Supplemental Information

Including five supplemental figures, seven supplemental tables, and Supplemental Experimental Procedures.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell Culture

Human neuroblastoma SK-*N*-SH and endometrial adenocarcinoma HEC-1B cells, and mouse neuro2A (N2A) cells were cultured in MEM (Gibco), supplemented with 10% (v/v) FBS (Gibco), 2 mM GlutaMAX (Gibco), 1 mM sodium pyruvate (Sigma), and 1% penicillin-streptomycin (Gibco). Human K562 and HEK293T cells were cultured in DMEM (Hyclone) supplemented with 10% (v/v) FBS and 1% penicillin-streptomycin. Cells were maintained at 37 °C in a humidified incubator containing 5% CO_2.

Recombinant CTCF Protein Production

The recombinant full-length CTCF proteins were prepared by a pTNT-CTCF plasmid through *in vitro* translation in the rabbit reticulocyte lysate as previously described (Guo et al., 2012; Jia et., 2014). A series of truncated CTCF proteins with sequential deletion of each zinc finger domain from either N- or C- terminus

were prepared similarly from a repertoire of 17 sequencing-confirmed plasmids constructed by PCR and subcloning. The primer sets used are listed in Table S7.

Western Blotting

The *in-vitro*-synthesized proteins were diluted with RIPA lysis buffer containing 1 mM PMSF. Proteins were separated by SDS–PAGE and transferred to nitrocellulose membranes. The membranes were then incubated with mouse anti-myc antibody (Millipore). Finally, the membranes were incubated with goat anti-mouse secondary antibody and scanned by using the Odyssey System (LI-COR Biosciences).

Electrophoretic Mobility Shift Assay (EMSA)

The sequences containing various CBS sites were cloned into the pGEM-T Easy plasmid (Promega). The mutations of each CBS site were constructed by PCR on the wild-type templates. Probes were amplified by PCR with high-fidelity DNA polymerase using 5’ biotin-labeled primers and were gel-purified. The DNA concentration of the probes was measured with a NanoDrop (Thermo). Each binding reaction contained equimolar amounts of the biotin-labeled probes. The primers used are listed in Table S7. EMSA was performed using LightShift Chemiluminescent EMSA reagents (Thermo) as described in

the manufacturer's manual. Briefly, the probes were incubated with *in-vitro*-synthesized proteins in binding buffer containing 10 mM Tris, 50 mM KCl, 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 1 mM dithiothreitol, 0.1% (v/v) Nonidet P-40 (NP-40), 50 ng/μl poly (dI-dC), and 2.5% (v/v) glycerol at room temperature for 20 min. The binding complex was electrophoresed on 5% nondenaturing polyacrylamide gels in ice-cold 0.5×TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH8.0). The gel was electrotransferred to a nylon membrane in ice-cold 0.5×TBE buffer. After crosslinking using UV-light for 10 min, the membrane was incubated with stabilized streptavidin-horseradish peroxidase conjugate and rinsed with the washing buffer. The biotin-labeled DNA was then detected by chemiluminescence using the ChemiDoc XRS+ system (Bio-Rad).

Chromosome Conformation Capture (3C)

Quantitative 3C was performed according to described procedures (Guo et al., 2012). Briefly, cells were cross-linked with 1% (v/v) formaldehyde for 10 min at 37 °C. After nuclear extraction, cross-linked DNA was digested overnight with 400 U of BglII or EcoRI at 37 °C while shaking at 900 rpm. After self-ligation, the DNA was purified and quantified using PicoGreen reagents (Invitrogen). The final quantitative PCR reactions were performed in triplicates by using SYBR Green (Roche) with 100 ng DNA as templates.

The 3C control experiments were performed according to the published method (Guo et al., 2012). BAC clones (CTD-3042N4, CTD-2506I6, CTD-2538C12, CTD-2527B17 and CTD-2371G16 from Invitrogen) were isolated and purified using a large-construct DNA isolation kit (Qiagen). The molar amount of these BAC clones was detected by qPCR titration using primers matching the BAC backbone sequences. Equimolar amounts of BAC clones were digested with 400 U BglII or EcoRI at 37 °C overnight. After purification, the DNA was then ligated with T4 DNA ligase at 16 °C overnight. The ligated BAC DNA was used to establish a standard PCR amplification curve by serial dilution.

To compare long-range DNA interaction frequencies in different cell lines, the PCR reactions were normalized to the ligation frequency of six restriction fragments of the tubulin, phosphoglycerate kinase 1, and 14-3-3 loci. The following six primer pairs were used for normalization between SK-*N*-SH and K562 cells: 3 pairs of tubulin (BglII-PGK-1 and BglII-PGK-3, BglII-PGK-2 and BglII-PGK-4, BglII-PGK-3 and BglII-PGK-5), and 3 pairs of 14-3-3 ζ/δ (BglII-YWHAZ-1 and BglII-YWHAZ-3, BglII-YWHAZ-2 and BglII-YWHAZ-3, BglII-YWHAZ-2 and BglII-YWHAZ-4) (Table S7). The following six primer pairs were used for normalization between inversion and wild-type HEC-1B cell lines: 3 pairs of tubulin (TUBB-1 and TUBB-2, TUBB-1 and TUBB-4, TUBB-2 and TUBB-4), and 3 pairs of 14-3-3 ζ/δ (YWHAZ-1 and YWHAZ-2, YWHAZ-1 and YWHAZ-5, YWHAZ-2 and YWHAZ-6). These 3C experiments were performed

at least three times. Data are means ± SEM. The significance of the differences was evaluated by the Student's t-test.

Circularized Chromosome Conformation Capture (4C)

The 4C-seq libraries were constructed as described (Guo et al., 2012; Jia et al., 2014; Simonis et al., 2006; Splinter et al., 2012) with some modifications. Briefly, mouse brain tissues were dispersed by collagenase (1.25 mg/ml, Sigma) treatment in DMEM supplemented with 10% (v/v) FBS for 45 min at 37 °C while shaking at 700 rpm. Cells were then filtered through a 40-μm cell strainer (BD Biosciences) to make a single-cell suspension. A total of 10^7 cells were cross-linked and then lysed to prepare cell nuclei. The cross-linked DNA in the nuclear preparations was digested with HindIII or EcoRI overnight while shaking at 900 rpm and then ligated with T4 DNA ligase. After purification, the DNA was digested with a second enzyme, DpnII or NlaIII, and was ligated again. The religated DNA was then purified using a High-Pure PCR Product Purification system (Roche). A series of 4C-seq libraries were generated by inverse PCR using a high-fidelity DNA polymerase with primer pairs containing Illumina adapter sequences (Table S7). High-throughput sequencing was performed using 49-bp single-end reads on an Illumina HiSeq 2000 platform. The sequenced reads were mapped to the mouse (NCBI37/mm9) or human (GRCh37/hg19) reference genomes using Bowtie (version 1.0.0) (Langmead et al., 2009).The r3Cseq program in the R/Bioconductor package (Thongjuea

et al., 2013) was used to detect statistically significant long-range chromatin-looping interactions. The sequencing data were visualized in the UCSC genome browser (Kent et al., 2002). All 4C-seq experiments were performed with at least two biological replicates.

Hi-C Data Generation and Analysis

Hi-C for SK-*N*-SH cells was performed as previously described (Dixon et al., 2012; 2015). We performed two biological replicates, each with roughly 2.5 x 10^8 cells. We obtained a total of more 200 million read pairs per replicate. We constructed normalized Hi-C contact matrices at 40-kb resolution after removing intrinsic biases in Hi-C data by using HiCNorm (Hu et al., 2012). Normalized contact matrices for the two replicates were highly correlated (Pearson correlation coefficient > 0.89). Topologically associated domains ("TADs" or "sub-TADs") were identified based on Directionality Index ("DI") as described (Dixon et al., 2012; 2015) with one exception: DI was calculated using a sliding window of 300 kb upstream/downstream of the anchor point. A smaller DI window will yield smaller TADs, while a larger window will yield larger TADs. Hi-C data in H1 human Embryonic Stem Cells and H1-derived Neural Precursor Cells was previously generated (Dixon et al., 2015).

CRISPR/Cas9 System

The DNA fragment inversion and deletion by CRISPR/Cas9 were performed as previously described (Li et al. 2015). The templates for producing target sgRNAs were constructed by PCR using pLKO.1 or pGL3-U6-sgRNA-PGK-Puro plasmid (Chang et al., 2013; Cong et al., 2013; Mali et al., 2013; Shen et al., 2014; Li et al., 2015) with appropriate primers (Table S7). All plasmids were confirmed by sequencing. To screen for inversion cell clones, HEC-1B or HEK293T cells at about 80% confluence were transfected with Lipofectamine 2000 reagents (Invitrogen) in a 6-well plate with plasmid DNA including pcDNA3.1-Cas9 and sgRNA constructs (2 μg each). The primers used for genotyping are listed in Table S7.

Reverse-transcriptase PCR

Total RNA was extracted from brain tissues or cultured cells using the Qiagen RNeasy system. Reverse-transcription was performed using reagents from Promega with 1 μg of total RNA. PCR was then performed as follows: 94 °C for 4 min; 35 cycles of 94 °C for 30 sec, 60 °C for 30 sec, 72 °C for 30-60 sec; and 72 °C for 5 min.

Genome-wide Computational Analyses

To identify putative CBSs and their orientations, we scanned ChIP-seq peak regions in human K562, MCF-7, H1-hESC, and IMR90 as well as mouse E14 pluripotent cells using the STORM program (CREAD-0.84) and the CTCF position weight matrices (PWM) (Schones et al., 2007; Schmidt el al., 2012). We defined CBS sequences with the highest PWM score on the forward (forward orientation) or on the reverse (reverse orientation) strands as a CTCF-occupied CBS by using the STORM program. To study the correlation between CBS orientation and CTCF-mediated chromatin looping, we first filtered CTCF ChIA-PET interactions for tethered DNA fragments in which both fragments contain CBSs (ENCODE Project Consortium, 2012; Handoko et al., 2011). ChIA-PET measures interactions of DNA fragments of paired-end tags (PETs) in a form of “tag-linker-tag”. Inter-ligation PETs refer to the reads from different DNA fragments. PET sequences that overlap at both ends form PET clusters. PET clusters of multiple PETs reflect the strength of chromatin interactions. Thus, inter-ligation PETs predict the chromatin interactions by clustering (Li et al., 2012). We then screened for CTCF/cohesin-mediated interactions in different combinations of orientation configuration of CBS pairs (i.e., forward-reverse, forward-forward, reverse-reverse, and reverse-forward) using a Python script. Clusters of overlapping chromatin-looping interactions with looping strength >300 (Li et al. 2010; Handoko et al., 2011) were merged to form a CTCF/cohesion-mediated chromatin domain (CCD). The orientation

configuration of CBS pairs between neighboring domains was quantified for each chromosome and combined to give the total number of domains in the whole human genome. The sources of the public data (ENCODE Project Consortium, 2012; Handoko et al., 2011) used: GEO accession numbers: GSM822297, GSM935379, GSM935404, GSM935624, GSM935407, GSM935310, GSM970216, GSM1022658, GSM1010791, and GSM970215.

Animals

Animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of Shanghai Jiao Tong University.

RNA sequencing (RNA-Seq)

RNA-Seq experiments were performed as previously described (Mortazavi et al., 2008; Shen et al., 2012). Briefly, total RNA was extracted from cultured cells (inversion and WT control in duplicates) using an RNeasy plus mini kit (Qiagen) according to the manufacturer's protocol. Messenger RNA was isolated from the total RNA by oligo (dT) magnetic beads, and fragmented under heating condition. After the first and second cDNA strand synthesis, as well as ends repairing, the 3' ends of cDNA fragments were added with a single 'A' nucleotide to facilitate adapter ligation. The sequencing libraries were then generated by PCR amplification of the cDNA products, and validated by an

Agilent 2100 Bioanalyzer before sequencing on the Illumina sequencing platform of HiSeq 2000. The resulting sequencing reads (49 bp, single-read) were mapped onto the human genome build GRCh37 using TopHat-v2.0.14 (Trapnell et al., 2009) with the setting of "-N 0 –g 1 –x 1". The output data were then averaged among biological replicates and normalized to reads per million (RPM) by the genomeCoverageBed program (Quinlan and Hall, 2010). The images were then generated using the University of California Santa Cruz (UCSC) Genome Browser (Kent et al., 2002). In order to identify genes that were changed in expression between two groups of data, the TopHat mapped reads were analyzed by the DEGseq program with the setting of "MARS" (Wang et al., 2010). A revised genome annotation file containing the clustered *Pcdh* genes (removing the three constant exons of *Pcdh*α and γ gene clusters which can affect the comparison of gene expression, as well as adding the annotation of *Pcdh*γ*b5* gene which is missing in the public genome annotation file) was used in the analysis.

SUPPLEMENTAL REFERENCES

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.W., and Xi, J.J. (2013). Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res. 23, 465-472.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823.

Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B., and Liu, J.S. (2012). HiCNorm: removing biases in Hi-C data via Poisson regression. Bioinformatics 28, 3131-3133.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. 12, 996-1006.

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

Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C., et al. (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol. 11, R22.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148, 84-98.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science 339, 823-826.

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

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

Schones, D.E., Smith, A.D., and Zhang, M.Q. (2007). Statistical significance of cis-regulatory modules. BMC Bioinformatics 8, 19.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat. Methods 11, 399-402.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. 38, 1348-1354.

Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58, 221-230.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105-1111.

Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res. 41, e132.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26, 136-138.