SUPPLEMENTAL MATERIALS

A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation

Anand Minajigi [1]†, John Froberg [1]†, Chunyao Wei [1], Hongjae Sunwoo [1], Barry Kesner [1], David Colognori [1], Derek Lessing [1], Bernhard Payer [1#], Myriam Boukhali [2], Wilhelm Haas [2], Jeannie T. Lee[1]*

[1] Howard Hughes Medical Institute; Department of Molecular Biology, Massachusetts General Hospital, Boston, MA USA; Department of Genetics, Harvard Medical School, Boston, MA, USA.

[2] Massachusetts General Hospital Cancer Center, Charlestown, Boston, MA; Department of Medicine, Harvard Medical School, Boston, MA, USA.

† Equal contribution

* Corresponding author: lee@molbio.mgh.harvard.edu

This PDF file includes:

Materials and Methods, Supplementary Text, Tables S1-S3 (as excel files), Figures S1-S15, References

SUPPLEMENTAL DISCUSSION

Here we discuss similarities and differences between our findings and those of two recent studies published while our manuscript was under review. One study performed Xist proteomic analysis on formaldehyde cross-linked chromatin of mouse cells and identified 30 Xist-specific proteins, via the ChIRP-MS methodology (*1*). A second study carried out Xist proteomic analysis on UV-crosslinked chromatin and identified 10 proteins, via the RAP-MS methodology (*2*). The proteins identified by the two studies are similar, with each highlighting SPEN. Our iDRiP method identified a larger set of interactors (80-250 proteins), largely inclusive of the proteins identified by the two groups, respectively. However, in addition, our study uncovered numerous other factors, such as multiple cohesin subunits, CTCF, multiple SWI/SNF subunits, three topoisomerases, condensins, RNA helicases, SUN2, AURKB, and macroH2A. Notably, our proteomics also isolated the known interactors, PRC2 and ATRX, whereas neither factor was identified using ChIRP-MS and RAP-MS (*1, 2*). The identification of these positive controls suggests that iDRiP is highly sensitive and further confirms a direct and specific interaction between Xist RNA, PRC2, and ATRX. The iDRiP-MS, ChIRP-MS, and RAP-MS methodologies have significant technical differences. Firstly, to minimize background due to DNA-bound proteins, a key step employed by iDRiP is inclusion of DNase I treatment before elution of complexes. We believe that this enhanced the specificity and sensitivity of iDRiP. Secondly, ChIRP-MS employed 3% formaldehyde crosslinking for 30 minutes, which could have resulted in large crosslinked networks of chromatin and favored identification of abundant factors (such as nuclear matrix factors and other general RNA-binding proteins) that do not necessarily directly interact with Xist. On the other hand, RAP-MS employed UV crosslinking, but a major difference is the use of a male cell line in which Xist is overexpressed from a doxycycline-inducible promoter. Our study was performed on a female cell line expressing endogenous, physiological levels of Xist. In our hands, doxycycline induction leads to >30-fold overexpression of Xist, which in turn could also favor identification of the highly abundant nuclear matrix factors and other general RNA-binding proteins over specific chromatin factors, such as cohesins, SWI/SNF, PRC2, ATRX, and topoisomerases. Here, we report all candidates obtained from the mass spectrometry, without applying any filters, and provide enrichment scores. The roles of numerous other interactors can be investigated in future.

MATERIALS AND METHODS

Identification of Direct RNA interacting Proteins (iDRiP)

Mouse Embryonic Fibroblasts (MEFs) were irradiated with UV light at 200 mJ energy (Stratagene 2400) after rinsing with PBS. The pellets were resuspended in CSKT-0.5% (10 mM PIPES, pH 6.8, 100 mM NaCl, 3 mM $MgCl_2$, 0.3 M sucrose, 0.5% Triton X-100, 1 mM PMSF) for 10 min at 4 ºC followed by a spin. The pellets were again resuspended in Nuclear Isolation Buffer (10 mM Tris pH 7.5, 10 mM KCl, 0.5% Nonidet-P 40, 1x protease inhibitors, 1 mM PMSF), and rotated at 4ºC for 10 min. The pellets were collected after a spin, weighed, flash frozen in liquid nitrogen, and stored at -80 ºC until use.

Approximately, equal amounts of female and male UV cross linked pellets were thawed and resuspended for treatment with Turbo DNase I in the DNase I digestion buffer (50 mM Tris pH 7.5, 0.5% Nonidet-P 40, 0.1% sodium lauroyl sarcosine, 1x protease inhibitors, SuperaseIn). The tubes were rotated at 37 ºC for 45 min. The nuclear lysates were further solubilized by adding 1% sodium lauroyl sarcosine, 0.3 M lithium chloride, 25 mM EDTA and 25 mM EGTA to final concentrations and continued incubation at 37 ºC for 15 min. The lysates were mixed with biotinylated DNA probes (Table S3) prebound to the streptavidin magnetic beads (MyOne streptavidin C1 Dyna beads, Invitrogen) and incubated at 55 ºC for 1 hr before overnight incubation at 37 ºC in the hybridization chamber. The beads were washed three times in Wash Buffer (10 mM Tris, pH 7.5, 0.3 M LiCl, 1% LDS, 0.5% Nonidet-P 40, 1x protease inhibitor) at room temperature followed by treatment with Turbo DNase I in DNase I digestion buffer with the addition of 0.3 M LiCl, protease inhibitors, and superaseIn at 37 ºC for 20 min. Then, beads were washed two more times in the Wash Buffer. For MS analysis, elution was done in Elution Buffer (10 mM Tris, pH 7.5, 1 mM EDTA) at 70 ºC for 4 min followed by brief sonication in Covaris. For the quantification of pulldown efficiency, MEFs, without crosslinking, were used and elution was done at 95 ºC. The elute was used for RNA isolation and RT-qPCR. When crosslinked MEFs were used, elute was subjected for proteinase-K treatment (50 mM Tris pH 7.5, 100 mM NaCl, 0.5% SDS, 10 µg proteiase K) for 1 hr at 55 ºC. RNA were isolated by Trizol and quantified with SYBR green qPCR. Input samples were used to make standard curve by 10 fold dilutions, to which the RNA pulldown efficiencies were compared and calculated. The efficiency of Xist pulldown was relatively lower after UV crosslinking, similar to (*3, 4*).

Quantitative proteomics

Proteins co-enriched with Xist from female or male cells were quantitatively analyzed either using a label-free approach based on spectral-counting (*5*) or by multiplexed quantitative proteomics using tandem-mass tag (TMT) reagents (*6, 7*) on an Orbitrap Fusion mass spectrometer (Thermo Scientific). Disulfide bonds were reduced with ditheiothreitol (DTT) and free thiols alkylated with iodoacetamide as described previously (*8*). Proteins were then precipitated with tricholoracetic acid, resuspended in 50 mM HEPES (pH 8.5) and 1 M urea and digested first with endoproteinase Lys-C (Wako) for 17 hours at room temperature and then with sequencing-grade trypsin (Promega) for 6 hours at 37 ˚C. Peptides were desalted over Sep-Pak C_{18} solid-phase extraction (SPE) cartridges (Waters), the peptide concentration was determined using a BCA assay (Thermo Scientific). For the label-free analysis peptides were then dried and re-suspended in 5 % formic acid (FA) and 5 % acetonitrile (ACN) and 5 µg of

peptides were analyzed by mass spectrometry as described below. For the multiplexed quantitative analysis a maximum of 50 µg of peptides were labeled with one out of the available TMT-10plex reagents (Thermo Scientific) (*6*). To achieve this, peptides were dried and resuspended in 50 µl of 200 mM HEPES (pH 8.5) and 30 % (ACN) and 10 µg of the TMT in reagent in 5 µl of anhydrous ACN was added to the solution, which was incubated at room temperature (RT) for one hour. The reaction was then quenched by adding 6 µl of 5 % (w/v) hydroxylamine in 200 mM HEPES (pH 8.5) and incubation for 15 min at RT. The labeled peptide mixture was then subjected to a fractionation using basic pH reversed phase liquid chromatography (bRPLC) on an Agilent 1260 Infinity HPLC system equipped with an Agilent Extend-C18 column (4.6x250 mm; particle size, 5 µm) basically as described previously (*9*). Peptides were fractionated using a gradient from 22-35 % ACN in 10 mM ammonium bicarbonate over 58 min at a flowrate of 0.5 ml/min. Fractions of 0.3 ml were collected into a 96-well plate to then be pooled into a total twelve fractions (A1-A12, B1-B12, etc.) that were dried and re-suspended in 8 µl of 5 % FA and 5 % ACN, 3 of which were analyzed by microcapillary liquid chromatography tandem mass spectrometry on an Orbitrap Fusion mass spectrometer and using a recently introduced multistage (MS3) method to provide highly accurate quantification (*10*).

The mass spectrometer was equipped with an EASY-nLC 1000 integrated autosampler and HPLC pump system. Peptides were separated over a 100 µm inner diameter microcapillary column in-house packed with first 0.5 cm of Magic C4 resin (5 µm, 100 Å, Michrom Bioresources), then with 0.5 cm of Maccel C_{18} resin (3 µm, 200 Å, Nest Group) and 29 cm of GP-C18 resin (1.8 µm, 120 Å, Sepax Technologies). Peptides were eluted applying a gradient of 8-27 % ACN in 0.125 % formic acid over 60 min (label-free) and 165 min (TMT) at a flow rate of 300 nl/min. For label-free analyses we applied a tandem-MS method where a full-MS spectrum (MS1; m/z 375-1500; resolution $6x10^4$; AGC target, $5x10^5$; maximum injection time, 100 ms) was acquired using the Orbitrap after which the most abundant peptide ions where selected for linear ion trap CID-MS2 in an automated fashion. MS2 scans were done in the linear ion trap using the following settings: quadrupole isolation at an isolation width of 0.5 Th; fragmentation method, CID; AGC target, $1x10^4$; maximum injection time, 35 ms; normalized collision energy, 30 %). The number of acquired MS2 spectra was defined by setting the maximum time of one experimental cycle of MS1 and MS2 spectra to 3 sec (Top Speed). To identify and quantify the TMT-labeled peptides we applied a synchronous precursor selection MS3 method (*8, 10, 11*) in a data dependent mode. The scan sequence was started with the acquisition of a full MS or MS1 one spectrum acquired in the Orbitrap (m/z range, 500-1200; other parameters were set as described above), and the most intense peptide ions from detected in the full MS spectrum were then subjected to MS2 and MS3 analysis, while the acquisition time was optimized in an automated fashion (Top Speed, 5 sec). MS2 scans were performed as described above. Using synchronous precursor selection the 10 most abundant fragment ions were selected for the MS3 experiment following each MS2 scan. The fragment ions were further fragmented using the HCD fragmentation (normalized collision energy, 50 %) and the MS3 spectrum was acquired in the Orbitrap (resolution, 60,000; AGC target, $5x10^4$; maximum injection time, 250 ms).

Data analysis was performed on an on an in-house generated SEQUEST-based (*12*) software platform. RAW files were converted into the mzXML format using a modified version of ReAdW.exe. MS2 spectra were searched against a protein sequence database containing all protein sequences in the mouse UniProt database (downloaded 02/04/2014) as well as that of known contaminants such as porcine trypsin. This target component of the database was followed by a decoy component containing the same protein sequences but in flipped (or reversed) order (*13*). MS2 spectra were matched against peptide sequences with both termini

consistent with trypsin specificity and allowing two missed trypsin cleavages. The precursor ion m/z tolerance was set to 50 ppm, TMT tags on the N-terminus and on lysine residues (229.162932 Da, only for TMT analyses) as well as carbamidomethylation (57.021464 Da) on cysteine residues were set as static modification, and oxidation (15.994915 Da) of methionines as variable modification. Using the target-decoy database search strategy (*13*) a spectra assignment false discovery rate of less than 1 % was achieved through using linear discriminant analysis with a single discriminant score calculated from the following SEQUEST search score and peptide sequence properties: mass deviation, XCorr, dCn, number of missed trypsin cleavages, and peptide length (*14*). The probability of a peptide assignment to be correct was calculated using a posterior error histogram and the probabilities for all peptides assigned to a protein were combined to filter the data set for a protein FDR of less than 1 %. Peptides with sequences that were contained in more than one protein sequence from the UniProt database were assigned to the protein with most matching peptides (*14*).

For a quantitative estimation of protein concentration using spectral-counts we simply counted the number of MS2 spectra assigned to a given protein (Table S1). TMT reporter ion intensities were extracted as that of the most intense ion within a 0.03 Th window around the predicted reporter ion intensities in the collected MS3 spectra. Only MS3 with an average signal-to-noise value of larger than 28 per reporter ion as well as with an isolation specificity (*8*) of larger than 0.75 were considered for quantification. Reporter ions from all peptides assigned to a protein were summed to define the protein intensity. A two-step normalization of the protein TMT-intensities was performed by first normalizing the protein intensities over all acquired TMT channels for each protein based to the median average protein intensity calculated for all proteins. To correct for slight mixing errors of the peptide mixture from each sample a median of the normalized intensities was calculated from all protein intensities in each TMT channel and the protein intensities were normalized to the median value of these median intensities.

UV RIP

The protocol followed is similar to the one described in (*15*). Briefly, MEFs were crosslinked with UV light at 200 mJ and collected by scraping in PBS. Cell pellets were resuspended in CSKT-0.5% for 10 min at 4 ºC followed by a spin. The nuclei were resuspended in the UV RIP buffer (PBS buffer containing 300 mM NaCl (total), 0.5% Nonidet-P 40, 0.5% sodium deoxycholate, and 1x protease inhibitors) with Turbo DNase I 30 U/IP for 30 min at 37 ºC. Supernatants were collected after a spin and incubated with 5 µg specific antibodies prebound to 40 µl protein-G magnetic beads (Invitrogen) at 4 ºC overnight. Beads were washed three times with cold UV RIP buffer. The beads were resuspended in 200 µl Turbo DNase I buffer with 20 U Turbo DNase, SuperaseIN, 1x protease inhibitors) for 30 min at 37 ºC. The beads were resuspended and washed three more times in the UV RIP washing buffer containing 10 mM EDTA. The final 3 washes were given after three fold dilution of UV RIP washing buffer. The beads were resuspended in 200 µl proteinase-K buffer with 10 µg proteinase-K and incubated at 55 ºC for 1 hr. RNA was isolated by Trizol and pulldown efficiencies were calculated by SYBR qPCR using input for the standard curve.

Generation of Xi-TgGFP clonal fibroblasts

Xi-TgGFP (68-5-11) tail-tip fibroblasts (TTF) were initially derived from a single female pup, a daughter of a cross between a *M. castaneus* male and a *M. musculus* female, homozygous for an X-linked GFP transgene driven by a strong, ubiquitous promoter (*16*). The fibroblasts were immortalized by SV40 transformation, and clonal lines were derived from individual GFP-negative cells selected by fluorescence-activated cell sorting. In our experience, occasional clones with undetectable GFP expression nevertheless have the transgene located on the active X chromosome. Thus, we confirmed the GFP transgene location on the inactive X for the particular clone used here, 68-5-11 (see fig. S2).

Generation of stable KD of Xi-TgGFP TTF and 16.7 ES cells

The protocol is as described in:

http://www.broadinstitute.org/rnai/public/resources/protocols

A cocktail of 3 shRNA viruses were used for infections (Table S2) followed with puromycin selection. In all the experiments, non-clonal knockdown cells were used.

Assay for the reactivation of Xi-TgGFP

Approximately, 125,000-150,000 Xi-TgGFP (68-5-11) cells were plated along with control (shNegative control, i.e., shNC) cells treated with DMSO or stable KD cells treated with 0.3 µM azacytidine and 0.3 µM Etoposide for 3 days in 6 well plates. RNA was isolated by Trizol twice, with an intermittent TurboDNase treatment after the first isolation for 30 min at 37 ºC. One µg RNA was used for each of the RT+ and RT- reactions (Superscript III, Invitrogen) followed by the SYBR green qPCR using the primers listed in supplementary Table S3, with annealing temperature of 60 ºC for 45 cycles. The relative efficiency of Xi-TgGFP reactivations was calculated by comparing to U1 snRNA as the internal control.

ImmunoFISH

Cells were grown on coverslips, rinsed in PBS, pre-extracted in 0.5% CSKT on ice, washed once in CSK, followed by fixation with 4% paraformaldehyde in PBS at room temperature. After blocking in 1% BSA in PBS for 20 min supplemented with 10 mM VRC (New England Biolabs) and RNase inhibitor (Roche), incubation was carried out with primary antibodies (Table S3) at room temperature for 1 hr. Cells were washed three times in PBST-0.02% Tween-20. After incubating with secondary antibody at room temperature for 30 min, cells were washed three times by PBS/0.02% Tween-20. Cells were fixed again in 4% paraformaldehyde and dehydrated in ethanol series. RNA FISH was performed using a pool of Cy3B or Alexa 568 labeled Xist oligonucleotides for 4-6 hours at 42 $^{\circ}$C in a humid chamber. Cells were washed three times in 2X SSC and nuclei were counter-stained by Hoechst 33342. Cells were observed under Nikon 90i microscope equipped with 60X/1.4 N.A. objective lens, Orca ER CCD camera (Hamamatsu), and Volocity software (Perkin Elmer). Xist RNA FISH probes, a set of total 37 oligonucleotides with 5' amine modification (IDT), were labeled with NHS-Cy3B (GE Healthcare) overnight at room temperature followed by ethanol precipitation. In the case of confirmation of Xi-TgGFP cells, probes were

made by nick-translation of a GFP PCR product with Cy3-dUTP and of a plasmid containing the first exon of the mouse Xist gene, with FITC-dUTP.

Allelic ChIP-seq

Allele-specific ChIP-seq was performed according to the method of Kung et al (*17*), in two biological replicates. To increase available read depth, we pooled together two technical replicates for $Xi^{\Delta xist}/Xa^{WT}$ Rad21 replicate 1 sequenced on a 2x50bp HiSeq2500 rapid run and we also pooled two technical replicates of wild-type Rad21 replicate 1, one sequenced on a HiSeq 2x50 bp run and one on a MiSeq 2x50 bp run. All other libraries were sequenced on using 2x50 bp HiSeq2500 rapid runs. To visualize ChIP binding signal, we generated fpm-normalized bigWig files from the raw ChIP read counts for all reads (comp), mus-specific (mus) and cas-specific reads separately. For Smc1a, CTCF and Rad21, peaks were called using macs2 with default settings. To generate consensus peak sets for all three epitopes, peaks for the two wild-type and $Xi^{\Delta xist}/Xa^{WT}$ replicates were pooled and peaks present in at least two experiments were used as the common peak set. To make comparisons between allelic read counts between different experiments, we defined a scaling factor as the ratio of the total read numbers for the two experiments and multiplied the allelic reads for each peak in the larger sample by the scaling factor. We plotted the number of reads on Xi vs Xa in wild-type for all peaks on the X-chromosome to determine if there is a general bias towards binding to the Xa or the Xi. To evaluate allelic skew on an autosome, we generated plots of mus read counts vs cas read counts for all peaks on chromosome 5 from 1-140,000,000. We used this particular region of chromosome 5 because $Xi^{\Delta xist}/Xa^{WT}$ is not fully hybrid, and this is a large region of an autosome that is fully hybrid based on even numbers of read counts from input and from our Hi-Cs over this region in $Xi^{\Delta xist}/Xa^{WT}$ (data not shown). To identify peaks that are highly Xa-skewed in wild-type but bind substantially to the Xi in $Xi^{\Delta xist}/Xa^{WT}$ (restored peaks), for Xa-skewed peaks in wild-type, we plotted normalized read counts on Xi in $Xi^{\Delta xist}/Xa^{WT}$ versus read counts on Xa in wild-type. We defined restored peaks as peaks that are 1.) more than 3X Xa-skewed in wild-type 2.) have at least 5 allelic reads in wild-type 3.) exhibit normalized read counts on Xi in $Xi^{\Delta xist}/Xa^{WT}$ that are at least half the level of Xa in wild-type. This threshold ensures that all restored peaks have at least a 2X increase in binding to the Xi in $Xi^{\Delta xist}/Xa^{WT}$ relative to wild-type. We identified restored peaks using these criteria in both replicates of Smc1a and Rad21 ChIP separately, and to merge these calls into a consensus set for each epitope, we took all peaks that met criteria for restoration in at least one replicate and had at least 50% wild-type Xa read counts on Xi in $Xi^{\Delta xist}/Xa^{WT}$ in both replicates.

Allele specific RNA-seq

Xi-TgGFP TTFs (68-5-11) with the stable knock down of candidates were treated with 5'-azacytidine and etoposide at 0.3 μM each for 3 days. Strand-specific RNA-seq, the library preparation, deep sequencing, and data analysis was followed as described in (*17*). Two biological replicates of each drug treatment were produced. All libraries were sequenced with Illumina Hiseq 2000 or 2500 using 50 cycles to obtain paired end reads. To determine the allelic origin of each sequencing read from the hybrid cells, reads were first depleted of adaptors dimers and PCR duplicates, followed by the alignment to custom mus/129 and cas genomes to separate mus and cas reads. After removal of PCR duplicates, ~90% of reads were mappable.

Discordant pairs and multi-mapped reads were discarded. Reads were then mapped back to reference mm9 genome using Tophat v2.0.10 (-g 1 --no-coverage-search --read-edit-dist 3 --read-mismatches 3 --read-gap-length 3 --b2-very-sensitive --mate-inner-dist 50 --mate-std-dev 50 --library-type fr-firststrand), as previously described (*17-19*). Following alignment, gene expression levels within each library were quantified using Homer v4.7 (rna mm9 -count genes -strand + -noadj -condenseGenes) (*18*) and the normalized differential expression analyses across samples were performed by using EdgeR (*20*).

HiC library preparation and analysis

HiC libraries were generated according to the protocol in Lieberman-Aiden et al., 2009 (*21*). Two biological replicate libraries were prepared for wild-type and $Xi^{\Delta xist}/Xa^{WT}$ fibroblasts each. We obtained 150-220 million 2x50bp paired-end reads per library. The individual ends of the read-pairs were aligned to the mus and cas reference genomes separately using novoalign with default parameters for single-end alignments, and the quality score of the alignment was used to determine whether each end could be assigned to either the mus or the cas haplotype (*22*). The single-end alignments were merged into a Hi-C summary file using custom scripts. Reads were filtered for self-ligation events and short fragments (less than 1.5X the estimated insert length) likely to be random shears using Homer (*18, 23*). Hi-C contact maps were generated using Homer. "Comp" maps were made from all reads. "Xi" and "Xa" reads were from reads where at least one read-end could be assigned to either the mus or cas haplotype, respectively. A small fraction of reads (~5% of all allelic reads) aligned such that one end aligned to mus, the other to cas. These "discordant" reads were excluded from further analysis, as they are likely to be noise arising due to random ligation events and/or improper SNP annotation (*24, 25*). All contact maps were normalized using the matrix balancing algorithm of Knight and Ruiz (*26*), similar to iterative correction (*24, 27*), using the MATLAB script provided at the end of their paper. We were able to generate robust contact maps using the comp reads in one replicate at 40kb resolution, but due to the fact that only ~44% of reads align allele-specifically, we were only able to generate contact maps for the cas and mus haplotypes at 200kb. To increase our resolution, we pooled together both biological replicates and analyzed the comp contact map at 40kb resolution and the mus and cas contact maps at 100kb. We called TADs at 40kb on chrX, chr5 and chr13 using the method of Dixon et al. (*28*). specifically, we processed the normalized comp 40kb contact maps separately into a vector of directionality indices using DI_from_matrix.pl with a bin size of 40000 and a window size of 200000. We used this vector of directionality indices as input for the HMM_calls.m script and following HMM_generation, we processed the HMM and generated TAD calls by passing the HMM output to file_ends_cleaner.pl, converter_7col.pl, hmm_probablity_correcter.pl, hmm-state_caller.pl and finally hmm-state_domains.pl. We used parameters of min=2, prob=0.99, binsize=40000 as input to the HMM probability correction script.

To create a general metric describing interaction frequencies within TADs at resolution available in the allele-specific interaction maps, for each TAD, on chrX and chr5 we averaged the normalized interaction scores for all bins within each TAD, excluding the main diagonal. To make comparisons between interaction frequency over TADs between the cas (Xa) and mus (Xi) haplotypes at the resolution available with our current sequencing depth, we defend the "fraction mus" as the average interaction score for a TAD in the mus contact map divided by the sum of the average interaction scores in the mus and cas contact maps.

To discover TADs that show significantly increased interaction frequency in $Xi^{\Delta xist}/Xa^{WT}$,

we generated a null distribution of changes in average normalized interaction scores for all TADs on chromosome 5, 1-140Mb using the cas and mus contact maps. We reasoned that there would be few changes in interaction frequency on an autosome between the mus or cas contact maps for wild-type and $Xi^{\Delta xist}/Xa^{WT}$, thus the distribution of fold changes in interaction score on an autosome constitutes a null distribution. Using this distribution of fold changes allowed us to calculate a threshold fold change for an empirical FDR of 0.05, and all TADs that had a greater increase in average normalized interaction score on Xi between wild-type and $Xi^{\Delta xist}/Xa^{WT}$ were considered restored TADs. We preformed this analysis of restored TADs separately in each biological replicate using the 200kb contact maps to generate interaction scores over TADs, and using the combined data at 100kb resolution.

RNA-seq, ChIP-seq, and HiC-seq data are deposited in GEO.

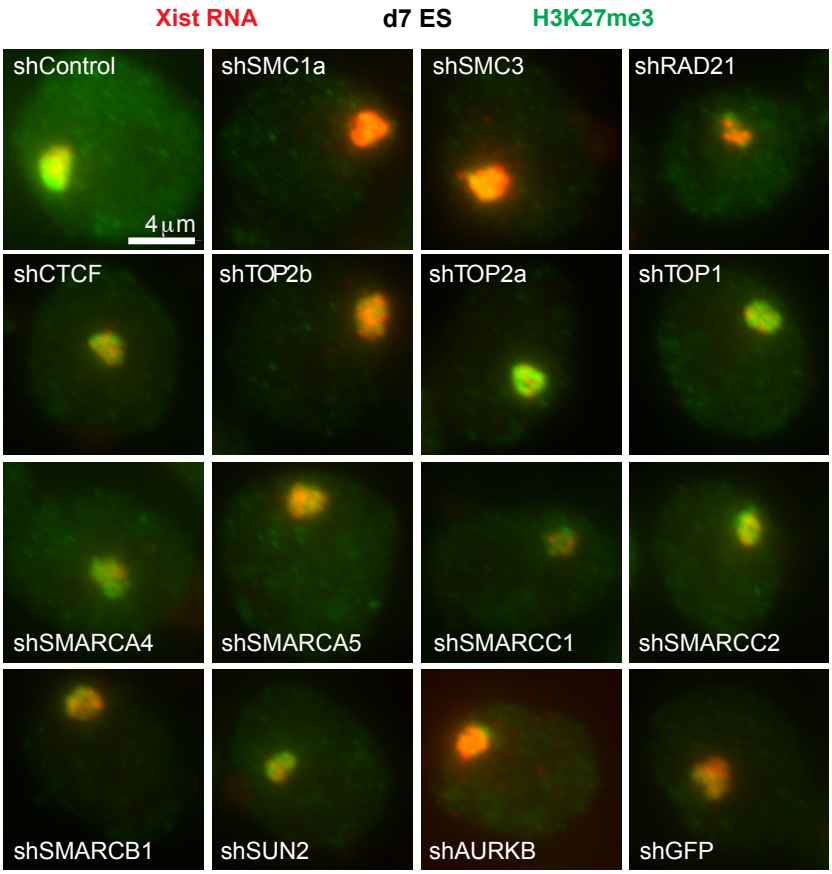


Figure S1. The microscopic images of knock down day 7 ESCs.

The stable knock down embryonic stem cells (ESCs) were differentiated after the withdrawal of LIF for seven days. On day 4, the cells were plated on the gelatin coated coverslips until day 7 of differentiation. The coverslips were prepared for immunoFISH, as described in methods, followed by imaging for Xi markers, Xist (Red) and H3K27me3 (Green).

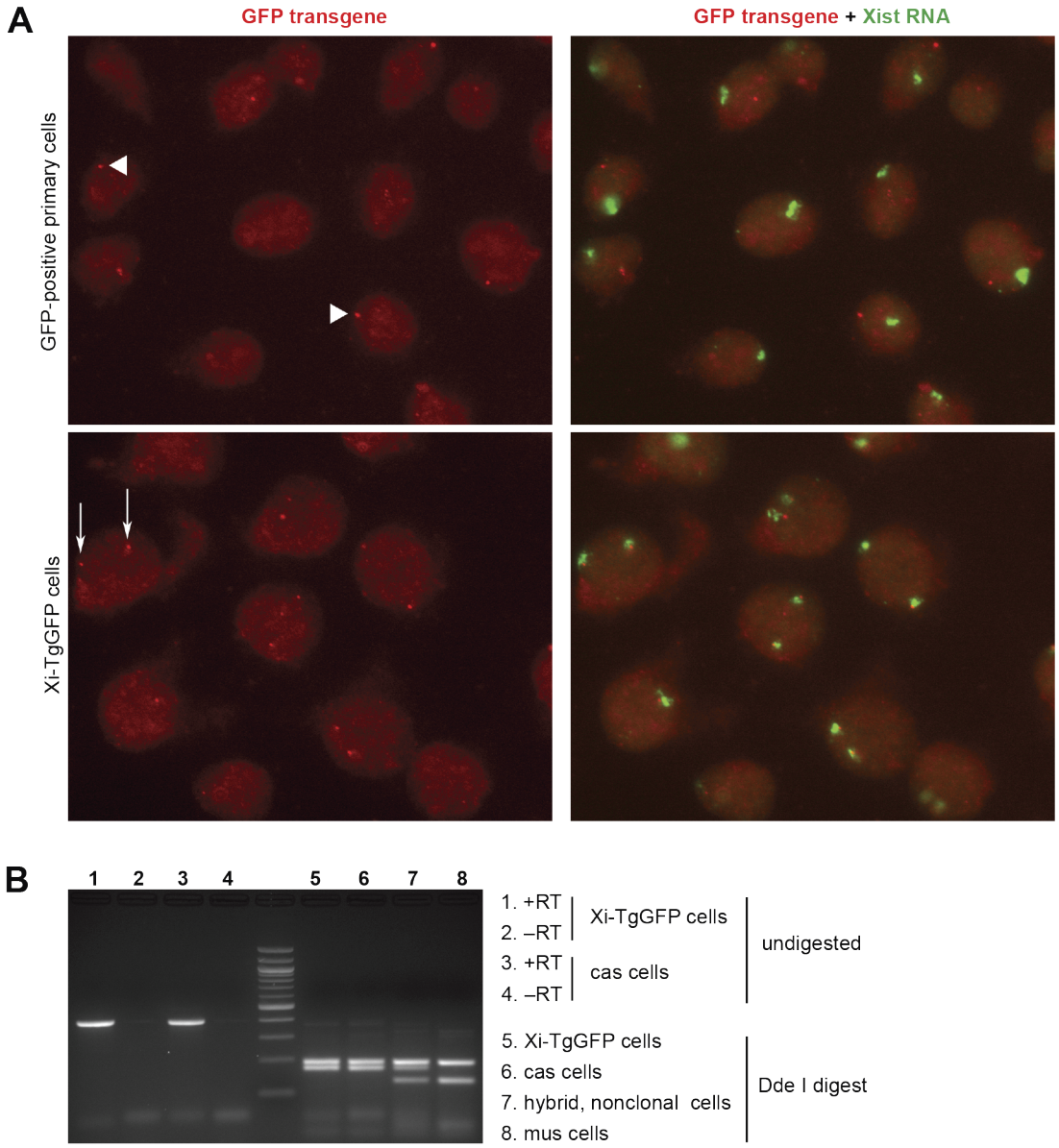


Figure S2. Confirmation that the GFP transgene of Xi-TgGFP cells is on the inactive X.

(**A)** Fluorescent In Situ Hybridization (FISH) indicates the location of the GFP transgene (DNA FISH, red) relative to the inactive X (characterized by a cloud of Xist RNA, identified by RNA FISH in green). In primary fibroblasts selected for high GFP expression (top panels), the transgene is on the active X and does not colocalize with the inactive X (examples indicated by white arrowheads). However, in Xi-TgGFP cells the GPF transgene does colocalize with the inactive X (bottom panels, arrowheads indicate one cell as an example. Xi-TgGFP cells are tetraploid; thus two inactive X chromosomes are seen per cell).

(**B)** Allele-specific expression of the X-linked gene *Mecp2* shown by RT-PCR. Hybrid Xi-TgGFP cells have one *M. musculus* (mus) X chromosome with the GFP transgene, and one *M. castaneus* (cas) X. A mus-cas single nucleotide polymorphism is detected by Dde I digest, yielding a 179-bp band for expression from the cas allele, or a 140-bp band for expression from the mus allele. A 200-bp band is common to both alleles. Only the expected cas allele of *Mecp2* is expressed in Xi-TgGFP cells (lanes 1, 2, 5), as for purely cas cells (lanes 3, 4, 6), and in contrast to cells of a pure mus background (lane 8), or from a non-clonal hybrid cell population with expression from both alleles (lane 7).

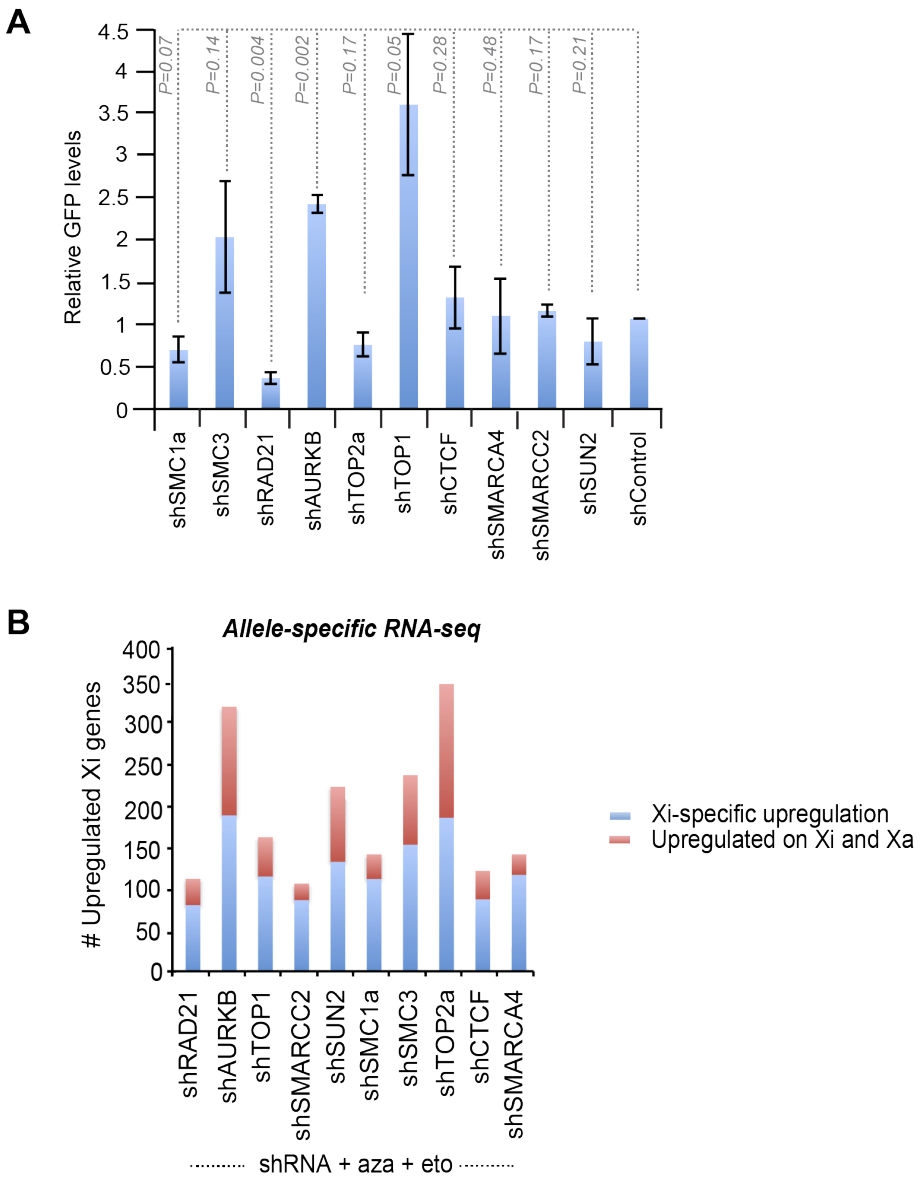
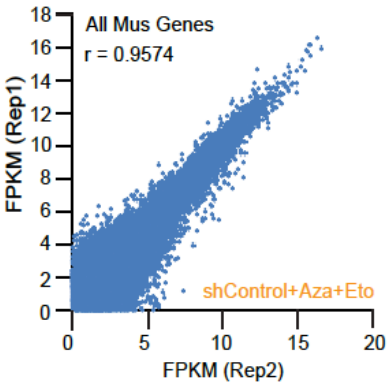
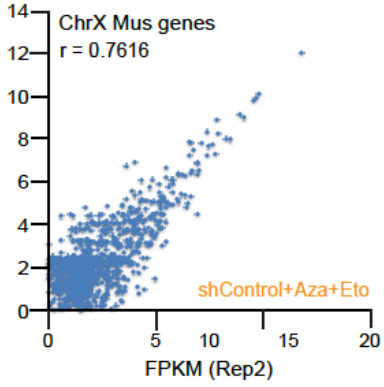


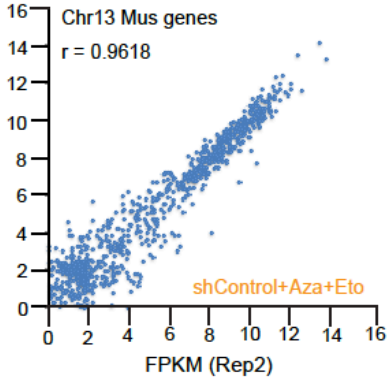
Figure S3. Xi reactivation by inhibiting single versus multiple Xist interactors.

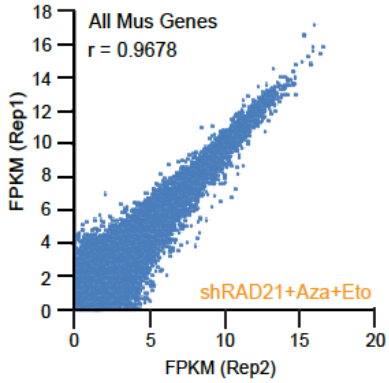
(**A**) Quantitative RT-PCR demonstrated that shRNA knockdown of single Xist interactors resulted in a maximum of 4-fold GFP upregulation.

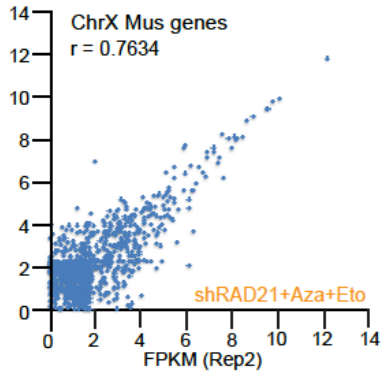
(**B**) Biological replicates for allele-specific RNA-seq analysis: Number of upregulated Xi genes for triple-drug treated cells (aza+eto+shRNA). Blue, genes specifically reactivated on Xi; red, genes also upregulated on Xa. There was a net increase in expression level (ΔFPKM) from the Xi in the triple-drug treated samples relative to the shControl+aza+eto, whereas the Xa and autosomes showed no obvious net increase, thereby suggesting direct effects on the Xi as a result of disrupting the Xist interactome.

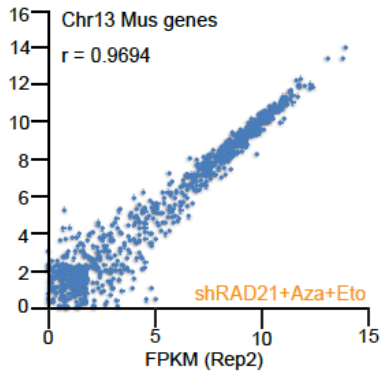

All Mus Genes
r = 0.9574
shControl+Aza+Eto
FPKM (Rep1)
FPKM (Rep2)

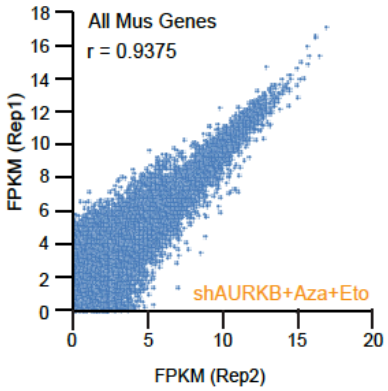

ChrX Mus genes
r = 0.7616
shControl+Aza+Eto
FPKM (Rep2)

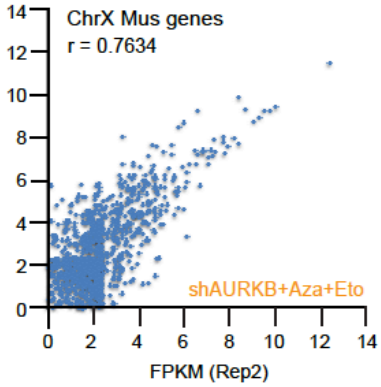

Chr13 Mus genes
r = 0.9618
shControl+Aza+Eto
FPKM (Rep2)

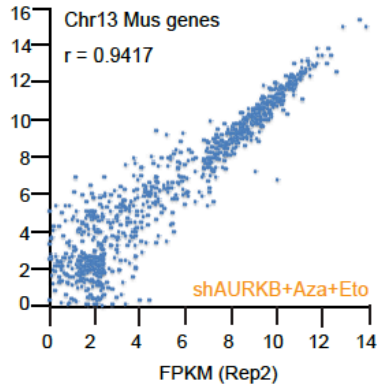

All Mus Genes
r = 0.9678
shRAD21+Aza+Eto
FPKM (Rep1)
FPKM (Rep2)

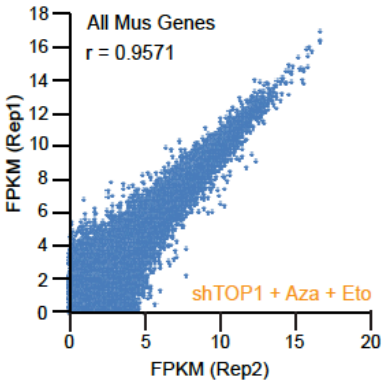

ChrX Mus genes
r = 0.7634
shRAD21+Aza+Eto
FPKM (Rep2)

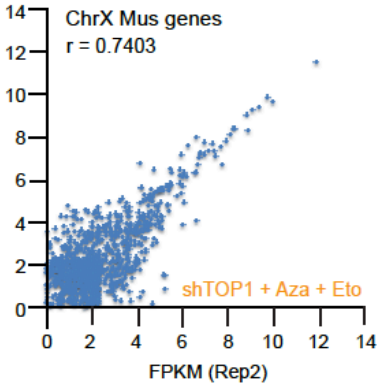

Chr13 Mus genes
r = 0.9694
shRAD21+Aza+Eto
FPKM (Rep2)


All Mus Genes
r = 0.9375
shAURKB+Aza+Eto
FPKM (Rep1)
FPKM (Rep2)


ChrX Mus genes
r = 0.7634
shAURKB+Aza+Eto
FPKM (Rep2)


Chr13 Mus genes
r = 0.9417
shAURKB+Aza+Eto
FPKM (Rep2)


All Mus Genes
r = 0.9571
shTOP1 + Aza + Eto
FPKM (Rep1)
FPKM (Rep2)


ChrX Mus genes
r = 0.7403
shTOP1 + Aza + Eto
FPKM (Rep2)

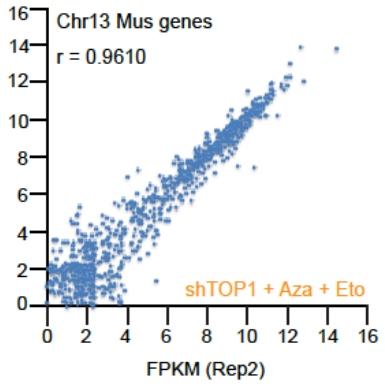
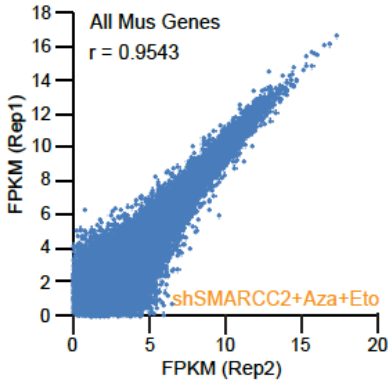
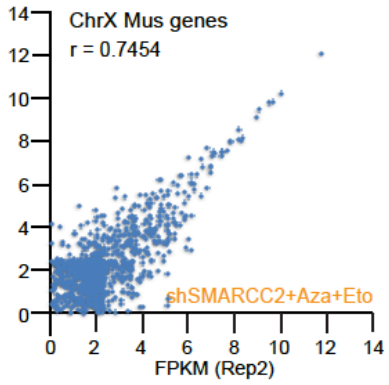
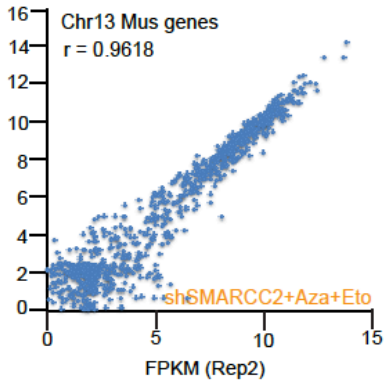

Chr13 Mus genes
r = 0.9610
shTOP1 + Aza + Eto
FPKM (Rep2)

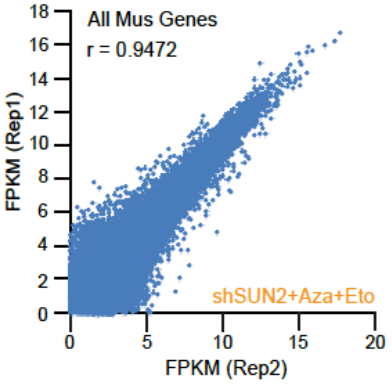
Figure S4. Correlations between biological replicates for allelic-specific RNA-seq analysis.

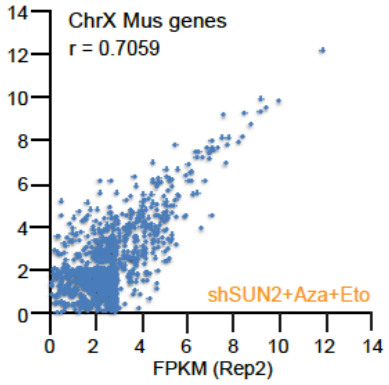
Shown are allelic (mus) FPKM values for replicate 1 (Rep1) and replicate 2 (Rep2) for indicated triple-drug treatment (orange text) for all genes, Xi genes, and Chr13 genes.

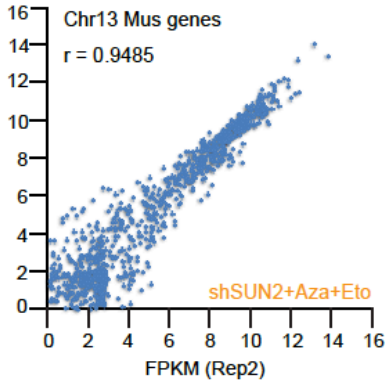

All Mus Genes
r = 0.9543
shSMARCC2+Aza+Eto
FPKM (Rep1)
FPKM (Rep2)

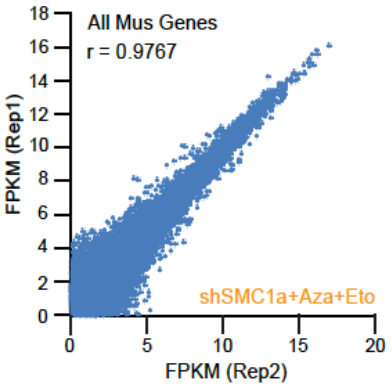

ChrX Mus genes
r = 0.7454
shSMARCC2+Aza+Eto
FPKM (Rep2)

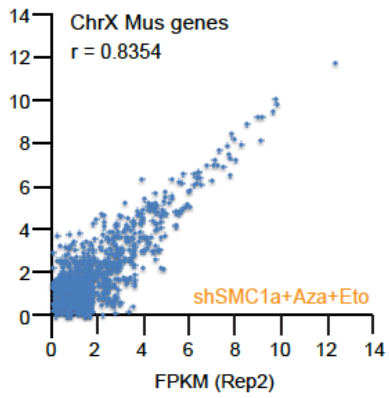

Chr13 Mus genes
r = 0.9618
shSMARCC2+Aza+Eto
FPKM (Rep2)

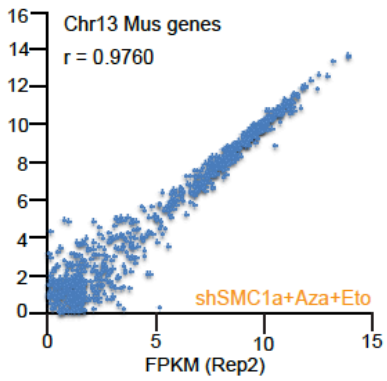

All Mus Genes
r = 0.9472
shSUN2+Aza+Eto
FPKM (Rep1)
FPKM (Rep2)

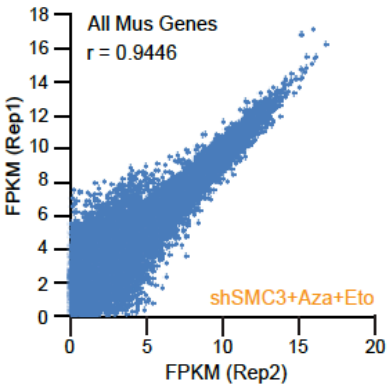

ChrX Mus genes
r = 0.7059
shSUN2+Aza+Eto
FPKM (Rep2)

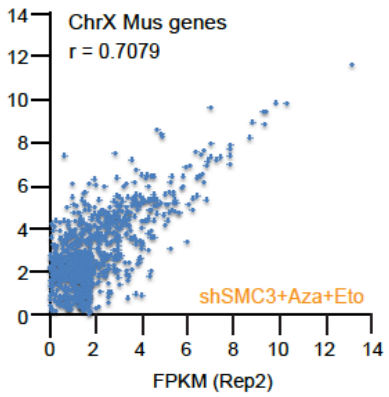

Chr13 Mus genes
r = 0.9485
shSUN2+Aza+Eto
FPKM (Rep2)


All Mus Genes
r = 0.9767
shSMC1a+Aza+Eto
FPKM (Rep1)
FPKM (Rep2)


ChrX Mus genes
r = 0.8354
shSMC1a+Aza+Eto
FPKM (Rep2)


Chr13 Mus genes
r = 0.9760
shSMC1a+Aza+Eto
FPKM (Rep2)


All Mus Genes
r = 0.9446
shSMC3+Aza+Eto
FPKM (Rep1)
FPKM (Rep2)


ChrX Mus genes
r = 0.7079
shSMC3+Aza+Eto
FPKM (Rep2)

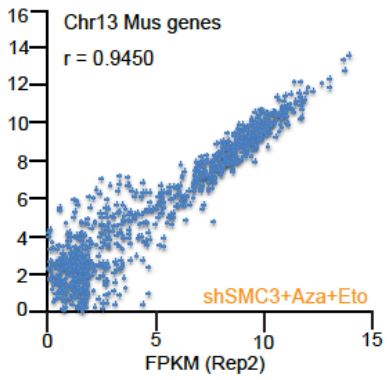

Chr13 Mus genes
r = 0.9450
shSMC3+Aza+Eto
FPKM (Rep2)

Figure S5. Correlations between biological replicates for allelic-specific RNA-seq analysis.

Shown are allelic (mus) FPKM values for replicate 1 (Rep1) and replicate 2 (Rep2) for indicated triple-drug treatment (orange text) for all genes, Xi genes, and Chr13 genes.

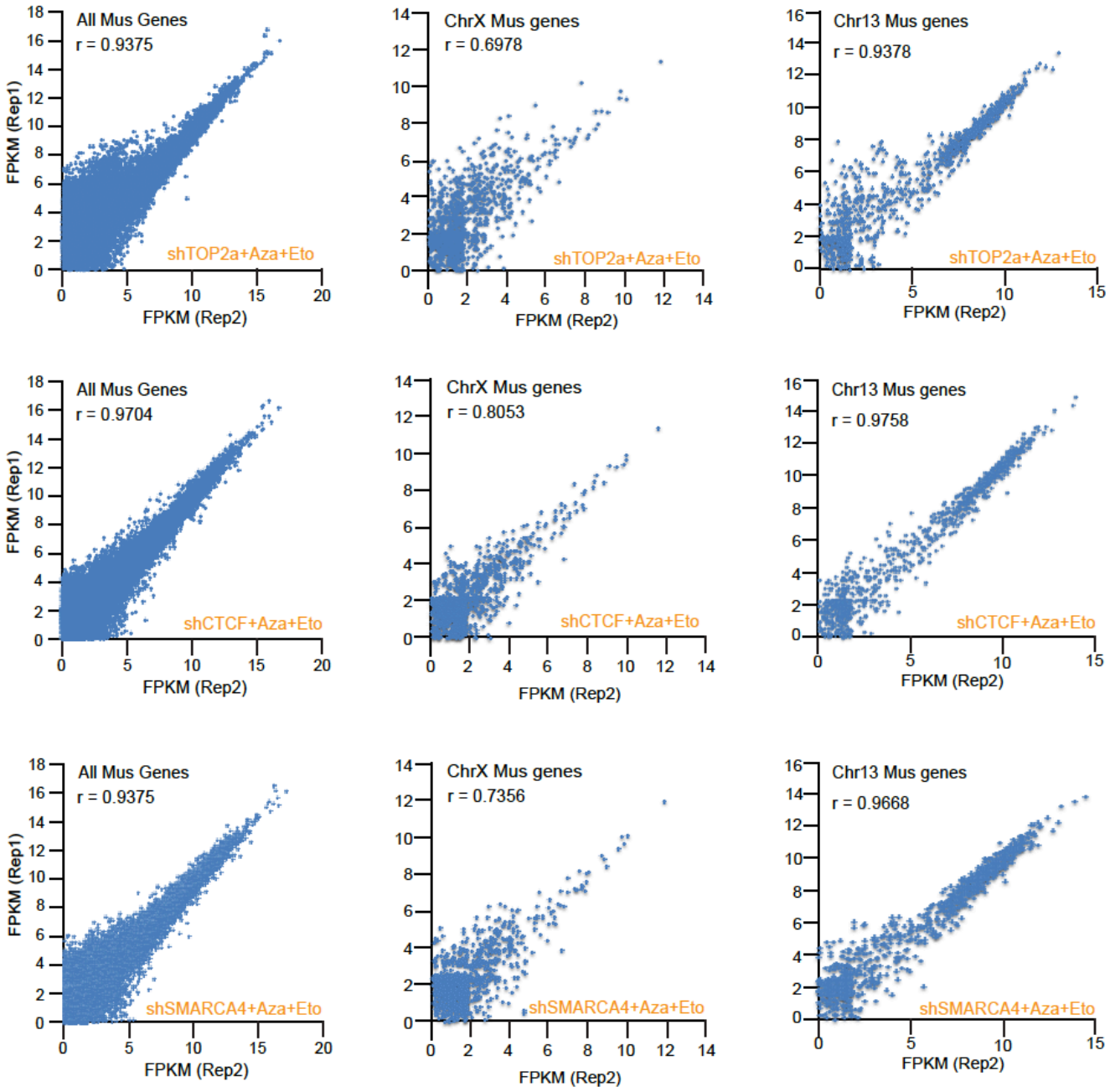


Figure S6. Correlations between biological replicates for allelic-specific RNA-seq analysis.

Shown are allelic (mus) FPKM values for replicate 1 (Rep1) and replicate 2 (Rep2) for indicated triple-drug treatment (orange text) for all genes, Xi genes, and Chr13 genes.

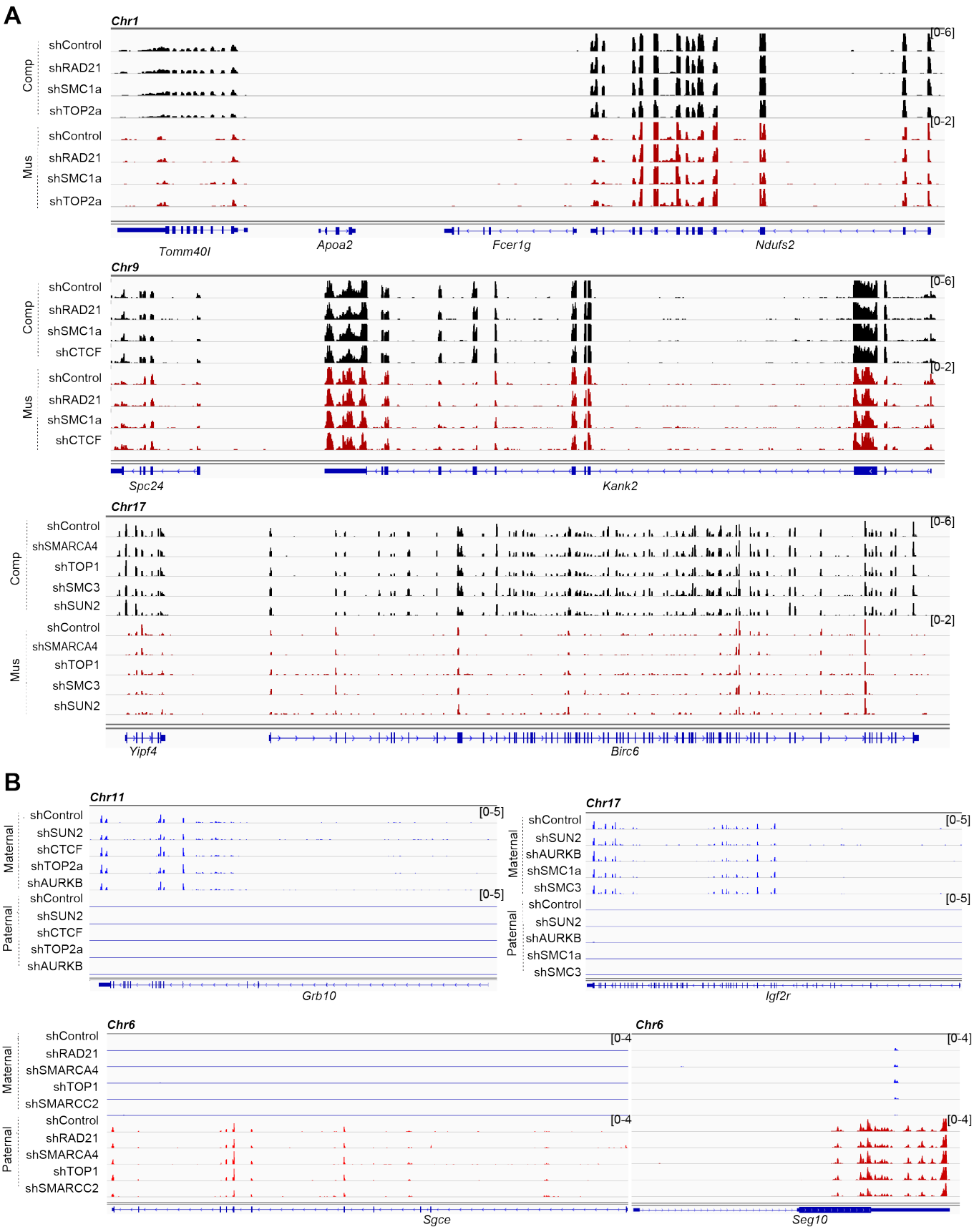

A
Chr1
Comp
shControl
shRAD21
shSMC1a
shTOP2a
Mus
shControl
shRAD21
shSMC1a
shTOP2a
[0-6]
[0-2]
Tomm40l
Apoa2
Fcer1g
Ndufs2
Chr9
Comp
shControl
shRAD21
shSMC1a
shCTCF
Mus
shControl
shRAD21
shSMC1a
shCTCF
[0-6]
[0-2]
Spc24
Kank2
Chr17
Comp
shControl
shSMARCA4
shTOP1
shSMC3
shSUN2
Mus
shControl
shSMARCA4
shTOP1
shSMC3
shSUN2
[0-6]
[0-2]
Yipf4
Birc6
B
Chr11
Maternal
shControl
shSUN2
shCTCF
shTOP2a
shAURKB
Paternal
shControl
shSUN2
shCTCF
shTOP2a
shAURKB
[0-5]
[0-5]
Grb10
Chr17
Maternal
shControl
shSUN2
shAURKB
shSMC1a
shSMC3
Paternal
shControl
shSUN2
shAURKB
shSMC1a
shSMC3
[0-5]
[0-5]
Igf2r
Chr6
Maternal
shControl
shRAD21
shSMARCA4
shTOP1
shSMARCC2
Paternal
shControl
shRAD21
shSMARCA4
shTOP1
shSMARCC2
[0-4]
[0-4]
Sgce
Chr6
[0-4]
[0-4]
Seg10

Figure S7. Allelic expression of autosomal genes, including imprinted genes, is not affected by the triple-drug treatments.

(**A,B**) Read coverages of three representative autosomal genes **(A)** and four representative imprinted genes **(B)** after triple-drug treatment. Mus, *Mus musculus* allele. Comp, total reads. Tracks are shown at the same scale within each grouping. Red tags appear only in exons with SNPs.

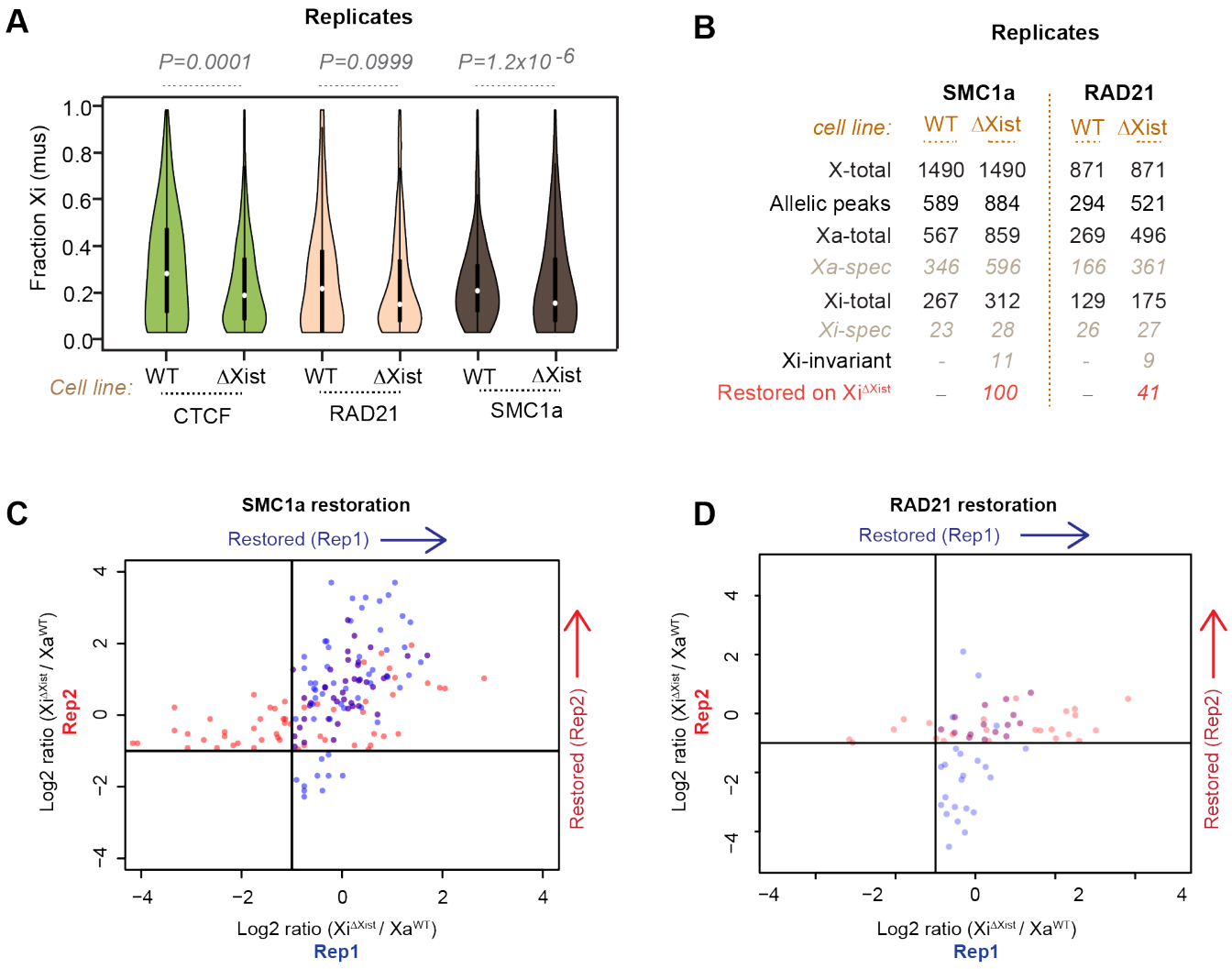


Figure S8. Analysis of CTCF and cohesin ChIP-seq replicates demonstrates similar allelic trends on ChrX.

(**A**) Allele-specific ChIP-seq results of biological replicates: Violin plots of allelic skew for CTCF, RAD21, SMC1a in wild-type (WT) and $Xi^{\Delta Xist}/Xa^{WT}$ (ΔXist) fibroblasts. Fraction of mus reads [mus/(mus+cas)] is plotted for every peak with ≥10 allelic reads. *P values* determined by the Kolmogorov-Smirnov (KS) test.

(**B**) Table of total, Xa-specific, and Xi-specific cohesin binding sites in WT versus ΔXist ($Xi^{\Delta Xist}/Xa^{WT}$) cells. Significant SMC1a and RAD21 allelic peaks with ≥5 reads were analyzed. Allele-specific skewing is defined as ≥3-fold skew towards Xa or Xi. Sites were considered "restored" if $Xi^{\Delta Xist}$'s read counts were ≥50% of Xa's. X-total, all X-linked binding sites. Allelic peaks, sites with allelic information. Xa-total, all Xa sites. Xi-total, all sites. Xa-spec, Xa-specific. Xi-spec, Xi-specific. Xi-invariant, Xi-specific in both WT and $Xi^{\Delta Xist}/Xa^{WT}$ cells. Note: The net gain of sites on the Xi in the mutant does not equal the number of restored sites. This difference is

due to defining restored peaks separately from calling ChIP peaks (macs2). Allele-specific skewing is defined as ≥3-fold skew towards either Xa or Xi.

(**C**) Correlation analysis showing Log2 $Xi^{\Delta Xist}$ to Xa^{WT} ratios of SMC1a coverage in replicates 1 and 2 (Rep1, Rep2). Rep1, blue dots. Rep2, red dots. Both, purple dots. Consensus, upper right quadrant.

(**D**) Correlation analysis showing Log2 $Xi^{\Delta Xist}$ to Xa^{WT} ratios of RAD21 coverage in replicates 1 and 2 (Rep1, Rep2). Rep1, blue dots. Rep2, red dots. Both, purple dots. Consensus, upper right quadrant.

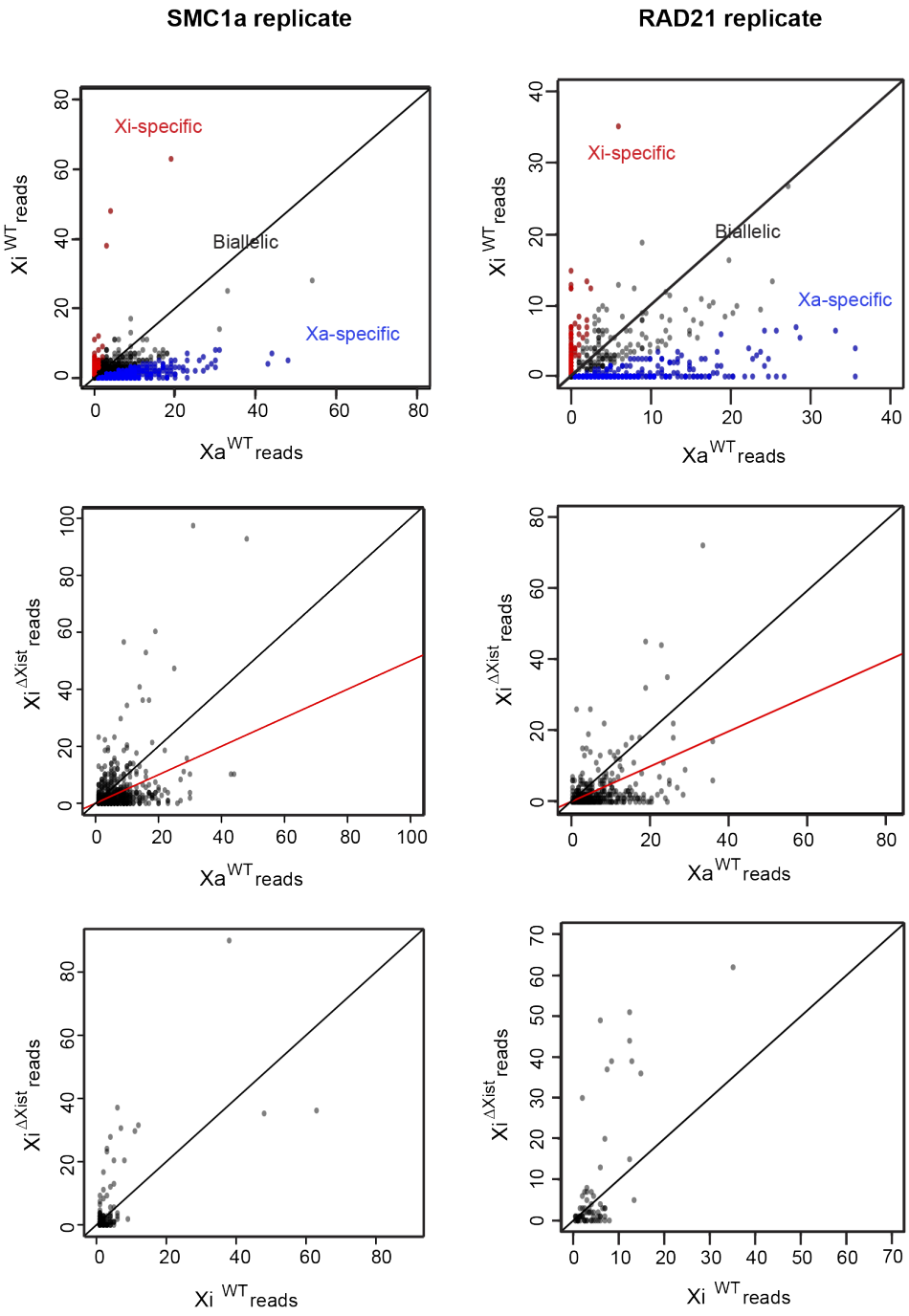
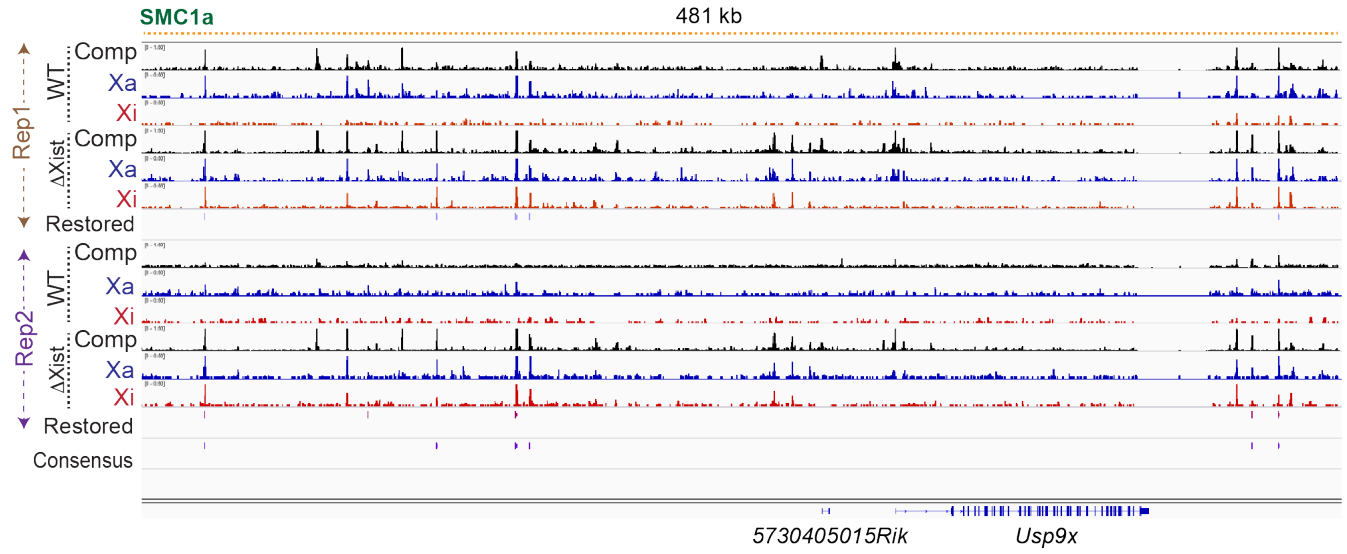
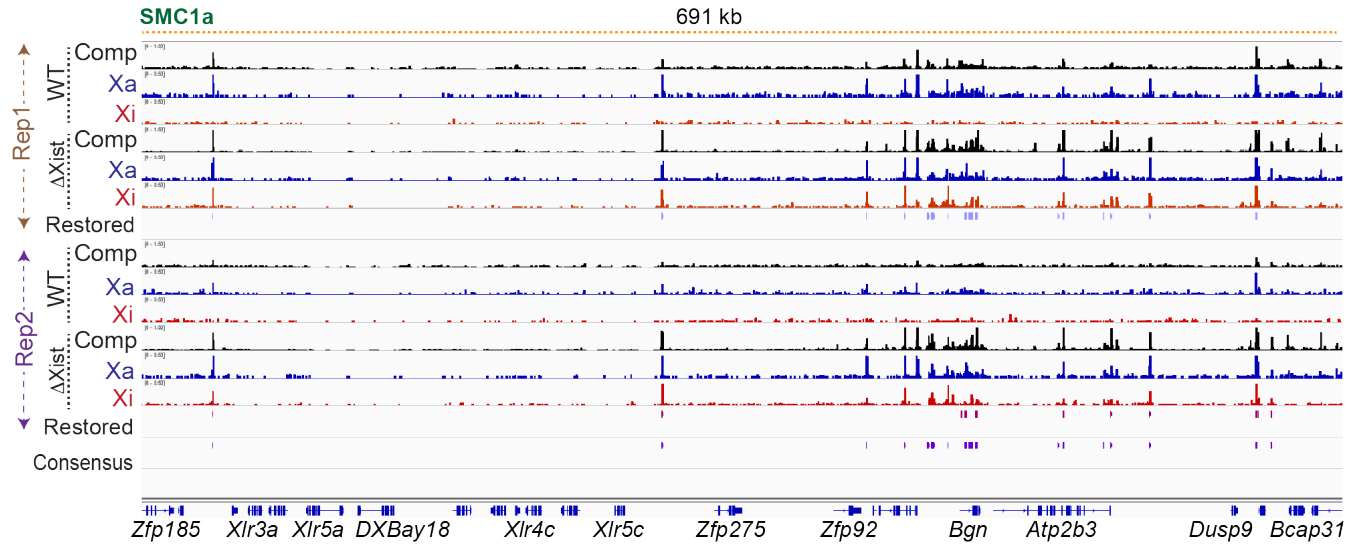


Figure S9. Analysis of biological replicates for cohesin ChIP-seq confirms cohesin restoration *in cis* when Xist is ablated.

Allele-specific ChIP-seq analysis of SMC1a and RAD21 biological replicates. Top panels: Differences between SMC1a or RAD21 peaks on the Xi^{WT} versus Xa^{WT}. Black diagonal, 1:1 ratio. Plotted are read counts for all SMC1a or RAD21 peaks. Allele-specific skewing is defined as ≥3-

fold skew towards either Xa (cas, blue dots) or Xi (mus, red dots). Biallelic peaks, grey dots. Middle panels: Partial restoration of SMC1a or RAD21 peaks on the $Xi^{\Delta Xist}$ to an Xa pattern. Plotted are peaks with read counts with ≥3-fold skew to Xa^{WT} ("Xa-specific"). x-axis, normalized Xa^{WT} read counts. y-axis, normalized $Xi^{\Delta Xist}$ read counts. Black diagonal, 1:1 $Xi^{\Delta Xist}/Xa^{WT}$ ratio; red diagonal, 1:2 ratio. Bottom panels: Xi-specific SMC1a or RAD21 peaks remained on $Xi^{\Delta Xist}$. Plotted are read counts for SMC1a or RAD21 peaks with ≥3-fold skew to Xi^{WT} ("Xi-specific").


SMC1a
481 kb
Rep1
WT
Comp
Xa
Xi
ΔXist
Comp
Xa
Xi
Restored
Rep2
WT
Comp
Xa
Xi
ΔXist
Comp
Xa
Xi
Restored
Consensus
5730405015Rik
Usp9x


SMC1a
691 kb
Rep1
WT
Comp
Xa
Xi
ΔXist
Comp
Xa
Xi
Restored
Rep2
WT
Comp
Xa
Xi
ΔXist
Comp
Xa
Xi
Restored
Consensus
Zfp185
Xlr3a
Xlr5a
DXBay18
Xlr4c
Xlr5c
Zfp275
Zfp92
Bgn
Atp2b3
Dusp9
Bcap31

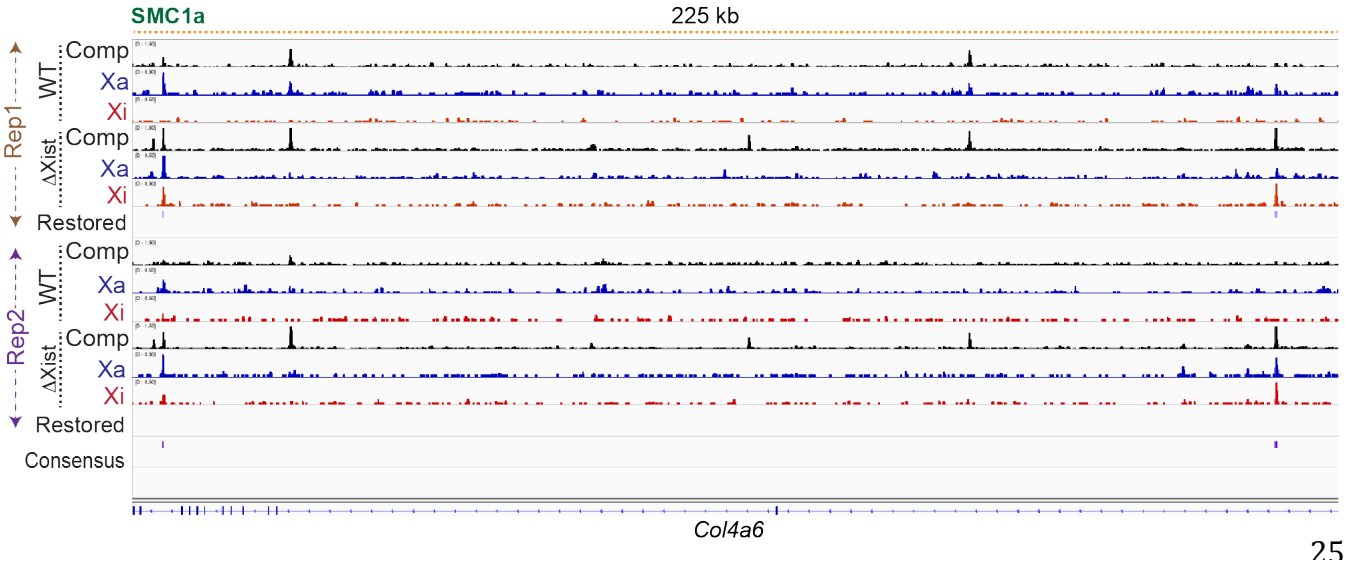

SMC1a
225 kb
Rep1
WT
Comp
Xa
Xi
ΔXist
Comp
Xa
Xi
Restored
Rep2
WT
Comp
Xa
Xi
ΔXist
Comp
Xa
Xi
Restored
Consensus
Col4a6

Figure S10. Restored SMC1a peaks are reproducible in biological replicates and occur throughout $Xi^{\Delta Xist}$ (Example set 1).

The representative examples of SMC1a restoration on $Xi^{\Delta Xist}$. "Restored" peaks shown as ticks under each biological replicate (Rep1, Rep2). The "consensus" restored peaks are shown in the last track of each grouping.

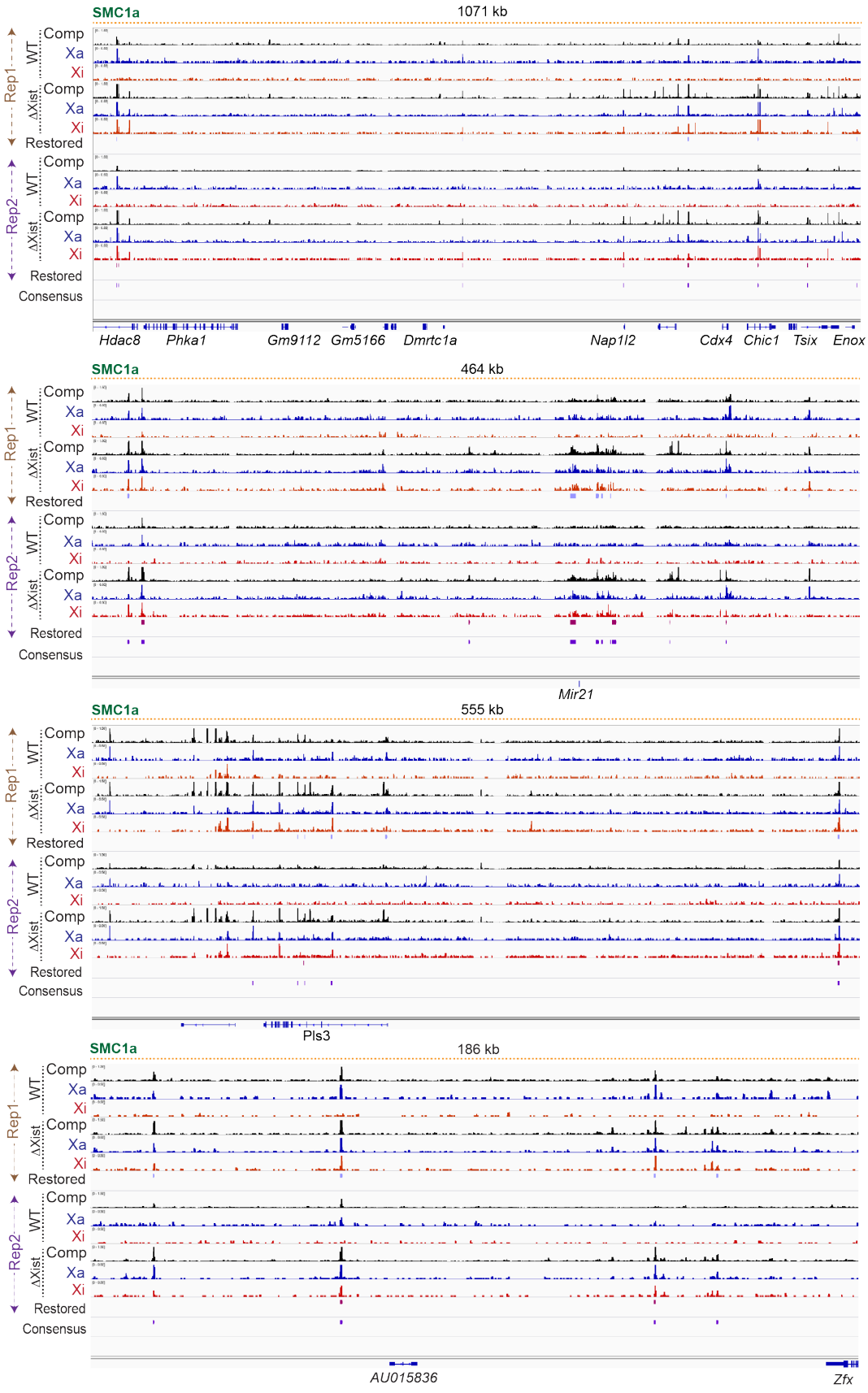
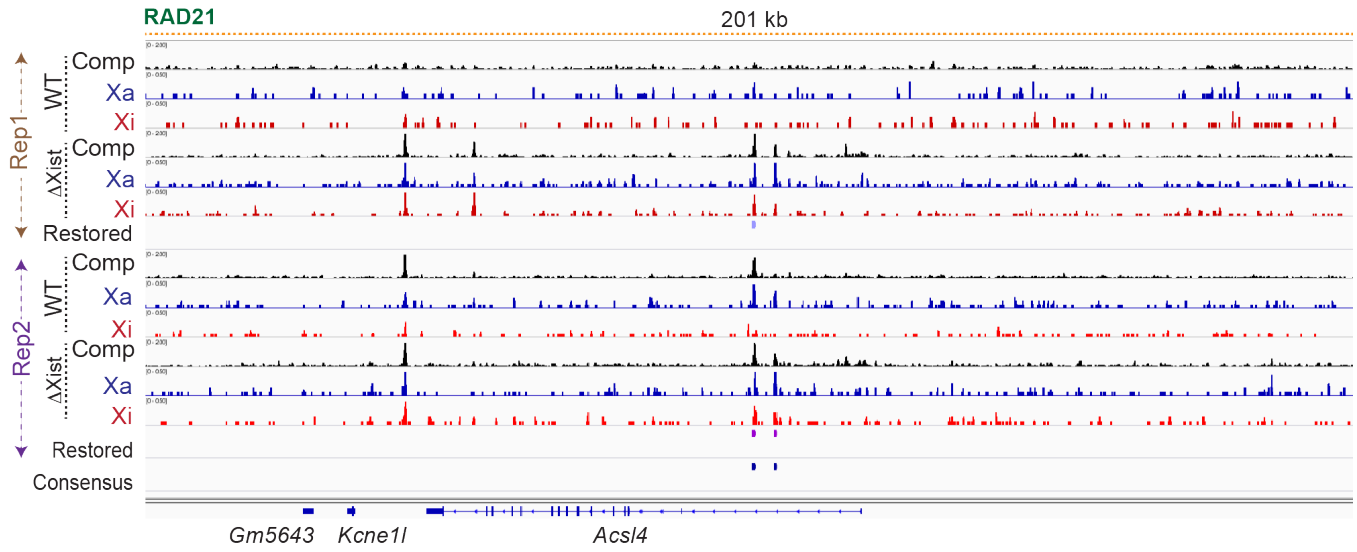
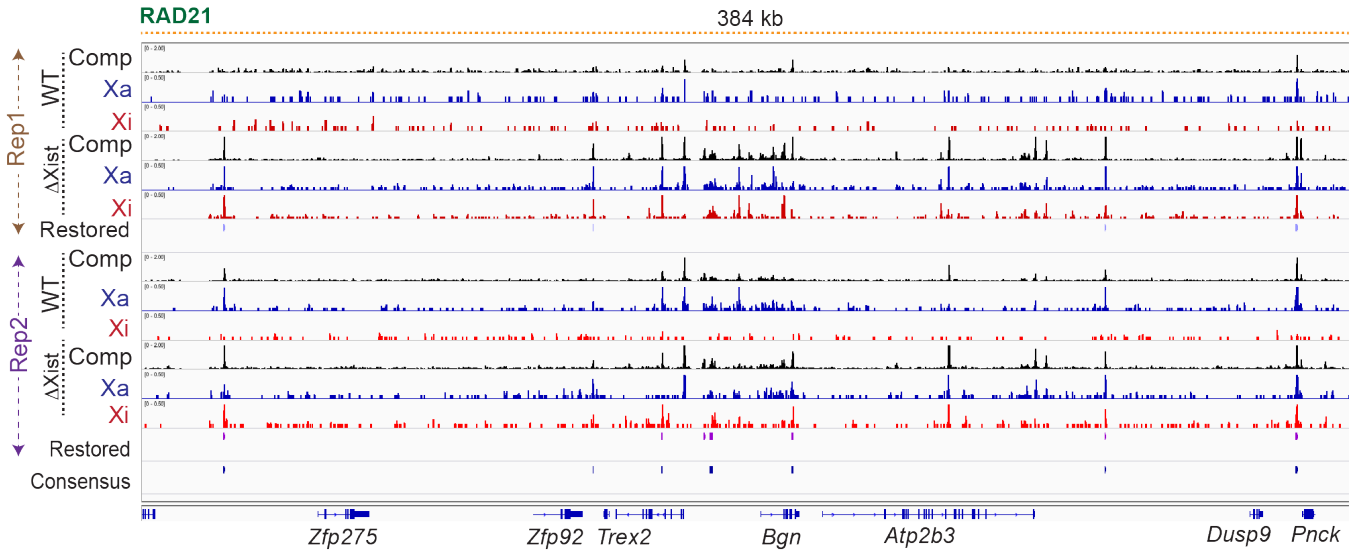

SMC1a
1071 kb
Rep1
Rep2
WT
ΔXist
Comp
Xa
Xi
Restored
Consensus
Hdac8
Phka1
Gm9112
Gm5166
Dmrtc1a
Nap1l2
Cdx4
Chic1
Tsix
Enox
SMC1a
464 kb
Mir21
SMC1a
555 kb
Pls3
SMC1a
186 kb
AU015836
Zfx

Figure S11. Restored SMC1a peaks are reproducible in biological replicates and occur throughout $Xi^{\Delta Xist}$ (Example set 2).

The representative examples of SMC1a restoration on $Xi^{\Delta Xist}$. "Restored" peaks shown as ticks under each biological replicate (Rep1, Rep2). The "consensus" restored peaks are shown in the last track of each grouping.


RAD21
201 kb
Rep1
Rep2
WT
ΔXist
Comp
Xa
Xi
Restored
Consensus
Gm5643
Kcne1l
Acsl4


RAD21
384 kb
Rep1
Rep2
WT
ΔXist
Comp
Xa
Xi
Restored
Consensus
Zfp275
Zfp92
Trex2
Bgn
Atp2b3
Dusp9
Pnck

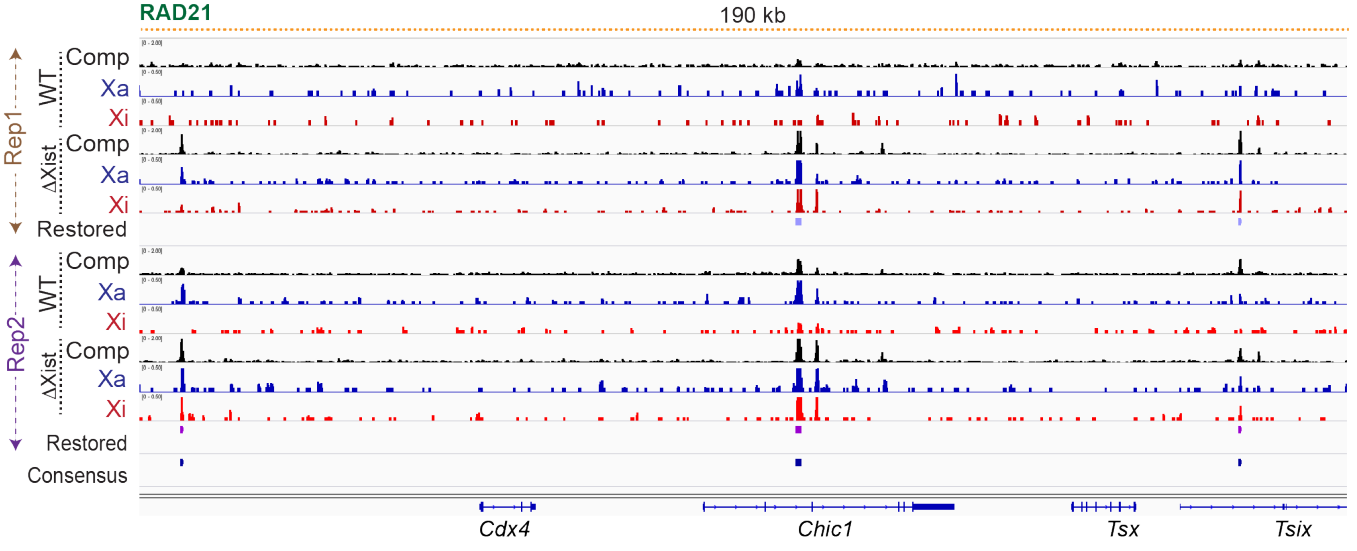

RAD21
190 kb
Rep1
Rep2
WT
ΔXist
Comp
Xa
Xi
Restored
Consensus
Cdx4
Chic1
Tsx
Tsix

Figure S12. Restored RAD21 peaks are reproducible in biological replicates and occur throughout $Xi^{\Delta Xist}$.

The representative examples of RAD21 restoration on $Xi^{\Delta Xist}$. "Restored" peaks shown as ticks under each biological replicate (Rep1, Rep2). The "consensus" restored peaks are shown in the last track of each grouping.

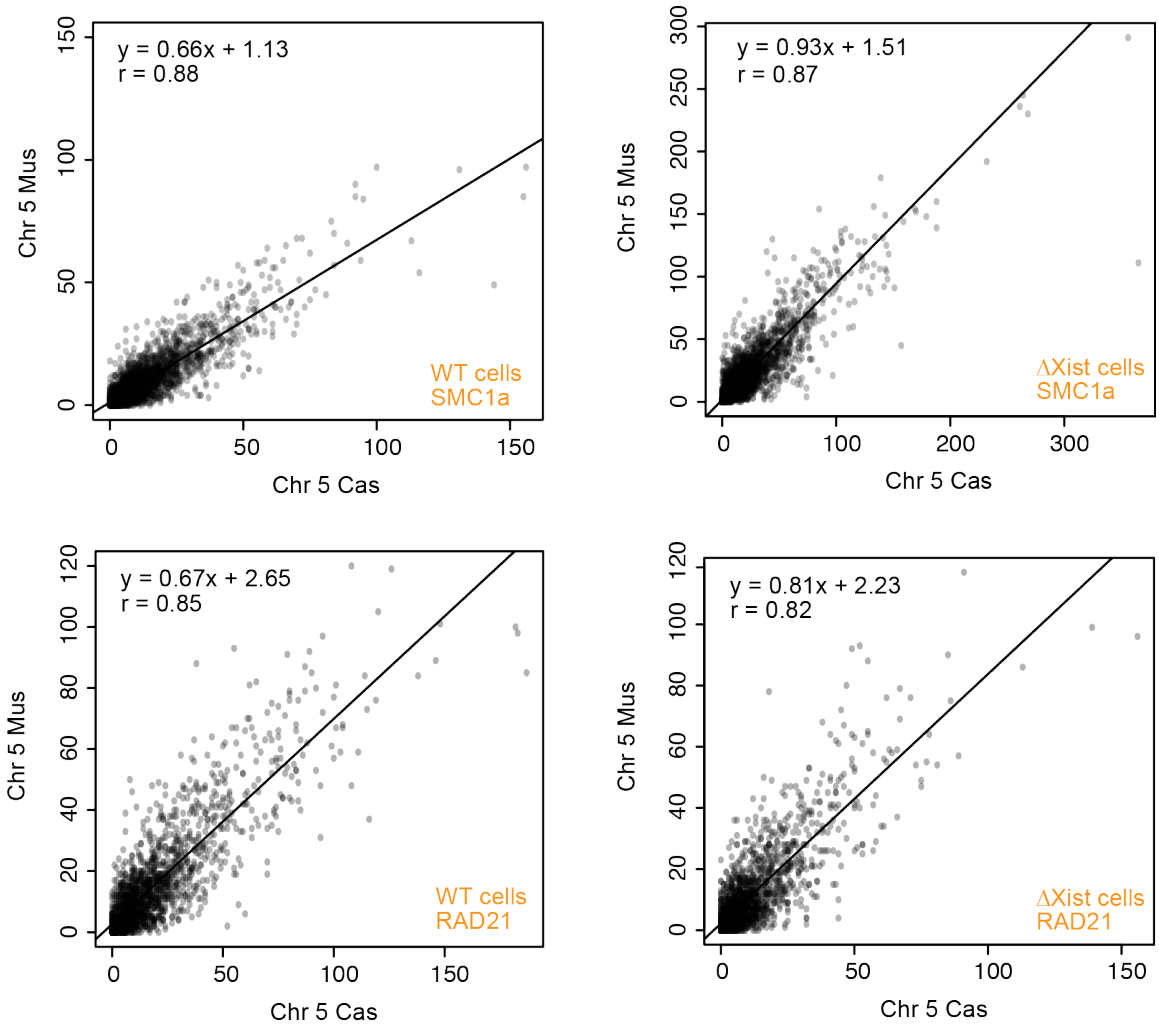


Figure S13. Cohesin restored in $Xi^{\Delta Xist}/Xa^{WT}$ fibroblasts was Xi-specific and did not occur on autosomes.

Correlation plots comparing SMC1a or RAD21 coverages on the mus versus cas alleles in wildtype fibroblasts (WT) versus $Xi^{\Delta Xist}/Xa^{WT}$ fibroblasts (ΔXist). Representative autosome, Chr5, is shown. Equation shows the slope and y-intercepts for the black diagonals as a measure of correlation. Pearson's *r* also shown.

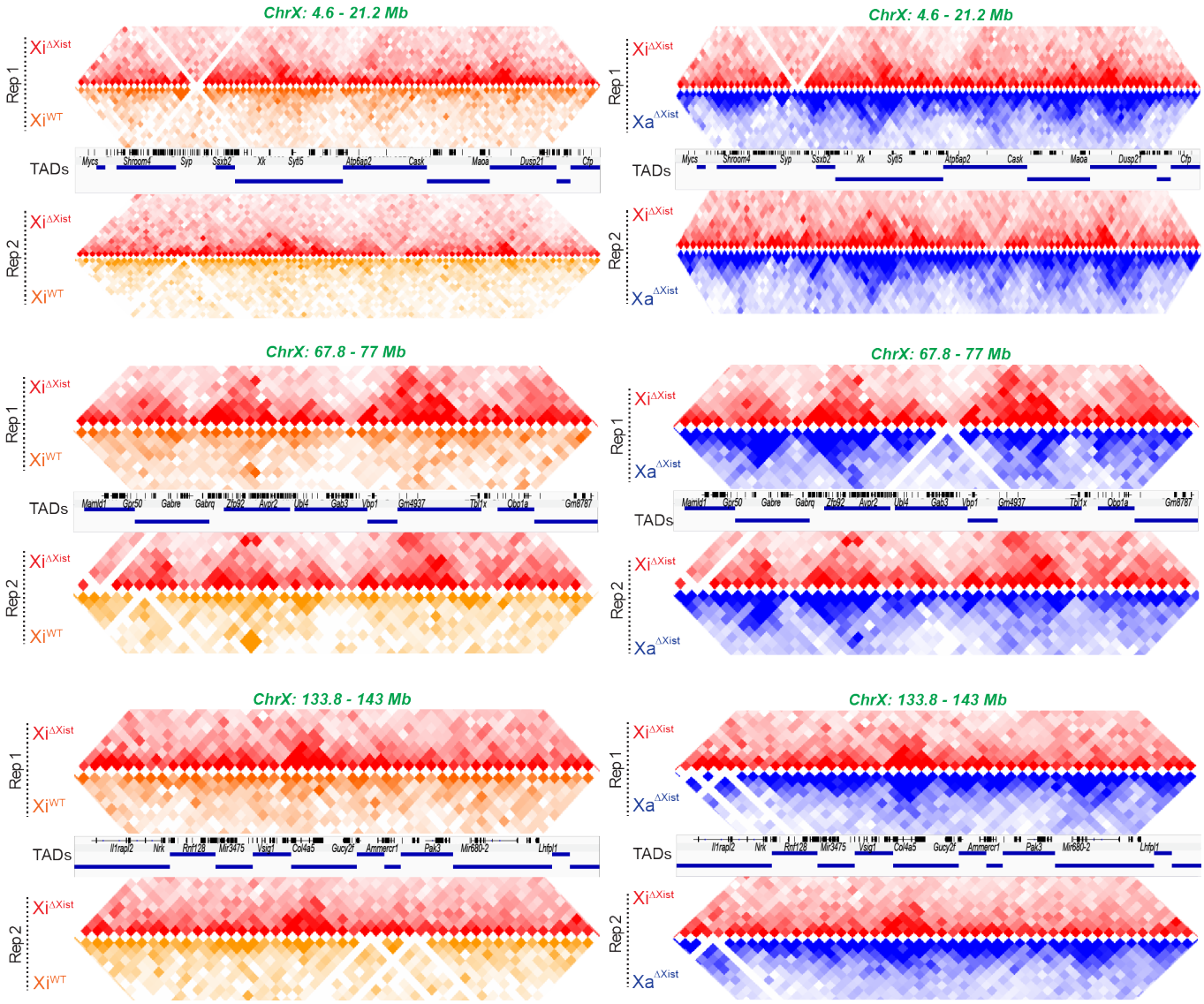


Figure S14. Biological replicates of HiC-seq analysis yield similar findings.

Contact maps for various ChrX regions at 40-kb resolution comparing $Xi^{\Delta Xist}$ (red) to Xi^{WT} (orange), and $Xi^{\Delta Xist}$ (red) versus Xa (blue) of the mutant cell line. Our Xa TAD calls are shown with RefSeq genes. Rep1 and Rep2 contact maps are shown side by side.

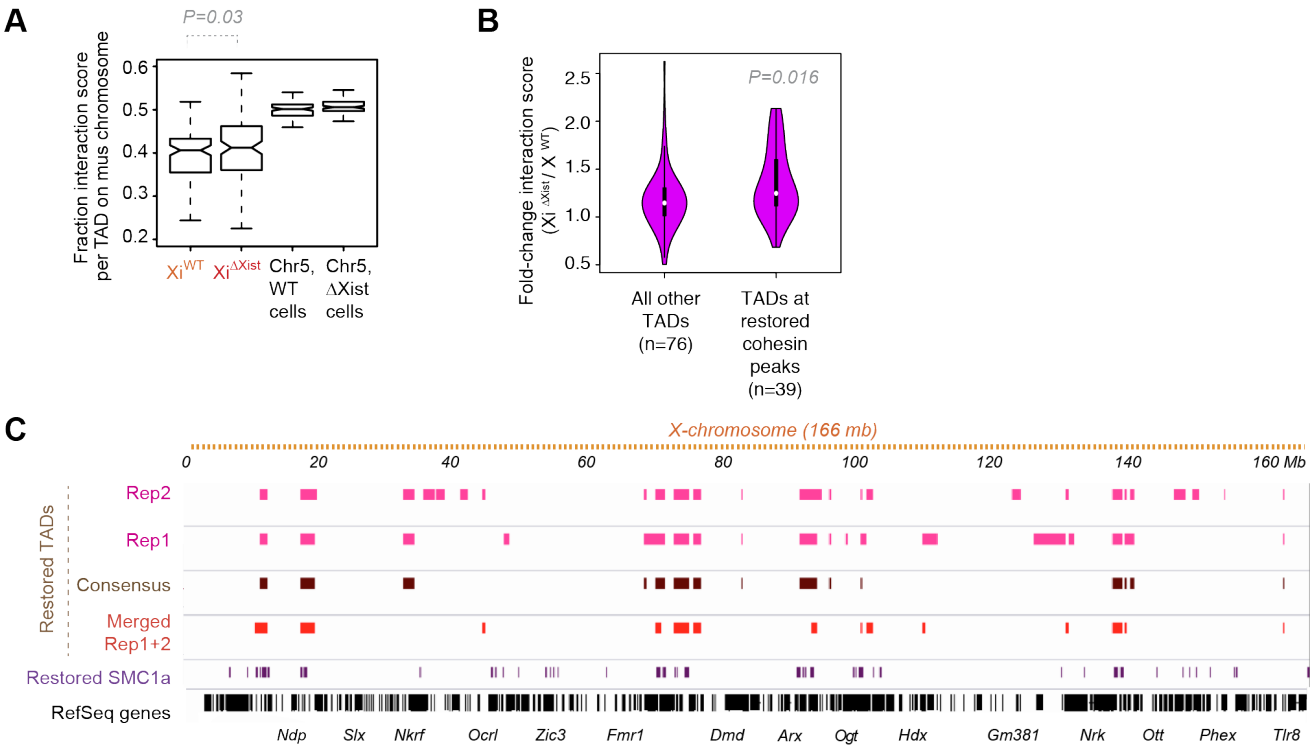


Figure S15. Restored TADs identified in Xi$^{\Delta Xist}$ using Xa TADs of Dixon et al. (*28*) as reference.

(**A**) Using TADs called by Dixon et al. (*28*) (rather than our own called TADs, as shown in Figure 5C) as a basis for identifying restored TADs, we calculated the fraction of interaction frequency per TAD on the Xi (mus) chromosome. Highly similar results were obtained. The positions of our Xa TAD borders were rounded to the nearest 100 kb and submatrices were generated from all pixels between the two endpoints of the TAD border for each TAD. We calculated the average interaction score for each TAD by summing the interaction scores for all pixels in the submatrix defined by a TAD and dividing by the total number of pixels in the TAD. We then averaged the normalized interaction scores across all bins in a TAD in the Xi (mus) and Xa (cas) contact maps, and computed the fraction of averaged interaction scores from mus chromosomes. ChrX and a representative autosome, Chr5, are shown for the WT cell line and the Xist$^{\Delta Xist}$/+ cell line. P value determined by KS test. P-value determined by paired Wilcoxon signed rank test.

(**B**) Using TADs called by Dixon et al. (*28*) (rather than our own called TADs, as shown in Figure 5C) as a basis for identifying restored TADs, violin plots also showed that TADs overlapping restored peaks have larger increases in interaction scores relative to all other TADs. We calculated the fold-change in average interaction scores on the Xi for all X-linked TADs and intersected the TADs with SMC1a sites (Xi$^{\Delta Xist}$/ Xi^{WT}). 32 TADs occurred at restored cohesin sites; 80 TADs did not overlap restored cohesin sites. Violin plot shows distributions of fold-change average interaction scores between Xi^{WT} and Xi$^{\Delta Xist}$. P-value determined by Wilcoxon ranked sum test.

(**C**) Using TADs called by Dixon et al. (*28*) (rather than our own called TADs, as shown in Figure 5C) as a basis for identifying restored TADs, we also found that restored TADs overlapped

regions with restored cohesins on across $Xi^{\Delta Xist}$. Note highly similar results obtained here relative to Figure 5E. Several datasets were used to identify restored TADs, each producing similar results. Restored TADs were called in two separate replicates (Rep1, Rep2) where the average interaction score was signficantly higher on $Xi^{\Delta Xist}$ than on Xi^{WT}. We also called restored TADs based on merged Rep1+Rep2 datasets. Finally, a consensus between Rep1 and Rep2 was derived. Method: We calculated the fold-change in mus or cas for all TADs on ChrX and on a control, Chr5; then defined a threshold for significant changes based on either the autosomes or the Xa. We treated Chr5 as a null distribution (few changes expected on autosomes) and found the fraction of TADs that crossed the threshold for several thesholds. These fractions corresponded to a false discovery rate (FDR) for each given threshold. An FDR of 0.05 was used.

TABLE LEGENDS

Table S1: iDRiP proteomics results.

Sheet 1: Spectral counts of proteins pulled down by iDRiP and identified by mass spectrometry.

Sheet 2: Multiplexed quantitation of proteins pulled down by iDRiP and identified by mass spectrometry.

Table S2: Lentiviral shRNA constructs used for stable knockdowns of candidate Xist interactors.

Table S3: Table of reagents.

Sheet 1: Primer pairs used for RT-qPCR of candidate proteins following stable knockdown.

Sheet 2: Various antibodies and small molecules used for immunofluorescence and X-reactivation studies.

SUPPLEMENTAL REFERENCES

1. C. Chu *et al.*, Systematic discovery of xist RNA binding proteins. *Cell* **161**, 404-416 (2015).
2. C. A. McHugh *et al.*, The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236 (2015).
3. A. Castello *et al.*, Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406 (2012).
4. S. C. Kwon *et al.*, The RNA-binding protein repertoire of embryonic stem cells. *Nature structural & molecular biology* **20**, 1122-1130 (2013).
5. D. H. Lundgren, S. I. Hwang, L. Wu, D. K. Han, Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics* **7**, 39-53 (2010).
6. G. C. McAlister *et al.*, Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem* **84**, 7469-7478 (2012).
7. A. Thompson *et al.*, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
8. L. Ting, R. Rad, S. P. Gygi, W. Haas, MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature methods* **8**, 937-940 (2011).
9. A. C. Tolonen, W. Haas, Quantitative proteomics using reductive dimethylation for stable isotope labeling. *J Vis Exp*, (2014).
10. G. C. McAlister *et al.*, MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem* **86**, 7150-7158 (2014).
11. M. P. Weekes *et al.*, Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* **157**, 1460-1472 (2014).
12. J. K. Eng, A. L. McCormack, J. R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989 (1994).
13. J. E. Elias, S. P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods* **4**, 207-214 (2007).
14. E. L. Huttlin *et al.*, A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189 (2010).
15. Y. Jeon, J. T. Lee, YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119-133 (2011).
16. A. K. Hadjantonakis, L. L. Cox, P. P. Tam, A. Nagy, An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta. *Genesis* **29**, 133-140 (2001).
17. J. T. Kung *et al.*, Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Molecular cell* **57**, 361-375 (2015).
18. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589 (2010).

19. S. F. Pinter *et al.*, Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res* **22**, 1864-1876 (2012).
20. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
21. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).
22. E. Yildirim, R. I. Sadreyev, S. F. Pinter, J. T. Lee, X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. *Nature structural & molecular biology* **19**, 56-61 (2012).
23. Y. C. Lin *et al.*, Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature immunology* **13**, 1196-1204 (2012).
24. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
25. S. Selvaraj, R. D. J, V. Bansal, B. Ren, Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat Biotechnol* **31**, 1111-1118 (2013).
26. P. A. Knight, D. Ruiz, A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis* **33**, 1029-1047 (2012).
27. M. Imakaev *et al.*, Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods* **9**, 999-1003 (2012).
28. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).