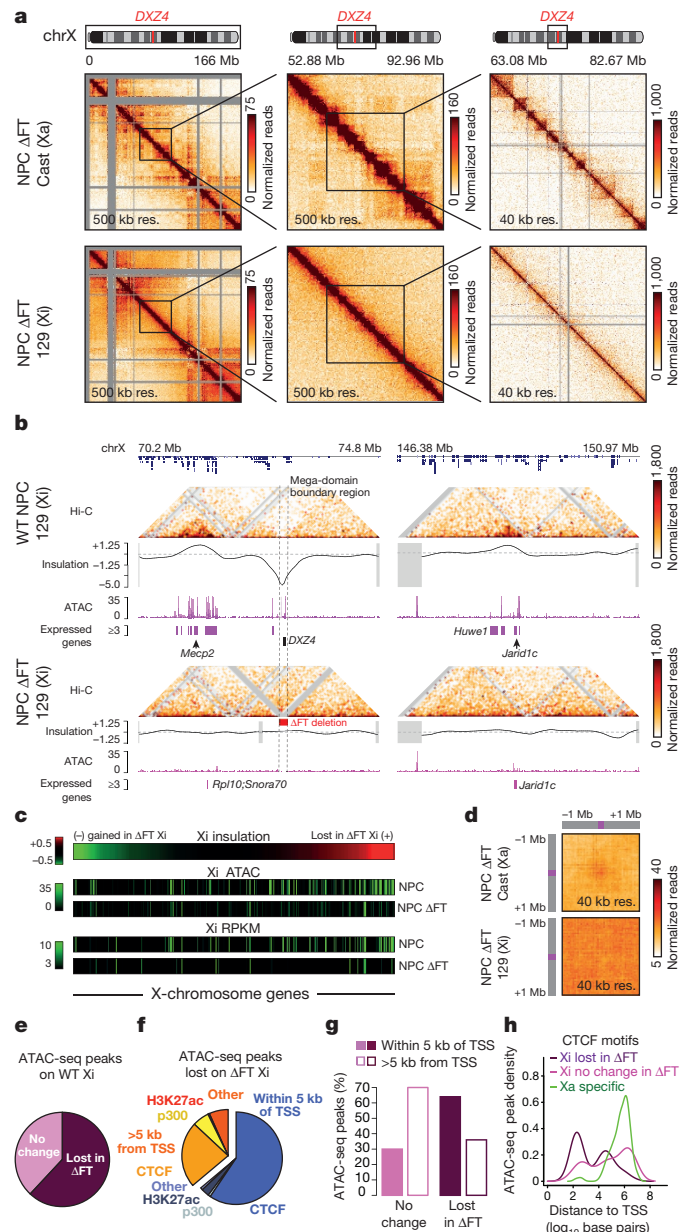


Figure 3 | Deletion of the mega-domain boundary leads to loss of bipartite folding. **a**, Hi-C contact maps for Xa (Cast) and Xi (129 Δ FT) chromosomes in mega-domain boundary mutant NPCs (left), and for increasingly smaller regions (centre and right). **b**, Two regions on the Δ FT Xi chromosome showing Hi-C, RNA-seq and ATAC-seq signal. The same regions as in Fig. 2a are shown. ATAC-seq and RNA-seq from the Xi chromosome in wild-type NPCs are included for reference (shown in Fig. 2a). **c**, Loss of TAD structure in Δ FT NPCs correlates with loss of accessibility (ATAC) and expression (RPKM). Each column represents a promoter. Heat maps are sorted by change in TAD strength (insulation score) from wild-type to Δ FT Xi in the 40-kb interval containing the promoter. ATAC counts are extracted from promoters (TSS \pm 500 bp TSS). **d**, Interaction pile-up map showing the 87 wild-type NPC Xi chromosome escapees. **e**, Quantification of ATAC-seq peaks in wild-type and Δ FT NPCs on the Xi chromosome. Of 224 Xi chromosome peaks in the wild type, 139 are lost in the mutant. **f**, Chromatin immunoprecipitation followed by sequencing (ChIP-seq) annotation of ATAC-seq peaks lost in Δ FT NPCs. **g**, Distance from TSS of ATAC-seq peaks that are lost, do not change, after deletion of the mega-domain boundary. **h**, Distance from escapee TSS of CTCF peaks that are lost in Δ FT NPCs, do not change in Δ FT NPCs, and peaks on the Xa chromosome.

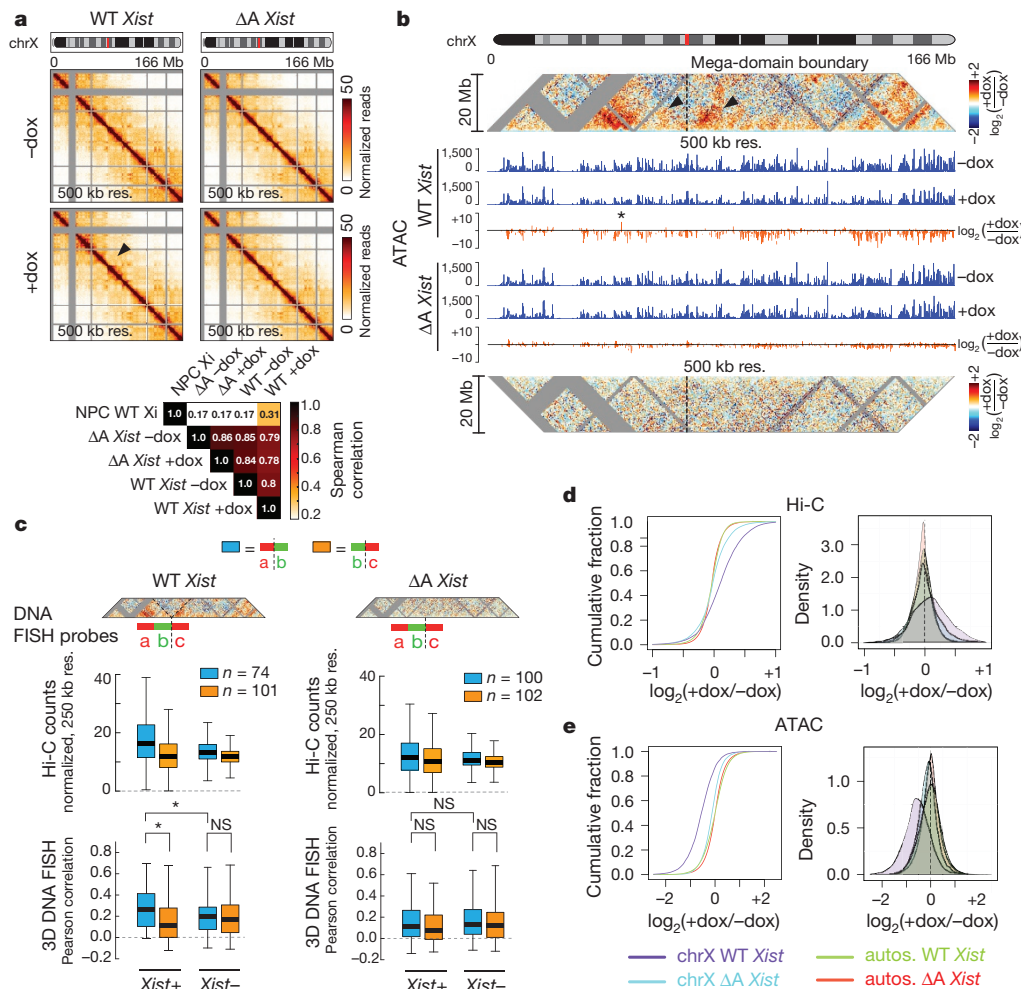


Figure 4 | *Xist*-mediated silencing is sufficient to generate a boundary at *DXZ4* in ES cells. **a**, Top, Hi-C analysis on the X chromosome in male ES cells expressing wild-type (WT) or A-repeat mutant (ΔA) *Xist*, before (–dox) and after (+dox) 48 h induction. dox, doxycycline. Bottom, correlation analysis of male X chromosome and NPC Xi chromosome Hi-C maps, showing increased similarity between the male *Xist*-coated X and Xi chromosomes after induction of wild-type, but not ΔA *Xist*. **b**, Alignment of structural changes detected in Hi-C and changes in chromatin accessibility measured by ATAC-seq after 48 h of wild-type or ΔA *Xist* expression. Dashed line: *DXZ4* position. Arrowheads: increased interactions on either side of the mega-domain boundary. Asterisk: genomic position of *Firre*. **c**, RNA/DNA FISH showing increased overlap of probes a–b on the *Xist*-coated X chromosome after 48 h of

wild-type, but not ΔA *Xist* induction. Probes b–c show lower overlap (indistinguishable from non-*Xist* coated chromosomes in cells where *Xist* expression was not induced after doxycycline treatment), in line with Hi-C predictions (see Methods). *Xist*– signals correspond to the –dox sample. * $P < 0.05$ (Wilcoxon's rank sum test corrected with Bonferroni). n denotes number of cells analysed in DNA FISH. Centre lines: medians. Boxes: middle 50% of data points. Two biological replicates were analysed. **d**, Changes in Hi-C contact probability after 48 h of wild-type or ΔA *Xist* expression, indicating a wild-type *Xist*-specific increase in contact probability on the X chromosome. **e**, Changes in the ATAC-seq signal at X chromosome and autosomal peaks, indicating wild-type *Xist*-specific loss of chromatin accessibility on the X chromosome.

Xist from one X chromosome in female ES cells²⁰ generated a boundary of comparable magnitude (Extended Data Fig. 9c). ATAC-seq showed that 48 h of wild-type (but not A-repeat mutant) *Xist* expression results in globally reduced (but not eliminated) accessibility on the X chromosome (Fig. 4b, e). Interestingly, a small number of loci showed increased chromatin accessibility after wild-type *Xist* coating including the *Firre* long noncoding RNA (Fig. 4b), which together with the *DXZ4* macrosatellite has been proposed to anchor the Xi chromosome to the nucleolus²¹. Regions of reduced chromatin accessibility tend to show increased interaction frequencies with loci in the surrounding ~20 Mb of genomic sequence (Extended Data Fig. 9d). We note that conformational changes on the *Xist*-coated X chromosome in male ES cells were relatively mild compared to the NPC Xi chromosome, probably for several reasons. First, *Xist* was induced in ~35–45% of cells (in two independent replicates). Detecting structural alterations in such a sub-population may be difficult by Hi-C, and this may explain our observation that TADs are still detected in the overall cell population.

Second, the *Xist*-coated Xi chromosome chromatin state in ES cells may not be comparable to that in differentiated NPCs²².

In conclusion, our study uncovers a complex molecular architecture of the Xi chromosome, and reveals new insights into TAD formation. The Xi chromosome is moderately compacted and organized into two mega-domains, with global loss of TADs, except at clusters of expressed genes. Thus, in contrast with the notion that TADs are highly stable across differentiation and do not require transcription for their presence or maintenance^{10,23}, our findings demonstrate that (1) TADs can be lost in the context of the Xi chromosome, at least in NPCs; and (2) transcription together with binding of factors such as CTCF may enable their maintenance or *de novo* re-creation.

The bipartite folding of the Xi chromosome into two mega-domains is evolutionarily conserved, pointing to a possible role for this peculiar organization in the XCI process. Deletion of the boundary region, leading to fusion of the two mega-domains, does not affect initiation of XCI. However, somewhat reduced rates of facultative,

but not constitutive, escape are observed. Facultative escapees are silenced during XCI and then re-expressed^{24,25} and may be more prone to stochastic events and/or the influence of long-range interactions with other parts of the X chromosome, such as the CTCF-rich *DXZ4* region²⁶. As promoter-proximal CTCF sites characterize both facultative and constitutive escape genes, we speculate that transient interactions of escapees with the boundary region during XCI may influence escape (Extended Data Fig. 9e). However, the highly variable facultative escape that we found between NPC clones, both with or without the *DXZ4* boundary, suggests that escape efficiency cannot be due just to the presence of the *DXZ4* region, but may be influenced by local folding and accessibility to CTCF binding. The exact timing and mechanism of facultative escape, and its relationship with TAD formation, will require further investigation. In conclusion, our study establishes the Xi chromosome as a powerful system for studying the mechanistic relationships between chromosome conformation and gene regulation, and points to key roles for gene activity and CTCF in the establishment of TAD structure in the context of facultative heterochromatin.

Note added in proof: Consistent with our observations in mouse, deletion of *DXZ4* from the human Xi chromosome results in loss of mega-domain structure (E. Lieberman Aiden & B. Chadwick, Deletion of the macrosatellite *DXZ4* on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl Acad. Sci USA* (in press)).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 July 2015; accepted 18 May 2016.

Published online 18 July 2016.

- Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nature Genet.* **30**, 167–174 (2002).
- Chu, C. *et al.* Systematic discovery of *Xist* RNA binding proteins. *Cell* **161**, 404–416 (2015).
- Peeters, S. B., Cotton, A. M. & Brown, C. J. Variable escape from X-chromosome inactivation: identifying factors that tip the scales towards expression. *BioEssays* **36**, 746–756 (2014).
- Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. 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 (2006).
- Splinter, E. *et al.* The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on *Xist* RNA. *Genes Dev.* **25**, 1371–1383 (2011).
- Engreitz, J. M. *et al.* The *Xist* lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
- Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **162**, 687–688 (2015).
- Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 152 (2015).
- Minajigi, A. *et al.* Chromosomes. A comprehensive *Xist* interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
- Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Giacalone, J., Friedes, J. & Francke, U. A novel GC-rich human macrosatellite VNTR in *Xq24* is differentially methylated on active and inactive X chromosomes. *Nature Genet.* **1**, 137–143 (1992).
- Chadwick, B. P. *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.* **18**, 1259–1269 (2008).
- Teller, K. *et al.* A top-down analysis of Xa- and Xi-territories reveals differences of higher order structure at ≥ 20 Mb genomic length scales. *Nucleus* **2**, 465–477 (2011).
- Naughton, C., Sproul, D., Hamilton, C. & Gilbert, N. Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Mol. Cell* **40**, 397–409 (2010).
- Eils, R. *et al.* Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *J. Cell Biol.* **135**, 1427–1440 (1996).
- Gendrel, A.-V. *et al.* Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* **28**, 366–380 (2014).
- Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996–1001 (2014).
- Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).
- Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports* **10**, 1297–1309 (2015).
- Schulz, E. G. *et al.* The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* **14**, 203–216 (2014).
- Yang, F. *et al.* The lncRNA *Firre* anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **16**, 52 (2015).
- Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5**, 695–705 (2000).
- Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Corbel, C., Diabougouaya, P., Gendrel, A.-V., Chow, J. C. & Heard, E. Unusual chromatin status and organization of the inactive X chromosome in murine trophoblast giant cells. *Development* **140**, 861–872 (2013).
- Patrat, C. *et al.* Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proc. Natl Acad. Sci. USA* **106**, 5198–5203 (2009).
- Horakova, A. H. *et al.* The mouse *DXZ4* homolog retains Ctf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. *Genome Biol.* **13**, R70 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Heard, Dekker, and Chang laboratories for their help and critical insights; PICT-IBISA@BDD (UMR3215/U934) Imaging facility of the Institut Curie. L.G. would like to thank L. Mirny for discussing gyration tensor analysis. Supported by grants from the National Institutes of Health (P50-HG007735) and Scleroderma Research Foundation (to H.Y.C.), from the National Human Genome Research Institute (R01 HG003143) and the National Institutes of Health Common Fund, National Institute of Diabetes and Digestive and Kidney Diseases (U54 DK107980) to J.D., the Human Frontier Science Program to N.K., an EMBO Fellowship to L.G., an ERC Advanced Investigator award (ERC-2014-AdG no. 671027), EU FP7 grants SYBOSS (EU 7th Framework G.A. no. 242129) and MODHEP (EU 7th Framework G.A. no. 259743), La Ligue, Fondation de France, Labex DEEP (ANR-11-LBX-0044) part of the IDEX Idex PSL (ANR-10-IDEX-0001-02 PSL) and ABS4NGS (ANR-11-BINF-0001) to E.H. J.D. is an investigator of the Howard Hughes Medical Institute.

Author Contributions E.H. and J.D. conceived the original strategy. For Hi-C, M.A. and L.G. prepared the ES cell and NPC samples, Y.Z. performed the Hi-C experiments; for ATAC-seq, H.Y.C., E.H. and J.D. designed the experiments; A.C.C. and J.X. prepared the samples and performed the experiments; for RNA-seq, M.A. and L.G. prepared the samples and performed the experiments and C.J.C. analysed the data. Integrated analysis of Hi-C sequencing, RNA-seq and ATAC-seq data was performed by B.R.L., J.X., L.G. and A.C.C., with assistance from C.J.C. and N.K. and input from J.D.; L.G. and E.H. designed the FISH experiments, L.G. and M.A. performed FISH experiment and L.G. analysed data. M.A., L.G. and E.H. designed the NPC and *DXZ4* mutant strategy, M.A. and L.G. performed the experiments and analysed them. L.G., B.R.L., A.C.C., E.H. and J.D. wrote the manuscript with input from H.Y.C.

Author Information Sequencing data have been deposited in the Gene Expression Omnibus (GEO) under accession numbers GSE72697 (Hi-C); GSE71156 (ATAC-seq); and GSE72697 (boundary deletion data). Code used in this work is available at (<https://github.com/dekkerlab/giorgetti-nature-2016>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.H. (edith.heard@curie.fr) or J.D. (job.dekker@umassmed.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture. The hybrid mouse ES cell line F121.6 (129Sv-Cast/Eij), a gift from J. Gribnau, was grown on mitomycin-C-inactivated mouse embryonic fibroblasts in ES cell media containing 15% FBS (Gibco), 10^{-4} M β -mercaptoethanol (Sigma), $1,000\text{ U ml}^{-1}$ leukaemia inhibitory factor (LIF, Chemicon). Male-inducible TXY and TXY: Δ A lines (a gift from A. Wutz, called *Xist-tetOP* and *Xist- Δ SX-tetOP*, respectively, in ref. 1) were cultured in the same conditions and treated for 48 h with $2\mu\text{g ml}^{-1}$ doxycycline. Differentiation of F121.6 ES cells into NPCs was performed as previously described¹⁶. Subcloning of NPCs was made by limiting dilution and manual colony picking. All cells used in this study were characterized for absence of mycoplasma contamination.

Boundary deletion. To generate the boundary region deletion, 5×10^6 ES cells were transfected with $5\mu\text{g}$ each of two plasmids (pX459) each expressing Cas9 and a chimaeric guide RNA (gRNA1: CATGTTGAGCATGGAAACCCGG, chrX:72823838–72823860; gRNA2: GGTTATGGCGGTCGGTTCCTGG, chrX:73025513–73025535). Subcloning of ES cells was made by limiting dilution. Cells were treated for 24 h with puromycin. As soon as visible, single colonies were picked under a microscope to be screened for deletion by PCR (forward primer: 5'-CGTAGACGCGGCAGTAGT-3', reverse primer: 5'-ACATAAACTCCTTTTCAGGACCA-3'). To identify the targeted allele, we performed a PCR using primers (forward: 5'-CTGTCCAAATGGAGGTGCTT-3', reverse: 5'-CCTAGTCCGCTCTCTATCG-3') that amplify a 203-bp amplicon specifically on the wild-type allele, which contains a single nucleotide polymorphism (SNP; rs29035891). After amplification, PCR products were gel-purified and sequenced using the forward or reverse primer used for PCR. Clones positive carrying the deletion were expanded and differentiated into NPC as previously described¹⁶ and subcloned by limiting dilution. NPC lines were maintained in N2B27 medium supplemented with EGF and FGF (10 ng ml^{-1} each), on 0.1% gelatin-coated flasks. Clones carrying the boundary deletion on the inactive X were identified by RNA FISH against *Xist* with the p510 plasmid probe and DNA FISH with a BAC hybridizing inside the deleted region (RP23-299L1).

RT-PCR and pyrosequencing. RNA extraction and cDNA synthesis were performed with the Cells-to-Ct kit (Ambion) following the manufacturer's instructions. Pyrosequencing primers were designed using the PyroMark Assay Design software. PCR products were purified and annealed with sequencing primers for pyrosequencing using the PyroMark q24 (Qiagen) (*Xist*: forward primer 5'-AGAGAGCCCAAAGGGACAAA-3', reverse primer 5'-TGTATAGGCTGCTGGCAGTCC-3', sequencing primer 5'-GCTGGCAGTCCCTGA-3'; *Mecp2*: forward primer 5'-CTGAAGGTTGAGTGGCTCATG-3', reverse primer 5'-ATGGTAGCTGGGATGTTAGGG-3', sequencing primer 5'-CAGAGACAAGCCACTGA-3'; *Huwei1*: forward primer 5'-GCAGGTGTCTGCAAATCCA-3', reverse primer 5'-GCCGATGTAAGGCTCCAAG-3', sequencing primer 5'-TGGGTTTCATGTGACAG-3'; *Jarid1c*: forward primer 5'-GCTGCCTCCTTTGCCTGAAAT-3', reverse primer 5'-TTCAGGGGCGCCTTACA-3', sequencing primer 5'-CTCCTTTCCTGAAAT-3').

Hi-C read mapping, binning, ICE correction. Hi-C was performed as previously described^{27,28}. To obtain allele-specific Hi-C interaction maps in female ES cells ($X^{\text{cast}}X^{\text{129}}$) and a derived clonal NPC line ($X^{\text{cast}}X^{\text{129}}$) (Methods; Extended Data Fig. 1)¹⁶, we first constructed an allelic genome using the reference mm9 genome and all 19,722,473 SNPs. The allelic (Cast and 129) genomes were then combined to create a reference diploid genome (consisting of 44 chromosomes; 1–19 X,Y,M). All reads were aligned to the diploid genome (as described in ref. 29), thus allowing for a competitive mapping strategy between the two alleles. All reads were trimmed to 50 bp and then aligned using the novoCraft novoalign (v.3.02.00) software package. Reads were aligned using the following options (-r all 5 -R 30 -q 2 -n 50, minimumReadDistance = 5). The best alignment was selected from the list of the top 5 alignments. The alignment was considered unique (allelic), if its alignment score was ≥ 5 from the second best alignment score (alignment score taken from the ZQ tag). Reads that aligned uniquely to an allele were classified as allelic (either Cast or 129) whereas reads that aligned to both alleles equally (≤ 5 distance) were classified as ambiguous (AMB) (Extended Data Fig. 1d). Uniquely aligned Hi-C interactions between loci located on the same chromosome were assigned to a specific parental chromosome in *cis* when at least one of the two reads contained a diagnostic SNP, and the other either contained a SNP from the same allele, or mapped to both alleles³⁰. We obtained the following paired-end read counts: For ES cells (GUR.2d), a total of 401,684,614 interactions could be aligned combining the two replicates, 372,272,389 of which were unique (after PCR duplicate filter), and 95,650,438 of which could be placed to either the Cast or 129 allele (25.69%). For NPCs

(GEL.72b), a total of 277,440,656 interactions could be aligned, 253,254,798 of which were unique (after PCR duplicate filter), and 82,323,031 of which could be placed to either the Cast or 129 allele (32.51%). For Δ FT NPCs (D9B2/B129T3), a total of 229,331,123 interactions could be aligned, 222,941,525 of which were unique (after PCR duplicate filter), and 85,331,870 of which could be placed to either the Cast or 129 allele (38.28%). The difference in percentage of reads assignable to either allele is probably due to differences in the percentage of *cis* interactions found in each sample (biological or technical variation). The 82–95 million read depth supported generation of allele-specific chromatin interaction maps at multiple resolutions (10 Mb, 2.5 Mb, 1 Mb, 500 kb, 250 kb, 100 kb and 40 kb).

Biological replicates were highly correlated. Pearson's correlation coefficients for 500 kb data on chrX were as follows: EHSNP-mF1216_R1R2_chrX-129S1, 0.992331; EHSNP-mF1216_R1R2_chrX-cast, 0.990373; EHSNP-mNPe-deltaRF_R1R2_chrX-129S1, 0.976562; EHSNP-mNPe-deltaRF_R1R2_chrX-cast, 0.983614; EHSNP-mNPe_R1R2_chrX-129S1, 0.990976; EHSNP-mNPe_R1R2_chrX-cast, 0.995202. Autosomes showed similar correlation values. Overall these numbers indicate that the produced Hi-C data was of high quality and well correlated between biological replicates. We pooled all biological replicates into a single Hi-C data set per sample and subsequently used the pooled data for all analyses.

Iterative mapping and error filtering/iterative correction of the chromatin interaction data were performed as previously described^{29,31}. Iterative correction was performed on the diploid (44 chromosomes) (replicate pooled) genome-wide matrix for all resolutions.

Hi-C for the TXY (male) samples was performed as previously described^{27,28}. Reads originating from the TXY (male) sample was aligned to the mm9 reference genome. Iterative mapping and error filtering/iterative correction of the chromatin interaction data were performed as previously described^{29,31}. Iterative correction was performed genome-wide (22 chromosomes) (replicate pooled) on the genome-wide matrix for all resolutions. Biological replicates were highly correlated. We pooled all biological replicates into a single Hi-C data set per sample and subsequently used the pooled data for all analyses.

We obtained the following paired-end read counts: For TXY +dox, a total of 277,191,448 interactions could be aligned, 267,007,192 of which were unique (after PCR duplicate filter). For TXY WT -dox, a total of 308,671,996 interactions could be aligned, 300,102,244 of which were unique (after PCR duplicate filter). For TXY: Δ A +dox, a total of 281,116,218 interactions could be aligned, 273,612,976 of which were unique (after PCR duplicate filter). For TXY: Δ A -dox, a total of 298,436,664 interactions could be aligned, 289,376,893 of which were unique (after PCR duplicate filter).

Allele-specific read mapping validation (Hi-C and ATAC-seq). To validate the accuracy of the allele-specific read alignment strategies used in this paper, we first constructed a set of validation reads tiled across all SNP locations between the Cast and 129 genomes. In brief, for each SNP location on the X chromosome, all overlapping 50-bp reads were extracted (50 total) for each of the 129 and Cast alleles (Extended Data Fig. 1i). All reads were then processed through the Hi-C and ATAC-seq mapping pipelines described in the methods to measure assignment accuracy. Encoded into each fastq readID, was the allelic genome that each read originated from, the relative-position (within the 50-bp read) of the SNP, the base-call of the SNP (A, C, T, G), the chromosome and position (start, end) of the read. The set of validation reads were then processed through the ATAC-seq allele-specific and Hi-C allele-specific pipelines with no modifications. Each processed read was then scored according to whether or not it was correctly placed to not only the correct chromosomal location, but to the correct allele as well. For the Hi-C allele-specific pipeline we found 0 reads assigned to the incorrect chromosomal location and 0 reads assigned to the incorrect allele. All processed validation reads were assigned to the correct coordinate and allele. Reads that span repetitive regions or are of low complexity were inherently filtered via the Hi-C pipeline and thus would be excluded from both the actual data and from the set of validation reads. For the ATAC-seq allele-specific pipeline, we found that for reads coming from the 129 X chromosome, only 0.09% mapped to the Cast chromosome (only 3 of these fell within ATAC-seq peaks). For reads coming from the Cast X chromosome, only 0.21% mapped to the 129 chromosome (5 of which fell within ATAC-seq peaks). Furthermore, we feel that our analysis may be even more accurate when using paired-end reads as we do for all ATAC-seq data analysis (Extended Data Fig. 1i).

Generation of Xist-positive Hi-C signal for comparison with DNA FISH. *Xist* RNA FISH performed in parallel with Hi-C on the same inducible *Xist* ES cell samples revealed that *Xist* expression (either wild-type or A-repeat mutant) was induced in ~ 35 –45% of dox-treated cells. Hence the Hi-C signal can be represented as the sum of ~ 35 –45% of reads coming from *Xist*-positive X chromosomes,

and ~55–65% of reads that are generated from non-coated X chromosomes. For example,

$$\text{Hi-C(+dox)} = 35\% \times \text{Hi-C}(Xist+) + 65\% \times \text{Hi-C}(Xist-)$$

We took advantage of the fact that the signal from $Xist-$ cells is measured independently in the non-induced ($-dox$) sample:

$$\text{Hi-C}(Xist-) = \text{Hi-C}(-dox)$$

to extract the signal of $Xist$ -positive cells:

$$\text{Hi-C(+dox)} = 35\% \times \text{Hi-C}(Xist+) + 65\% \times \text{Hi-C}(-dox)$$

and hence

$$\text{Hi-C}(Xist+) = [\text{Hi-C(+dox)} - 65\% \times \text{Hi-C}(-dox)] \times (1/0.35).$$

Negative values were assigned to 'not analysed' (NAs). The $\text{Hi-C}(Xist+)$ signal was then used to compare Hi-C data with RNA/DNA FISH experiments in Fig. 4b, which allow to discriminate $Xist$ -coated and non-coated chromosomes visually by the presence of an $Xist$ cloud. In Fig. 4c, $Xist-$ signals correspond to the $-dox$ sample.

Hi-C SNP density filter. To remove potential biases in the Hi-C data related to the density of SNPs in each bin, we calculated the number of SNPs residing in each genomic interval (bin) for all Hi-C bins across all bin sizes. We then calculated the median number of SNPs per bin, and produced a minimum required SNP density cutoff defined as the (median $- 1.5 \times \text{IQR}$). Any bins with less SNPs than the cutoff were removed from all analyses. The SNP density cutoffs used for each bin size were: 40 kb, 43 SNPs; 100 kb, 216 SNPs; 250 kb, 776.5 SNPs; 500 kb, 1,767.25 SNPs. The non-SNP-density-filtered data was only used for visualization purposes (figure heat maps). Throughout the manuscript, we refer to Hi-C as data that has been iteratively corrected³¹ and run through the SNP-density filter.

Compartment analysis. The presence and location of the A/B compartments were calculated as previously described³². Compartments were derived from the 250-kb iteratively corrected Hi-C data for each chromosome separately using the CIS maps for each sample/allele. The code used to generate the compartments (PC1 from PCA analysis) is publicly available on Github (matrix2compartment.pl): <https://github.com/dekkerlab/giorgetti-nature-2016>. Compartments were generated all default options except the (cis alpha) option, set to (-ca 0.005).

Insulation and boundary calculation. TAD structure (insulation/boundaries) was defined via the insulation method as previously described with minor modifications³². The code used to calculate the insulation score is publicly available on Github (matrix2insulation.pl): <https://github.com/dekkerlab/giorgetti-nature-2016>. Insulation vectors were detected using the following options: (-is 480000 -ids 320000 -im iqrMean -nt 0 -ss 160000 -yb 1.5 -nt 0 -bmoe 0). The output of the insulation script is a vector of insulation scores, and a list of minima along the insulation vector (inferred as TAD boundaries). The TAD boundaries were not used in this study.

Interaction pile-up maps. Interaction pile-up maps were constructed from all pairwise interactions between either the list of 87 wild-type NPC Xi chromosome escapees or the 29 ΔFT NPC Xi chromosome escapees. Using the 40 kb Hi-C data, a 2-Mb window centred around each pairwise interaction (pixel) was taken (25 bins in each direction, yielding 51×51 sub-matrix). Any resulting sub-megabase that overlapped the ($y=x$) diagonal in the matrix was excluded from the analysis (effectively excluding all interactions < 2 Mb). All sub-matrices were then averaged to produce the final (mean) pile-up map. A strong signal at the centre suggests that the elements used tend to contact one another in 3D space.

The Xi chromosome is as accessible and detectable in Hi-C as the Xa chromosome and autosomes. The number of RAW reads observed for both the Xa and Xi chromosomes were very similar for all chromosomes, thus demonstrating that the Xi chromosome is not simply less accessible/visible to the Hi-C methodology. ES-cell-chrX-129S1, 1,118,327; ES-cell-chrX-Cast, 1,104,709; NPC-chrX-129S1, 1,147,072; NPC-chrX-Cast, 1,148,128; ΔFT NPC-chrX-129S1, 1,314,476; ΔFT NPC-chrX-Cast, 1,288,802. Bias in read directional due to partial digestion is typically observed up to ~10 kb. For interactions between fragments separated by over 10 kb this bias is negligible, indicating at least one digestion occurring between them in every cell. This genomic distance is therefore a measure for digestion efficiency²⁷. For both the Xa and Xi chromosomes, this genomic distance is ~6–10 kb, indicating that digestion efficiency of chromatin on the Xa and Xi chromosomes are comparable. Thus, the unique conformation of the Xi chromosome does not affect Hi-C analysis, as was also found for condensed mitotic chromosomes²⁸.

Correlation analysis of Hi-C matrices. We compared X chromosome interaction matrices of $Xist$ -inducible lines (pre/post- $Xist$ induction in wild-type and A-repeat mutant samples) with that of the NPC Xi chromosome, at 500-kb resolution. As

distance-dependent decay of interaction frequency causes all interaction matrices to be highly correlated, we first compensated for this effect by multiplying the read count in each bin by its respective genomic distance. We then calculated the Spearman correlation between each pair of matrices. After wild-type $Xist$ induction in TXY cells, the interaction map becomes more similar to that of the NPC Xi chromosome ($\rho = 0.17$ to 0.31), while no change is observed in the repeat-A mutant ($\rho = 0.17$ to 0.17).

RNA and 3D-DNA FISH. FISH was performed as previously described³³. ES cells and NPCs were cultured on gelatin-coated coverslips #1.5 (1 mm) and fixed in 3% paraformaldehyde for 10 min at room temperature. Cells were permeabilized on ice for 5 min in $1 \times$ PBS, 0.5% Triton X-100 and 2 mM vanadyl-ribonucleoside complex (VRC, New England Biolabs), and coverslips were stored in 70% ethanol at -20°C . Before FISH, samples were dehydrated through an ethanol series (80%, 95%, 100% twice) and air-dried briefly. For RNA FISH, cells were directly hybridized with denatured probes. For DNA FISH, samples were first denatured in 50% formamide/ $2 \times$ SSC (pH = 7.3) at 80°C for 37 (ES cell) and 35 (NPC) min, immediately placed on ice and washed twice with ice-cold $2 \times$ SSC. After overnight hybridization at 37°C for RNA FISH or 42°C for DNA FISH, coverslips were washed at 42°C for RNA or 45°C for DNA, three times for 5 min in 50% formamide/ $2 \times$ SSC at pH 7.3, and three times for 5 min in $2 \times$ SSC. Nuclei were counterstained with 0.2 mg ml^{-1} DAPI (2 mg ml^{-1} for structured illumination microscopy), further washed twice for 5 min in $2 \times$ SSC at room temperature and finally mounted with 90% glycerol, $0.1 \times$ PBS, 0.1% *p*-phenylenediamine at pH9 (Sigma).

RNA FISH probes. We used the p510 plasmid coupled with Cy5 to detect $Xist$. For RNA FISH on escape genes, we used the following BAC and fosmid probes: RP23-436K3, RP23-328M22, RP24-436K3, WI1-1269O10 (*Mecp2*), RP24-157H12 (*Huwei1*), RP23-13D21 (*G6pdx*), RP24-148H21 (*Jarid1c*).

DNA FISH probes. In experiments to detect the mega-domain boundary, fluorescent oligonucleotides (average length 45 bp, 5'-modified with Atto 448 or Atto 550, average density: one oligonucleotide every 3 kb) were obtained from MYcroarray Inc. Oligonucleotides were designed to tile the following consecutive 18-Mb regions: chrX:35,000,000–53,000,000, chrX:53,000,000–72,000,000, and chrX:72,000,000–90,000,000. To detect the *DXZ4* region we used the RP23-299L1 BAC.

Imaging and quantification of 3D DNA FISH. Three-dimensional image stacks (200 nm distance between consecutive *xy* planes) were acquired on a DeltaVision Core wide-field microscope (Applied Precision) equipped with a CoolSNAP HQ2 camera operated at 2X binning, and a $100 \times$ PlanApo oil immersion objective (the effective pixel size was $129 \times 129 \text{ nm}$). Xi chromosome signals were identified via the presence of an $Xist$ mRNA cloud in the far-red channel (p510-Cy5 probe). Pearson correlation between red and green signals was calculated using custom-made ImageJ macros as follows. After subtracting the background from each *xy* plane (generated by morphological opening the image with a circle of 5 pixels in radius), Pearson correlation between red and green pixel intensities was measured inside a fixed-size region of $40 \times 40 \times 20$ pixels ($5.16 \times 5.16 \times 4 \mu\text{m}^3$) centred on each FISH signal. The significance of Xi versus Xa chromosome differences in correlation was assessed by Wilcoxon's rank sum test. Random nuclear positions were used to estimate the background correlation that could be observed due to non-specific probe hybridization.

The gyration tensor of a greyscale image is defined as

$$S_{ab} = \sum_k I_k (r_a^k - r_a^{CM})(r_b^k - r_b^{CM}) / \sum_k I_k,$$

where k is an index running over voxels, I_k is the greyscale intensity of voxel k , and r_a^k and r_b^{CM} are the a -th components (x , y , or z) of the xyz position of voxel k , and of the centre of mass of the image, respectively. The gyration tensor was valuated in a region of interest of $3.8 \times 3.8 \times 4 \mu\text{m}^3$ centred on each FISH signal and the gyration radius was calculated as

$$R_g = \sqrt{\lambda_1 + \lambda_2 + \lambda_3}$$

where $\lambda_{1,2,3}$ are the eigenvalues of S_{ab} .

RNA-seq. RNA-seq data for the ES cell (GUR.2d) and NPC (GEL.72b) was obtained from previously published work (PMID 24576422)¹⁶. RNA-seq data for the mutant NPC (D9B2/B129T3) was obtained and processed as previously described¹⁶.

RNA-seq 'expressed/escapee' classification. The allelic RPKM values were derived for each gene by splitting the RPKM value by the 129 ratio. $129 \text{ RPKM} = (\text{RPKM} \times 129 \text{ ratio})$; $\text{Cast RPKM} = (\text{RPKM} \times (1 - 129 \text{ ratio}))$. Any gene with an allelic RPKM value ≥ 3 RPKM was classified as being expressed. Any gene expressed on the Xi chromosome was classified as being an escapee.

ATAC-seq. ATAC-seq library preparation was performed exactly as previously described³⁴. Sequencing was carried out on an Illumina NextSeq 500 generating 2×75 bp paired-end reads. Libraries were sequenced to a depth of 25–35 million reads per sample. Reads were trimmed using CutAdapt and aligned using Bowtie2.

Reads were aligned to a custom 129/CastEiJ genome in which SNP sites were replaced by 'N'. Approximately 52–58% of reads per line contained 'N's and were assigned to the 129 or Cast allele based on the identity of the base at that location. Reads containing non-concordant SNPs were rare and were discarded. Reads not containing SNP sites were included in overall peaks but not were excluded from allele-specific tracks. ATAC-seq peaks were called using MACS2 with no shifting model. For TXY *Xist*-inducible male cell lines, which contain only one X chromosome, peaks were called after normalizing all samples by read count on the autosomes. The set of X chromosome peaks was defined as the merge of peaks from all samples, and bedtools was used to calculate coverage within peaks.

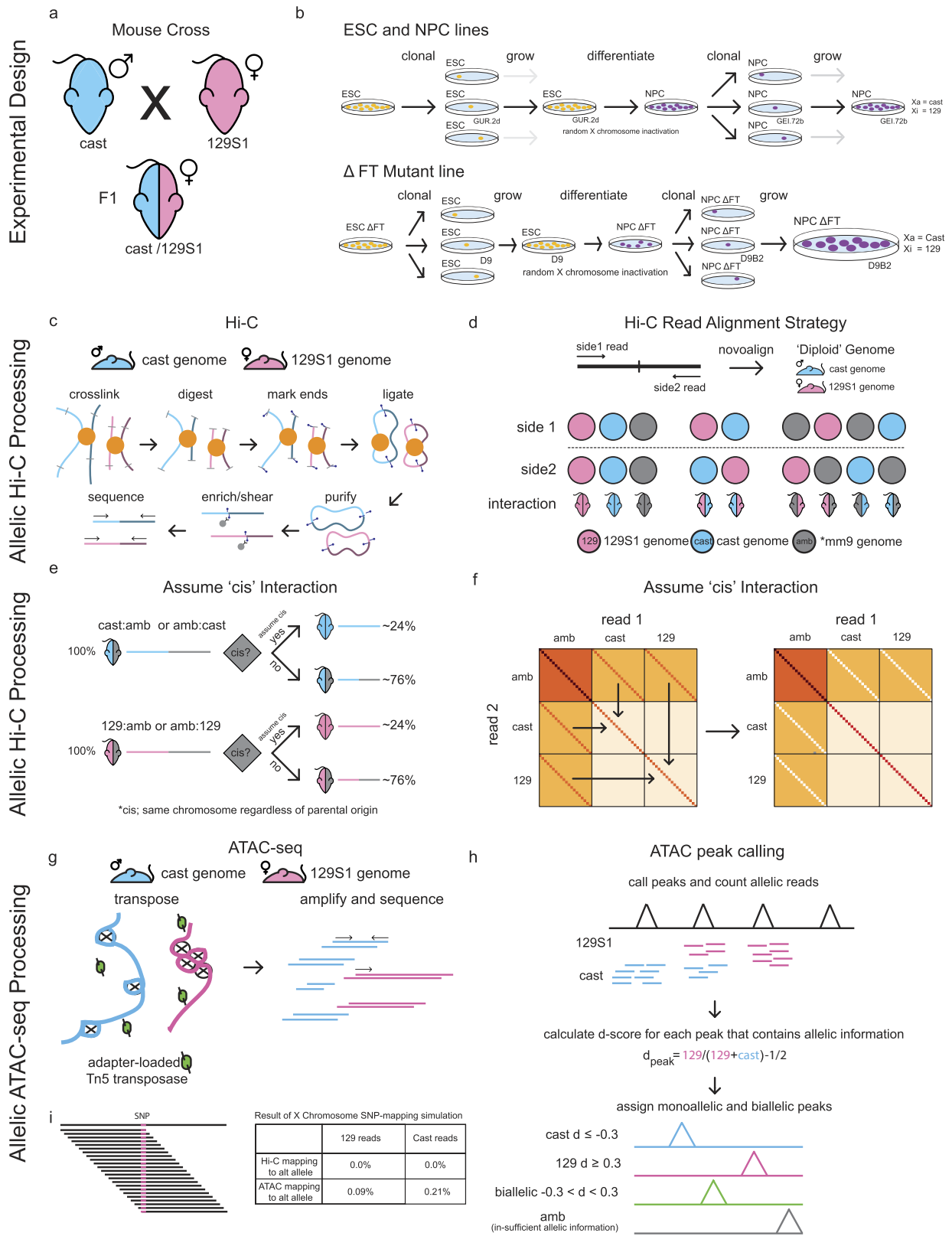
Assigning allele-specific ATAC-seq peaks. For each ATAC-seq peak, all N-containing reads were counted and assigned to 129 or Cast alleles based on SNP at the N-containing position. To assign monoallelic and biallelically accessible peaks, a d-score was calculated as a measure of allelic imbalance³⁵. In brief, for a given peak the d-score was calculated as the ratio of 129 reads to total number of reads $- 1/2$. A peak with a d-score ≥ 0.3 was assigned as a 129-specific peak. A peak with a d-score ≤ -0.3 was assigned as a Cast-specific peak. Any peak with a d-score > -0.3 was assigned as a peak in 129 (monoallelic or biallelic). Any peak with a d-score $< +0.3$ was assigned as a peak in Cast (monoallelic or biallelic).

Annotating ATAC-seq peaks using ChIP-seq data. ATAC-seq peaks were annotated using existing published ChIP-seq data sets. CTCF ChIP-seq came from whole female mouse brain³⁶. Called CTCF binding sites were used and extended ± 300 bp before overlapping with ATAC-seq peaks. H3K27ac and p300 ChIP-seq are from mouse NPCs³⁷. For H3K27ac and p300 ChIP-seq data, peaks were called using MACS2 and then overlapped with ATAC-seq peak locations.

Integrating Hi-C, ATAC-seq and RNA-seq data. Integrative analysis of Hi-C insulation (TAD structure), ATAC-seq counts, and RNA-seq RPKM was performed as follows. A promoter region was defined for each gene as ± 500 bp from the TSS. ATAC peaks were assigned to a gene if they overlapped with the promoter region. In the event that >1 ATAC peak overlapped with the promoter, the closer

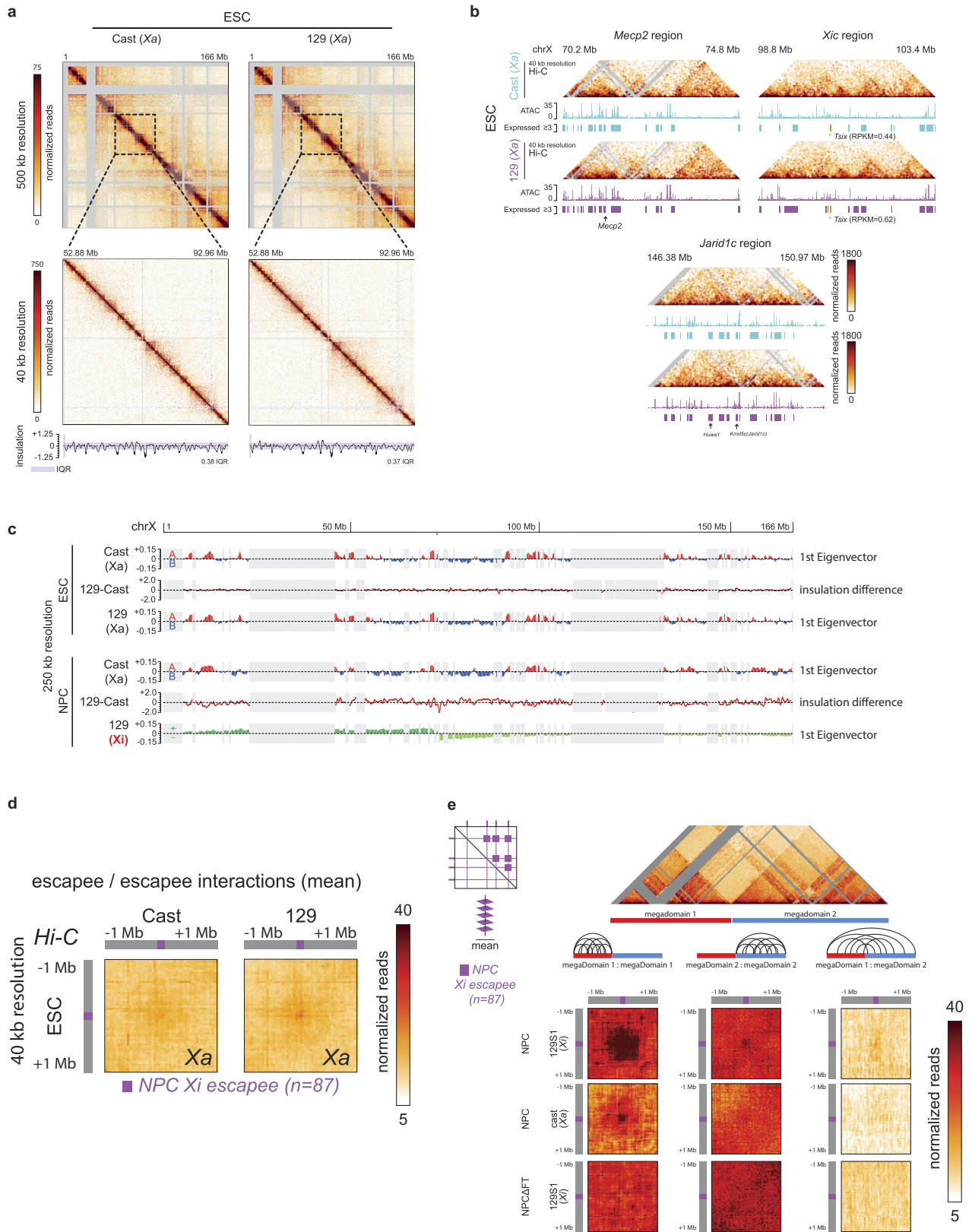
ATAC peak was chosen. An ATAC count of 0 was assigned to each promoter, if it did not contain an ATAC peak. If the ATAC allelic counts overlapping the promoter were < 10 , then the ATAC count was set to 'NA'. The 40-kb bin overlapping the promoter region was used to display the insulation and insulation-difference value.

27. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
28. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).
29. Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* **72**, 65–75 (2015).
30. Selvaraj, S., R Dixon, J., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* **31**, 1111–1118 (2013).
31. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
32. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
33. Chaumeil, J., Augui, S., Chow, J. C. & Heard, E. Combined immunofluorescence, RNA fluorescent *in situ* hybridization, and DNA fluorescent *in situ* hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. *The Nucleus* **463**, 297–308 (2008).
34. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
35. Eckersley-Maslin, M. A. *et al.* Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* **28**, 351–365 (2014).
36. Berletch, J. B. *et al.* Escape from X inactivation varies in mouse tissues. *PLoS Genet.* **11**, e1005079 (2015).
37. Martynoga, B. *et al.* Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev.* **27**, 1769–1786 (2013).
38. McHugh, C. A. *et al.* The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).



Extended Data Figure 1 | See next page for caption.

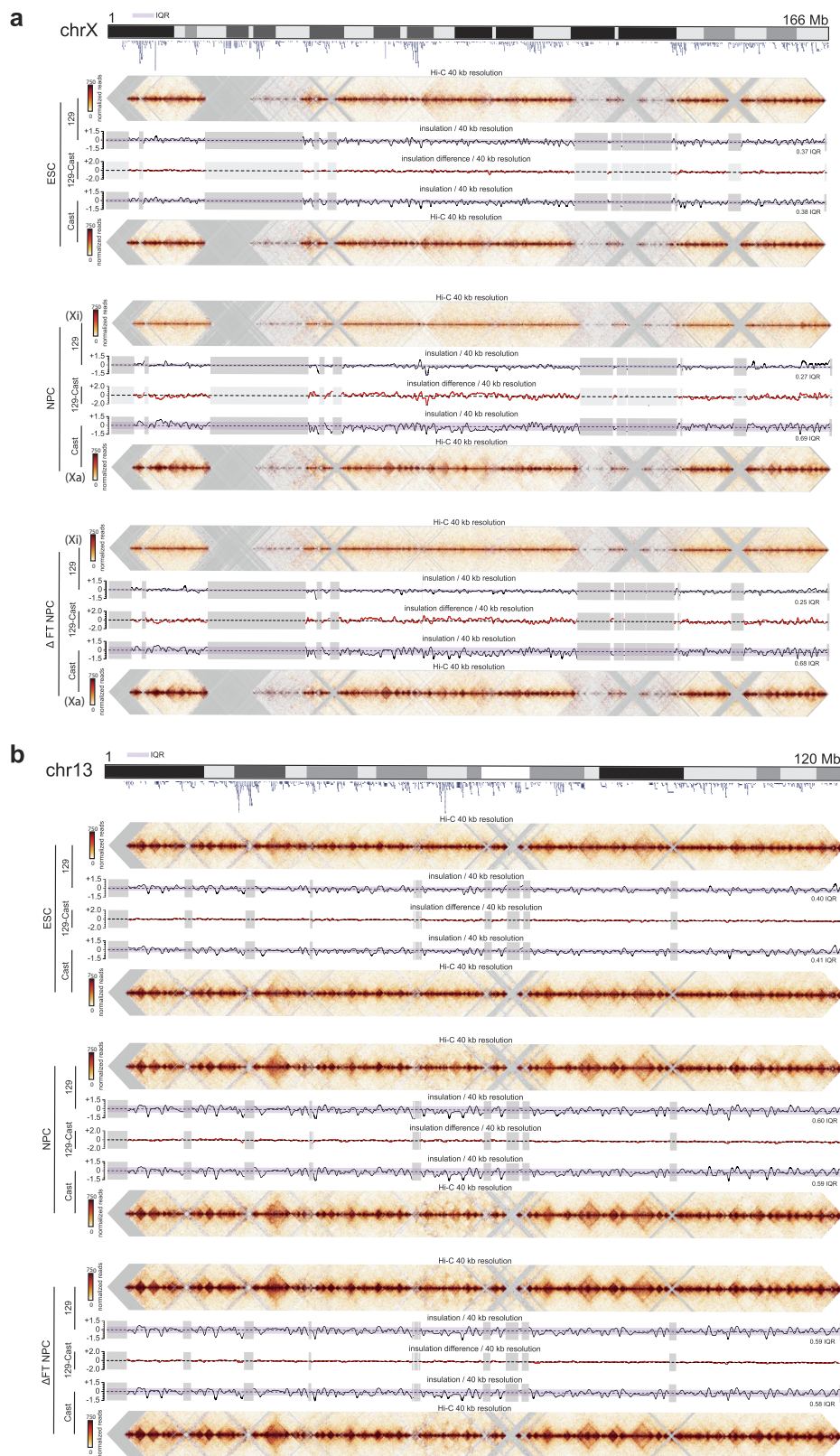
Extended Data Figure 1 | Description of allele-specific Hi-C and**ATAC-seq. a**, Schematic of hybrid mouse strains used for all experiments.**b**, Top, scheme outlining differentiation of ES cells to NPCs and picking of clones. Bottom, scheme outlining CRISPR deletion of the mega-domain boundary in ES cells, differentiation to NPCs and the picking of clones.**c**, Schematic of Hi-C library generation. **d**, Schematic of the Hi-C alignment strategy. Paired-end reads are aligned to a 'diploid' genome consisting of 22 chromosomes from Cast, and 22 chromosomes from 129 (1–19 X, Y, M). The interaction row shows all possible paired-end read combinations between the 129, Cast and ambiguous (AMB) genomes.**e**, Schematic showing the re-assignment of a subset of 'cis' interactions. Paired-end reads in which one side uniquely aligned to an allele and the other side aligned equally to both alleles (AMB), were re-classifiedas allelic reads, only if both reads aligned to the same chromosome (*cis*). **f**, Cartoon explaining the re-assignment of 129:amb or cast:amb *cis* interactions. **g**, Scheme for ATAC-seq library preparation. Cells are lysed followed by incubation with adaptor-loaded hyperactive Tn5 transposase. The transposase integrates into accessible DNA, and these fragments are then directly amplified and sequenced. **h**, Scheme for allele-specific ATAC-seq data analysis. **i**, SNP-mapping simulation. For each SNP location on the X chromosome, all overlapping 50 bp reads were extracted (50 total) for each of the 129 and Cast alleles. All reads were then processed through the Hi-C and ATAC-seq mapping pipelines described in the Methods to measure assignment accuracy. Results are shown in the table.



Extended Data Figure 2 | See next page for caption.

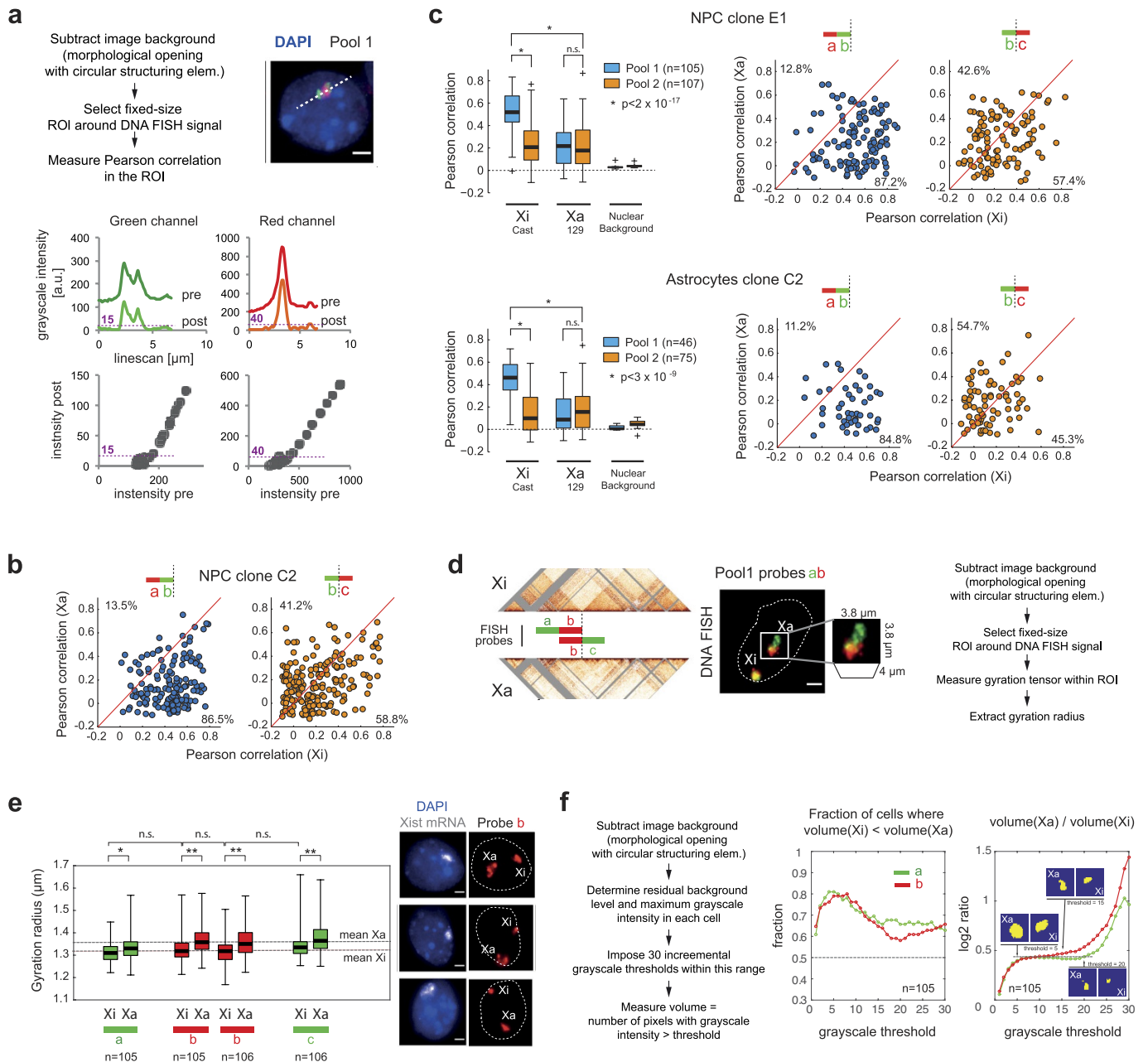
Extended Data Figure 2 | Structure of the Xa chromosome in ES cells, compartment analysis and characterization of interactions between genes that escape XCI. **a**, Allele-specific Hi-C contact maps for X chromosome in ES cells at 500-kb resolution (top), and for a ~40-Mb region centred around the *DXZ4*-containing locus at 40-kb resolution (bottom). The insulation score is plotted at the bottom of each 40-kb heat map as in Fig. 1a. Purple shaded areas indicate the IQR of insulation scores along the chromosome. **b**, Zoomed in view of three regions showing Hi-C interactions, RNA-seq and ATAC-seq signal on the Cast and 129 Xa chromosomes in ES cells. **c**, Compartment profiles of X chromosome in ES cells and NPCs. The first eigenvector (PC1) of each allele-specific Hi-C contact map, obtained with principal component analysis, is shown, together with the difference in chromosome-wide insulation score between the 129 and Cast allele. A/B-compartments are evident in

ES cells and NPCs along both Xa (red and blue signal), whereas the first eigenvector corresponds to the two mega-domains for the Xi chromosome in NPCs. In ES cells, both Xa chromosomes display comparable insulation profiles (difference is close to 0 along the chromosome), whereas in NPCs large differences are observed (difference in insulation fluctuates along the chromosome). Grey areas indicate regions with low SNP density that were excluded from analysis. **d**, Same plots as in Fig. 2e for the same set of genes that escape XCI in NPCs, but using Hi-C data obtained in ES cells. **e**, As in **d**, where the interactions between loci that escape XCI on the wild-type Xi chromosome were divided in three different groups, according to whether pairs of loci belong to the same or different mega-domains. Interactions across the mega-domain boundary only occur on the wild-type Xi chromosome and are lost on the Δ FT Xi chromosome concomitant with loss of transcription of constitutive escapees.



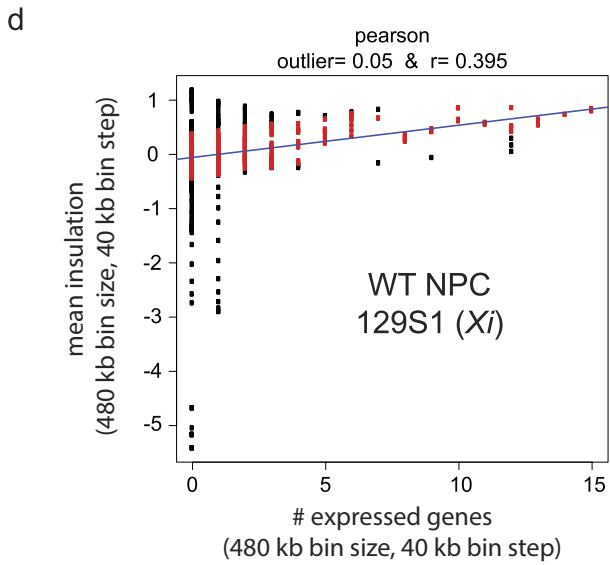
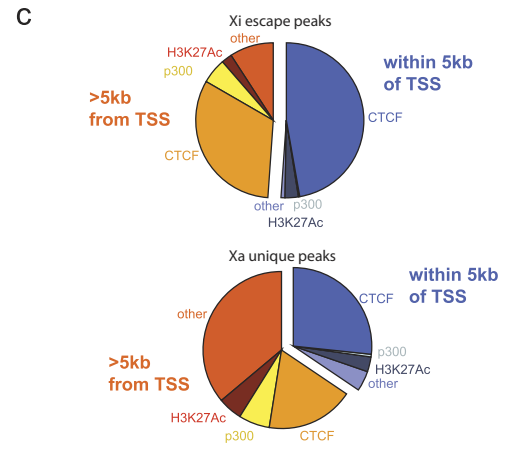
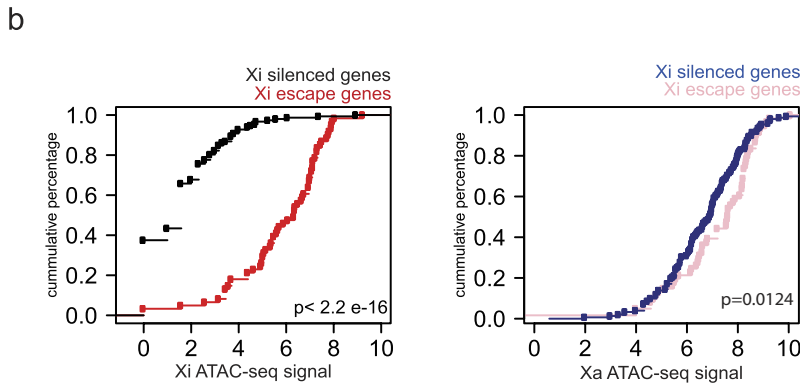
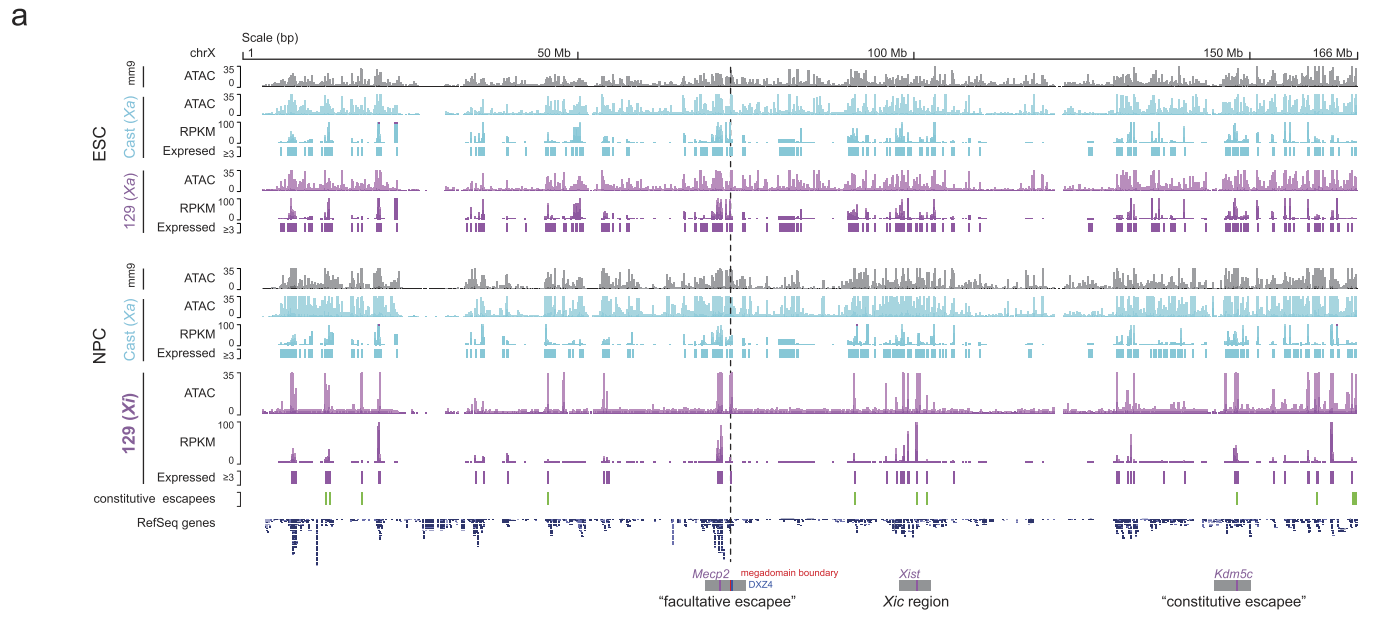
Extended Data Figure 3 | Comparison of allele-specific Hi-C maps of chromosomes X and 13 in ES cells, NPCs and boundary-mutant Δ FT NPCs. a. Hi-C data, insulation scores, and the difference in insulation scores (129 Cast) are shown for ES cells, NPCs and mutant Δ FT NPCs for both alleles (Cast and 129) for the X chromosome. Large dips in the insulation vector are found at TAD boundaries. Peaks in the insulation

vector are found towards the centre of each TAD. The insulation difference plot highlights areas of differential TAD structure between the alleles (many differences along the X chromosome in NPCs as compared to the allelic differences along autosomes, see b). b. Same as a for chromosome 13. The insulation difference plot highlights areas of differential TAD structure between the alleles (rare).



Extended Data Figure 4 | Quantitative analysis of 3D DNA FISH on the Xa and Xi chromosomes. **a**, Top left, scheme of the procedure used to quantify the Pearson correlation. A background is generated for each xy plane in a three-dimensional z -stack by morphological opening the image with a circle of 5 pixels in radius, and subtracted from it. Pearson correlation between red and green pixel intensities is measured inside a fixed-size region of $40 \times 40 \times 20$ pixels ($5.16 \times 5.16 \times 4 \mu\text{m}$) centred on each FISH signal. To demonstrate that background subtraction does not affect the measured correlations, we show here a line-scan of $10 \mu\text{m}$ across a typical DNA FISH signal (top right). The shape of the signals along the line scan, as well as their relative intensities, is not affected by background subtraction (bottom). Pre- and post- refer to line scans taken on images before and after background subtraction, respectively. **b**, In more than 80% of nuclei in NPCs, Pearson correlations are higher on the Xi than on the Xa chromosome. Shown is NPC clone C2 (the same where Hi-C was performed). **c**, Same quantification as in Fig. 1c (and b) for an independent NPC clone (E1) where the active X is on the 129 allele and the inactive X chromosome on the Cast, and in astrocytes derived from NPC clone C2. n denotes the number of cells analysed in DNA FISH. Centre lines represent medians; crosses denote experimental points that were considered as outliers (larger than $(q3 + 1.5 \times (q3 - q1))$ or smaller than $q1 - w(q3 - q1)$, where $q1$ and $q3$ are the twenty-fifth and seventy-fifth percentiles of the data distribution, respectively). **d**, Scheme of the gyration tensor based analysis

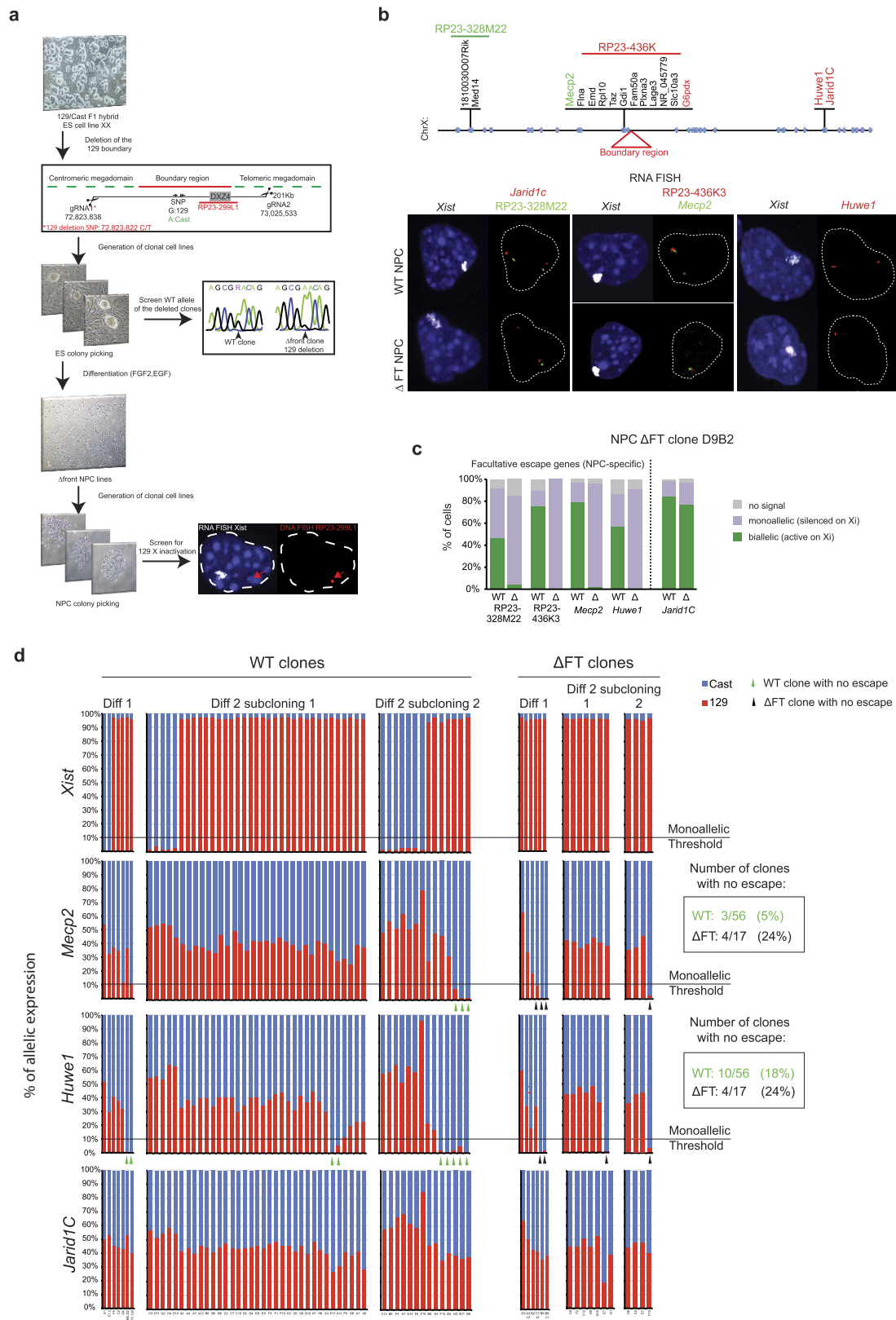
of FISH volumes (see Methods). **e**, Left, gyration radii of DNA FISH signals from probes a, b and c. Probe b was used in combination with both probes a and c separately in two independent experiments. Statistical significance was assessed by Wilcoxon's rank sum test ($*P < 0.05$, $**P < 1 \times 10^{-5}$). The mean gyration radii for Xa and Xi chromosome signals are indicated by dotted lines as a guide for the eye. On the Xi chromosome, the mean gyration radius is approximately 6% smaller than on the Xa chromosome corresponding to approximately 20% smaller volume. Right, representative images of probe a, showing smaller size and increased roundness of the Xi chromosome signals. **f**, Left, scheme of the thresholding-based method for volume quantification. Thirty increasing threshold levels were imposed, starting from the residual grayscale background level surrounding the signal, up to the minimum between the red and green channel grayscale maxima. For each of these thresholds we determined the number of voxels in each channel, where the grayscale intensity was higher than the threshold. Centre, the fraction of cells where the Xa chromosome signal is larger than the Xi chromosome is between 60% and 80% in the entire threshold range. Right, in a wide range of thresholds, the volume of Xa chromosome signals is approximately 25% bigger than Xi chromosome signals. Results are shown here for probes a and b; the same holds for probe c (not shown). n denotes the number of cells analysed in DNA FISH. Centre lines: medians. Boxes: middle 50% of data points. Two biological replicates were analysed for each experiment.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Integrative analysis of allele-specific Hi-C, RNA-seq and ATAC-seq. **a**, X-chromosome-wide ATAC-seq and RNA-seq in ES cells and NPCs. ATAC shows signal for ambiguous, 129- and Cast-specific reads in ES cells and NPCs. RNA-seq shows total signal as well as expressed gene calls. ATAC-seq shows global loss of chromatin accessibility and expression on the Xi chromosome, except at specific locations that mostly overlap with escape genes. Dotted line denotes mega-domain boundary. Position of constitutive escapees was adapted from ref. 3. **b**, Escape genes on the Xi chromosome (as determined by RNA-seq) fall within regions with high ATAC-seq signal (Kolmogorov-Smirnov test $P < 2.2 \times 10^{-16}$). **c**, Pie charts showing the distribution of peaks that escape

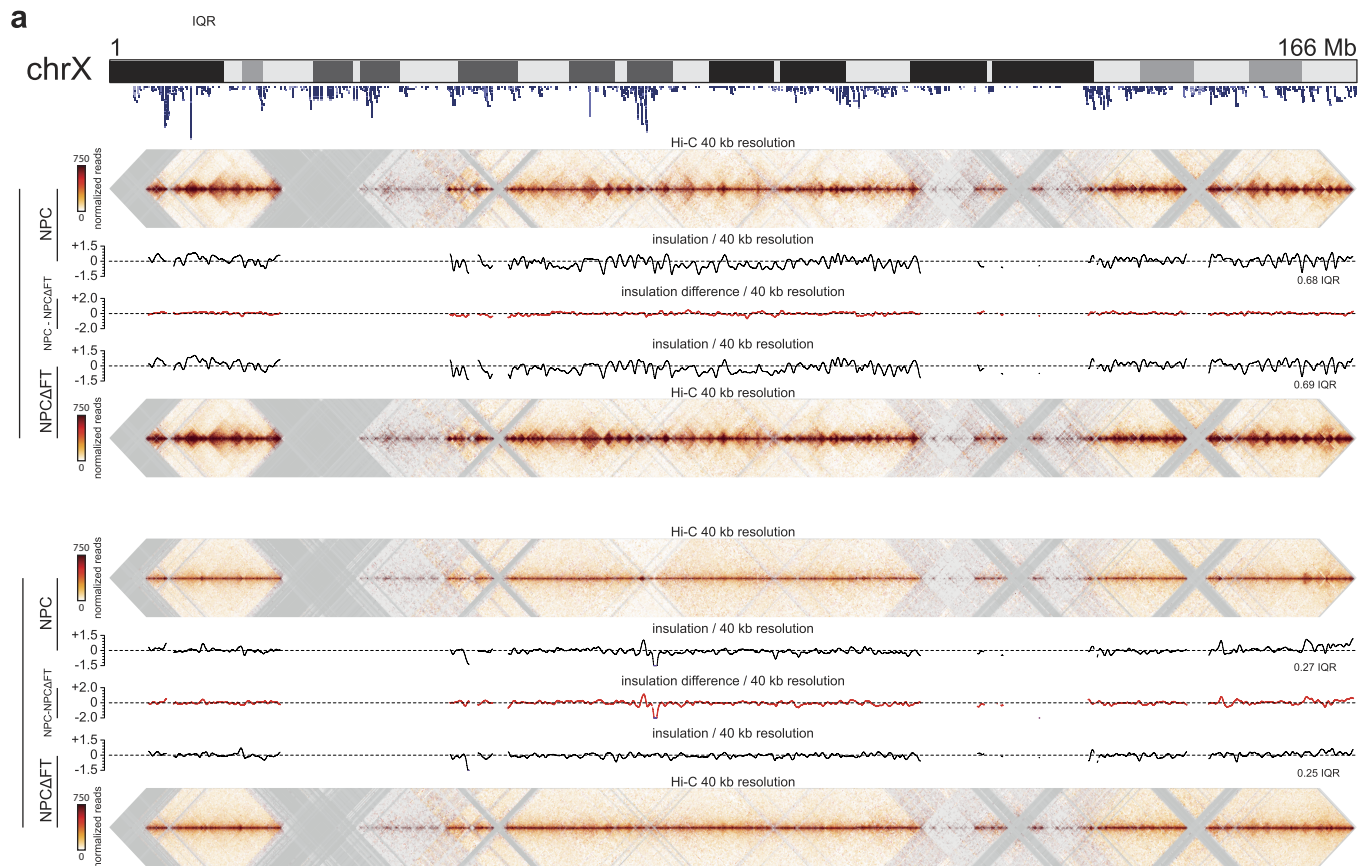
XCI versus the peaks that are unique to the Xa chromosome. Peaks are classified into those that are promoter-proximal (within 5 kb of TSS) and distal (>5 kb from TSS). Annotations are based on binding sites identified by ChIP-seq^{37,38}. **d**, Scatter plot showing the relationship between TAD signal (insulation score) and number of expressed genes of the wild-type NPC 129 (Xi) Hi-C data set. The y axis shows the mean insulation score in a 480-kb window with a 40-kb step size. The x axis shows the number of expressed genes in a 480-kb window with a 40-kb step size. The Pearson R value (0.395) is shown above and is calculated on only the middle 90% (red points) of the residuals. Outliers are shown in black (bottom 5% and top 5% of residuals)



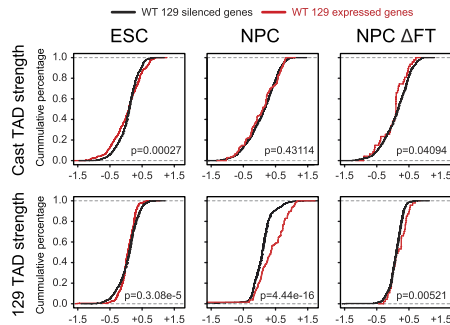
Extended Data Figure 6 | Analysis of wild-type and boundary-mutant Δ FT NPCs by FISH and RT-PCR followed by pyrosequencing.

a, Scheme of the strategy used to delete the mega-domain boundary region in ES cells and to derive Δ FT NPCs. **b**, RNA FISH against constitutive and facultative escapees confirms RNA-seq and ATAC-seq results in the Δ FT NPC clone D9B2. Top, the positions of BAC probes (RP23-328M22 and RP23-436K) are shown relative to the escape genes that they span. Coloured gene names correspond to transcripts that were detected with specific fosmid probes. Bottom, sample RNA FISH images showing that

expression of facultative (*Mecp2* and BAC probes) but not constitutive (*Jarid1c*) escapees is lost on the Δ FT Xi chromosome. **c**, Quantification of the RNA FISH experiment in **d**. **d**, RT-PCR followed by pyrosequencing of *Xist*, two facultative escape genes (*Huwe1* and *Mecp2*) and the constitutive escapee *Jarid1c* indicate the percentage of expression from the 129 allele (red) or the Cast allele (blue) in 56 wild-type and 17 Δ FT NPC clones. Absence of *Mecp2* and *Huwe1* escape is observed in a moderately higher proportion of Δ FT than wild-type clones. Absence of *Jarid1c* escape is never observed.



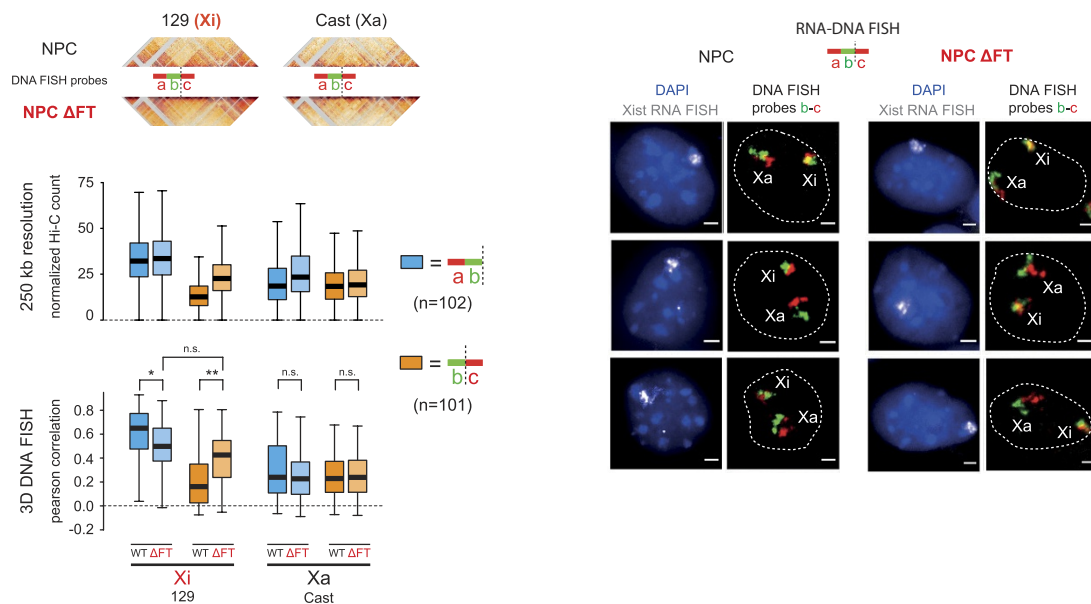
b



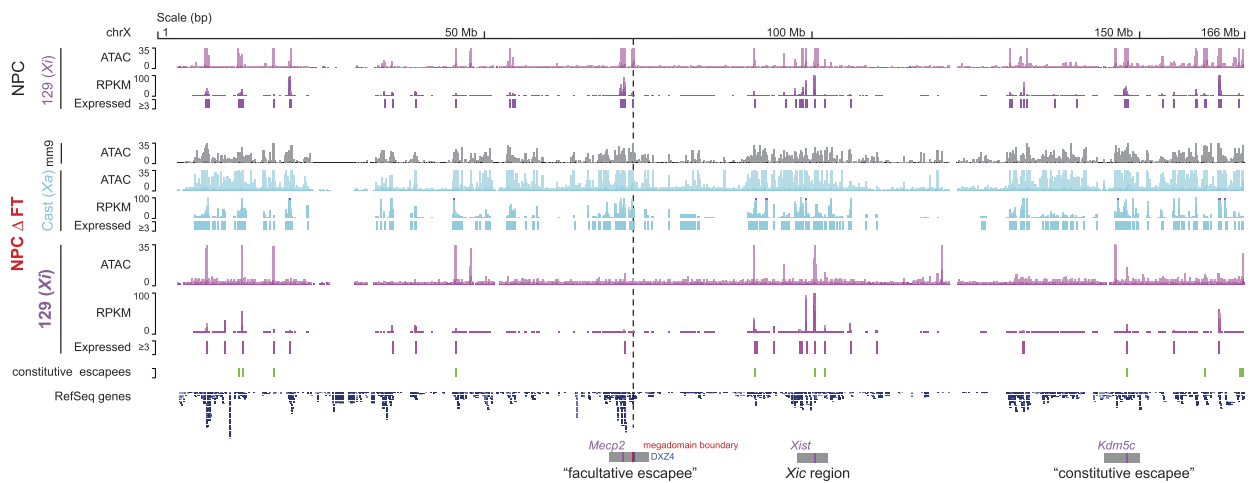
Extended Data Figure 7 | Allele-specific Hi-C analysis of wild-type and boundary-mutant Δ FT NPCs. **a**, Hi-C data, insulation scores, and the difference in insulation scores are shown to compare the wild-type Xi chromosome (NPC 129) and the Δ FT Xi chromosome (NPC 129). Top, the Cast allele (Xa chromosome) for both samples. Bottom, the 129 allele (Xi chromosome) for both samples. Large dips in the insulation vector are indicative of TAD boundaries. Peaks in the insulation vector are found

towards the centre of each TAD. The insulation difference plot highlights areas of differential TAD structure between the wild-type and Δ FT NPCs. **b**, Cumulative plots of TAD strength at the wild-type expressed versus the wild-type silenced genes on Cast and 129 chromosomes, for ES cells, wild-type and Δ FT NPCs. Escapee genes on the Xi chromosome (NPC 129, Δ FT NPC 129) show higher insulation scores as compared to silenced genes.

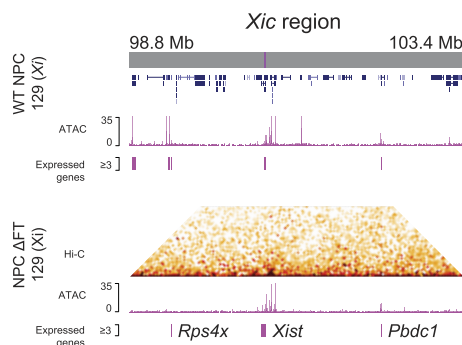
a



b

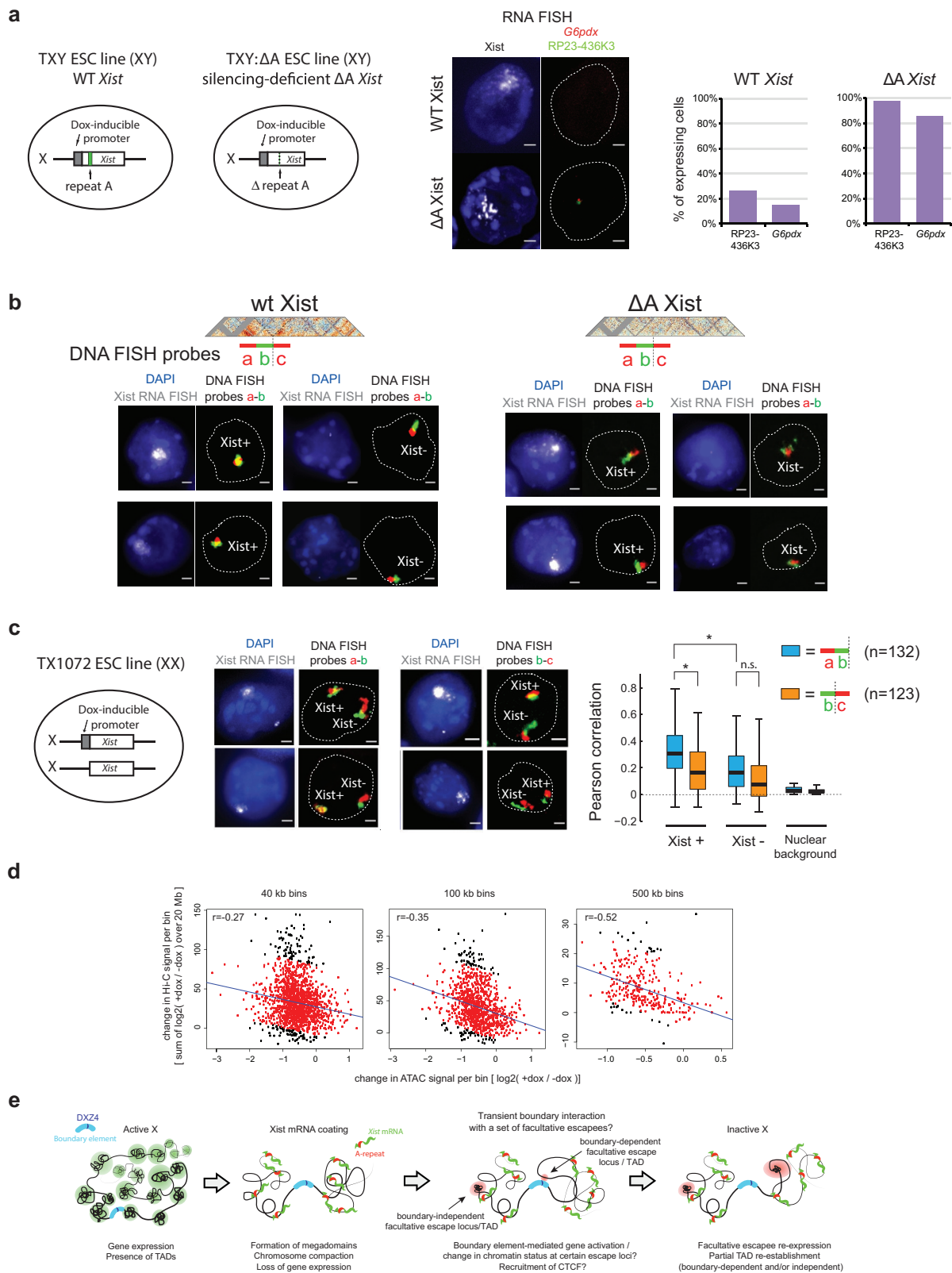


c



Extended Data Figure 8 | 3D DNA FISH and integrative ATAC-seq/RNA-seq analysis in Δ FT NPCs. **a**, Top left, scheme of the DNA FISH probe sets (a-b: inside the same mega-domain, b-c: across the boundary). Bottom, loci detected by probe set b-c are more interacting in the Δ FT than in the wild-type Xi chromosome both in Hi-C (left) and in 3D-DNA FISH (right), showing loss of mega-domain boundary. $*P < 2 \times 10^{-4}$, $**P < 1 \times 10^{-5}$ (Wilcoxon's rank sum test corrected with Bonferroni for multiple hypothesis testing). Right, sample RNA/DNA FISH images showing that signals from probe set b-c are more overlapping on the Δ FT Xi than on the wild-type Xi chromosome. Two biological replicates were

analysed. **b**, Chromosome-wide ATAC-seq and RNA-seq signal generated with ambiguous (mm9), 129- and Cast-specific reads in wild-type NPCs and Δ FT NPCs, showing global loss of chromatin accessibility on the Δ FT Xi chromosome except at the X-inactivation centre (Xic) region and constitutive escape genes. Dotted line indicates the position of the mega-domain boundary. **c**, Zoomed-in view of a region on the Δ FT Xi chromosome encompassing the Xic showing Hi-C interactions, RNA-seq and ATAC-seq signal. ATAC-seq from wild-type NPCs is included for reference (previously shown in Fig. 1).



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | 3D DNA FISH and integrative Hi-C/ATAC-seq analysis of inducible *Xist* ES cell lines. **a**, Left, schematic representation of TXY and TXY:ΔA male cell lines, carrying a tetracycline-inducible wild-type and A-repeat mutant *Xist*, respectively, at the endogenous *Xist* locus. Centre right, RNA FISH against *G6pdx* and a group of genes recognized by the RP23-436K BAC (see Extended Data Fig. 6b) showing that expression of X-chromosome-linked genes is lost upon induction of wild-type but not A-repeat mutant *Xist* in male ES cells. TXY and TXY:ΔA were treated with doxycycline for 48 h. **b**, Representative images from RNA/DNA FISH experiments in TXY and TXY:ΔA male ES cell lines, showing increased overlap of probes a–b on the *Xist*-coated X chromosome after 48 h of wild-type, but not ΔA *Xist* induction. **c**, Left, schematic representation of TX1072 female ES cells in which *Xist* expression can be induced via a tetracycline-responsive promoter at one of the two endogenous *Xist* loci. Center right, RNA/DNA FISH in TX1072 cells treated for 3 days with doxycycline. Probes a–b overlap more on the *Xist*-coated than on the wild-type X chromosome, whereas signals from b–c show lower overlap and partitioning of the *Xist*-coated chromosome

into two separate domains. $*P < 1 \times 10^{-7}$ (Wilcoxon's rank sum test corrected with Bonferroni for multiple hypothesis testing). n denotes the number of cells analysed in DNA FISH. Centre lines: medians. Boxes: middle 50% of data points. Two biological replicates were analysed. **d**, Correlation between changes in Hi-C interactions and ATAC-seq signal following 48-h wild-type *Xist* induction in male ES cells (TXY cell line). $\log_2(+\text{dox}/-\text{dox})$ was calculated for bins of 40 kb (left), 100 kb (middle), and 500 kb (right). **e**, Model of mega-domain boundary-mediated control of chromosome folding and facultative escape. *Xist* coating causes gene silencing and initiates chromosome-wide conformational changes in a manner dependent on its A-repeat region, which result in formation of mega-domains, overall compaction of chromosome folding, and loss of TADs. During differentiation, transient interactions with the mega-domain boundary may occur and may contribute to facultative escape and re-establishment of TADs at escapee loci, although the presence of residual TADs at escapee loci on the Xi chromosome is not necessarily due to the mega-domain boundary.