Molecular Cell, Volume *76*

Supplemental Information

Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF

Anders S. Hansen, Tsung-Han S. Hsieh, Claudia Cattoglio, Iryna Pustova, Ricardo Saldaña-Meyer, Danny Reinberg, Xavier Darzacq, and Robert Tjian

Supplemental Figures and Figure Legends

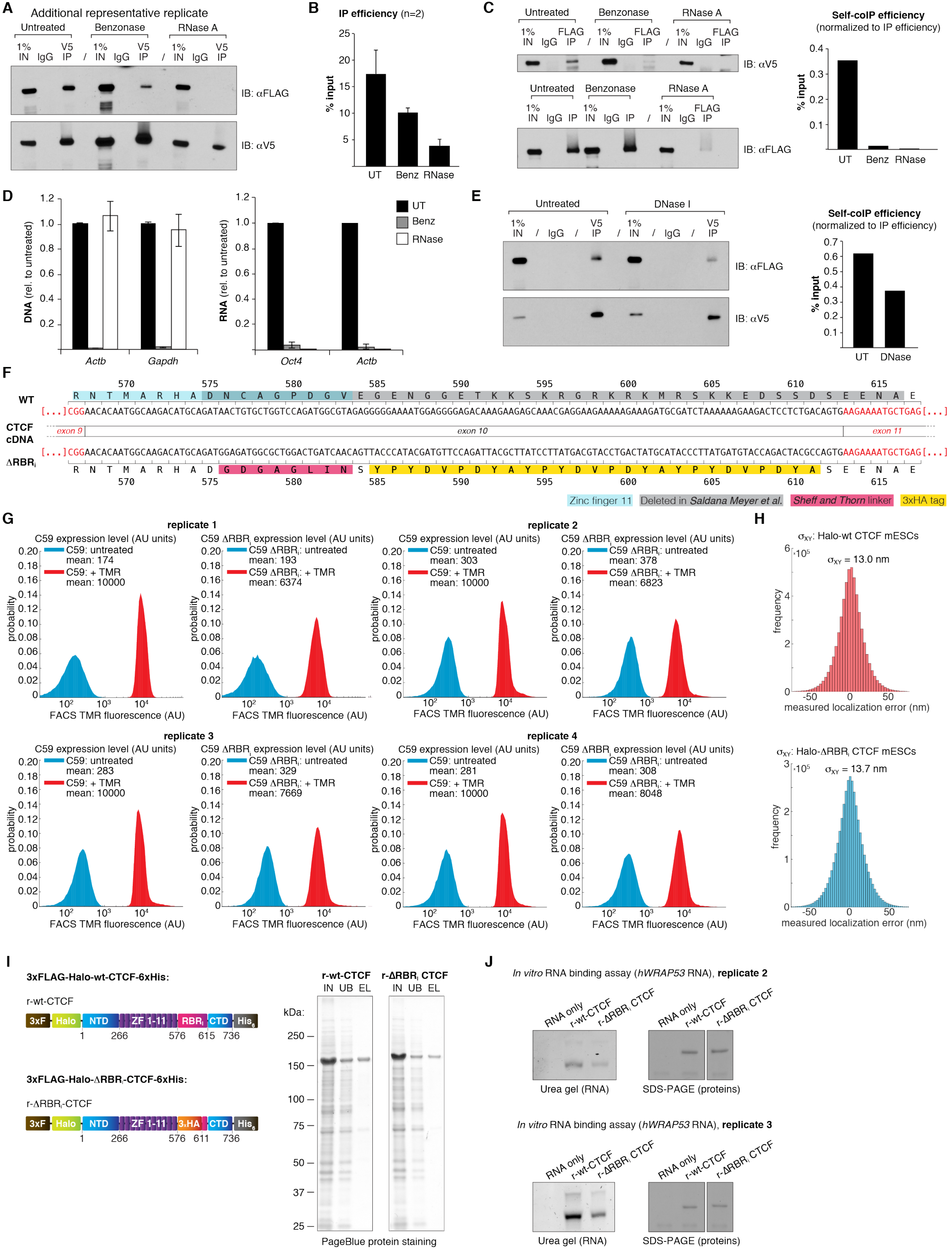


Figure S1. Related to Figure 1 and 2.

Additional coIP experiments and controls. (**A**) Representative coIP experiment (V5 IP) indicating RNA-dependent CTCF self-interaction (additional replicate). Top: V5 IP followed by FLAG immunoblotting measures self-coIP efficiency (90% of total IP material loaded); bottom: V5 IP followed by V5 immunoblotting controls for IP efficiency (remaining 10% of IP sample loaded). (**B**) IP efficiency for the V5 coIP experiments of Figure 1C and S1A, used to calculate the self-coIP efficiency of Figure 1D. Error bars are SD, n=2. (**C**) Reciprocal coIP experiment (left) and quantification (right). Top left: FLAG IP followed by V5 immunoblotting measures self-coIP efficiency (45% of total IP material loaded); bottom left: FLAG IP followed by FLAG immunoblotting controls for IP efficiency (remaining 45% of IP sample loaded). (**D**) Effective nucleic acid digestion by Benzonase (Benz) and RNase A during CTCF coIP experiments. DNA (left) and RNA (right) were extracted from coIP lysates and quantified by qPCR and RT-qPCR, respectively, using primers specific to *Actb* and *Gapdh* gene/mRNA. Error bars are SD, n=3. (**E**) CoIP experiment with DNase I treatment (left) and quantification (right). Top left: V5 IP followed by FLAG immunoblotting measures self-coIP efficiency (80% of total IP material loaded); bottom left: V5 IP followed by V5 immunoblotting controls for IP efficiency (remaining 20% of IP sample loaded). (**F**) Detailed view of CTCF mRNA exon 10 in wild type and ΔRBR$_i$-CTCF mESCs. Both the DNA and the protein sequences (one letter code) are provided. Relevant features are highlighted in colors. Amino acid numbers are based on the NCBI Reference Protein NP_851839.1. (**G**) Quantification of Halo-wt CTCF and C59 Halo-ΔRBR$_i$-CTCF expression levels in the mESC clones C59 and C59 ΔRBR$_i$, respectively. Cells were labeled with 500 nM Halo-TMR for 30 min and their background-subtracted fluorescence measured on a LSR Fortessa Cytometer, exciting fluorescence with a 561 nm laser and collecting fluorescence through a 610/20 bandpass emission filter. All 4 biological replicates are shown. (**H**) Quantification of localization error/uncertainty in PALM experiments. Localization error (defined as the standard deviation) was measured from single-molecule localizations that appeared for at least 20 frames. We then calculated the mean X,Y coordinates and took the difference between the measured X,Y coordinates in a given frame from the overall mean to be the localization error. (**I**) Recombinant 3xFLAG-Halo-wt-CTCF-6xHis (r-wt-CTCF) and 3xFLAG-Halo-ΔRBR$_i$-CTCF-6xHis (r-ΔRBR$_i$-CTCF) were purified from insect cells with a two-step affinity purification scheme and their purity checked by SDS-PAGE and PageBlue protein staining. Shown are eluates from the first Ni-NTA column that served as inputs (IN) to the FLAG pulldown used as a second purification step. UB: fraction not bound to the FLAG resin, EL: eluate from the FLAG pulldown, which was then used for *in vitro* RNA binding assays. (**J**) Replicates of the *in vitro* RNA binding assay shown in Figure 2E. A fragment of human *WRAP53* mRNA (*hWRAP53*, nucleotides 1 to 167) was transcribed *in vitro* and incubated with a 5 molar excess of recombinant (r-) wt- or ΔRBR$_i$-CTCF protein (see STAR methods for details). Recovered RNA was run on urea denaturing gels and stained with SYBR Gold, while recovered proteins were run on SDS-PAGE and stained with PageBlue (proteins were run on the same gel, which is here cropped for clarity to remove lanes irrelevant to this study).

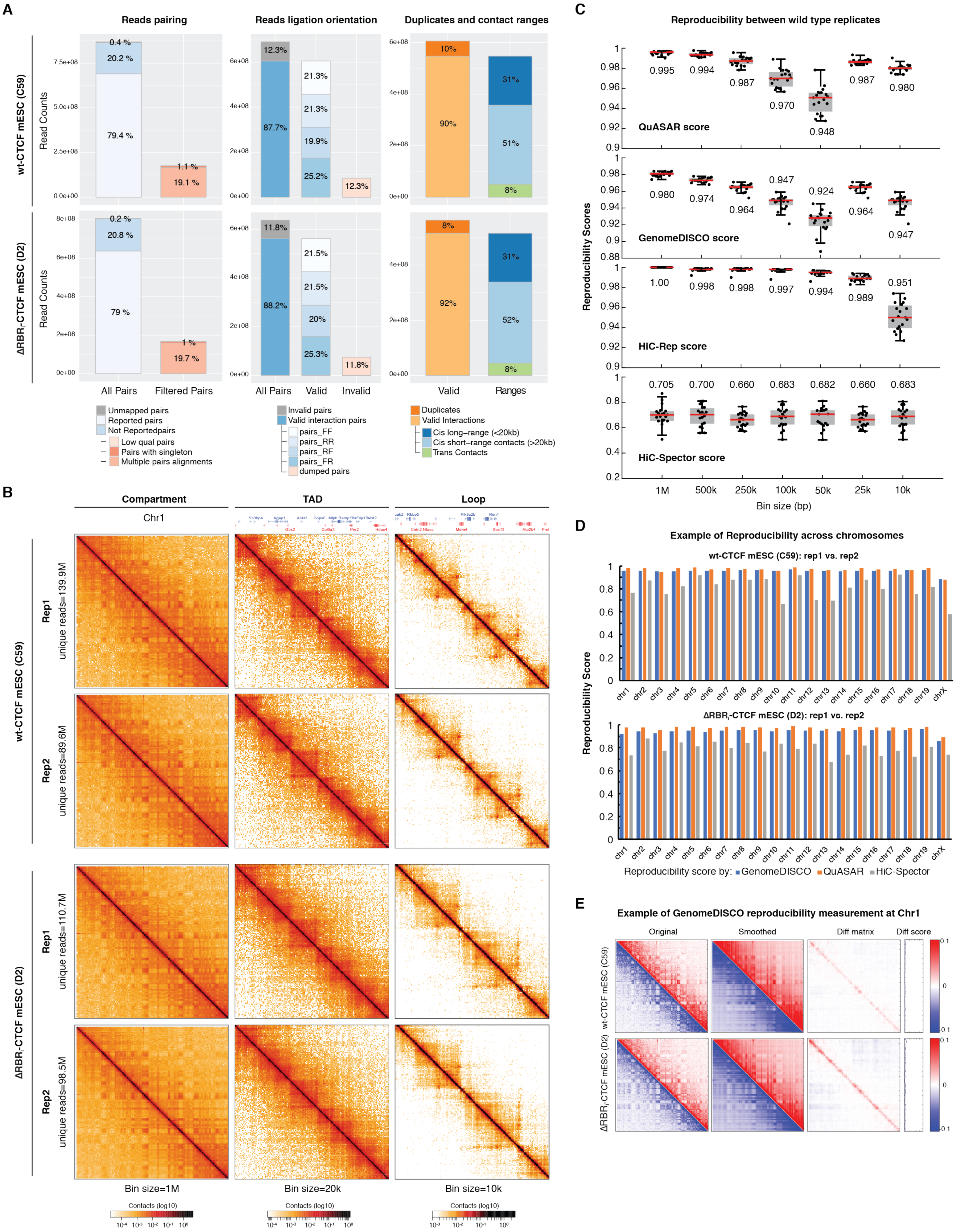

A
Reads pairing
Reads ligation orientation
Duplicates and contact ranges
wt-CTCF mESC (C59)
ΔRBR$_i$-CTCF mESC (D2)
Read Counts
0.4 %
20.2 %
79.4 %
1.1 %
19.1 %
12.3%
87.7%
21.3%
21.3%
19.9%
25.2%
12.3%
10%
90%
31%
51%
8%
0.2 %
20.8 %
79 %
1 %
19.7 %
11.8%
88.2%
21.5%
21.5%
20%
25.3%
11.8%
8%
92%
31%
52%
8%
All Pairs
Filtered Pairs
All Pairs
Valid
Invalid
Valid
Ranges
Unmapped pairs
Reported pairs
Not Reportedpairs
Low qual pairs
Pairs with singleton
Multiple pairs alignments
Invalid pairs
Valid interaction pairs
pairs_FF
pairs_RR
pairs_RF
pairs_FR
dumped pairs
Duplicates
Valid Interactions
Cis long–range (<20kb)
Cis short–range contacts (>20kb)
Trans Contacts
B
Compartment
TAD
Loop
Chr1
Rep1
unique reads=139.9M
Rep2
unique reads=89.6M
Rep1
unique reads=110.7M
Rep2
unique reads=98.5M
wt-CTCF mESC (C59)
ΔRBR$_i$-CTCF mESC (D2)
Bin size=1M
Bin size=20k
Bin size=10k
Contacts (log10)
C
Reproducibility between wild type replicates
Reproducibility Scores
QuASAR score
0.995
0.994
0.987
0.970
0.948
0.987
0.980
GenomeDISCO score
0.980
0.974
0.964
0.947
0.924
0.964
0.947
HiC-Rep score
1.00
0.998
0.998
0.997
0.994
0.989
0.951
HiC-Spector score
0.705
0.700
0.660
0.683
0.682
0.660
0.683
1M
500k
250k
100k
50k
25k
10k
Bin size (bp)
D
Example of Reproducibility across chromosomes
wt-CTCF mESC (C59): rep1 vs. rep2
ΔRBR$_i$-CTCF mESC (D2): rep1 vs. rep2
Reproducibility Score
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chrX
Reproducibility score by: ■GenomeDISCO ■QuASAR ■HiC-Spector
E
Example of GenomeDISCO reproducibility measurement at Chr1
Original
Smoothed
Diff matrix
Diff score
wt-CTCF mESC (C59)
ΔRBR$_i$-CTCF mESC (D2)

Figure S2. Related to Figure 3

Micro-C mapping summary and reproducibility. (**A**) Statistics of Micro-C assays in wt-CTCF and ∆RBR_i-CTCF mESCs. About 80% of reads were successfully aligned and paired. Interaction orientations are equally distributed in four directionalities: forward-forward, reverse-reverse, forward-reverse, or reverse-forward. Among all valid pairs, ~50% of pairs are cis interactions shorter than 20 kb, ~30% of pairs are cis interactions longer than 20 kb, and 8% are inter-chromosomal interactions. The statistics are highly similar to those of high quality Hi-C data, except that Micro-C captures more short-range of interactions. (**B**) Micro-C recapitulates chromatin structures including compartments, TADs, and loops in two biological replicates with ~100 M reads per sample. (**C**) Micro-C reproducibility analysis for two replicates of wild type samples. We chose to use 4 algorithms to cross-validate the reproducibility of Micro-C data. 1) QuASAR calculates the correlation of values in two distance-based transformed matrices. 2) GenomeDISCO measures differences in two smoothed contact maps. 3) Hi-Rep calculates reproducibility by weighted sum of correlation coefficients. 4) HiC-Spector measures weighted difference of eigenvectors. All methods reported high reproducibility rate at the resolution from 10 kb to 1 Mb. (**D**) An example of reproducibility rate across chromosomes. (**E**) An example of reproducibility measurement by GenomeDISCO. The original contact matrices in chr1 were smoothed by graph diffusion. Reproducibility scores can be obtained by calculating the difference of subtraction of two smoothed matrices.

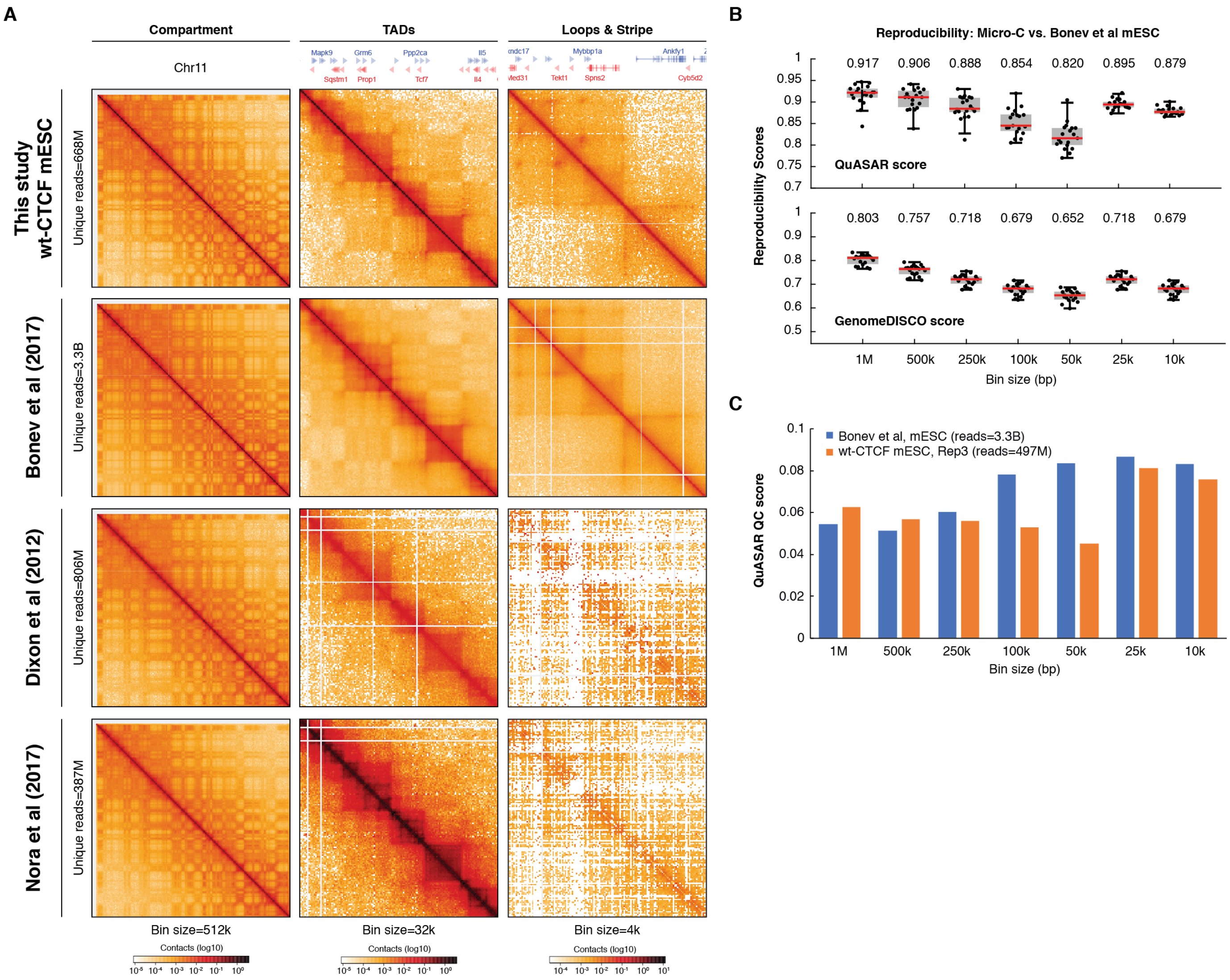


Figure S3. Related to Figure 3

Comparison of Micro-C and published Hi-C datasets. (**A**) Snapshots of chromatin organization in wt-CTCF mESC mapped by Micro-C, Bonev et al (2017), Dixon et al (2012), and Nora et al (2017). Micro-C and Hi-C successfully identify large-scale chromatin structures like compartments and TADs with similar signal strength. However, standard Hi-C requires many more reads to capture chromatin structures at finer scales, as the datasets from Dixon et al and Nora et al have no enrichment in loops and stripe structures. Micro-C robustly reveals all kinds of chromatin signatures with significant fewer reads. (**B**) Reproducibility analysis of Micro-C and Hi-C data from Bonev et al (2017). The reproducibility was calculated by the same approaches as Figure S2. The reproducibility score ranges from 0.8 to 0.9 by QuASAR and from 0.65 to 0.8 by GenomeDISCO at 10 kb to 1 Mb resolutions. (**C**) Data quality analysis. Quality scores were calculated by sequencing coverage and background noise across multiple resolutions. Higher score means higher read depth and lower random noise.

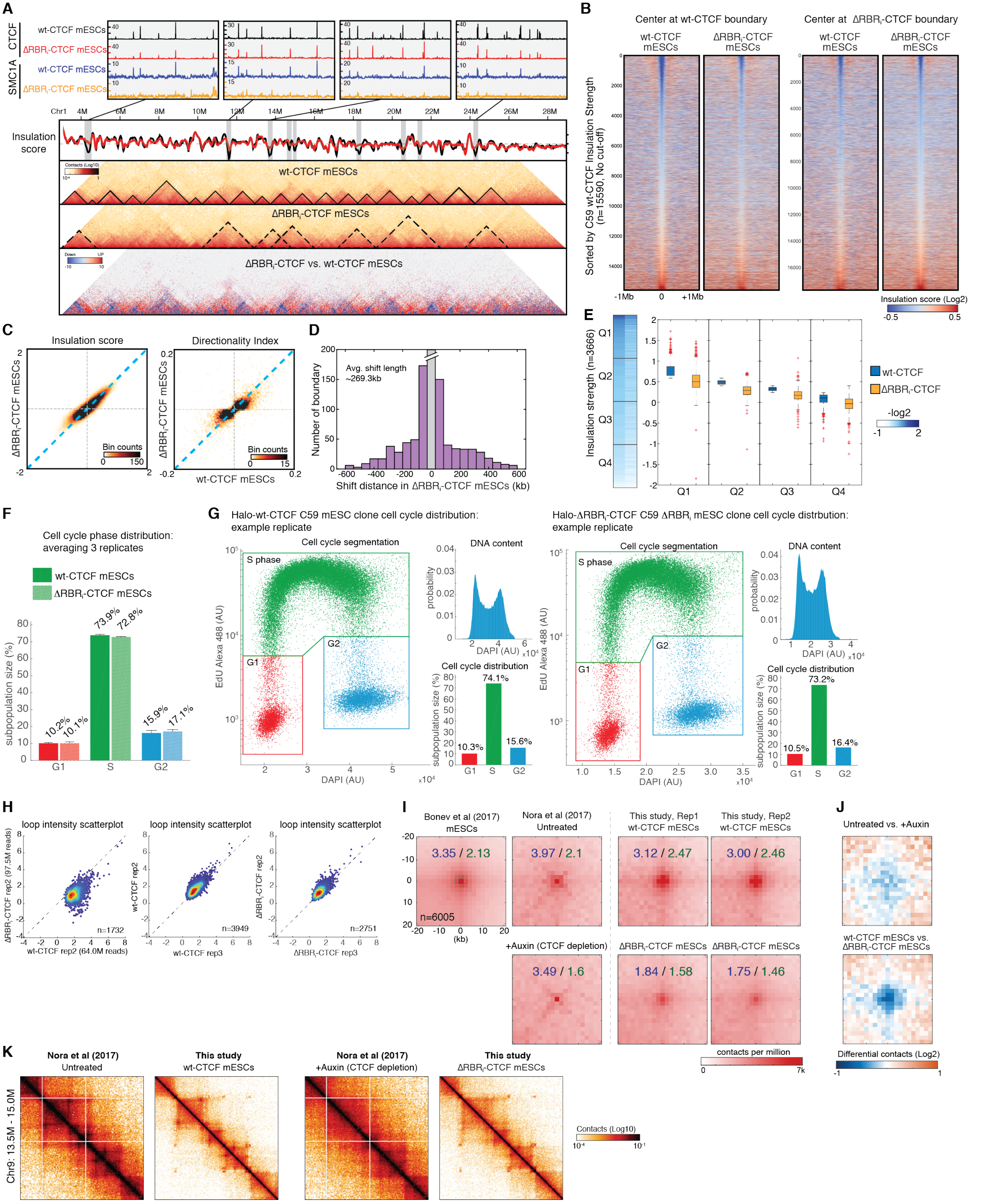

A
CTCF
wt-CTCF mESCs
ΔRBRi-CTCF mESCs
SMC1A
wt-CTCF mESCs
ΔRBRi-CTCF mESCs
Chr1 4M 6M 8M 10M 12M 14M 16M 18M 20M 22M 24M 26M 28M
Insulation score
Contacts (Log10)
wt-CTCF mESCs
ΔRBRi-CTCF mESCs
Down UP
ΔRBRi-CTCF vs. wt-CTCF mESCs
B
Center at wt-CTCF boundary
Center at ΔRBRi-CTCF boundary
wt-CTCF mESCs
ΔRBRi-CTCF mESCs
Sorted by C59 wt-CTCF Insulation Strength (n=15590, No cut-off)
-1Mb 0 +1Mb
Insulation score (Log2)
-0.5 0.5
C
Insulation score
Directionality Index
ΔRBRi-CTCF mESCs
wt-CTCF mESCs
Bin counts
0 150
0 15
D
Avg. shift length ~269.3kb
Number of boundary
Shift distance in ΔRBRi-CTCF mESCs (kb)
E
Insulation strength (n=3666)
Q1 Q2 Q3 Q4
wt-CTCF
ΔRBRi-CTCF
-log2
-1 2
F
Cell cycle phase distribution: averaging 3 replicates
wt-CTCF mESCs
ΔRBRi-CTCF mESCs
subpopulation size (%)
10.2% 10.1% 73.9% 72.8% 15.9% 17.1%
G1 S G2
G
Halo-wt-CTCF C59 mESC clone cell cycle distribution: example replicate
Halo-ΔRBRi-CTCF C59 ΔRBRi mESC clone cell cycle distrbution: example replicate
Cell cycle segmentation
S phase
G1
G2
EdU Alexa 488 (AU)
DAPI (AU)
DNA content
probability
Cell cycle distribution
10.3% 74.1% 15.6%
10.5% 73.2% 16.4%
H
loop intensity scatterplot
ΔRBRi-CTCF rep2 (97.5M reads)
wt-CTCF rep2 (64.0M reads)
n=1732
wt-CTCF rep2
wt-CTCF rep3
n=3949
ΔRBRi-CTCF rep2
ΔRBRi-CTCF rep3
n=2751
I
Bonev et al (2017) mESCs
Nora et al (2017) Untreated
This study, Rep1 wt-CTCF mESCs
This study, Rep2 wt-CTCF mESCs
3.35 / 2.13
3.97 / 2.1
3.12 / 2.47
3.00 / 2.46
n=6005
(kb)
+Auxin (CTCF depletion)
ΔRBRi-CTCF mESCs
ΔRBRi-CTCF mESCs
3.49 / 1.6
1.84 / 1.58
1.75 / 1.46
contacts per million
0 7k
J
Untreated vs. +Auxin
wt-CTCF mESCs vs. ΔRBRi-CTCF mESCs
Differential contacts (Log2)
-1 1
K
Nora et al (2017) Untreated
This study wt-CTCF mESCs
Nora et al (2017) +Auxin (CTCF depletion)
This study ΔRBRi-CTCF mESCs
Chr9: 13.5M - 15.0M
Contacts (Log10)

Figure S4. Related to Figure 4 and 5.

TAD/Insulator analysis, cell cycle distribution and loop analysis and controls (**A**) Additional example of TAD/Insulator disruption in ΔRBR$_i$-CTCF mESCs. Top panel shows the browser tracks of CTCF and Smc1a ChIP-Seq data zoomed-in on the regions where insulators were disrupted. Smc1a signal is largely reduced in ΔRBR$_i$-CTCF mESCs in all the cases shown here. Bottom panel shows a snapshot for insulation score, contact maps for wt-CTCF and ΔRBR$_i$-CTCF mESCs, and differential contact maps at Chr1:4M – 28M. Insulation scores were analyzed as described in the Methods section. A lower insulation score represents a strong insulator activity. ΔRBR$_i$-CTCF mESCs (red) lose some peaks comparing to wt-CTCF mESCs (black). Contact maps shown in the same region highlight larger TADs, which corresponds to insulation disruption in ΔRBR$_i$-CTCF mESCs. Gain of contacts between domains in ΔRBR$_i$-CTCF mESCs also supports this observation. (**B**) Heatmaps of insulation strength centered at the called insulators (n=15590) in wt-CTCF or ΔRBR$_i$-CTCF mESCs flanked by ±1Mb and plotted by the rank of the insulation score from low to high (strong to weak insulators) without a cutoff value. The insulation strength is decreased in ΔRBR$_i$-CTCF mESCs when centering at the wt-CTCF insulators, with a more evident effect for the strongest insulators. However, there is no significant change in insulation strength between wt-CTCF and ΔRBR$_i$-CTCF mESCs when centering at ΔRBR$_i$-CTCF insulators. (**C**) Binned scatter plots of insulation score and directionality index. Scatter plots were binned to 250 bins and the quantity of each bin was shown as the color bar. Insulation score was analyzed as described in the Methods section and directionality index was analyzed as described in (Dixon et al., 2012). Two independent approaches confirmed that insulation strength is reduced in ΔRBR$_i$-CTCF mESCs. (**D**) Histogram of shift distance of ΔRBR$_i$-CTCF insulators. The insulators unique to ΔRBR$_i$-CTCF mESCs were taken to calculate one's distance to the closest insulator in wt-CTCF mESCs. The majority of insulators unique to ΔRBR$_i$-CTCF mESCs are shifted ~50 to 300 kb from the original site (average length ~269.3 kb). Insulators with shifted distance shorter than 40 kb were excluded from the analysis.

(**E**) Quantification of insulation strength. Insulation strength was analyzed by using insulation score algorithm (see detail in STAR Methods). Heatmap plot was sorted by wt-CTCF mESC insulation strengths (-log2) and grouped into quartiles (Q1-Q4). The distribution of insulation strengths for each quartile was plotted as box plots. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points; '+' symbols indicate individual outliers.

(**F-G**) No significant change to cell cycle distribution in ΔRBR$_i$-CTCF mESCs. (**F**) Distribution of cell cycle phases in wt-CTCF and ΔRBR$_i$-CTCF mESCs (from 3 biological replicates). (**G**) Representative replicate (1 of 3). Cell cycle analysis was performed using the Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (ThermoFisher Scientific Cat. # C10425) according to manufacturer's instructions. DNA content was inferred from DAPI staining and G1, S and G2 phases were gated as illustrated in red, green and blue in the plot.

(**H-J**) Controls for loop analysis.

(**H**) Analysis of loop strength reproducibility between replicates. To validate the reproducibility of loop analysis between biological replicates, we quantified loop intensity (n=14732) for each replicate. Approximate ~3949 loops in wt-CTCF mESC (unique reads=64M) and ~2751 loops in ΔRBRi-CTCF mESC (unique reads=97.5M) meet our stringent filtering criteria (FDR < 0.1). Although replicates with low-sequencing depth result in fewer high-quality loops, the differences of loop intensity between wt-CTCF and ΔRBRi-CTCF mESC nevertheless robustly recapitulate the finding in Figure 5A. Also, loop intensities within the same cell types are highly correlated (Spearman's=~0.7), while the correlation coefficient value drops to ~0.3 between wt-CTCF and ΔRBRi-CTCF mESC for both replicates. Hence, the loop analysis used in this study detects the difference of loop intensity reproducibly across replicates.

(**I**) Aggregate peak analysis for published Hi-C datasets and Micro-C. 6005 loops were called by using (Bonev et al., 2017) mESC dataset with FDR < 0.1. Datasets include mESCs in Bonev et al, wt-CTCF and AID-CTCF mESCs in (Nora et al., 2017), and two replicates of wt-CTCF and ΔRBR$_i$-CTCF mESCs generated by Micro-C. Loops were aggregated and plotted at the center of a 20 kb x 20 kb window. The enrichments denoted on the plot were calculated by dividing one pixel at the center (blue) or five pixels around the center (green) by the average of bottom-left pixels. (**J**) Differential contact maps. Differential contact matrix was calculated by the log2 change between untreated vs. +auxin (top) or wt-CTCF vs. ΔRBR$_i$-CTCF mESCs (bottom). (**K**) Snapshots of the same region shown in Nora et al. Loops are largely disrupted in CTCF-depleted mESCs (left panel) and in ΔRBR$_i$-CTCF mESCs (right panel). This further confirms our genome-wide analysis that both CTCF-depletion and ΔRBR$_i$-CTCF affect loop formation/stability, although high-resolution Micro-C data may not be directly comparable to the standard Hi-C data.

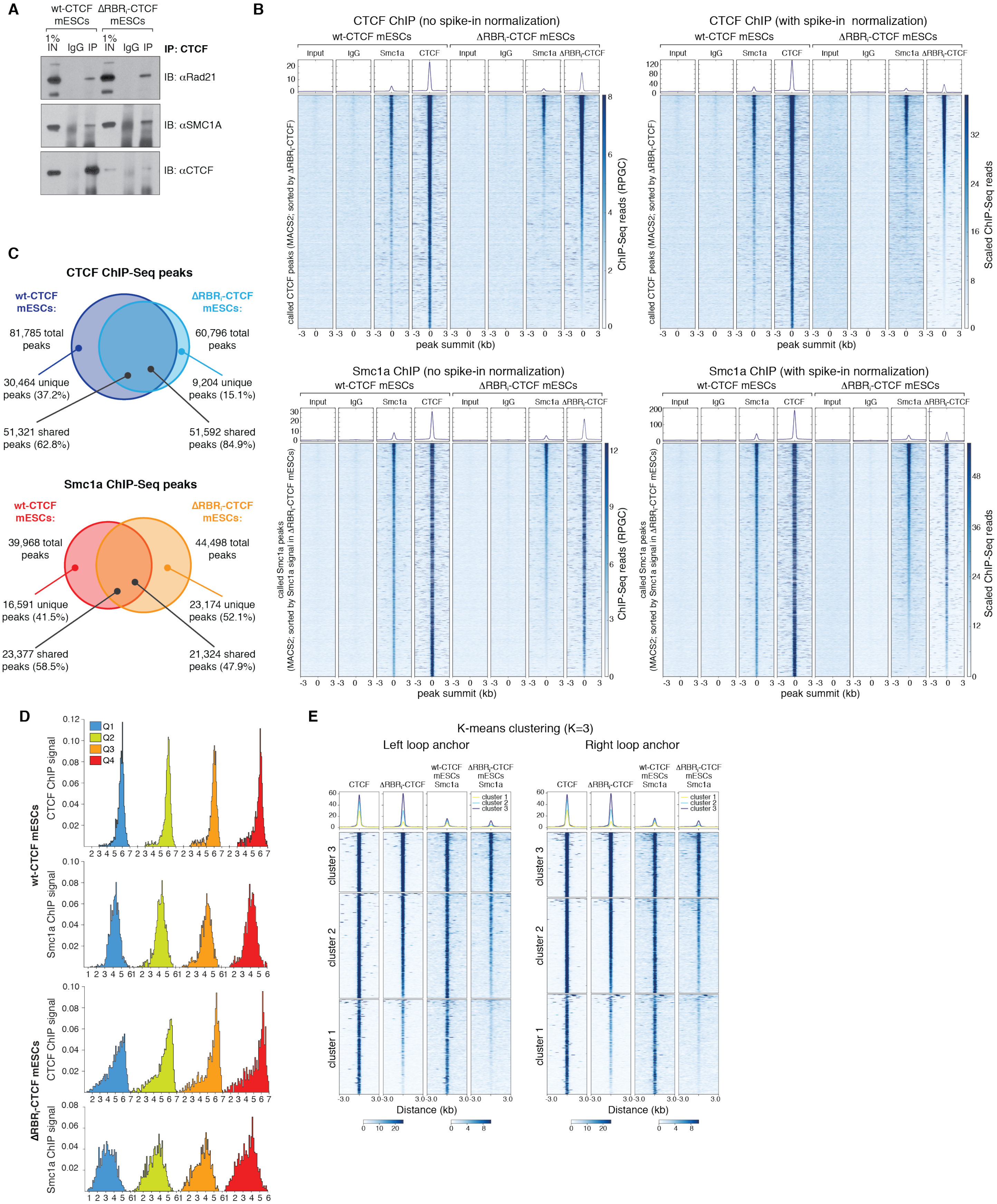
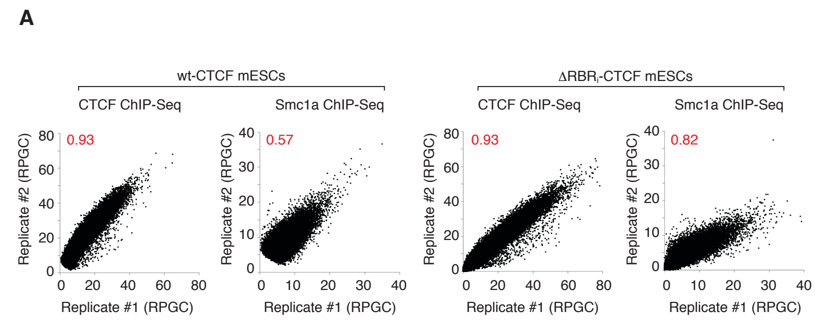


Figure S5. Related to Figure 6.

Additional coIP experiment and ChIP-Seq / Micro-C analyses. (**A**) Additional coIP experiment to the one shown in Figure 6A, demonstrating that CTCF interaction with cohesin is preserved upon deletion of the RBR$_i$. CTCF antibodies can pull down Rad21 and Smc1a cohesin subunits in both wt- and ΔRBR$_i$-CTCF mESCs. (**B**) Heatmaps of CTCF and Smc1a ChIP-Seq signal around wt-CTCF (top 2 panels) and Smc1a in wt-CTCF mESCs (bottom 2 panels) peaks as called by MACS2. We show results after normalization by sequencing depth (left panels; deepTools RPGC: reads per genomic content) or after normalization by the spike-in yeast DNA (right panels; scaled number of reads; see STAR Methods on how we computed the scale factor). Heatmaps are sorted by the peak intensity of ∆RBR$_i$-CTCF (top) and Smc1a in ∆RBR$_i$-CTCF mESCs (bottom). Above the heatmaps are summary plots with mean signal intensities. (**C**) Venn diagrams showing overlap of CTCF and Smc1a ChIP-Seq peaks called by MACS2 in wt-CTCF and ∆RBR$_i$-CTCF mESCs. Peaks were considered overlapping if sharing at least 1 bp.

(**D**) Histograms of ChIP signal distribution for Q1-Q4 loop anchors (also see CDF in Figure 6D). Loop anchors were identified as described in the Method section. ChIP signal enrichments were quantified in a 500 bp window centered at the anchors and plotted as a function of probability distribution in 50 bins. (**E**) Heatmaps of k-means clustering for the Q1 loop anchors (also see K-S probability density curve in Figure 6E). ChIP signals at ± 3 kb of the left and right loop anchors were clustered by k-means analysis with k=3 and sorted by the sum of regions in CTCF and Smc1a data in ∆RBR$_i$-CTCF mESCs. We also performed k-mean clustering analysis with k=2 and k=4 (not shown). Consistent to our conclusion, both results indicate two major subclasses of RBR$_i$-dependent structures with complete/partial loss (clusters 1 and 2, type 1 loop in Figure 5D, 6E) and no loss (cluster 3, type 2 loop in Figure 5D, 6E) of CTCF and/or cohesin in ∆RBR$_i$-CTCF mESCs.



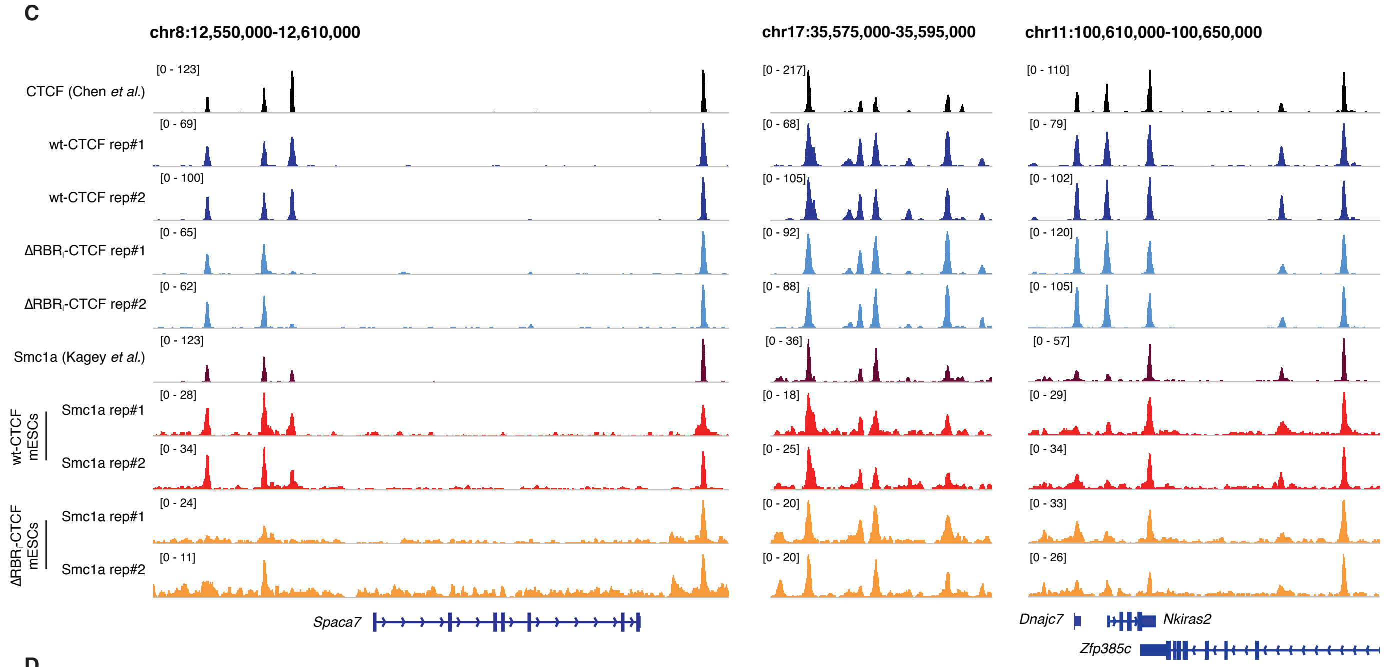
B

wt-CTCF mESCs ChIP-Seq peaks

CTCF ChIP-Seq	Unique	Shared	Total	% Unique	% Shared
Replicate #1	860	58740	59600	1.4%	98.6%
Replicate #2	22223	58531	80754	27.5%	72.5%
Smc1a ChIP-Seq	Unique	Shared	Total	% Unique	% Shared
Replicate #1	1624	14942	16566	9.8%	90.2%
Replicate #2	18530	14974	33504	55.3%	44.7%

ΔRBR$_i$-CTCF mESCs ChIP-Seq peaks

CTCF ChIP-Seq	Unique	Shared	Total	% Unique	% Shared
Replicate #1	5039	40465	45504	11.1%	88.9%
Replicate #2	6994	40562	47556	14.7%	85.3%
Smc1a ChIP-Seq	Unique	Shared	Total	% Unique	% Shared
Replicate #1	12099	14841	26940	44.9%	55.1%
Replicate #2	7067	15092	22159	31.9%	68.1%



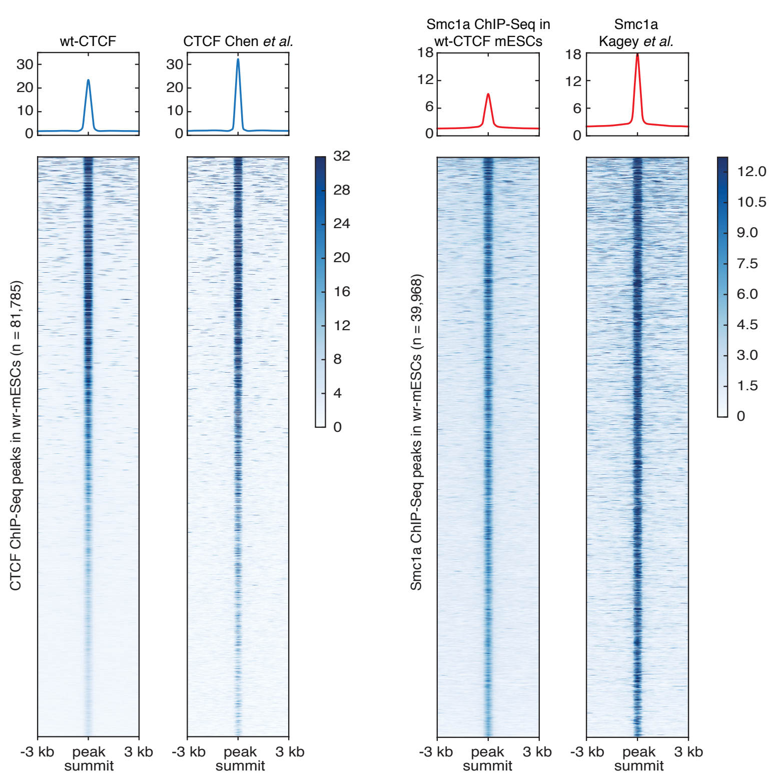


Figure S6. Related to Figure 4-6.

Reproducibility of ChIP-Seq data. We here assess both the reproducibility between ChIP-Seq replicates (A-C) and the consistency between our datasets and those generated by others in mouse embryonic stem cells in the past (C-D). (**A**) Scatterplots of ChIP-Seq signal across the two biological replicates used in this study. The Spearman correlation coefficient for each plot is indicated in red. Each dot is a CTCF or Smc1a ChIP-Seq peak called by MACS2 in wt-CTCF mESCs. Scale is reads per genomic content (deepTools RPGC: number of reads per bin / scaling factor for 1x average coverage). See STAR Methods for details. (**B**) Table reporting the number of unique and shared peaks called by MACS2 across the two biological replicates used in this study for wt-CTCF and ∆RBR$_i$-CTCF mESCs. Peaks are called "shared" when overlapping for at least 1 bp. (**C**) Representative genome-browser views of biological replicates of CTCF and Smc1a ChIP-Seq in wt-CTCF and ∆RBR$_i$-CTCF mESCs. As can been seen, the binding pattern of both CTCF and Smc1a is nearly identical across the two replicates. CTCF data from Chen *et al.* (Chen et al., 2008) and Smc1a data from Kagey *et al.* (Kagey et al., 2010) are also plotted to show the high consistency of our data with previously published ChIP-Seq datasets in mESCs. Scale is reads per genomic content as in (A). (**D**) Heatmaps of ChIP-Seq signals around wt-CTCF (left panels) and Smc1a (right panels) peaks called by MACS2 in wt-CTCF mESCs provide a genome-wide comparison between ChIP-Seq data obtained in this study with previously published datasets (CTCF data from Chen *et al.* (Chen et al., 2008) and Smc1a data from Kagey *et al.* (Kagey et al., 2010)). The side-by-side comparison clearly indicates a high level of correlation between our data and those generated by others. Scale is reads per genomic content as in (A). Heatmaps are sorted by the mean signal intensity of CTCF (leftmost heatmaps) and Smc1a (rightmost heatmaps) datasets generated by this study, calculated across the entire 6 kb interval around peak summits. The plots on top of the heatmaps summarize mean values ± 3 kb around peak summits.

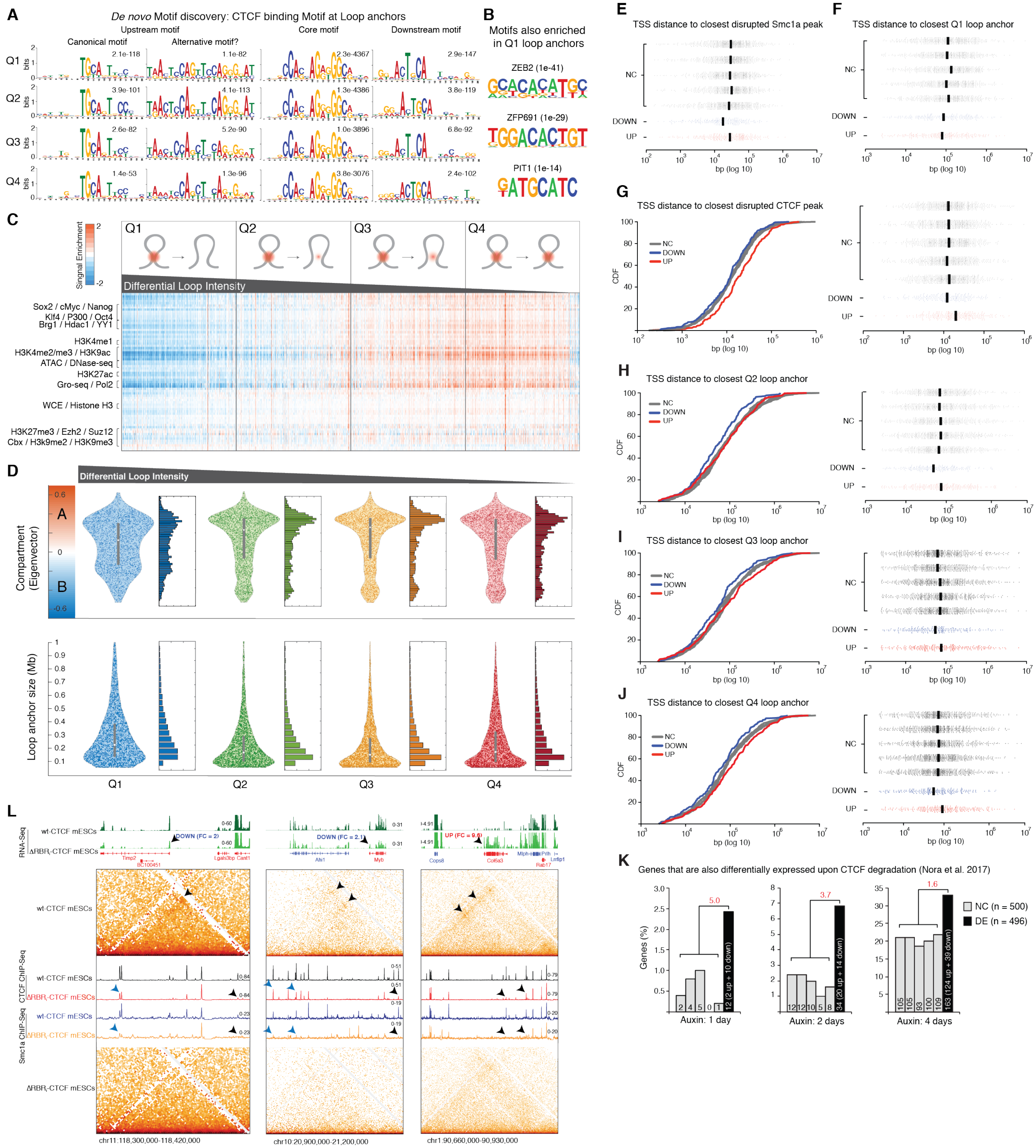


Figure S7. Related to Figure 6 and 7.

DNA determinants of RBR$_i$-dependent and RBR$_i$-independent loops and potential factors in regulating loop formation/stability and Supplemental analyses of RNA-Seq. (**A**) *De novo* motif discovery at loop anchors. Loop anchors were identified as described in the Method section. Motif analysis was performed to identify potential binding sequences at the core loop anchors and ± 20 bp upstream and downstream of them. Motif searching was set as zero or one occurrence per sequence (zoops) mode by using MEME algorithm. Note that an alternative upstream motif was strongly enriched in addition to the CTCF canonical upstream motif, with an even greater significance (lower E-value) in Q2-Q4. The CTCF core motif was strongly and almost equally enriched at the anchors of chromatin loops Q1 through Q4 (Q1 > Q2 > Q3 > Q4), while the canonical CTCF upstream motif was particularly abundant at Q1 loops anchors. This is notable, since the upstream motif is the one recognized by CTCF Zn fingers 9-11 (Nakahashi et al., 2013), which are adjacent to the RBR_i domain. On the contrary, Q1 loop anchors, the least affected by RBR_i deletion, scored less significant for the canonical upstream motif, and were instead enriched in the alternative upstream motif. (**B**) Additional motifs discovered at Q1 loop anchors. Differential motif analysis was performed to identify potential factors involved in loop formation in addition to CTCF. Motifs were searched within a 5-kb window at loop anchors in Q1, and in Q4 as a control. Many prior uncharacterized zinc-finger proteins were shown with a high enrichment at loop anchors. These candidates can be subjects of future investigation. (**C**) Signal enrichments of 70 published genome-wide datasets at loop anchors were quantified as rlog variances by DEseq2. The heatmap was sorted by the changes in loop intensity between wt-CTCF and ΔRBR_i-CTCF mESCs. As a control, whole cell extract (WCE) and histone H3 do not change significantly from Q1 to Q4. Most transcription-related factors (e.g. pluripotent factors, Brg1, YY1, etc.), active chromatin marks (e.g. H3K4me3, H3K9ac, etc.), enhancer marks (e.g. H3K4me1, H3K27ac, P300, etc.), and chromatin accessibility (e.g. ATAC-seq, DNase-seq) are strongly depleted in the Q1 loop anchors but enriched in the Q4 loop anchors. Constitutive heterochromatin marks (e.g. H3K9me2/3) and facultative heterochromatin marks (e.g. H3K27me3) have no preferential enrichment in any of the loop quartiles. (**D**) Distributions of loop anchors. Eigenvalues were ranked at the y-axis in top panel, in which positive EV_1 represents compartment A (active chromatin) and negative EV2 represents compartment B (inactive chromatin). More Q1 loops were found in inactive chromatin or at the border of two compartments but Q2 to Q4 loops were largely in active chromatin. Loop size distribution is similar across the four loop quartiles, ranging from 50 to 500 kb, although larger loops (> 500 kb) are enriched in Q1. This is consistent with a previous report that the average size of TADs/loops is larger in inactive chromatin.

(**E**) Scatterplots of single data points (with median) underlying the cumulative distribution functions shown in Figure 7B for genes not changed (NC; n = 500 for each of the 5 groups), downregulated (DOWN; n = 221) and upregulated (UP; n = 275) in ΔRBR_i-CTCF mESCs compared to wt-CTCF mESCs. (**F**) Scatterplots of single data points (with median) underlying the cumulative distribution functions shown in Figure 7C for genes not changed (NC; n = 500 for each of the 5 groups), downregulated (DOWN; n = 221) and upregulated (UP; n = 275) in ΔRBR_i-CTCF mESCs compared to wt-CTCF mESCs. (**G**) Left: for each gene deregulated (DOWN or UP) in ΔRBR_i-CTCF mESCs, we measured the distance in base pairs (bp) from its transcription start site (TSS) to the closest disrupted CTCF ChIP-Seq peak (i.e., a peak called by MACS2 in wt-CTCF mESCs but not in ΔRBR_i-CTCF mESCs), and plotted the results as a cumulative distribution function (CDF). As controls, we randomly selected five groups of ~ 500 unaltered genes each (not changed, NC). Right: scatter plots with single data points and median value. (**H-J**) Same as (G), but plotting the distance to the closest Q2 (H), Q3 (**I**) and Q4 (J) loop anchor. (**K**) Comparison between genes differentially expressed (DE) in ΔRBR_i-CTCF mESCs *vs.* wt-CTCF mESCs and gene expression changes measured by Nora *et al.* upon acute auxin-induced degradation of CTCF. We counted how many DE genes were also deregulated in the same direction in the CTCF degron system after 1, 2 and 4 days of auxin treatment and plotted it as a percent of total deregulated genes. Actual numbers are also indicated inside the bars. Five groups of 500 genes each randomly selected among those that do not change (NC) expression in ΔRBR_i-CTCF mESCs serve as controls. The red number quantifies the enrichment above averaged controls. Full gene list in Table S2. (**L**) Additional examples of genes differentially expressed in ΔRBR_i-CTCF mESCs compared to wt-CTCF mESCs. Snapshots of three genomic regions showing two genes (*Timp2*, *Myb*) downregulated and one gene (*Col6a3*) upregulated in ΔRBR_i-CTCF mESCs. RNA-Seq tracks are plotted at the very top, and for each deregulated gene (black arrowhead) the direction of deregulation (UP or DOWN), as well as the fold change (FC) in ΔRBR_i-CTCF mESCs *vs.* wt-CTCF mESCs are specified. Blue genes are transcribed from the "plus" strand, red genes from the "minus" strand. Zoomed-in, 1-kb resolution contact maps are plotted on the top and bottom panels for wt-CTCF and ΔRBR_i-CTCF mESCs, respectively. Arrowheads highlight loops disrupted following CTCF RBR_i deletion. CTCF and cohesin (Smc1a) ChIP-Seq data is overlaid, with blue/black arrowheads pointing at the disrupted left/right loop anchors in ΔRBR_i-CTCF mESCs. ChIP-Seq and RNA-Seq scale is reads per genomic content (deepTools RPGC).