Supplementary Information

COLA Protocol

Lysis and Restriction Digest

1) Resuspend 10M formaldehyde-crosslinked cells in 500µl of ice-cold Hi-C lysis buffer (10mM Tris-HCl pH8.0, 10mM NaCl, 0.2% Igepal CA630) with added 100µl of protease inhibitors (Sigma, P8340).

2) Incubate cell suspension on ice for >15 minutes. Centrifuge at 2500xG for 5 minutes. Discard the supernatant.

3) Wash pelleted nuclei once with 500µl of ice-cold Hi-C lysis buffer.

4) Gently resuspend pellet in 50µl of 0.5% sodium dodecyl sulfate (SDS) and incubate at 62°C for 5-10 minutes.

5) After heating is over, add 145µl of water and 25µl of 10% Triton X-100 (Sigma, 93443) to quench the SDS. Mix well and incubate at 37°C for 15 minutes.

6) Add 25µl of 10X CviJI reaction buffer and dimethyl sulfoxide (DMSO) to 10% final concentration.

7) Digest chromatin with 200U of CviJI restriction enzyme (Chimerx, 2125) for 4 hours at 37°C.

Proximity Ligation and Crosslink Reversal

8) Incubate at 65°C for 10 minutes to inactivate CviJI, then cool to room temperature.

9) Set aside 20% of the library to be used as pre-ligation control and store it at 4°C.

10) Add 900µl of ligation master mix:

    663µl of water

    120µl of 10X NEB T4 DNA ligase buffer (NEB, B0202)

    100µl of 10% Triton X-100

    12µl of 10mg/ml Bovine Serum Albumin (100X BSA)

    5µl of 400 U/ µl T4 DNA Ligase (NEB, M0202)

11) After ligation is over, pellet nuclei by spinning at 2500xG for 5 min. Discard the supernatant and re-suspend the pellet in 250-500µl of 1% SDS in 1X Tris buffer (10 mM Tris-HCl, pH 8).

12) Degrade proteins by adding 50µl of 20mg/ml proteinase K (NEB, P8102) and sodium chloride to 0.5M final concentration. Incubate at 55°C for 15-30 min with mixing.

13) Process in parallel the restricted "pre-ligation" control (step 14), adjusting all reagent volumes used in step #19 to the same final concentration as for the ligated chromatin.

14) Reverse crosslinks by incubating at 68°C overnight.

15) Cool tubes at room temperature. To precipitate DNA, add to each tube 2X volumes of pure ethanol. Mix by inverting and incubate at -80°C for 15 minutes.

16) Centrifuge at max speed, 2°C for 15 minutes. Keep the tubes on ice after spinning and carefully remove the supernatant by pipetting.

17) Resuspend pellet in 800µl of 70% ethanol. Centrifuge at max speed for 5 minutes.

18) Remove all supernatant and dissolve the pelleted pre-ligation control and ligated library in 10µl and 50µl respectively of 1X Tris buffer with added 1µl RNaseA (Life Technologies, 12091-039). Incubate at 37°C for 15min.

Concatemer Selection

19) Prepare 1% agarose gel in 1X Tris-acetate EDTA (TAE) buffer using low-melting point SeaPlaque GTG agarose (Lonza, 50111)

20) Mix DNA samples with gel-loading buffer in the recommended ratio. Load 1kb NEB Quickload ladder (NEB, N0468); pre-ligation control in a single well; and the entire volume of ligated library split in 4-5 wells.

21) Run gel in 1X TEA buffer for 2 hours at 100V.

22) Stain the gel with SYBR green for 20 min and visualize DNA. Restricted chromatin should be under 3kb in size.

23) Under UV light, excise fragments in the range 5-10kb from the gel lanes containing ligated library.

24) Transfer excised gel pieces to labeled tubes and equilibrate DNA-containing agarose slices by submerging them in 2 volumes of 1X β-Agarase I buffer (NEB, M0392) for 30 minutes at room temperature.

25) Remove the buffer and melt the agarose by incubation at 68°C for 10 minutes. Reduce temperature to 42°C and equilibrate liquid agarose for 15 min.

26) Digest agarose with 1 unit of β-Agarase I per 100ul of molten agarose at 42°C for 1 hour.

27) Adjust the salt concentration of the β-Agarase I treated solution for ethanol precipitation of DNA by adding 0.1X volumes of 3M sodium acetate, pH 5.2.

28) Incubate on ice for 10 minutes and centrifuge at max speed for 15 minutes to pellet any remaining undigested carbohydrates.

29) Transfer DNA-containing supernatant to a new tube. Precipitate DNA by addition of 2X volumes of ethanol.

30) Mix thoroughly, incubate at -80C for 15min and centrifuge at max speed for 15 minutes. Discard the supernatant.

31) Resuspend pellet in 800µl of 70% ethanol. Centrifuge at max speed for 5 minutes and discard the supernatant.

32) Dissolve the pellet in 130µl of 1X Tris buffer (10 mM Tris-HCl, pH 8)

33) To make the DNA suitable for high-throughput sequencing using Illumina sequencers, shear to a size of 300-500bp using the following parameters:

    Instrument: Covaris LE220 (Covaris, Woburn, MA)

    Volume of Library: 130µl in a Covaris microTUBE

    Fill Level: 10

Duty Cycle: 15

PIP: 500

Cycles/Burst: 200

Time: 58 seconds

34) Transfer sheared DNA to a fresh tube. Run a 1:5 dilution of DNA on a 2% agarose gel to verify successful shearing.

35) Prepare library for Illumina sequencing following Illumina's TruSeq® DNA Sample Preparation Guide[1] or comparable protocol.

TALENs.

TALENs were assembled using FLASH essentially as described (1). Activity at the proximal and distal inverted repeat sequence was assessed using the Surveyor assay (Transgenomic) (2) with oligonucleotide sequences CAACCCCACAGGGAGAACT and ACTTTCATTTTCAGGTCAGAC for the proximal side, and CTTTTCTTCTCAGGTTACAGT and TCAATCACACAAGGGAAATGC for the distal side.

CRISPR/Cas9.

Suitable gRNA targets were identified using Optimized CRISPR Design (http://crispr.mit.edu) and cloned into pSpCas9(BB)-2A-GFP (PX458) as directed (3). pSpCas9(BB)-2A-GFP (PX458) was a gift from Feng Zhang (Addgene plasmid # 48138). Oligonucleotides were obtained from Eurofins Genomics.

[1] Illumina, Inc. TruSeq DNA Sample Preparation Guide. 2012. Part # 15026486 Rev. C. www.illumina.com

Isolation and characterization of RPE1-ΔDXZ4a and RPE1-ΔDXZ4i.

TALEN expression constructs were introduced into hTERT-RPE1 cells (Clontech Laboratories, No. C4000-1) using a 4D-Nucleofector™ (Lonza Group Ltd.). Twenty-four hours post-Nucleofection, cells were seeded to 24-well plates at a density of 25 cells/well. At 50% confluence, 50% of cells from each row were pooled and genomic DNA isolated before screening for deletion of *DXZ4* using oligonucleotides CAACAGCAATTTCAGTAAGGTG and GATCTGGTCAAAATCAGAGAT. Genomic DNA was isolated from cells in individual wells of positive rows and re-screened for *DXZ4* deletions. Cells from positive wells were then seeded into 96-well plates at a density of 0.25 cells/well. Screening of pooled rows was performed as above. TA cloning and sequencing the deletion PCR product determined the exact nature of the *DXZ4* deletion for RPE1-ΔDXZ4a and RPE1-ΔDXZ4i. Assignment of the deletion to Xa or Xi was achieved by FISH analysis on metaphase chromosomes using a direct-labeled *DXZ4* probe (Spectrum Red, Abbott Molecular) in combination with a second X-linked direct-labeled BAC probe (Spectrum Green, Abbott Molecular). hTERT-RPE1 has an X:10 translocation on Xa at Xq28 making Xq substantially longer on Xa compared to Xi. The karyotypically normal X is Xi in all cells (4). Metaphase chromosome preparation and FISH were performed essentially as described previously (5). A pair of gRNAs (CCAGCGAAGACCTTACTCAA and GCCCTTTAAGCACCTACGAC) were cloned into pX458 and introduced into hTERT-RPE1 as above. Seventy-two hours later, GFP positive cells were collected using a BD FACSAria SORP flow cytometer and manually seeded at a density of 1 cell every 3-wells in 96-well plates. Screening for deletion of *DXZ4* used oligonucleotides CGTCAGGCATACAGATGATAC and GATTCTACCGAGCAACTGGC. Assignment of deletion to Xa or Xi took advantage of dbSNP

rs2301953 that resides within the deletion. DNA from deletion clones was amplified using oligonucleotides ATTCTTTCAGCCCTCACCTC and ACAAACCCTGATCTCGTGTC, and PCR products sequenced. Xa allele = G, Xi allele = C. Metaphase chromosomes were prepared as described previously (6). EdU was added to cell culture media to 10μM for 3 hours before addition of 25ng/ml KaryoMAX Colcemid for an additional hour. Metaphase chromosome preparation was performed as described previously (5). Detection of EdU incorporation used the Click-iT EdU Alexa Fluor 555 imaging kit from Thermo Fisher Scientific according to the manufacturer's recommendations.

***In Situ* Hi-C Protocol.**

We performed 20 *in situ* Hi-C experiments (Supplementary Table S1) as described previously (7).

DNA & RNA FISH.

Three-color DNA FISH was performed on B-Lymphocyte cell lines GM12878 and GM06992 that were obtained from the Coriell Institute for Medical Research. Direct-labeled probes were prepared by nick translation and included BAC clones 2272M5 for *DXZ4* (Alexa Fluor 647, Invitrogen Corporation), RP11-754H22 for *FIRRE* (Spectrum Green, Abbott Molecular) and RP11-818I17 for *ICCE* (Spectrum Red, Abbott Molecular). DNA FISH was performed essentially as described (8). Imaging was performed on an Olympus IX71 operated by the DeltaVision pDV, deconvolved with softWoRx 5.5.1 (DeltaVision), and compiled using Adobe Photoshop CS6 (Adobe Systems). Deconvolved images were corrected for chromatic aberration through the use of 100nm TetraSpeck Microspheres according to the manufacturer's instructions

(ThermoFisher Scientific). In order to score for overlap, signal levels for each channel were reduced to a single pixel and distance between pixels was determined using the measure function in softWoRx. Contact was scored positive if two signals were within 300nm in a projected image and within 3 z-stacks based on the range of chromatic aberration determined with the TetraSpeck Microspheres. Contact between two pairs was considered positive for 3-way contact. RNA FISH was performed using a direct-labeled *XIST* probe (Spectrum Red, Abbott Molecular) combined with direct-labeled BAC clones (Spectrum Green, Abbott Molecular) as described previously (8).

RNA-Seq.

Total RNA was isolated using the NucleoSpin RNA kit according to the manufacturers recommendations (Machery-Nagel). RNA quality was assessed on a Bioanalyzer 2100 (Agilent Technologies) and quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). RNA-Seq libraries were constructed in triplicate using 500ng of total RNA with the NEBNext Ultra Directional RNA Library Kit for Illumina, the NEBNext poly(A) mRNA Magnetic Isolation Module was used for isolation of polyadenylated mRNA and multiplexing used the NEBNext Multiplex Oligo for Illumina set (New England Biolabs). Libraries were assessed with a Bioanalyzer 2100 (Agilent Technologies) and qPCR using the Illumina Library Quantification kit (Illumina). Pooled libraries were loaded onto two lanes of an Illumina HiSeq 2000 for Rapid Run 2x100bp paired-end sequencing.

To perform allele-specific analyses, fastqs were first aligned to hg19 using *STAR*. The aligned reads were then split into maternal or paternal reads using our standard human diploid pipeline. We used *kallisto* (indexed to GRCh37 cdna) to calculate transcript abundances on the maternal

and paternal allele of each sample. Differential expression of transcripts was evaluated using *EdgeR*, using all three samples from the two cell lines. In Supplementary Table S3, we report significant transcriptional changes on the Xi between RPE1-WT and RPE1-ΔDXZ4i (significance required a fold change of at least 2 and an FDR less than 0.05.)

Immunofluorescence.

Indirect immunofluorescence was performed essentially as described previously (5). Rabbit anti-H3K27me3 (07-449) was obtained from EMD Millipore. Rabbit anti-H3K9me3 used for metaphase analysis (07-523) was obtained from EMD Millipore, whereas interphase analysis used rabbit anti-H3K9me3 from Active Motif (39161). Direct immunofluorescence utilized the Zenon Alexa Fluor 555 Rabbit IgG labeling kit from Thermo Fisher Scientific according to the manufacturer's recommendations. Conjugate secondary antibodies (Alexa-Fluor®) were obtained from Life Technologies Corporation. DNA was counterstained using the VECTASHIELD® mounting medium with DAPI from VECTOR Laboratories. Imaging was performed as described above. Metaphase chromosome immunofluorescence was performed on samples prepared by Cytospinning as described previously (6).

***DXZ4* Xa and Xi allele size determination.**

Pulsed field gel electrophoresis, Southern blotting, hybridization and detection were performed essentially as described previously (9). Determining the size of Xa and Xi *DXZ4* alleles was achieved by comparing the size of hybridizing bands obtained for *Xba*I digested plugs from parental hTERT-RPE1 and a somatic cell hybrid that contains hTERT-RPE1 Xa as its only human chromosome.

Hi-C Data Pipeline.

Read files were processed as described (7). Reads for the human, mouse, and macaque cell lines were aligned to hg19, mm10, and rheMac2, respectively.

Diploid Hi-C Pipeline.

Phased data for GM12878 was obtained from (7).

To phase chrX in RPE1, we first used the Hi-C reads from all three libraries (RPE1-WT, RPE1-ΔDXZ4a, and RPE1-ΔDXZ4i) to call SNPs using *GATK*. We then obtained unpublished whole genome amplification data of two cells which both were monosomic for chrX (courtesy of Cheng-Zhong Zhang, (10)). The aligned reads from these cells were used as input to *GATK*'s *Unified Genotyper*, along with our list of SNPs. All SNPs which were found in the monosomic cells were phased to the first allele, and the alternate variant at these locations was assigned to the second allele.

The Patski cell line is a female mouse embryonic kidney fibroblast obtained from a *Mus spretus* x C57BL/6J hybrid mouse (acquired courtesy of the Disteche Lab, UW). The maternal (C57BL/6J) X chromosome is always the inactive X. To create two diploid maps, we followed the procedure described previously (7), using SNPs obtained from the Wellcome Trust Sanger Institute to distinguish reads on the different alleles. Of 780,701,963 total reads, we were able to phase 117,412,620.

To ensure that differences between the Patski maps and human maps were not due to differences in SNP density, we randomly subsampled the Patski SNPs from their original density (1/2000 bp) to the density we had in our human lines (1/ 90 bp). SNPs that overlapped MboI restrictions site were also removed. These maps displayed similar characteristics as the full map, with a large superdomain score at *Dxz4* (3.2, p-value 6 $*10^{-4}$) and a large half-correlation length (49 Mb).

In order to ensure that our results were robust to allelic mapping bias, we also created alternative phased Patski maps using a different algorithm. We first created a pseudo genome for *Mus spretus* by taking the mm10 genome and changing SNPs at the appropriate locations to their *Mus spretus* variants. We then aligned the Hi-C reads to this genome as well. Each Hi-C contact was then phased to its respective map if it aligned unambiguously better to one genome over the other. The criterion for phasing was: if the contact only aligned to one of the genomes, it was phased to that genome's map. If the contact aligned to both genomes, we examined if the mapq was significantly different to warrant phasing. (We required the mapq of one of the two reads in a contact to be 5 or more than the mapq of the corresponding read in the other genome, and the mapq of the second read to be equal to or greater than the mapq of that read in the other genome). Both reads had to have a mapq greater than or equal to 10. While this approach did not exhibit significant allelic mapping bias (Xa map had 88% reads as the Xi map), it was also very similar to our original map, and exhibited a large superdomain score at *Dxz4* (3.0, p-value 1.5 $*10^{-3}$) and a large half-correlation length (11.5 Mb). We therefore used our original Patski map for the figures and analyses.

Higher-order Contact Analysis Pipeline.

In total, we examined 78 published Hi-C libraries from 6 different female cell lines, in addition to the GM12878 COLA library. As described previously (7), the Hi-C data pipeline annotates Hi-C contacts that map to three or more places in the genome as ambiguously chimeric, and these are usually discarded from 2-d contact map analysis. For the current higher order contact analysis, we examined these chimeric contacts to create a list of triples, quadruples, and quintuples in the genome. Due to the overwhelming contribution of triples as compared to other chimeras, further analysis was limited to triples.

In order to ensure that no duplicate triple reads entered the analysis, we first created a 'pseudo genome', concatenating all of the triple reads. We then used *bwa* to align each of the individual chimeric triples to the pseudo genome. Reads which aligned to multiple places (alpha <= 0.04) were discarded as duplicates.

For the most stringent filtering of higher order contacts, we required that all alignments were high-quality (MAPQ >10), all loci were on the same chromosome, and that the distance between any pair of loci was at least 20 kb. Since *DXZ4* and *ICCE* comprise tandem repeats, this filtering strategy threw out most of the relevant reads at these superloop anchors, so we developed a second filtering strategy for the superloop collocation analysis. For this analysis, reads with MAPQ=0 were included. However, the reads that did not align uniquely but landed on superloop anchors were then manually examined to ensure that all alignments fell within the identified superloop anchor.

The enrichment analysis of superloop triple data was performed with respect to a global and a local expected model, using a Poisson distribution for significance testing. In the global model, the expected number of reads was calculated based on the number of triples with anchors separated by the same distances as the superloop anchors in question. Since the global model does not account for possible biases due to pairwise contacts (for example contacts between any three loci will be enriched with respect to a global expected if two loci form a loop), we also calculated an expected value based on the local neighborhood of the superloop triple (local model), closely following the HiCCUPs strategy described previously (7).

The local model assesses whether the observed enrichment in a 3D voxel can be explained solely by a combination of 1D coverage biases and 2D looping biases (increased pairwise interaction frequency). This model computes the bias enrichment at site *(A,B,C)* as f_{ABC} such that the expected number of reads, E_{ABC}, is $f_{ABC}*M$, where M is the mean number of reads in the local neighborhood. This model assumes that f_{ABC} is composed only of independent 1D and 2D biases:

$$f_{ABC} = f_A*f_B*f_C*f_{AB}*f_{BC}*f_{AC}$$

$$E_{ABC} = f_{ABC}*M$$

We estimate each of the terms on the right hand side as follows.

f_A is the 1D bias of A, i.e. the fold enrichment of contacts involving locus A with respect to the local mean; this can be estimated by computing the average number of reads found in the plane located at A, which we denote as $\langle A \rangle$, and dividing this by the local mean. Thus $f_A \approx \langle A \rangle / M$. (And similarly for 1D biases for B and C.)

f_{AB} represents the 2D bias of *AB*, i.e. the fold enrichment in contacts between pairs of loci *A* and *B*, above what would be expected based on 1D biases. One is tempted to estimate this by simply computing the average enrichment of *AB* with respect to the mean, which would amount to taking the average in the *AB* line, which we denote as $<AB>$, and dividing by the local mean. However, the enrichment in $<AB>$ is due to a combination of 1D biases of *A* and *B*, and an independent 2D bias. Similar to the expected number of reads in a voxel based on known biases, we consider the expected number of reads for two loci (in the *AB* line) as

$E_{AB} = f_{AB} * f_A * f_B * M$

We estimate $E_{AB} \approx <AB>$ and so $f_{AB} \approx <AB> / (f_A * f_B * M) \approx <AB> * M / (<A> * <B>)$.

Combining all terms we find (see Supplementary Figure S4C for visual explanation):

$E_{ABC} = f_{ABC} * M = <AB> * <BC> * <AC> / (<A> * <B> * <C>) * M$.

To compute *M*, the local mean, we choose either the mean number of reads in the entire cube, or the number of reads in the cube's corner that is closest to the diagonal, based on whichever yields a more conservative (higher) value for the expected enrichment. Using this conservative estimate we compute an expected number of reads of 0.114 and 0.09 for the *ICCE-DXZ4-FIRRE* and *ICCE-x75-DXZ4* triples respectively, amounting to a more than 40-fold enrichment at the centers of these local 3D contact tensors (see Supplementary Figure S4D,E).

Contact Map Analyses.

Map Similarities: To measure map similarities, normalized contact maps were first transformed into observed/expected matrices, where the number of contacts in each entry is divided by the average number of reads at that distance. Rows which were more than 90% sparse were excluded. The two maps were then flattened to a single dimension and log transformed. Pearson correlation coefficients were then computed. Using this method we found the RPE1-WT and RPE1-ΔDXZ4a maps at 2.5 Mb resolution were highly similar (GW, Pearson r=.99; chrX, Pearson r=.98), while RPE1-WT and RPE1-ΔDXZ4i maps at 2.5 Mb resolution exhibited significant differences, especially on chrX (GW, Pearson r=.86; chrX, Pearson r=.73).

Compartments: To assign loci to compartments, normalized contact maps were transformed into observed/expected matrices. Rows that were more than 90% sparse were masked. We then computed the first principal component of this matrix, and mean normalized it. The PC of a map which is highly compartmentalized will exhibit entries which are mostly bimodally distributed, with loci in the same compartment having similar values in the PC. By using only the sign of the entries, each locus can be assigned to one of two compartments.

Compartmentalization Strength: Typically, compartmentalization is observed in maps via the appearance of a plaid pattern – a stretch of loci will have one pattern, and then the next stretch will have the opposite pattern, and the two patterns will alternate across the length of the chromosome. To measure the compartmentalization strength of a map, we examined its principal component (computed as described above).

The PC for most maps switches rapidly between bimodal values, while the PC for the Patski Xi decays slowly over the superdomains. As one measure of this, we compute the autocorrelation of the PC as a function of genomic distance and measure the half correlation distance: the distance at which the autocorrelation drops below half of the autocorrelation it achieves at distance zero. The PCs of the Xa and Xi in GM12878, RPE1-WT, RPE1-ΔDXZ4a and RPE1-ΔDXZ4i have half correlation lengths between 1 and 4 Mb. Similarly in the Patski Xa map, this distance is 1.5 Mb. However, in the Patski Xi map, the PC has a half correlation length of 30.5 Mb (Supplementary Table S5).

Superdomain Strength: To measure superdomain strength, we created the following simple "*S*" metric. A putative superdomain locus defines two superdomains, one to the locus' left and one to the locus' right, which should have strong boundaries along the chromosome. For a given distance *i* off the diagonal, we measure $S_i(k)$, the strength of the *k*th locus' two putative superdomain boundaries at distance *i*, by taking the average log-fold enrichment across the two (left and right) boundaries, values found in d_i, the *i*th diagonal of the contact matrix:

$$L_i = log(d_i[k-i-1]/d_i[k-i])$$
$$R_i = log(d_i[k]/d_i[k-1])$$
$$S_i(k) = (L_i + R_i)/2$$

We obtain the mean superdomain strength for locus *k*, S*(k)*, by averaging over all distances *i.* (To be robust to sparsity, we examine 500 kb matrices, and exclude rows that are more than 90% sparse, which only affects a small percentage of rows. Furthermore, any infinite or *nan* enrichment values are excluded from the mean value calculation.) For a randomly chosen locus, the expected value of *S(k)* is 0, since there should be no enrichment across the putative superdomain boundaries. We quantify the statistical enrichment of *S* at *DXZ4* by computing its

z-score and p-value, with the mean and standard deviation derived from the *S* values along the remainder of the chromosome. As expected from visual inspection, *S* is highly enriched in Xi maps at *DXZ4*/*Dxz4* relative to other loci, except for in RPE1-ΔDXZ4i, where no superdomain enrichment is seen (Supplementary Table S6).

Extrusion Simulations.

Molecular dynamics simulations were performed using *LAMMPS* as described in (12) with minor modifications. To simulate the entire Xi chromosome, 3400 beads were used, with each bead representing 50 kb. A total of 8 extrusion complexes were active in each simulation, and the parameters of temperature = 2.0 and t_{damp}=10 were used. Collapsed structures were simulated in 300,000 time steps and extrusion was simulated for 800,000 time steps or equivalently 4,000 extrusion steps. A total of 100 replicates were run.

The probability for an extrusion complex to halt at any given bead was calculated directly from Patski Ctcf ChIP-seq data from Berletch et al.,(11) by binning, thresholding, and rescaling as described in (12). Note that, because *Dxz4* is a variable repeat region, stringent map quality filtering will remove ChIP-seq reads at the *Dxz4* locus; however, these are correctly mapped to the *Dxz4* locus, even though their exact position within the locus may be uncertain. To obtain accurate statistics for Ctcf binding at *Dxz4*, we used a stringent map quality filter of 30 for the whole X chromosome, except at *Dxz4* where no map quality filtering was used. Simulations of the *Dxz4* deletion were performed with the exact same parameters with the extrusion complex halting probability set to 0 at the *Dxz4* locus. Because 50 kb regions typically contain multiple

Ctcf motifs oriented in both a forward and reverse fashion, binding of the extrusion complex at each monomer was allowed in a direction-independent manner.

While our original model assumed that extrusion complexes bind indefinitely at CTCF sites, a model in which extrusion complexes instead undergo partial arrest would explain the presence of large numbers of CTCF binding sites at superloop anchors. To test this, we ran simulations in which each extrusion subunit bound to a CTCF site would have a small probability $P_{continue}$ at each extrusion step to unbind and continue sliding along the chromatin. When binding was sufficiently stable (e.g. $P_{continue} < 0.0002$), the simulated contact maps accurately reproduced contact maps from Hi-C experiments (Supplementary Figure S11, (12)).

1. Reyon D*, et al.* (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30(5):460-465.
2. Guschin DY*, et al.* (2010) A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* 649:247-256.
3. Ran FA*, et al.* (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8(11):2281-2308.
4. Chadwick BP & Willard HF (2003) Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. *Hum Mol Genet* 12(17):2167-2178.
5. McLaughlin CR & Chadwick BP (2011) Characterization of DXZ4 conservation in primates implies important functional roles for CTCF binding, array expression and tandem repeat organization on the X chromosome. *Genome Biology* 12:R37.

6. Chadwick BP & Willard HF (2004) Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc Natl Acad Sci U S A* 101(50):17450-17455.

7. Rao SS*, et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665-1680.

8. Horakova AH*, et al.* (2012) The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. *Genome Biol* 13(8):R70.

9. Tremblay DC, Alexander G, Jr., Moseley S, & Chadwick BP (2010) Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. *BMC Genomics* 11:632.

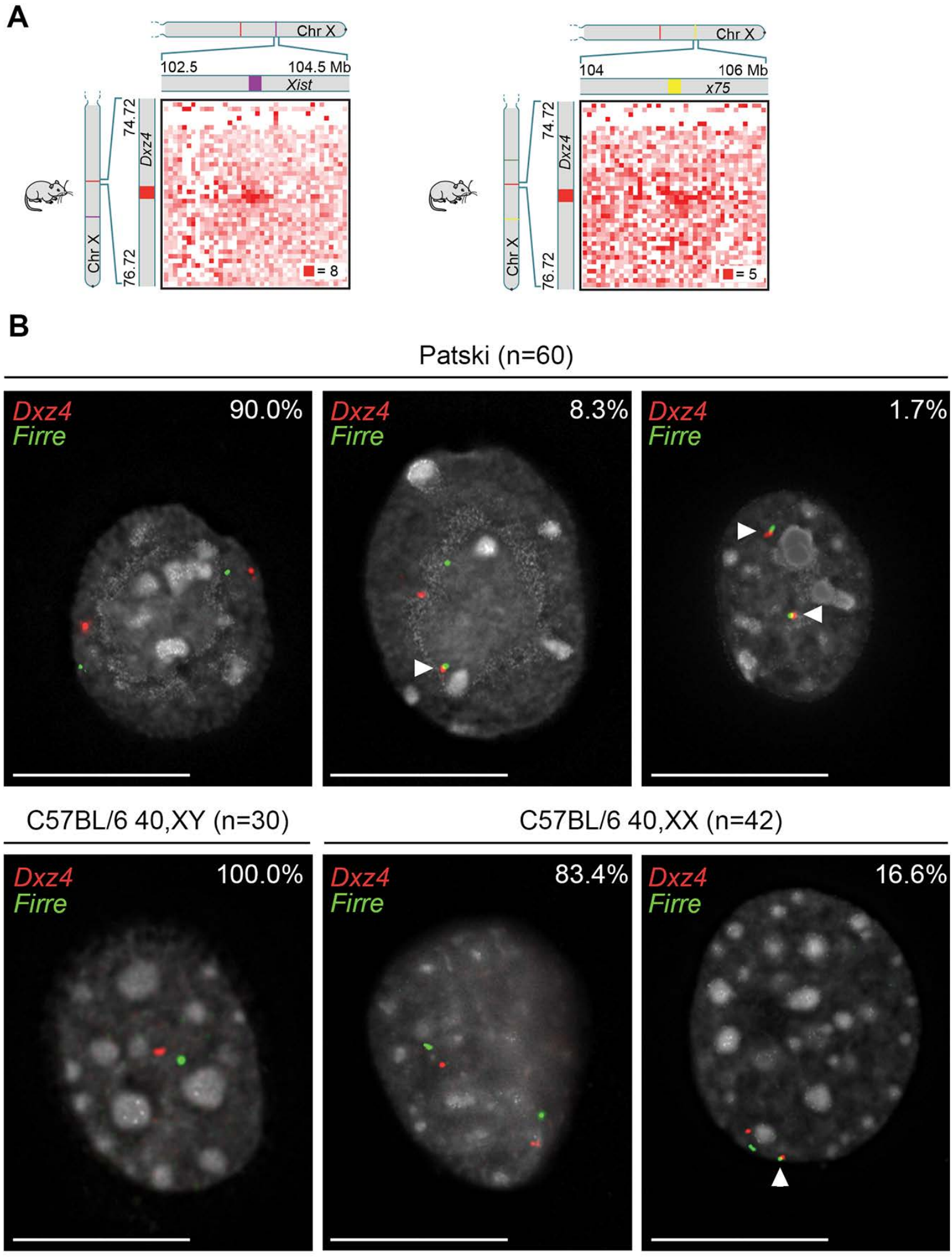
10. Zhang CZ*, et al.* (2015) Chromothripsis from DNA damage in micronuclei. *Nature* 522(7555):179-184.

11. Berletch JB*, et al.* (2015) Escape from X inactivation varies in mouse tissues. *PLoS Genet* 11(3):e1005079.

12. Sanborn AL*, et al.* (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* 112(47):E6456-6465.

Supplementary Fig. S1: Mouse cells exhibit superloops.

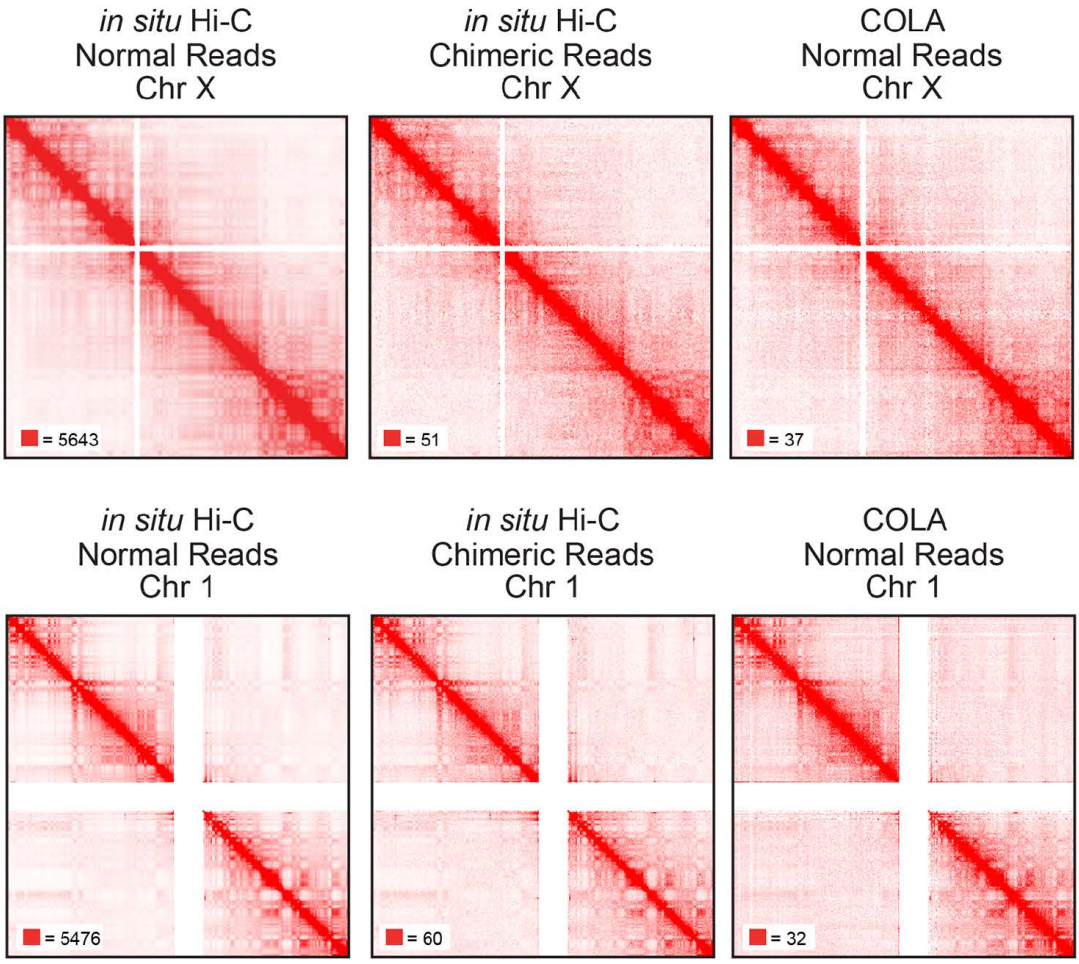
(*A*) A superloop between *Dxz4* and *Xist* and between *Dxz4* and the mouse ortholog of *x75* were observed in Patski cells. (*B*) Examples of direct-labeled DNA FISH for *Dxz4* (red) and *Firre* (green) merged with DAPI (grey) are shown for the female Patski cell line as well as male and female diploid primary fibroblasts. White bar represents 10 μm. The percentage of nuclei showing no, single or double-overlap (where appropriate) of *Dxz4* and *Firre* loci is given in the top right of each panel and the location of overlapping FISH signals is indicated by the white arrowhead.


A
Chr X
102.5
104.5 Mb
Xist
74.72
Dxz4
76.72
Chr X
= 8
Chr X
104
106 Mb
x75
74.72
Dxz4
76.72
Chr X
= 5
B
Patski (n=60)
Dxz4
Firre
90.0%
Dxz4
Firre
8.3%
Dxz4
Firre
1.7%
C57BL/6 40,XY (n=30)
C57BL/6 40,XX (n=42)
Dxz4
Firre
100.0%
Dxz4
Firre
83.4%
Dxz4
Firre
16.6%

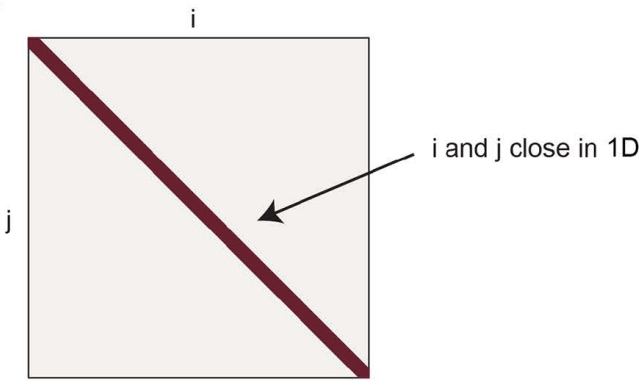
Supplementary Fig. S2: COLA can be used to map higher order contacts.

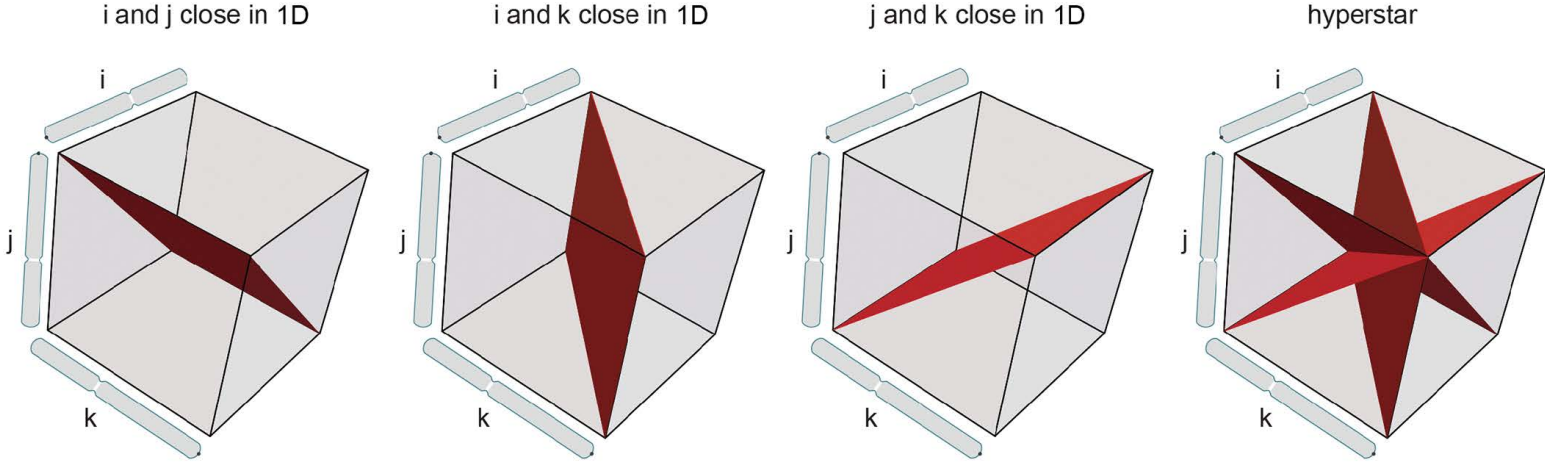
(*A*) 2D contact maps for *in situ* Hi-C (left), for inferred contacts from chimeric reads from an *in situ* Hi-C library (middle), and for normal reads from COLA (right). Maps shown for Chr X (top) and Chr 1 (bottom) confirm that the three procedures yield comparable data. (Resolution: 500 kb). (*B*) In a 2D contact map, a dark stripe appears along the diagonal because loci that are close together in 1D tend to form contacts inside the nucleus. In a 3D contact tensor, a similar phenomenon occurs. Triple contacts are indexed by (i, j, k). Since there is a preference for loci that are near to each other in 1D to form contacts, there is enrichment in the tensor any time two of the three loci (i and j, i and k, or j and k) are nearby in 1D. This enrichment manifests as a diagonal line in a 2D contact map; analogously, in a 3D contact tensor, the enrichment manifests as a diagonal plane. Since there are three types of enrichments (i and j, i and k, or j and k), there are three diagonally enriched planes, the superposition of which creates the hyperstar.

A


in situ Hi-C
Normal Reads
Chr X
in situ Hi-C
Chimeric Reads
Chr X
COLA
Normal Reads
Chr X
= 5643
= 51
= 37
in situ Hi-C
Normal Reads
Chr 1
in situ Hi-C
Chimeric Reads
Chr 1
COLA
Normal Reads
Chr 1
= 5476
= 60
= 32

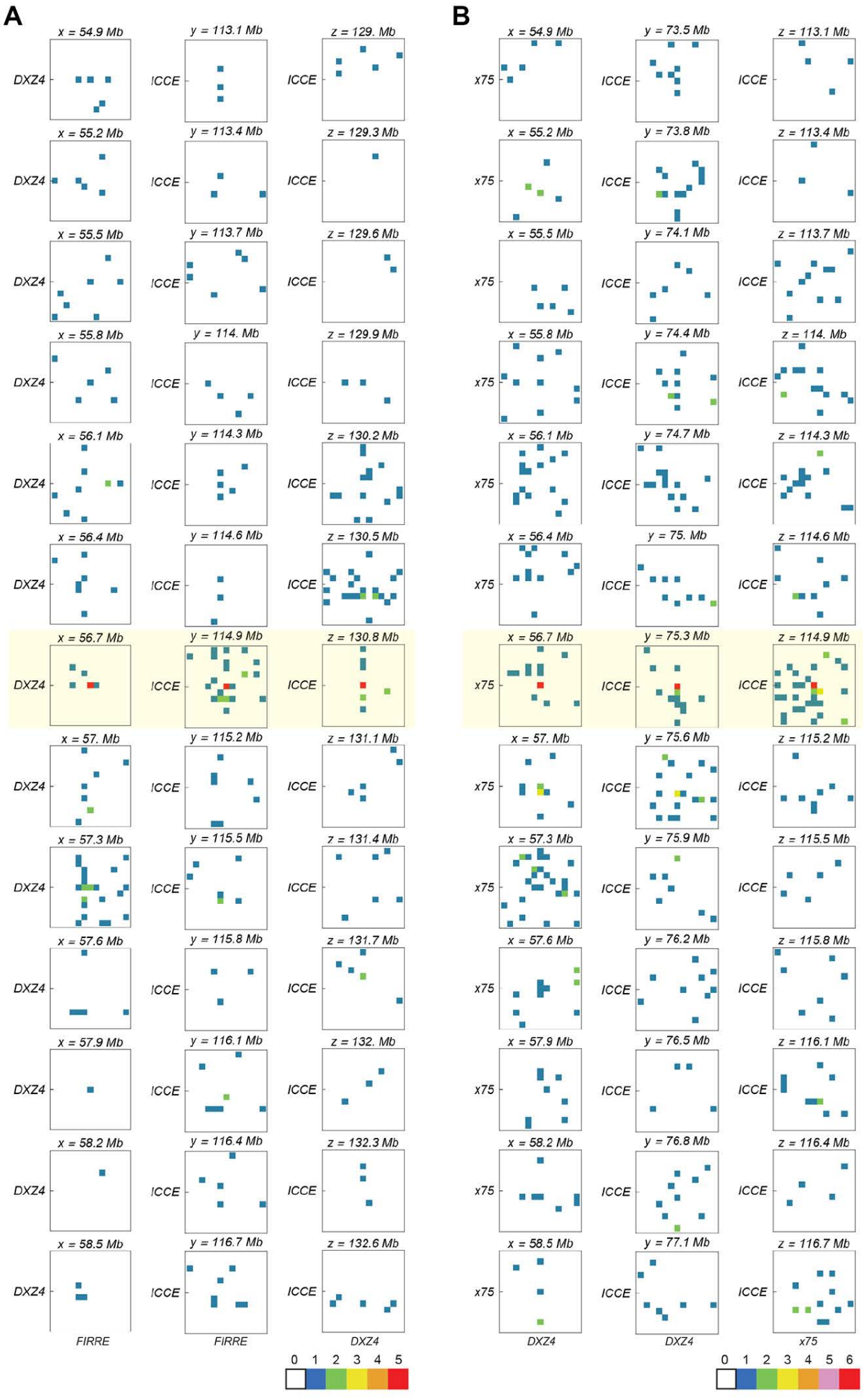
B


i
j
i and j close in 1D


i and j close in 1D
i and k close in 1D
j and k close in 1D
hyperstar
i
j
k

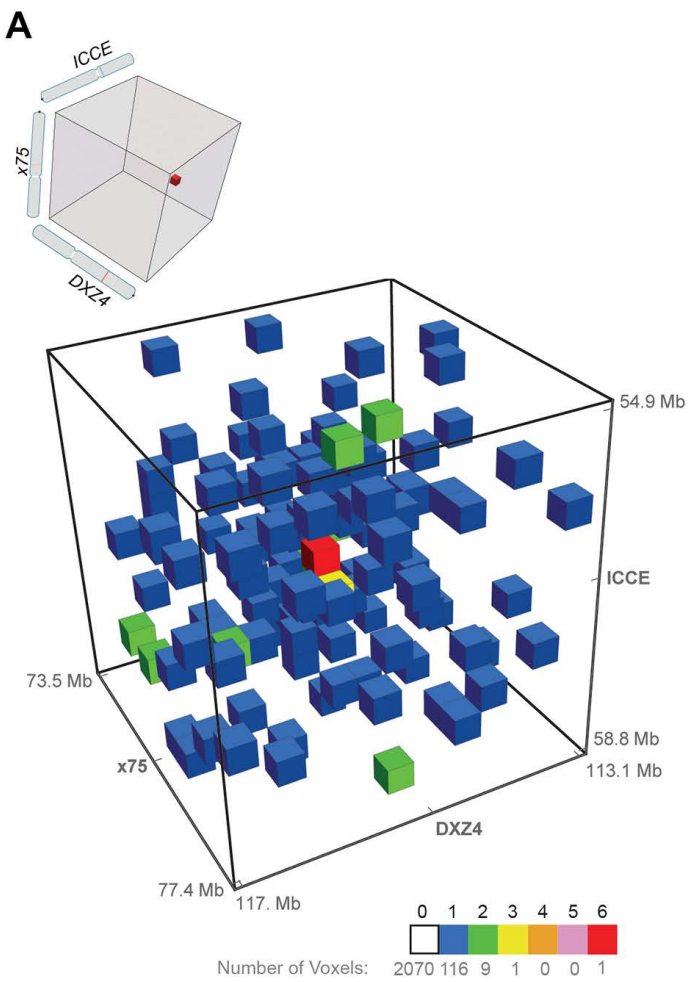
Supplementary Fig. S3: Superloop trios are locally enriched.

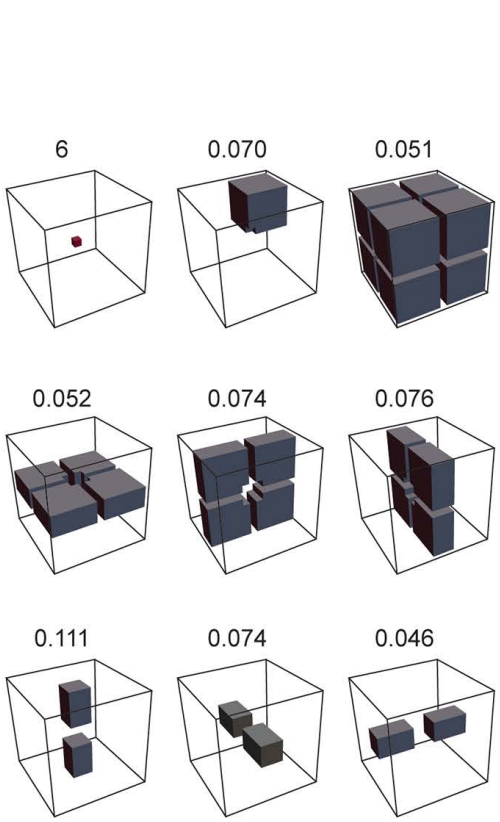
(*A*) All of the individual 2D slices of the cube in Fig. 3A, a small region from the 3D contact tensor of the X chromosome centered on *ICCE*, *DXZ4*, and *FIRRE*. Only slices containing the superloop trio show enrichment (highlight). (*B*) All of the individual 2D slices of the cube in Supplementary Fig. 4A, a small region from the 3D contact tensor of the X chromosome centered on *ICCE*, *x75,* and *DXZ4*. Only slices containing the superloop trio show enrichment (highlight).

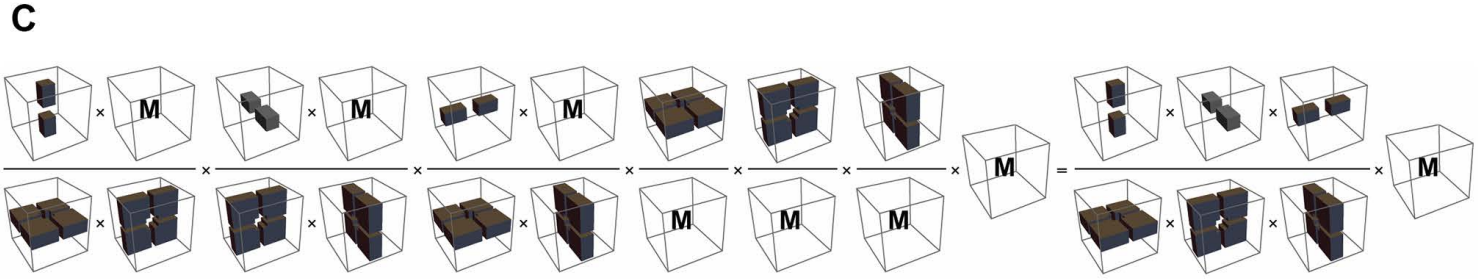

A
x = 54.9 Mb
y = 113.1 Mb
z = 129. Mb
DXZ4
ICCE
ICCE
x = 55.2 Mb
y = 113.4 Mb
z = 129.3 Mb
DXZ4
ICCE
ICCE
x = 55.5 Mb
y = 113.7 Mb
z = 129.6 Mb
DXZ4
ICCE
ICCE
x = 55.8 Mb
y = 114. Mb
z = 129.9 Mb
DXZ4
ICCE
ICCE
x = 56.1 Mb
y = 114.3 Mb
z = 130.2 Mb
DXZ4
ICCE
ICCE
x = 56.4 Mb
y = 114.6 Mb
z = 130.5 Mb
DXZ4
ICCE
ICCE
x = 56.7 Mb
y = 114.9 Mb
z = 130.8 Mb
DXZ4
ICCE
ICCE
x = 57. Mb
y = 115.2 Mb
z = 131.1 Mb
DXZ4
ICCE
ICCE
x = 57.3 Mb
y = 115.5 Mb
z = 131.4 Mb
DXZ4
ICCE
ICCE
x = 57.6 Mb
y = 115.8 Mb
z = 131.7 Mb
DXZ4
ICCE
ICCE
x = 57.9 Mb
y = 116.1 Mb
z = 132. Mb
DXZ4
ICCE
ICCE
x = 58.2 Mb
y = 116.4 Mb
z = 132.3 Mb
DXZ4
ICCE
ICCE
x = 58.5 Mb
y = 116.7 Mb
z = 132.6 Mb
DXZ4
ICCE
ICCE
FIRRE
FIRRE
DXZ4
0 1 2 3 4 5
B
x = 54.9 Mb
y = 73.5 Mb
z = 113.1 Mb
x75
ICCE
ICCE
x = 55.2 Mb
y = 73.8 Mb
z = 113.4 Mb
x75
ICCE
ICCE
x = 55.5 Mb
y = 74.1 Mb
z = 113.7 Mb
x75
ICCE
ICCE
x = 55.8 Mb
y = 74.4 Mb
z = 114. Mb
x75
ICCE
ICCE
x = 56.1 Mb
y = 74.7 Mb
z = 114.3 Mb
x75
ICCE
ICCE
x = 56.4 Mb
y = 75. Mb
z = 114.6 Mb
x75
ICCE
ICCE
x = 56.7 Mb
y = 75.3 Mb
z = 114.9 Mb
x75
ICCE
ICCE
x = 57. Mb
y = 75.6 Mb
z = 115.2 Mb
x75
ICCE
ICCE
x = 57.3 Mb
y = 75.9 Mb
z = 115.5 Mb
x75
ICCE
ICCE
x = 57.6 Mb
y = 76.2 Mb
z = 115.8 Mb
x75
ICCE
ICCE
x = 57.9 Mb
y = 76.5 Mb
z = 116.1 Mb
x75
ICCE
ICCE
x = 58.2 Mb
y = 76.8 Mb
z = 116.4 Mb
x75
ICCE
ICCE
x = 58.5 Mb
y = 77.1 Mb
z = 116.7 Mb
x75
ICCE
ICCE
DXZ4
DXZ4
x75
0 1 2 3 4 5 6

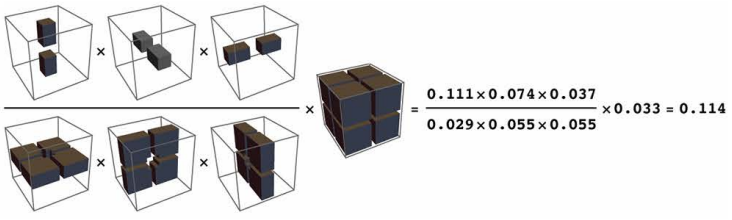
Supplementary Fig. S4: The *ICCE-DXZ4* and *x75-DXZ4* superloops tend to occur simultaneously, forming a hub on Xi.

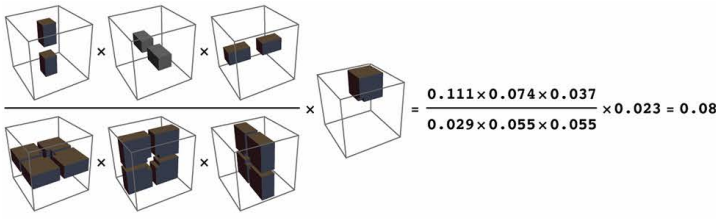
(*A*) Examination of a small region from the 3D contact tensor of the X chromosome, centered on *ICCE , x75*, and *DXZ4*, reveals a peak relative to the local neighborhood. Six contacts are seen in the $(300\ kb)^3$ voxel (i.e., 3D pixel) corresponding to simultaneous co-location of all three loci. There are over 2000 other $(300\ kb)^3$ voxels in the region shown; the number of contacts in each is indicated by the color. Of these voxels, one contains 3 contacts, 9 contain 2 contacts each, 116 contain 1 contact each, and over 2000 voxels contain no contacts. (*B*) The average frequency of contact in various local neighborhoods surrounding the *ICCE-x75-DXZ4* peak. The peak is strongly enriched with respect to every model. (*C*) A schematic derivation of the expected number of reads in the center of a 3D tensor according to the local model. (*D* and *E*) The expected number of reads for *ICCE-DXZ4-FIRRE (D)* and *ICCE-x75-DXZ4 (E)* triples based on the local model. Both donut and corner mean-based estimates are shown. For further analysis we choose the more conservative (the higher) of the two values for each respective triple.

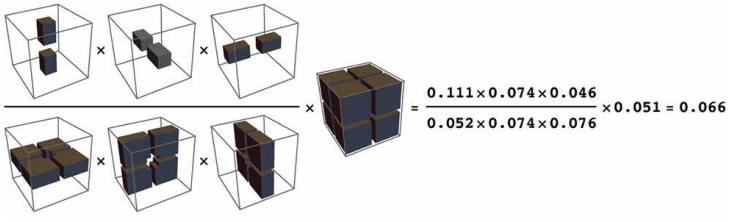

A
ICCE
x75
DXZ4
54.9 Mb
ICCE
73.5 Mb
x75
58.8 Mb
113.1 Mb
DXZ4
77.4 Mb
117. Mb
0 1 2 3 4 5 6
Number of Voxels: 2070 116 9 1 0 0 1

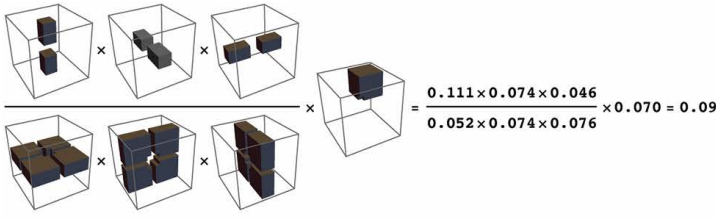

B
6
0.070
0.051
0.052
0.074
0.076
0.111
0.074
0.046


C
M


D
$\frac{0.111\times0.074\times0.037}{0.029\times0.055\times0.055}\times0.033 = 0.114$

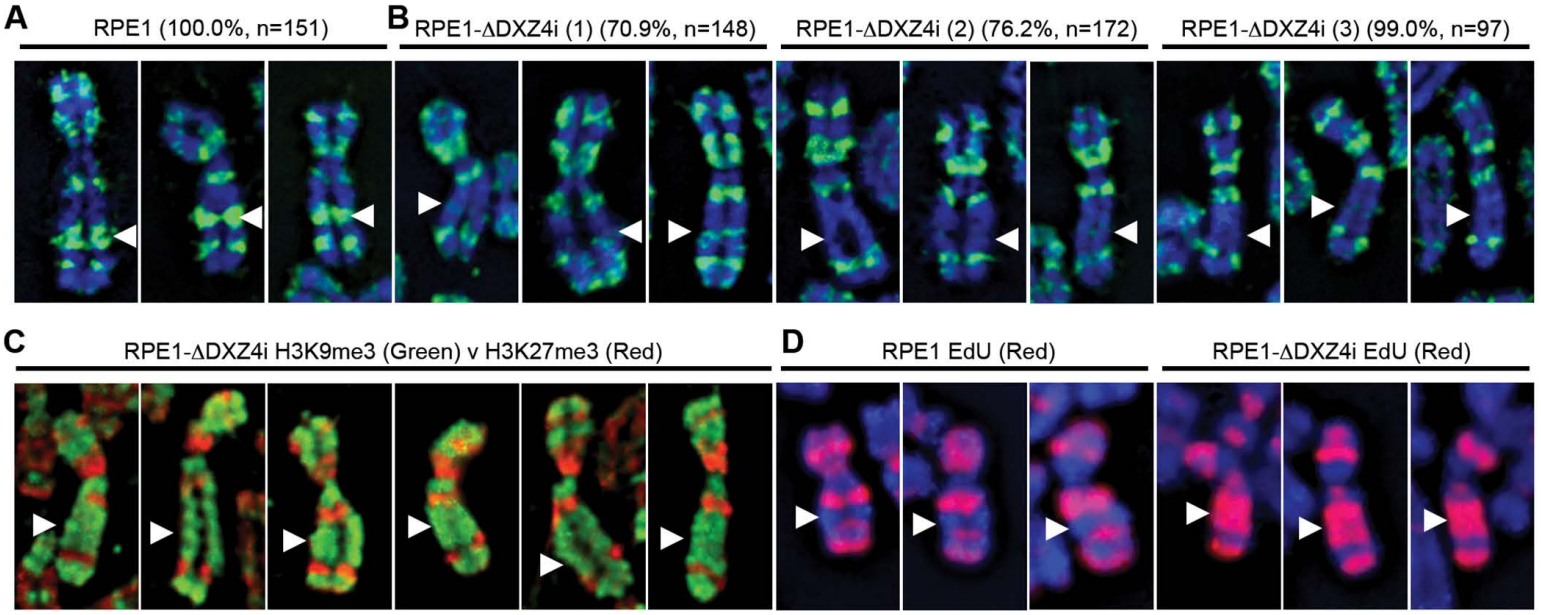

$\frac{0.111\times0.074\times0.037}{0.029\times0.055\times0.055}\times0.023 = 0.08$


E
$\frac{0.111\times0.074\times0.046}{0.052\times0.074\times0.076}\times0.051 = 0.066$


$\frac{0.111\times0.074\times0.046}{0.052\times0.074\times0.076}\times0.070 = 0.09$

Supplementary Fig. S5: Deletion of *DXZ4* interferes with chromatin organization on Xi and replication timing of the underlying DNA.

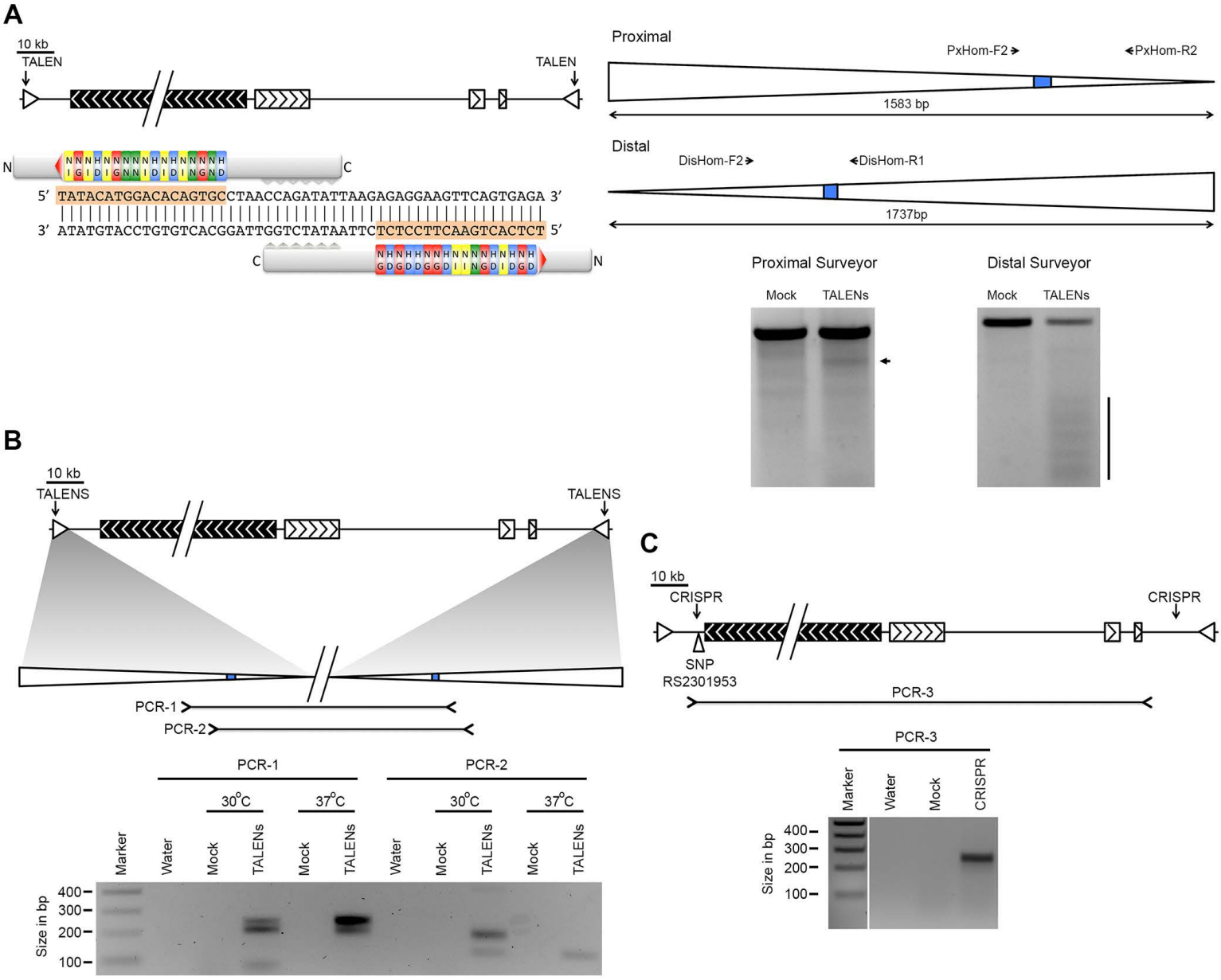
(*A*) Indirect immunofluorescence showing the distribution of H3K27me3 (green) at metaphase Xi in RPE1; DAPI shown in blue. (*B*) Indirect immunofluorescence showing the distribution of H3K27me3 (green) at metaphase Xi in three independent RPE1-ΔDXZ4i clones (1-3); DAPI shown in blue. The white arrowhead indicates the location of the *DXZ4*-proximal H3K27me3 band in RPE1 cells and its absence in the *DXZ4* mutants. (*C*) Indirect immunofluorescence (H3K9me3, green) combined with direct immunofluorescence (H3K27me3, red) showing H3K9me3 occupation of the former *DXZ4*-proximal H3K27me3 band on RPE1-ΔDXZ4i Xi. (*D*) Metaphase Xi from RPE1 and RPE1-ΔDXZ4i showing the pattern of EdU incorporation. The white arrowhead indicates the location of the *DXZ4*-proximal H3K27me3 band that is labeled by EdU in RPE1-ΔDXZ4i but not RPE1.


A
RPE1 (100.0%, n=151)
B
RPE1-ΔDXZ4i (1) (70.9%, n=148)
RPE1-ΔDXZ4i (2) (76.2%, n=172)
RPE1-ΔDXZ4i (3) (99.0%, n=97)
C
RPE1-ΔDXZ4i H3K9me3 (Green) v H3K27me3 (Red)
D
RPE1 EdU (Red)
RPE1-ΔDXZ4i EdU (Red)

Supplementary Fig. S6: A single TALEN pair targeting an inverted repeat that flanks *DXZ4* can excise the DNA between the two sites, as can a pair of gRNAs that flank DXZ4.

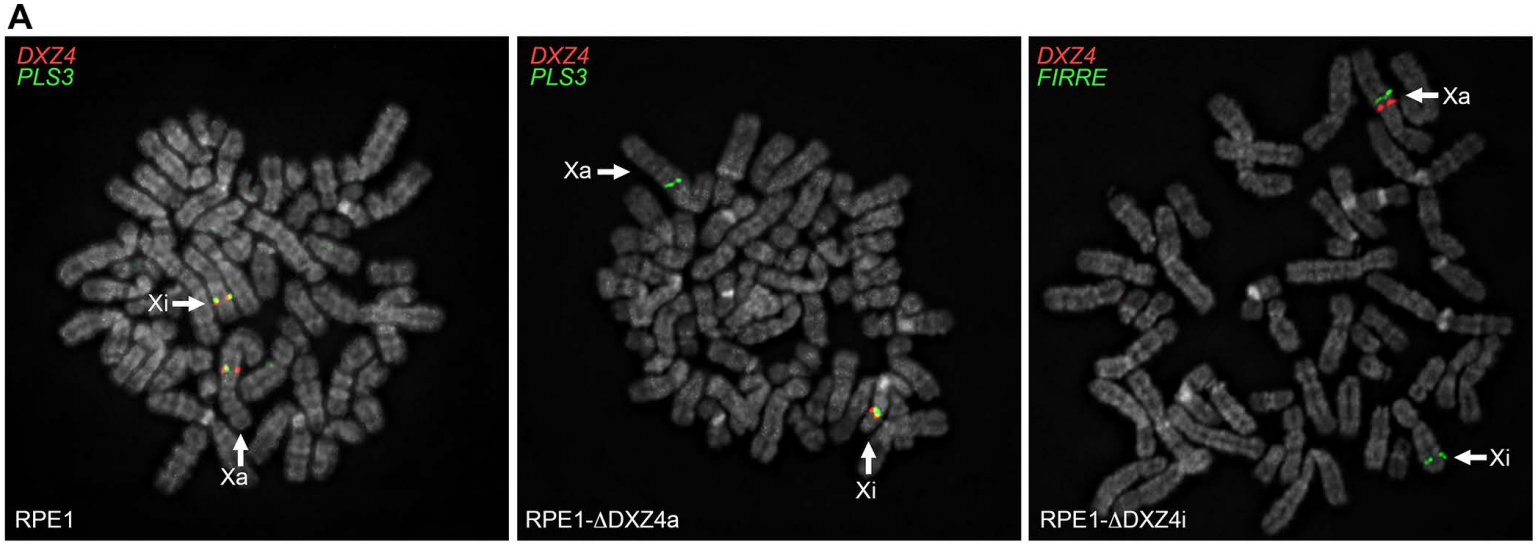
(*A*) At the top is a schematic representation of the genomic interval including the *DXZ4* locus. The main *DXZ4* array is represented by the black box with left facing arrows. The break within the array represents copy number variation between *DXZ4* alleles. The white boxes with right facing arrows indicate the location of sequences with inverted homology to *DXZ4*. The location of the inverted repeat to which the TALENs were designed is represented by the inward facing arrowheads at either end of the schematic. The inverted repeat shares 80% DNA sequence identity over 1486 bp. TALEN pairs were designed to an identical sequence within the proximal and distal inverted repeat and the target sequence and repeat variable domain composition of the TALEN pair is shown. The DNA sequence recognized by the TALENs is highlighted by the peach color. The alignment of the carboxyl-terminal FokI nuclease domain within the spacer region is represented by the jagged-teeth. Immediately to the right the TALEN schematic are maps of the proximal and distal repeat showing the location of where the TALENs bind (blue-box) and the location of primers that flank the TALEN binding sites that are unique to the proximal or distal copy of the inverted repeat. Confirmation of activity for the TALENs at the proximal and distal inverted repeat is shown in an inverted image of an ethidium bromide stained agarose gel using the Surveyor heteroduplex assay. Insertions/deletions are indicated by the arrowhead and black bar to the right of each image. Samples include mock-transfected cells and cells transiently transfected with the TALEN pair. (*B*) Schematic map of the *DXZ4* interval as in part-*A*, showing the location of the primer pairs used in the deletion assay (PCR-1 and PCR-2). At the bottom is an inverted image of an ethidium bromide stained agarose gel showing the results of PCR-1 and PCR-2 in cells mock-transfected or transfected with the TALEN pair and

grown at 30^{o}C or 37^{o}C before harvesting cells and isolating DNA. Heterogeneity in PCR product size for *DXZ4* excision represents multiple independent cutting events generating different sized repaired chromosomal intervals by non-homologous end joining. PCR products were TA cloned and sequence verified (data not shown). (*C*) Schematic map of the *DXZ4* interval as in part-*A*, showing the location of the CRISPR gRNA targets as well as SNP rs2301953. The location of the primer pair used in the deletion assay (PCR-3) is indicated. At the bottom is an inverted image of an ethidium bromide stained agarose gel showing the result of PCR-3 in cells mock-transfected or transfected with the CRISPR gRNA pair.

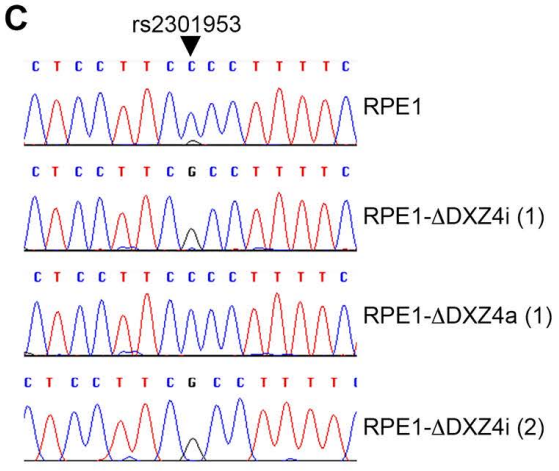

A
10 kb
TALEN
TALEN
N
C
5′ TATACATGGACACAGTGCCTAACCAGATATTAAGAGAGGAAGTTCAGTGAGA 3′
3′ ATATGTACCTGTGTCACGGATTGGTCTATAATTCTCTCCTTCAAGTCACTCT 5′
C
N
Proximal
PxHom-F2
PxHom-R2
1583 bp
Distal
DisHom-F2
DisHom-R1
1737bp
Proximal Surveyor
Distal Surveyor
Mock
TALENs
Mock
TALENs
B
10 kb
TALENS
TALENS
PCR-1
PCR-2
PCR-1
PCR-2
30°C
37°C
30°C
37°C
Marker
Water
Mock
TALENs
Mock
TALENs
Water
Mock
TALENs
Mock
TALENs
Size in bp
400
300
200
100
C
10 kb
CRISPR
CRISPR
SNP
RS2301953
PCR-3
PCR-3
Marker
Water
Mock
CRISPR
Size in bp
400
300
200
100

Supplementary Fig. S7: FISH validation of DXZ4 deletion clones.

(*A*) Metaphase chromosome spreads of RPE1 (left), RPE1-ΔDXZ4a (middle) and RPE1-ΔDXZ4i (right) stained with DAPI (grey) merged with direct-labeled DNA FISH signals for *DXZ4* (red) and either the *PLS3* or *FIRRE* locus (green). Xa is readily identifiable by the X:10 translocation at the tip of Xq28 substantially increasing the length of the long arm. Xa and Xi are indicated. (*B*) Schematic map of human X (coordinates hg19: 114,957,076-114,960,075) showing the location of SNP rs2301953 relative to the proximal CRISPR gRNA and the edge of *DXZ4*. The region PCR amplified is indicated. (*C*) DNA sequence traces for RPE1 as well as TALEN treated (1) and CRISPR treated (2) RPE1 clones. SNP rs2301953 is indicated: C/G in RPE1, C on Xa and G on Xi.


A
DXZ4
PLS3
Xi
Xa
RPE1
DXZ4
PLS3
Xa
Xi
RPE1-ΔDXZ4a
DXZ4
FIRRE
Xa
Xi
RPE1-ΔDXZ4i

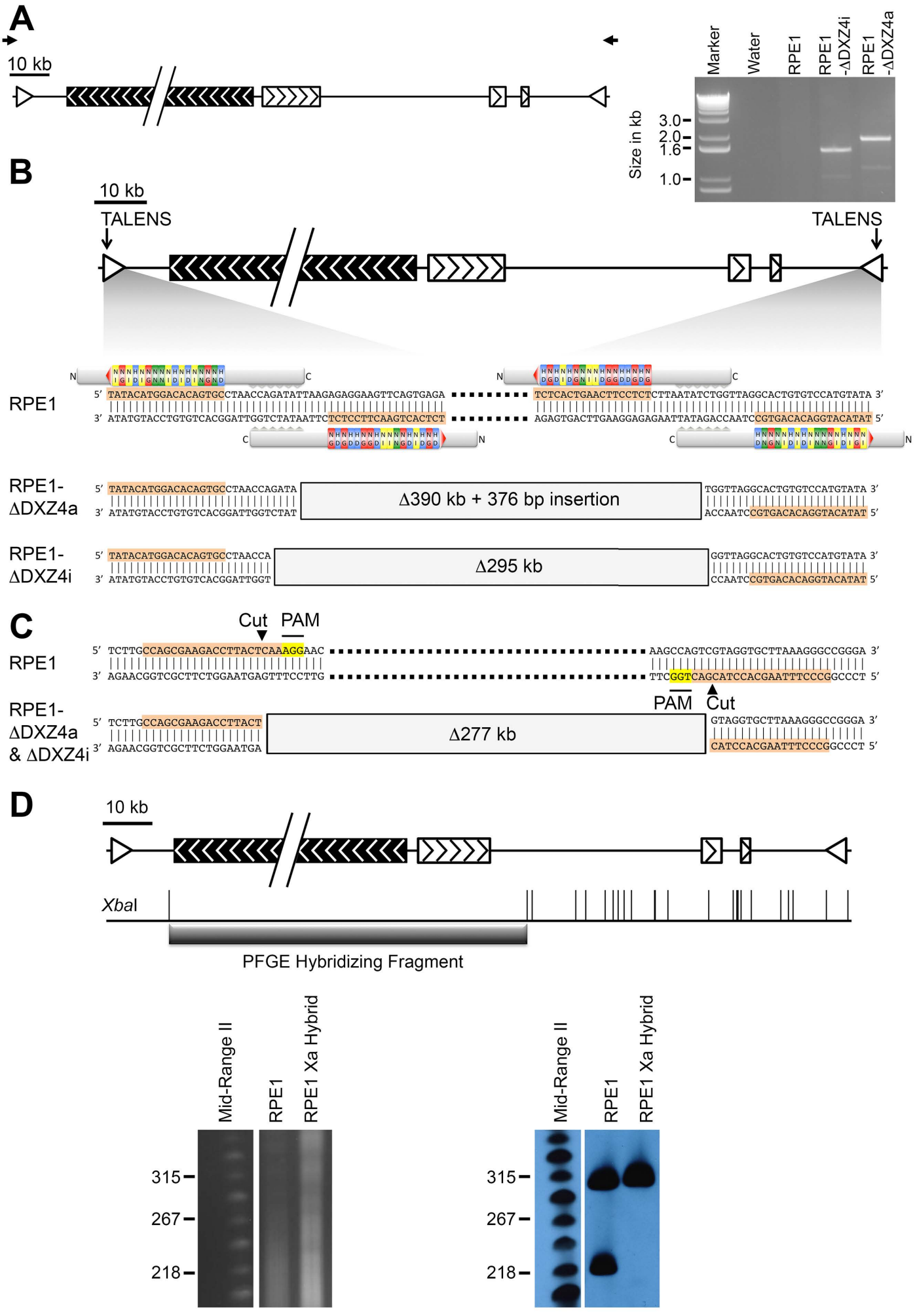

B
1kb
CRISPR
rs2301953
PCR
Proximal edge
DXZ4


C
rs2301953
C T C C T T C C C C T T T T C
RPE1
C T C C T T C G C C T T T T C
RPE1-ΔDXZ4i (1)
C T C C T T C C C C T T T T C
RPE1-ΔDXZ4a (1)
C T C C T T C G C C T T T T C
RPE1-ΔDXZ4i (2)

Supplementary Fig. S8: Characterization of Xa and Xi RPE1 deletion clones.

(*A*) Schematic representation of the *DXZ4* genomic interval as in Supplementary Figure S6. Above the interval map is the location of the two PCR primers (black inward facing arrows) used to amplify across the deletion. To the right is an ethidium bromide stained agarose gel image showing the result of the PCR for parental RPE1 and the Xi and Xa deletion clone. (*B*) The nature of Xa and Xi deletion mutations generated using TALENs. Beneath the schematic the inverted repeat regions are expanded to show the simultaneous binding of the TALEN pair at the proximal and distal target sites on the uncut RPE1 genomic locus. Beneath this is the sequence of the deletion break point for the Xa and Xi mutant and the extent of the deletion based on the *DXZ4* allele size. For the Xa mutant, the deletion was accompanied by a 376 bp insertion of part of the TALEN expression vector. (*C*) The nature of Xa and Xi deletion mutations generated by CRISPR. At the top is the sequence immediately surrounding the proximal and distal gRNA target sites (highlighted sequence). Highlighted is the location of the PAM sequence, as is the Cas9 cutting site. Immediately below this is the extent of the deletion. The exact same deletion was identified in 2 independent Xi deletion clones and 3 independent Xa deletion clones. (*D*) Sizing of the *DXZ4* alleles on Xa and Xi of RPE1 by pulsed field gel electrophoresis (PFGE). Below the schematic map is a restriction map indicating the location of all *Xba*I restriction endonuclease recognition sequences throughout the *DXZ4* genomic interval. The location of the probe used for hybridization to the PFGE Southern blot is indicated by the black bar. *Xba*I does not cut within the array and therefore it is excised intact as a single migrating fragment of a size that corresponds to the copy number of the individual 3.0 kb *DXZ4* repeat units. At the bottom is an image of an ethidium bromide stained agarose PFGE gel showing *Xba*I digested DNA for parental RPE1 and a somatic cell hybrid that contains the RPE1 Xa as its only human DNA

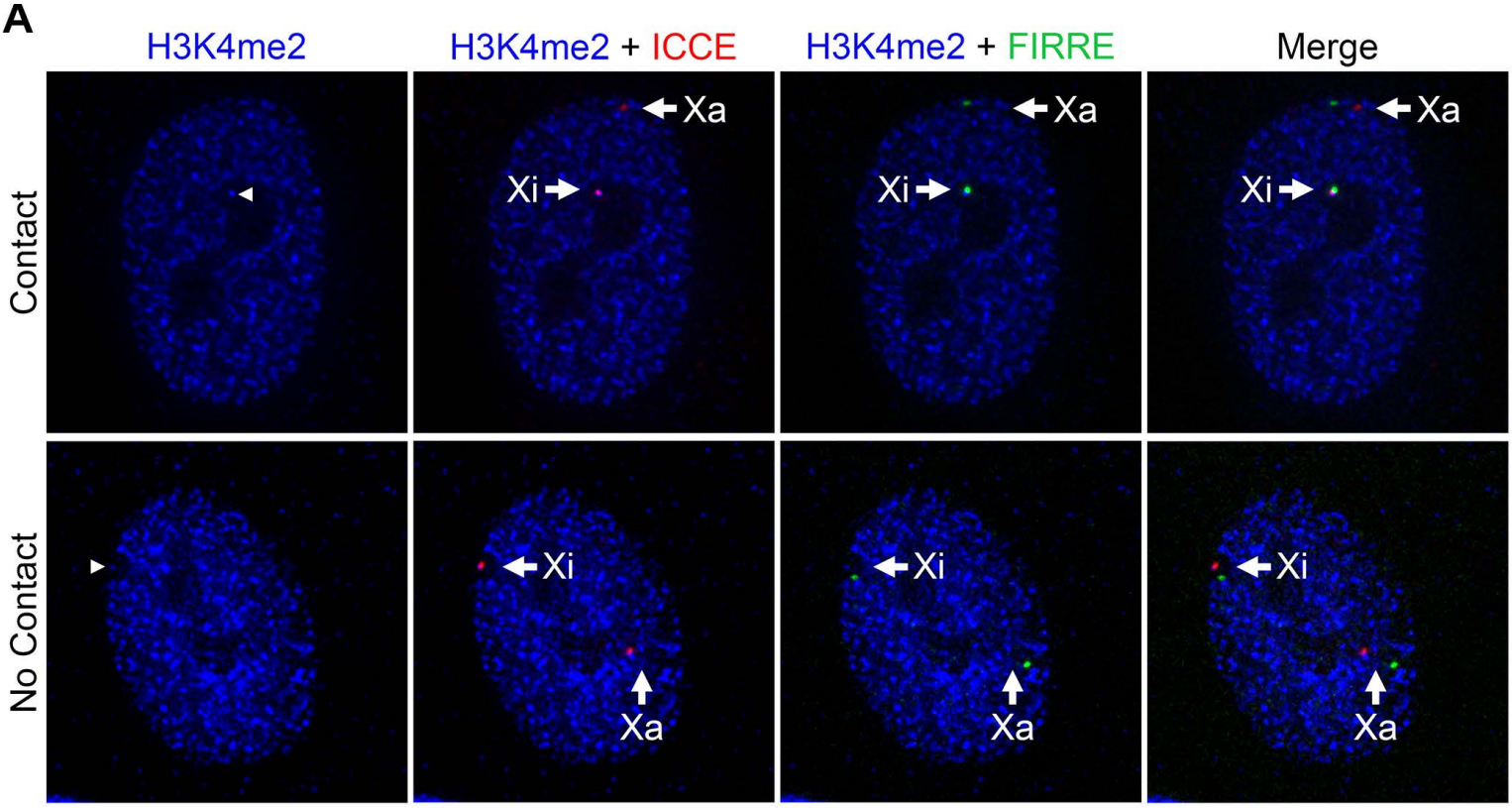
component. To the right is a Southern blot indicating the size of the *DXZ4*-positive fragment/s in the samples indicated.


A
10 kb
Marker
Water
RPE1
RPE1 -ΔDXZ4i
RPE1 -ΔDXZ4a
Size in kb
3.0
2.0
1.6
1.0
B
10 kb
TALENS
TALENS
RPE1
5′ TATACATGGACACAGTGCCTAACCAGATATTAAGAGAGGAAGTTCAGTGAGA
3′ ATATGTACCTGTGTCACGGATTGGTCTATAATTCTCTCCTTCAAGTCACTCT
TCTCACTGAACTTCCTCTCTTTAATATCTGGTTAGGCACTGTGTCCATGTATA 3′
AGAGTGACTTGAAGGAGAGAATTATAGACCAATCCGTGACACAGGTACATAT 5′
RPE1-ΔDXZ4a
5′ TATACATGGACACAGTGCCTAACCAGATA
3′ ATATGTACCTGTGTCACGGATTGGTCTAT
Δ390 kb + 376 bp insertion
TGGTTAGGCACTGTGTCCATGTATA 3′
ACCAATCCGTGACACAGGTACATAT 5′
RPE1-ΔDXZ4i
5′ TATACATGGACACAGTGCCTAACCA
3′ ATATGTACCTGTGTCACGGATTGGT
Δ295 kb
GGTTAGGCACTGTGTCCATGTATA 3′
CCAATCCGTGACACAGGTACATAT 5′
C
Cut
PAM
RPE1
5′ TCTTGCCAGCGAAGACCTTACTCAAAGGAAC
3′ AGAACGGTCGCTTCTGGAATGAGTTTCCTTG
AAGCCAGTCGTAGGTGCTTAAAGGGCCGGGA 3′
TTCGGTCAGCATCCACGAATTTCCCGGCCCT 5′
PAM
Cut
RPE1-ΔDXZ4a & ΔDXZ4i
5′ TCTTGCCAGCGAAGACCTTACT
3′ AGAACGGTCGCTTCTGGAATGA
Δ277 kb
GTAGGTGCTTAAAGGGCCGGGA 3′
CATCCACGAATTTCCCGGCCCT 5′
D
10 kb
XbaI
PFGE Hybridizing Fragment
Mid-Range II
RPE1
RPE1 Xa Hybrid
315
267
218
Mid-Range II
RPE1
RPE1 Xa Hybrid
315
267
218

Supplementary Fig. S9: ICCE-FIRRE contact in RPE1 and cells lacking *DXZ4* on Xa or Xi.

(*A*) Examples of *ICCE-FIRRE* contact or no contact at Xi in RPE1 cells. Images show anti-H3K4me2 immunofluorescence (blue) combined with *ICCE* (red) and *FIRRE* (green) direct-labeled FISH. The top panel shows examples of Xi-specific contact and the lower panel showing no contact. The location of the *DXZ4* signal within the Xi territory is indicated in the first image with the white arrowhead. The location of Xi is determined by the hypo-H3K4me2 pattern. (*B*) Summary of Xa v Xi *ICCE-FIRRE* contact in replicate experiments on RPE1, RPE1-ΔDXZ4a and RPE1-ΔDXZ4i cells. (*C*) We performed Aggregate Peak Analysis (7) at 50 kb resolution on the list of Xi superloops from (7), but excluded all superloops which involved *DXZ4*. The enrichment aggregated over all loops is not diminished after deletion of *DXZ4* from Xi.

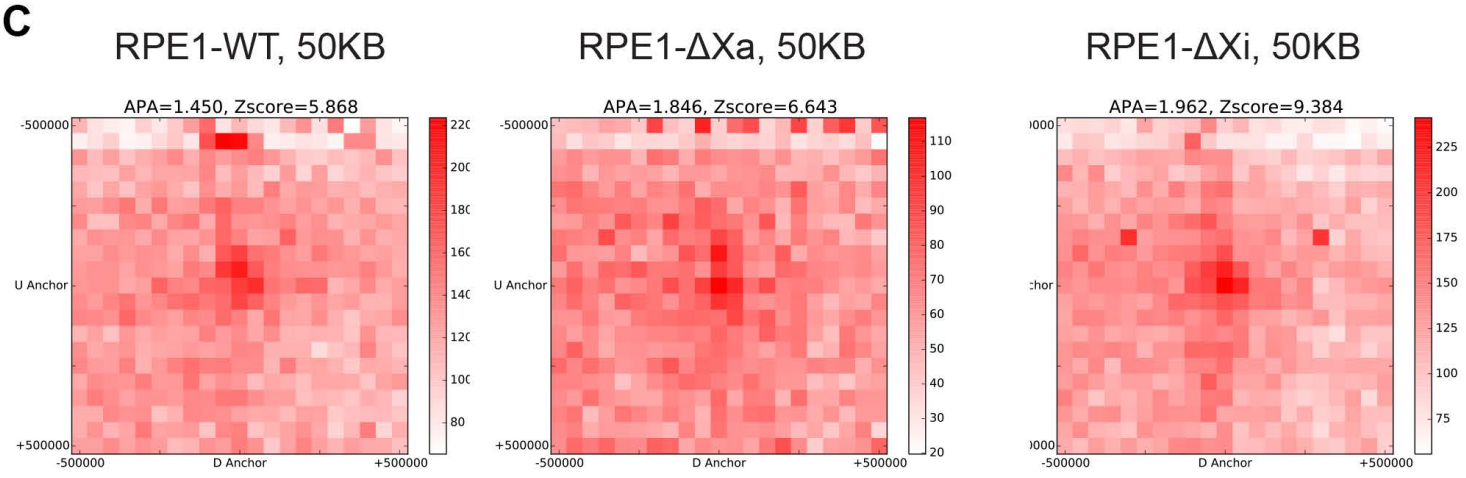
A



B

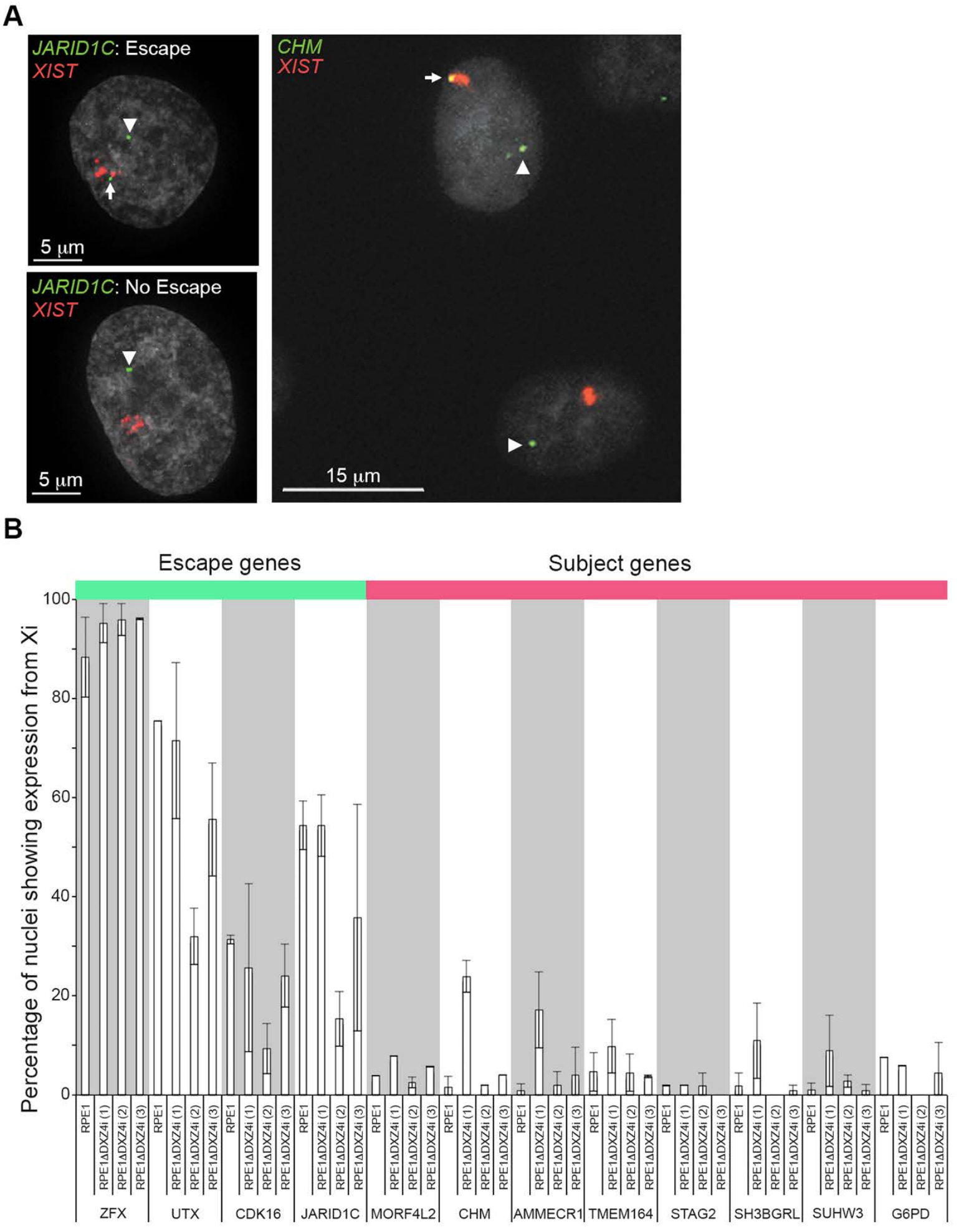
Cells	Xa FIRRE-ICCE Contact	Xi FIRRE-ICCE Contact	
RPE1 Replicate-1	0/62	5/62	
RPE1 Replicate-2	0/52	2/52	
Total	0% (n=114)	6% (n=114)	
RPE1-ΔDXZ4a Replicate-1	0/56	6/56	
RPE1-ΔDXZ4a Replicate-2	0/51	1/51	
Total	0% (n=107)	7% (n=107)	p = 1.000
RPE1-ΔDXZ4i Replicate-1	0/50	0/50	
RPE1-ΔDXZ4i Replicate-2	0/52	2/52	
Total	0% (n=102)	2% (n=102)	p = 0.1713

C



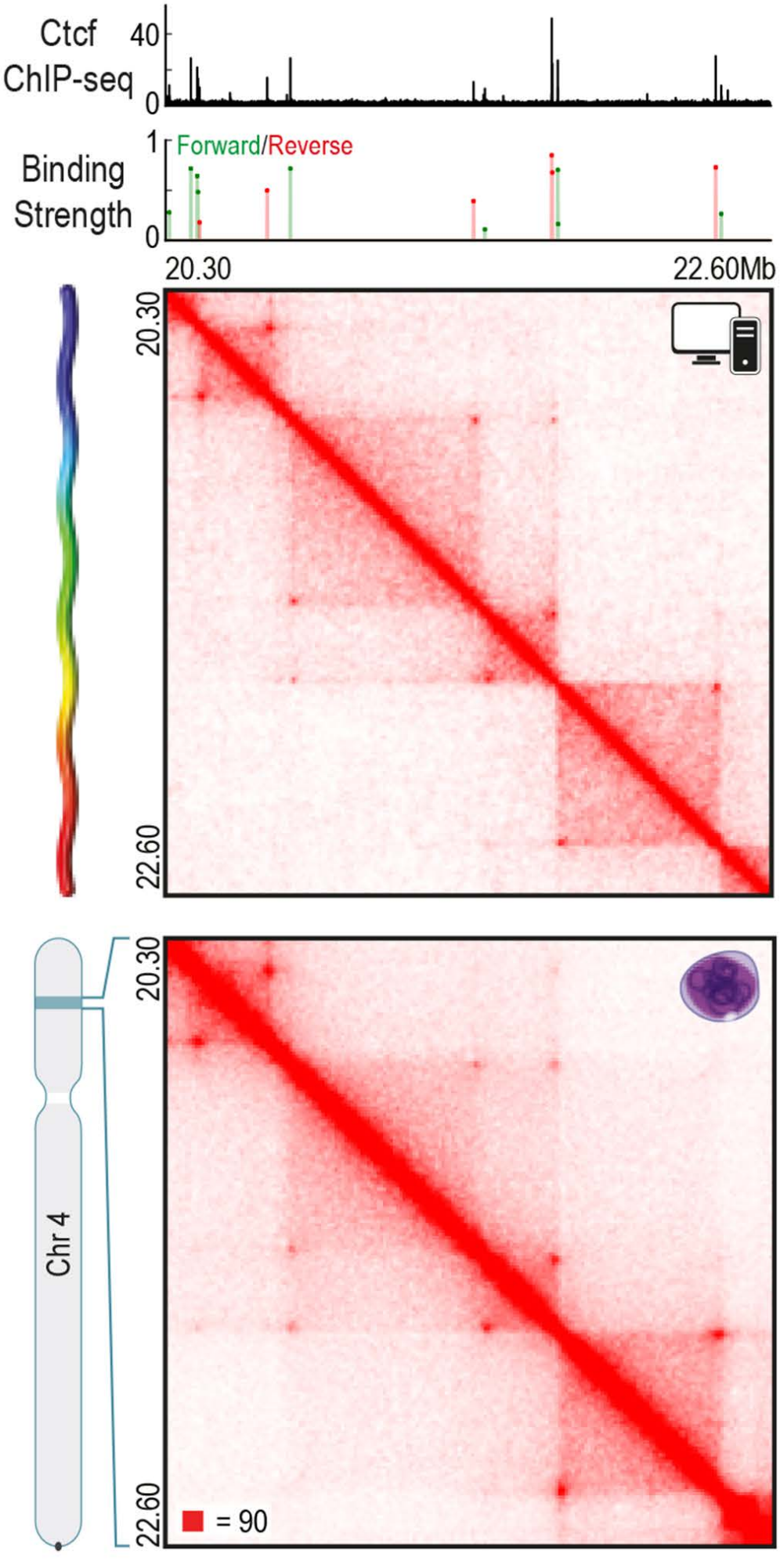
Supplementary Fig. S10: Analysis of Xi transcription in RPE1 deletion clones.

(*A*) Examples of scoring procedure. Escape or non-escape is determined by comparing the RNA-FISH signal from a gene of interest (green) to the signal from XIST (red), which coats Xi. White arrows indicate the Xi allele; white arrowheads indicate Xa. (*B*) Percentage of cells with expression of the indicated gene from Xi. Each data point is derived from between 100-140 nuclei comprising two replicate RNA-FISH experiments (Supplementary Table S4). Standard deviations are shown.


A
JARID1C: Escape
XIST
5 μm
JARID1C: No Escape
XIST
5 μm
CHM
XIST
15 μm
B
Escape genes
Subject genes
Percentage of nuclei showing expression from Xi
100
80
60
40
20
0
RPE1
RPE1ΔDXZ4i (1)
RPE1ΔDXZ4i (2)
RPE1ΔDXZ4i (3)
ZFX
UTX
CDK16
JARID1C
MORF4L2
CHM
AMMECR1
TMEM164
STAG2
SH3BGRL
SUHW3
G6PD

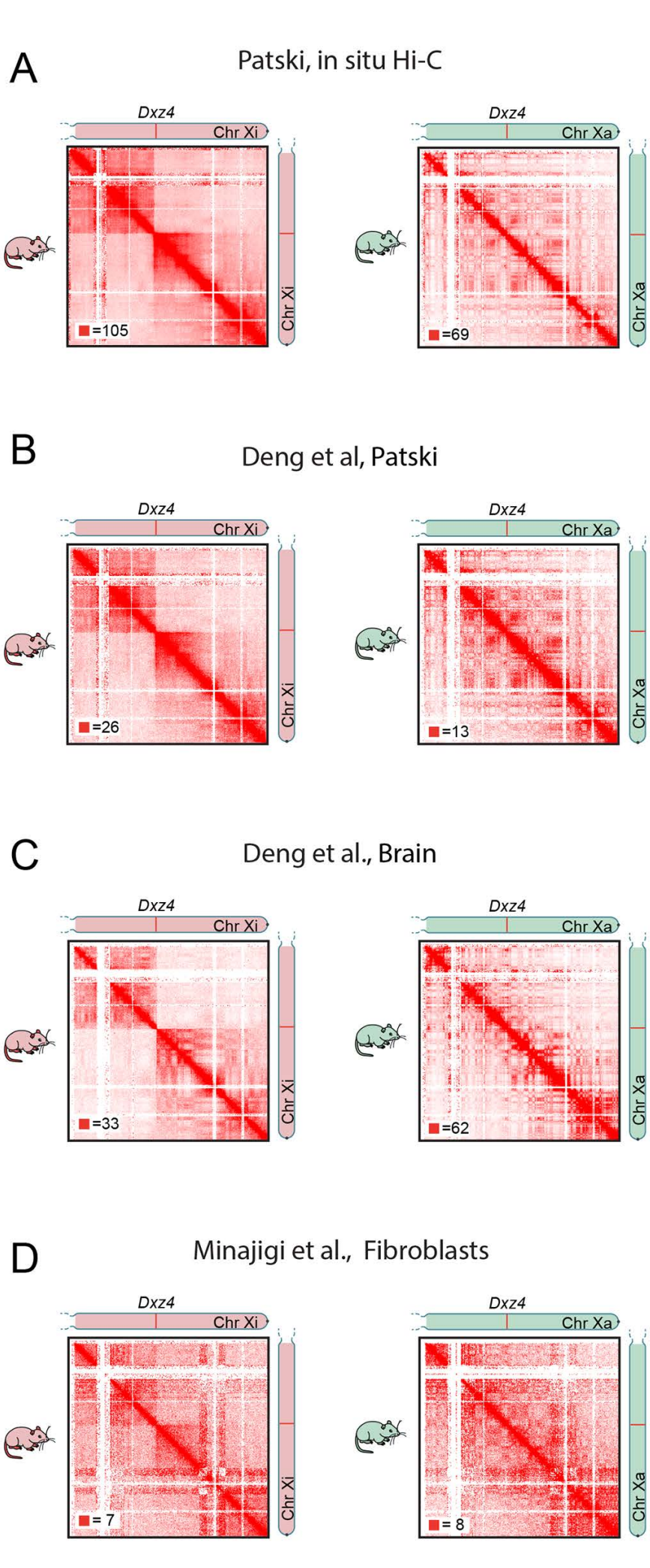
Supplementary Fig. S11: Partially arresting extrusion complexes recapitulate contact maps.

Contact map for molecular dynamics simulations of a region in GM12878 (top) accurately recapitulates the experimental Hi-C contact map of the same region (bottom). In these simulations, based on (12), extrusion complexes undergo partial arrest at CTCF binding sites, allowing a small probability of unbinding (~1 in 5000) at each step. Extrusion binding probabilities are derived directly from GM12878 CTCF ChIP-seq data (top). The small symbol in the upper right corner of the top map indicates that this map was generated in silico, whereas the symbol in the top right corner of the bottom image indicates that this map was generated from experimental data.


Ctcf
ChIP-seq
40
0
Binding
Strength
1
0
Forward/Reverse
20.30
22.60Mb
20.30
22.60
Chr 4
20.30
22.60
= 90

Supplementary Fig. S12: Comparison of phased murine Hi-C maps.

Xi (left) and Xa (right) maps are shown at 500 kb resolution for 4 different murine libraries: *(A)* our Patski data, *(B)* Deng et al. Patksi, *(C)* Deng et al. brain, and *(D)* Minajigi et al. fibroblast. The Xi in Patski and fibroblast cells do not exhibit plaid patterning, while weak compartmentalization is seen in the brain cells.


A
Patski, in situ Hi-C
Dxz4
Chr Xi
Chr Xi
=105
Dxz4
Chr Xa
Chr Xa
=69
B
Deng et al, Patski
Dxz4
Chr Xi
Chr Xi
=26
Dxz4
Chr Xa
Chr Xa
=13
C
Deng et al., Brain
Dxz4
Chr Xi
Chr Xi
=33
Dxz4
Chr Xa
Chr Xa
=62
D
Minajigi et al., Fibroblasts
Dxz4
Chr Xi
Chr Xi
= 7
Dxz4
Chr Xa
Chr Xa
= 8

Supplementary Table S1: Hi-C Experiments

Supplementary Table S2: 3-color DNA FISH data set

Supplementary Table S3: X-linked gene expression changes by RNA-Seq

Supplementary Table S4: RNA FISH data set

Supplementary Table S5: PC half-correlation lengths

Supplementary Table S6: Superdomain strengths at *DXZ4*/*Dxz4*

Table S1

Library	Sequenced Reads	Unmapped	PCR Duplicates	Intra-fragment	Total Contacts	Interchromosomal	Intra long range (>20Kb)
RPE1-ΔDXZ4a							
DarrowHuntley-2015-HIC001	640,677,346	7,929,585 (1%)	47,675,392 (7%)	6,198,590 (1%)	488,579,188 (76%)	125,103,723 (20%)	265,142,718 (41%)
RPE1 WT							
DarrowHuntley-2015-HIC002	269,646,003	3,283,687 (1%)	18,785,033 (7%)	3,503,295 (1%)	199,575,990 (74%)	40,643,062 (15%)	106,121,987 (39%)
DarrowHuntley-2015-HIC003	163,093,437	1,901,120 (1%)	10,634,414 (7%)	2,225,864 (1%)	120,497,188 (74%)	24,397,386 (15%)	63,317,615 (39%)
DarrowHuntley-2015-HIC004	193,268,170	2,363,811 (1%)	14,947,381 (8%)	1,815,977 (1%)	141,054,168 (73%)	19,972,916 (10%)	77,471,365 (40%)
DarrowHuntley-2015-HIC005	384,774,119	4,675,234 (1%)	25,961,916 (7%)	3,896,350 (1%)	291,186,776 (76%)	40,468,922 (11%)	159,578,063 (41%)
DarrowHuntley-2015-HIC006	281,285,484	3,186,267 (1%)	17,952,059 (6%)	2,110,447 (1%)	217,141,784 (77%)	54,979,689 (20%)	120,248,269 (43%)
DarrowHuntley-2015-HIC007	129,291,537	1,527,151 (1%)	5,562,000 (4%)	1,420,989 (1%)	103,121,703 (80%)	26,130,463 (20%)	56,744,431 (44%)
RPE1-ΔDXZ4i							
DarrowHuntley-2015-HIC008	598,757,212	7,630,581 (1%)	65,182,635 (11%)	5,221,337 (1%)	413,149,350 (69%)	65,775,478 (11%)	220,146,985 (37%)
DarrowHuntley-2015-HIC009	209,474,896	2,571,493 (1%)	39,953,702 (19%)	957,183 (0%)	128,932,280 (62%)	22,009,860 (11%)	70,113,812 (33%)
DarrowHuntley-2015-HIC010	151,205,505	1,981,157 (1%)	35,206,635 (23%)	493,152 (0%)	84,669,422 (56%)	12,133,457 (8%)	44,924,221 (30%)
DarrowHuntley-2015-HIC011	549,920,815	6,242,996 (1%)	63,225,183 (12%)	7,213,649 (1%)	378,401,417 (69%)	66,915,846 (12%)	200,802,938 (37%)
DarrowHuntley-2015-HIC012	127,330,455	1,404,587 (1%)	5,183,321 (4%)	1,043,767 (1%)	99,237,262 (78%)	22,530,466 (18%)	55,451,926 (44%)
DarrowHuntley-2015-HIC013	113,453,546	1,158,170 (1%)	4,449,773 (4%)	1,114,844 (1%)	88,182,315 (78%)	19,650,947 (17%)	48,949,756 (43%)
Patski							
DarrowHuntley-2015-HIC014	261,758,706	63,661,403 (24%)	1,589,580 (0%)	4,882,056 (2%)	128,034,628 (49%)	28,536,046 (11%)	75,586,489 (29%)
DarrowHuntley-2015-HIC015	101,205,975	24,980,401 (25%)	1,942,030 (2%)	637,268 (1%)	48,227,963 (48%)	12,056,115 (12%)	28,265,501 (28%)
DarrowHuntley-2015-HIC016	133,721,027	94,209,333 (70%)	545,642 (0%)	964,389 (1%)	22,840,199 (17%)	6,714,998 (5%)	13,095,451 (10%)
DarrowHuntley-2015-HIC017	123,373,502	85,983,485 (70%)	518,553 (0%)	1,129,254 (1%)	21,635,093 (18%)	6,371,978 (5%)	12,063,874 (10%)
DarrowHuntley-2015-HIC018	144,732,259	99,639,774 (69%)	629,972 (0%)	1,172,046 (1%)	26,078,552 (18%)	7,583,542 (5%)	14,778,898 (10%)
DarrowHuntley-2015-HIC019	318,324,722	77,354,162 (24%)	4,784,593 (2%)	11,304,198 (4%)	144,817,923 (45%)	28,739,176 (9%)	85,875,447 (27%)
Rhesus							
DarrowHuntley-2015-HIC020	701,157,628	26,796,423 (4%)	17,576,371 (3%)	11,528,034 (2%)	449,055,675 (64%)	115,956,251 (17%)	231,858,308 (33%)
GM12878 COLA							
DarrowHuntley-2015-HIC021	388,565,036	41,316,595 (11%)	35,938,830 (9%)	3,419,444 (1%)	247,990,915 (64%)	60,101,331 (15%)	35,805,138 (9%)
TOTAL	5,985,017,380	491,684,397 (8%)	364,729,814 (6%)	57,304,655 (1%)	3,145,363,201 (53%)	630,714,070 (11%)	1,718,679,746 (29%)

Table S2: 3-color DNA FISH data set

GM12878 (46,XX)

Slide	No allele contacts	1x allele *FIRRE* & *ICCE* only	Both alleles *FIRRE* & *ICCE*	1x allele *DXZ4* & *ICCE* only	Both alleles of *DXZ4* & *ICCE*	1x allele of *DXZ4* & *FIRRE* Only	Both alleles of *DXZ4* & *FIRRE*	1x allele each of *DXZ4*, *FIRRE* & *ICCE*	Both alleles of *DXZ4*, *FIRRE* & *ICCE*
1	29 (56%)	0 (0%)	0 (0%)	6 (11%)	0 (0%)	5 (10%)	0 (0%)	12 (23%)	0 (0%)
2	39 (66%)	0 (0%)	0 (0%)	0 (0%)	1 (<2%)	13 (22%)	1 (<2%)	5 (8%)	0 (0%)

GM6992 (46,XY)

Slide	No contacts	Contact *FIRRE* & *ICCE* only	Contact *DXZ4* & *ICCE* only	Contact *DXZ4* & *FIRRE* Only	Contact of *DXZ4*, *FIRRE* & *ICCE*
1	107 (96%)	1 (<1%)	2 (2%)	2 (2%)	0 (0%)
2	102 (95%)	0 (0%)	0 (0%)	5 (5%)	0 (0%)

Table S3: X-linked gene expression changes by RNA-Seq

Gene	Coordinates (hg19)	Ensembl identifier	Log fold change
PLCXD1	200401-207424	ENST00000448477	-2.927
PPP2R3B	308346-334779	ENST00000381625	-6.255
ZBED1	2404455-2419008	ENST00000381223	4.862
FRMPD4	12156585-12742642	ENST00000380682	9.148
TRAPPC2	13733656-13752664	ENST00000519382	-5.061
TRAPPC2	13730363-13752754	ENST00000359680	-6.547
SPANXC	140335596-140336629	ENST00000358993	7.603
SPANXD	140785568-140786896	ENST00000370515	6.323
MAGEC1	140991680-140997174	ENST00000285879	6.226
HAUS7	152713124-152736089	ENST00000370211	-5.171
TREX2	152713336-152736045	ENST00000370212	5.639

Supplementary Table S4: RNA FISH data set

	Escape genes				Subject genes							
Gene	ZFX	UTX	CDK16	JARID1C	MORF4L2	CHM	AMMECR1	TMEM164	STAG2	SH3BGRL	SUHW3	G6PD
RPE1	48/51	40/53	16/52	26/51	2/51	0/49	0/53	1/52	1/50	2/54	1/51	4/53
	43/52	40/53	16/50	33/67	4/55	2/64	1/55	4/54	1/53	0/51	0/50	4/53
RPE1 ΔDXZ4i (1)	49/50	32/53	20/53	30/51	4/51	13/60	6/51	7/51	1/51	3/54	7/50	3/50
	49/53	43/52	7/51	25/50	5/60	21/80	12/53	3/50	1/51	9/55	2/52	3/51
RPE1 ΔDXZ4i (2)	59/63	14/50	7/54	7/61	2/61	1/50	2/51	4/56	2/54	0/58	1/53	0/60
	56/57	18/50	3/52	11/57	1/56	1/50	0/55	1/55	0/57	0/61	2/55	0/54
RPE1 ΔDXZ4i (3)	48/50	29/61	16/56	10/51	3/53	2/50	4/50	2/51	0/53	0/50	0/54	0/57
	52/54	35/55	10/51	27/52	3/52	2/50	0/50	2/56	0/57	1/59	1/57	5/57

Table S5: PC half-correlation lengths

Phased Map (500kb Resolution)	Half Correlation Length (Mb)
GM12878 Xa	2
GM12878 Xi	1
RPE1-WT Xa	1
RPE1-WT Xi	4
RPE1-ΔDXZ4a Xa	1
RPE1-ΔDXZ4a Xi	4
RPE1-ΔDXZ4i Xa	1
RPE1-ΔDXZ4i Xi	2.5
Patski Xa	1.5
Patski Xi	30.5
*Deng et al., Patski Xa	1.5
*Deng et al., Patski Xi	11.5
*Deng et al., Brain Xa	1
*Deng et al., Brain Xi	10
*Minajigi et al., Fibroblast Xa	14
*Minajigi et al., Fibroblast Xi	17.5

* External Dataset with fewer reads

Table S6: Superdomain strengths at *DXZ4/Dxz4*

Map (500kb resolution)	zscore S(DXZ4)	p-value
GM128787 paternal (Xi)	3.335042138	0.000426432
GM128787 maternal (Xa)	0.185232799	0.426523222
RPE- WT maternal (Xi)	5.383512771	3.65E-08
RPE1-WT paternal (Xa)	1.056132642	0.14545381
RPE1-delXa maternal (Xi)	5.702490939	5.90E-09
RPE1-delXa paternal (Xa)	1.603290905	0.054435222
RPE1-delXi maternal (Xi)	-0.851431816	0.802735238
RPE1-delXi paternal (Xa)	-1.515800698	0.935215122
Patski maternal (Xi)	4.990979361	3.00E-07
Patski (alt) maternal (Xi)	2.959590171	0.001540243
Patski paternal (Xa)	0.488609493	0.312559095
Patski (alt) paternal (Xa)	0.484738465	0.313930939
rhesus macaque	3.173150546	0.000753972