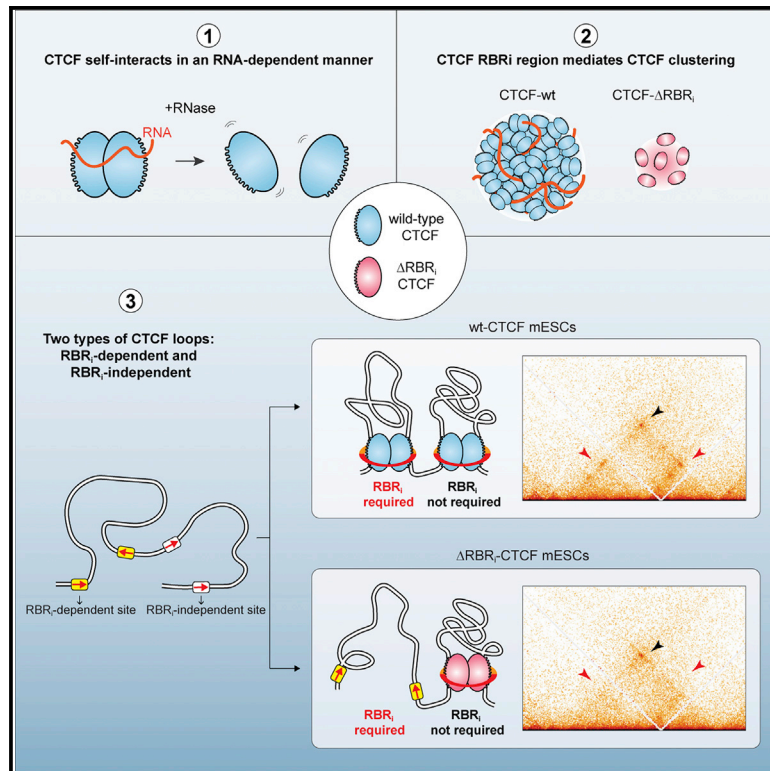


# Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF

## Graphical Abstract



## Authors

Anders S. Hansen, Tsung-Han S. Hsieh, Claudia Cattoglio, ..., Danny Reinberg, Xavier Darzacq, Robert Tjian

## Correspondence

darzacq@berkeley.edu (X.D.),  
jmlim@berkeley.edu (R.T.)

## In Brief

CTCF is an architectural protein that mediates chromatin looping. Here, Hansen et al. demonstrate that an internal RNA-binding region (RBR<sub>i</sub>) in CTCF mediates CTCF clustering and that deletion of the RBR<sub>i</sub> causes disruption of about half of all chromatin loops in mouse embryonic stem cells.

## Highlights

- An RNA-binding region (RBR<sub>i</sub>) in CTCF mediates self-association and clustering
- Reorganization of TADs, loops, and stripes in  $\Delta$ RBR<sub>i</sub> mutant cells
- About half of all CTCF loops are disrupted in  $\Delta$ RBR<sub>i</sub> mutant cells
- CTCF loops fall into two classes: RBR<sub>i</sub> dependent and RBR<sub>i</sub> independent



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

Anders S. Hansen,<sup>1,2,3,4,5,8</sup> Tsung-Han S. Hsieh,<sup>1,2,3,4,8</sup> Claudia Cattoglio,<sup>1,2,3,4,8</sup> Iryna Pustova,<sup>1,2,3,4</sup> Ricardo Saldaña-Meyer,<sup>6,7</sup> Danny Reinberg,<sup>6,7</sup> Xavier Darzacq,<sup>1,2,3,\*</sup> and Robert Tjian<sup>1,2,3,4,9,\*</sup>

<sup>1</sup>Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>2</sup>Li Ka Shing Center for Biomedical and Health Sciences, Berkeley, CA 94720, USA

<sup>3</sup>CIRM Center of Excellence, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>4</sup>Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>5</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

<sup>6</sup>Department of Biochemistry and Molecular Pharmacology, NYU School of Medicine, New York, NY 10016, USA

<sup>7</sup>Howard Hughes Medical Institute, NYU Langone Health, New York, NY 10016, USA

<sup>8</sup>These authors contributed equally

<sup>9</sup>Lead Contact

\*Correspondence: darzacq@berkeley.edu (X.D.), jmlim@berkeley.edu (R.T.)

<https://doi.org/10.1016/j.molcel.2019.07.039>

## SUMMARY

Mammalian genomes are folded into topologically associating domains (TADs), consisting of chromatin loops anchored by CTCF and cohesin. Some loops are cell-type specific. Here we asked whether CTCF loops are established by a universal or locus-specific mechanism. Investigating the molecular determinants of CTCF clustering, we found that CTCF self-association *in vitro* is RNase sensitive and that an internal RNA-binding region (RBR<sub>i</sub>) mediates CTCF clustering and RNA interaction *in vivo*. Strikingly, deleting the RBR<sub>i</sub> impairs about half of all chromatin loops in mESCs and causes deregulation of gene expression. Disrupted loop formation correlates with diminished clustering and chromatin binding of RBR<sub>i</sub> mutant CTCF, which in turn results in a failure to halt cohesin-mediated extrusion. Thus, CTCF loops fall into at least two classes: RBR<sub>i</sub>-independent and RBR<sub>i</sub>-dependent loops. We speculate that evidence for RBR<sub>i</sub>-dependent loops may provide a molecular mechanism for establishing cell-specific CTCF loops, potentially regulated by RNA(s) or other RBR<sub>i</sub>-interacting partners.

## INTRODUCTION

Mammalian genomes are organized at multiple scales ranging from nucleosomes (hundreds of base pairs) to chromosome territories (hundreds of megabases) (Hansen et al., 2018a). At the intermediate scale of kilobases to megabases, mammalian interphase chromosomes are organized into local units known as topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012). TADs are characterized by the feature that two loci within the same TAD contact each other more frequently,

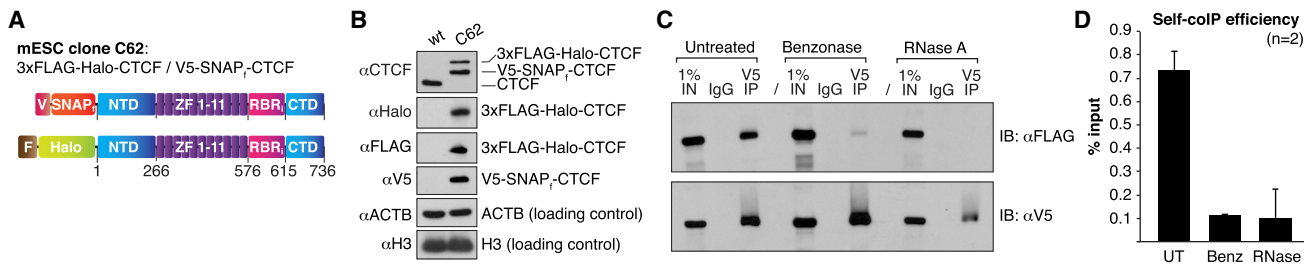
whereas two equidistant loci in adjacent TADs contact each other less frequently. Thus, TADs are thought to regulate contact probability between enhancers and promoters and therefore influence gene expression (Dekker and Mirny, 2016; Merkschlager and Nora, 2016; Rowley and Corces, 2018; Symmons et al., 2014).

Mechanistically, CCCTC-binding factor (CTCF) and the cohesin complex are hypothesized to form TADs through a loop extrusion mechanism: the cohesin ring complex entraps chromatin and extrudes intra-chromosomal chromatin loops until encountering convergently oriented chromatin-bound CTCF molecules on both arms of the loop, halting cohesin-mediated extrusion (Alipour and Marko, 2012; Fudenberg et al., 2016, 2017; Ganji et al., 2018; Sanborn et al., 2015). CTCF and cohesin then hold together a TAD as a chromatin loop until these loop anchor proteins dissociate from chromatin. Thus, both loop extrusion and chromatin loop maintenance are likely dynamic processes (Fudenberg et al., 2016; Hansen et al., 2017, 2018a). Consistent with a key role for CTCF and cohesin, TADs and chromatin loops largely disappear after acute depletion of CTCF and cohesin (Gassler et al., 2017; Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). Moreover, CTCF and several cohesin subunits are among the most frequently mutated proteins in cancer (Hnisz et al., 2017; Lawrence et al., 2014), while disruption of TAD boundaries can cause developmental defects (Lupiáñez et al., 2015).

However, despite their critical role in shaping the three-dimensional (3D) genome organization, we know surprisingly little mechanistically about CTCF and cohesin. Although it is clear that CTCF binds DNA through its 11-ZF domain, the function of CTCF's largely unstructured N- and C-terminal domains remain mostly unknown (Martinez and Miranda, 2010; Merkschlager and Nora, 2016). For example, it is not clear which domain(s) in CTCF are required for its interaction with cohesin and for loop formation. These observations motivated us to investigate whether a universal molecular mechanism controls CTCF and cohesin-anchored loops, or whether distinct classes of CTCF-loops exist.

Along these lines, we and others have recently shown that CTCF forms clusters and foci in cells (Hansen et al., 2017; Zirkel et al.,





**Figure 1. CTCF Self-Interacts in an RNA-Dependent Manner**

(A) Overview of CTCF domains in the endogenously dual-tagged mESC clone C62.

(B) Western blot of total cell lysates from WT mESCs and C62 line. 3xFLAG-Halo-CTCF and V5-SNAP<sub>f</sub>-CTCF are similarly expressed and together roughly equal to CTCF levels in WT cells.

(C) Representative coIP experiment indicating RNA-dependent CTCF self-interaction. 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).

(D) CTCF self-coIP efficiency after normalization for V5 IP efficiency. Error bars indicate SDs; n = 2.

See also [Figures S1A–S1E](#).

2018), and TADs are often demarcated by multiple CTCF binding sites (Kentepozidou et al., 2019). Beyond CTCF, recent work has clearly shown that many proteins are non-homogeneously distributed in the nucleus and dynamically exchanging between regions of local high concentration, termed clusters, condensates, or hubs (Boehning et al., 2018; Cho et al., 2018; Chong et al., 2018). Although in some cases weak and transient protein-protein interactions are sufficient to form and maintain clusters, several examples exist in which nucleic acids can nucleate and/or stabilize protein clusters or hubs (Banani et al., 2017; Chong et al., 2018; McSwiggen et al., 2019; Shin and Brangwynne, 2017). However, the functional role of clustering is poorly understood. We have previously shown that both CTCF and cohesin are clustered in mammalian nuclei (Hansen et al., 2017) and recently that protein-protein interactions play a dominant role in cohesin self-association (Cattoglio et al., 2019). We therefore chose to investigate the molecular determinants of CTCF clustering in cells and their role in regulating 3D genome organization and chromatin looping.

Here, through an integrated approach combining genome editing, single-molecule and super-resolution imaging, *in vitro* biochemistry, PAR-CLIP, ChIP-seq, RNA sequencing (RNA-seq) and Micro-C, we identify critical functions of an RNA-interaction domain C-terminal to CTCF's ZF 11 (RBR<sub>i</sub>). Specifically, we show that the RBR<sub>i</sub> mediates CTCF clustering and that loss of the RBR<sub>i</sub> disrupts only a subset of CTCF-mediated chromatin loops and affects the expression of ~500 genes. Our genome-wide analyses suggest that CTCF boundaries can be classified into at least two sub-classes: RBR<sub>i</sub> dependent and RBR<sub>i</sub> independent. More generally, our work reveals a potential mechanism for establishing and maintaining specific CTCF loops, which may direct the establishment of cell type-specific chromatin topology during development.

## RESULTS

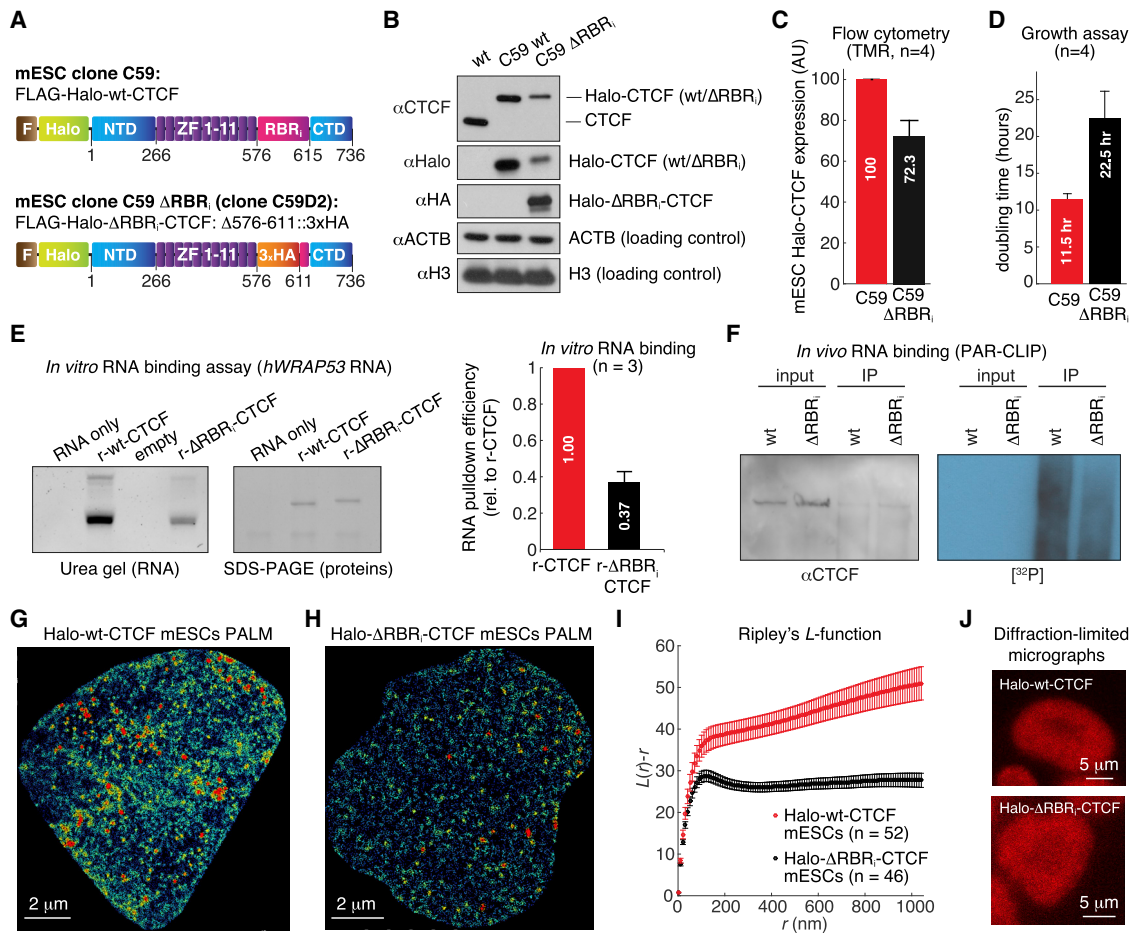
### CTCF Self-Associates in an RNA-Dependent Manner

We have previously shown that CTCF forms clusters in mouse embryonic stem cells (mESCs) and human U2OS cells (Hansen

et al., 2017), and others have reported that CTCF forms larger foci in senescent cells (Zirkel et al., 2018). But what is the mechanisms underlying CTCF cluster formation? Because clusters necessarily arise through direct or indirect self-association, we took a biochemical approach to probe if and how CTCF self-associates. Because CTCF overexpression causes artifacts and alters cell physiology (Hansen et al., 2017; Rasko et al., 2001), we used CRISPR/Cas9-mediated genome editing to generate a mESC line in which one CTCF allele was 3xFLAG-Halo tagged and the other allele was V5-SNAP<sub>f</sub> tagged (C62; [Figures 1A](#) and [1B](#)). Consistent with CTCF clustering, when we immunoprecipitated V5-tagged CTCF, FLAG-tagged CTCF was pulled down along with it (co-immunoprecipitation [coIP]; [Figure 1C](#); additional replicate and quantifications in [Figures S1A](#) and [S1B](#)). Conversely, immunoprecipitation of FLAG-tagged CTCF also co-precipitated significant amounts of V5-tagged CTCF ([Figure S1C](#)). This observation using endogenously tagged CTCF confirms and extends earlier studies that observed CTCF self-association using exogenously expressed CTCF (Pant et al., 2004; Saldaña-Meyer et al., 2014; Yusufzai et al., 2004). But what is the mechanism of CTCF self-interaction? Benzonzase treatment, which degrades both DNA and RNA ([Figure S1D](#)), strongly reduced the coIP efficiency ([Figures 1C](#), [1D](#), and [S1A–S1C](#)) whereas treatment with DNaseI had a significantly weaker effect on the CTCF self-coIP efficiency ([Figure S1E](#)). By contrast, treatment with RNase A alone severely impaired CTCF self-interaction ([Figures 1C](#), [1D](#), and [S1A–S1C](#)). We conclude that CTCF self-associates in a biochemically stable manner *in vitro* that is largely RNA dependent and largely DNA independent.

### An RNA-Binding Region (RBR<sub>i</sub>) in CTCF Mediates RNA Binding and Clustering

Our finding that CTCF self-association is predominantly RNA mediated is perhaps surprising, as CTCF is generally thought of as a DNA-binding protein. However, it confirms studies by Saldaña-Meyer et al. (2014), who also showed that CTCF self-association depends on RNA but not DNA. Importantly, Saldaña-Meyer et al. (2014) described an RNA-binding region



**Figure 2. CTCF RBR<sub>i</sub> Region Mediates CTCF Clustering**

(A) CTCF domains in the mESC clones C59 (Halo-WT CTCF) and C59  $\Delta$ RBR<sub>i</sub> (Halo- $\Delta$ RBR<sub>i</sub> CTCF).  
 (B) Western blot of total cell lysates from JM8.N4 WT mESCs, C59, and C59  $\Delta$ RBR<sub>i</sub>. WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF have the same number of amino acids, but  $\Delta$ RBR<sub>i</sub>-CTCF runs slightly slower in BisTris SDS-PAGE.  
 (C) Flow cytometry measurement of Halo-CTCF abundance in live C59 Halo-WT CTCF and C59  $\Delta$ RBR<sub>i</sub> mESCs after TMR labeling.  
 (D) Growth assay for C59 Halo-WT CTCF and C59  $\Delta$ RBR<sub>i</sub> mESCs. In (C) and (D), error bars indicate mean and SE (n = 4).  
 (E) *In vitro* RNA-binding assay. An *in vitro*-transcribed fragment of human *WRAP53* mRNA (*hWRAP53*, nucleotides 1–167) was incubated with recombinant (r-) WT- or  $\Delta$ RBR<sub>i</sub>-CTCF protein (see STAR Methods). Recovered RNA was run on urea denaturing gels and stained with SYBR Gold; recovered proteins were run on SDS-PAGE and stained with PageBlue. Left: representative experiment (replicates in Figure S1J). Right: RNA binding efficiency of WT- versus  $\Delta$ RBR<sub>i</sub>-CTCF averaged across three experiments, normalized by recovered proteins.  
 (F) PAR-CLIP of WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. Left: western blot of input and CTCF-IP. Right: autoradiography for <sup>32</sup>P-labeled RNA for input and CTCF-IP.  
 (G and H) Representative PALM reconstructions for Halo-WT CTCF (G) and Halo- $\Delta$ RBR<sub>i</sub> CTCF (H).  
 (I) Ripley's *L* function for WT-CTCF (52 cells) and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (46 cells) (mean and SE).  
 (J) Representative confocal micrographs of mESC colonies. Halo-WT CTCF and Halo- $\Delta$ RBR<sub>i</sub> CTCF mESCs were labeled with 500 nM Halo-TMR dye and visualized using a Zeiss LSM 710 laser scanning confocal microscope.  
 See also Figure S1.

(RBR) spanning ZFs 10 and 11 and the entire C terminus, and within this region identified 38 amino acids C-terminal to CTCF's ZF 11 that are necessary for RNA binding and for CTCF multimerization *in vitro* (Figure 2A). We refer henceforth to this required internal region in the RBR as the RBR<sub>i</sub>. We therefore asked whether CTCF clustering in cells is also RBR<sub>i</sub> dependent. The RBR<sub>i</sub> largely corresponds to mouse CTCF exon 10, which we endogenously and homozygously replaced with a 3xHA tag in C59 Halo-CTCF mESCs (Hansen et al.,

2017) to generate clone C59D2  $\Delta$ RBR<sub>i</sub> (Halo- $\Delta$ RBR<sub>i</sub>-CTCF = Halo-CTCF <sub>$\Delta$ 576–611</sub>); Figures 2A, 2B, and S1F).  $\Delta$ RBR<sub>i</sub>-CTCF mESCs express a full-length CTCF in which most of the RBR<sub>i</sub> (36 amino acids: N576–D611) have been substituted with a short linker (GDGAGLINS) followed by a 3xHA tag, preserving the original exon 10 structure and length. Interestingly, while Halo- $\Delta$ RBR<sub>i</sub>-CTCF protein levels are only mildly reduced compared with Halo-WT-CTCF, as measured by flow cytometry in live cells (Figures 2C and S1G),  $\Delta$ RBR<sub>i</sub>-CTCF mESCs



showed a ~2-fold growth defect, suggesting that the RBR<sub>i</sub> plays an important physiological role (Figure 2D).

First, we sought to confirm if the RBR<sub>i</sub> is required for RNA binding. Because CTCF was previously shown to bind the anti-sense transcript of human p53, *hWRAP53* RNA (Saldaña-Meyer et al., 2014), we purified recombinant WT-CTCF (r-WT-CTCF) or  $\Delta$ RBR<sub>i</sub>-CTCF (r- $\Delta$ RBR<sub>i</sub>-CTCF) from insect cells (Figure S1I) and tested binding to *hWRAP53* RNA *in vitro*. We observed ~3-fold reduction in *hWRAP53* RNA for  $\Delta$ RBR<sub>i</sub>-CTCF compared with WT-CTCF *in vitro* (Figure 2E; additional replicates in Figure S1J). Thus, the RBR<sub>i</sub> mediates RNA binding but is not absolutely required for it. Next, we tested if the RBR<sub>i</sub> also mediates RNA binding in cells using photo-activatable ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP) (Hafner et al., 2010).  $\Delta$ RBR<sub>i</sub>-CTCF mESCs showed substantially lower RNA binding, using <sup>32</sup>P-radiolabeled RNA as the readout, compared with WT-CTCF mESCs (Figure 2F). Consistent with our *in vitro* experiments (Figure 2E),  $\Delta$ RBR<sub>i</sub>-CTCF mESCs showed reduced, but not abolished, RNA binding. Taken together, we conclude that CTCF directly interacts with RNA and that the RBR<sub>i</sub> significantly contributes to RNA binding by CTCF but that some RNA binding remains after RBR<sub>i</sub> loss. This is consistent with CTCF bearing multiple, perhaps partially redundant, RNA-binding regions (Saldaña-Meyer et al., 2019, in this issue of *Molecular Cell*).

To test if the RBR<sub>i</sub> also mediates CTCF clustering, we performed super-resolution photoactivated localization microscopy (PALM) imaging in fixed mESCs. We labeled Halo-CTCF with the PA-JF<sub>549</sub> dye (Grimm et al., 2016), localized individual CTCF molecules inside the nucleus with a precision of ~13 nm (Figure S1H), and reconstructed CTCF nuclear organization. Indeed, WT-CTCF (Figure 2G) showed noticeably higher clustering than  $\Delta$ RBR<sub>i</sub>-CTCF (Figure 2H), which we further verified and quantified using Ripley's *L* function (Besag, 1977; Boehning et al., 2018; Ripley, 1976) (Figure 2I;  $L[r]-r$  values above 0 indicate clustering). We note that Ripley's *L* function is normalized by abundance such that lower clustering for  $\Delta$ RBR<sub>i</sub>-CTCF is not due simply to lower protein levels. These results suggest that CTCF largely self-associates in an RBR<sub>i</sub>-dependent manner and that CTCF clustering is significantly reduced, though not entirely abolished, in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs.

Because our RNA-binding experiments suggest that CTCF directly interacts with at least some RNA(s) (Figures 2E and 2F), it is tempting to speculate that RNA(s) directly bind CTCF and hold together CTCF clusters *in vivo*. However, our PALM and colP experiments cannot distinguish between a mechanism in which several CTCF proteins directly bind RNA from a model in which CTCF indirectly interacts with an unknown factor, which then mediates CTCF self-association in an RNase-sensitive manner. We also note that the RBR<sub>i</sub> region has been reported to be regulated by CK2-mediated phosphorylation (El-Kady and Klenova, 2005; Klenova et al., 2001). Although the RBR<sub>i</sub> contains a putative nuclear localization signal (NLS),  $\Delta$ RBR<sub>i</sub>-CTCF is still nuclear (Figures 2G, 2H, and 2J), consistent with prior work showing that nuclear localization and DNA binding are unaffected upon mutating (Klenova et al., 2001) or deleting (Saldaña-Meyer et al., 2014) the RBR<sub>i</sub>. Finally, although CTCF is clearly not generally misfolded in our  $\Delta$ RBR<sub>i</sub>-CTCF mESCs,

we cannot exclude slight effects on adjacent protein regions (e.g., ZF10–11 and the C-terminal regions), which could also contribute to the effects observed here.

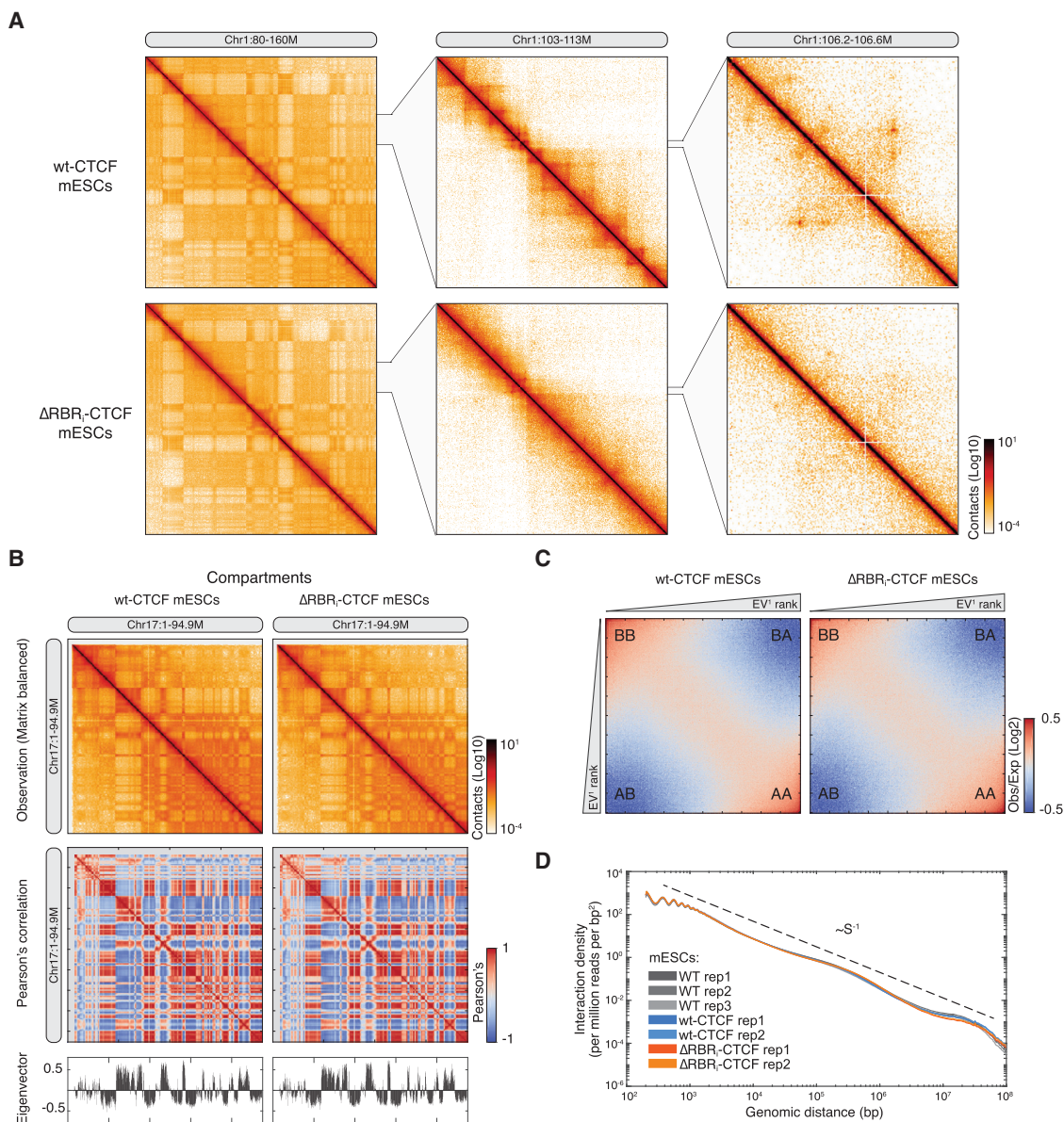
### The CTCF RBR<sub>i</sub> Regulates 3D Genome Organization, but Not Compartments

CTCF plays a major role in regulating 3D genome organization. We therefore next investigated whether impaired CTCF clustering (Figures 2G–2I), self-association (Figures 1C and 1D), RNA interaction (Figures 2E and 2F), and target searching of  $\Delta$ RBR<sub>i</sub>-CTCF (Hansen et al., 2018b) might also affect 3D genome organization, using a high-resolution genome-wide chromosomal conformation capture (3C) assay, Micro-C. Unlike Hi-C, which uses restriction enzymes, Micro-C fragments chromatin to single nucleosomes using micrococcal nuclease and generates 3D contact maps of the genome at all biologically relevant resolutions (Hsieh et al., 2015, 2016). Originally developed for analyzing the small yeast genome, here we have adapted a Micro-C protocol for large-genome organisms. Micro-C successfully recapitulates all the 3D genome features previously identified by Hi-C (Figures S2 and S3; see Data S1 for the protocol). We applied this Micro-C protocol to C59 (WT-CTCF) and C59D2 ( $\Delta$ RBR<sub>i</sub>-CTCF) mESCs (Figure 2A) over three replicates and generated ~668 million and ~694 million unique contacts, respectively. To test Micro-C, we assayed both reproducibility and consistency. Our Micro-C contact maps were highly reproducible between replicates (Figure S2), and the contact maps in WT-CTCF mESCs were consistent with Hi-C maps in mESCs, though notably, Micro-C reached “loop resolution” at substantially lower sequencing depth (Figure S3A). We also performed CTCF and cohesin (Smc1a) chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) in two replicates for WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (see below). We then surveyed 3D genome organization and analyzed features across several scales (Figure 3A) including compartments, TADs, loops, and stripes (Fudenberg et al., 2017), and began our analysis at the large end of the scale: compartments.

Mammalian chromosomes can be divided into two major compartments (Lieberman-aiden et al., 2009): A compartments, composed mainly of active euchromatin, and B compartments, composed mainly of inactive and gene-poor heterochromatin and lamina-associated domains (van Steensel and Belmont, 2017). We observed no significant change in compartmentalization when comparing WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figure 3B), nor did we observe significant changes in A-A, A-B, B-A, or B-B contact frequency (Figure 3C). Moreover, averaged over the whole genome, we observed the same contact probability scaling with genomic distance for WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figure 3D). We conclude that the CTCF RBR<sub>i</sub> does not affect the global distribution of active and inactive chromatin, consistent with compartments being largely unaltered after near-complete CTCF degradation (Nora et al., 2017; Wutz et al., 2017).

### Loss of CTCF RBR<sub>i</sub> Disrupts a Subset of TADs

Having analyzed compartments, we next zoomed in and analyzed TADs. TADs are demarcated by a pair of strong boundaries, or insulators, which are frequently bound by the



**Figure 3. Compartments Are Largely Unchanged in  $\Delta$ RBR<sub>1</sub>-CTCF mESCs**

(A) Overview of Micro-C contact matrices or maps at multiple resolutions in WT-CTCF and  $\Delta$ RBR<sub>1</sub>-CTCF mESCs. Contact matrix normalization: iterative correction and eigenvector decomposition (ICE); color scale:  $\log_{10}$ .

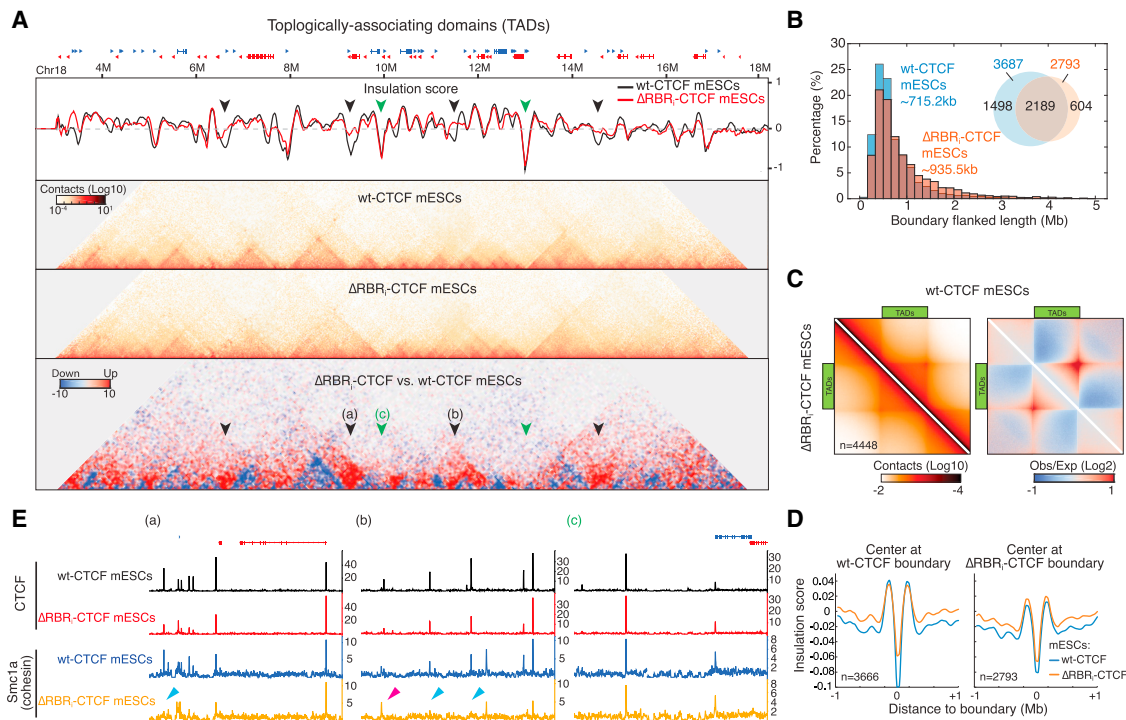
(B) Comparison of chromosome compartments. An example of plaid-like chromosome compartments at Chr17 is shown as ICE balanced contact maps, Pearson's correlation matrices, and eigenvector analysis for the first principle component at 100 kb resolution, showing no significant difference.

(C) Saddle plot for compartmentalization strength. Shows average distance-normalized contact frequencies between 100 kb bins in *cis* with ascending eigenvector values ( $\log_2$ ). Upper left and bottom right: contact frequency between B-B and A-A compartments. Upper right and bottom left: frequency of inter-compartment interactions.

(D) Genome-wide contact probability scaling plot, showing interaction density (per million reads per  $\text{bp}^2$ ) against genomic distance from 100 bp to 100 Mb. Biological replicates of WT-CTCF and  $\Delta$ RBR<sub>1</sub>-CTCF mESCs overlap and decay at slope of  $-1$ , as in Lieberman-aiden et al. (2009). Because of potential artifacts introduced by fragment self-ligation, we did not consider reads below 100 bp.

architectural proteins CTCF and cohesin and typically span lengths of  $\sim 100$  kb to  $\sim 1$  Mb in mouse and human genomes (Merkenschlager and Nora, 2016; Rowley and Corces, 2018). TADs are characterized by the feature that two loci inside the same TAD contact each other more frequently than two equidis-

tant loci in different TADs (Dixon et al., 2012; Nora et al., 2012). We defined TADs using either arrowhead or insulation score (Crane et al., 2015; Rao et al., 2014) and arbitrarily chose a cut-off value to obtain  $\sim 3,500$  TADs in WT-CTCF mESCs, corresponding to the previously reported TAD size and number



**Figure 4. TAD Organization Is Changed in  $\Delta RBR_i$ -CTCF mESCs**

(A) Example of TAD boundary disruption in  $\Delta RBR_i$ -CTCF mESCs. Snapshot of insulation score curves, 45°-rotated contact maps, and differential contact matrix (from top to bottom) plotted for Chr18: 3M–18M. Insulation scores were calculated using a 200 kb sliding window at 20 kb resolution. A lower value of insulation score means stronger insulation strength. Black arrows: examples of loss of insulation in  $\Delta RBR_i$ -CTCF mESCs. Green arrows: unaffected insulators. Differential contact matrices were generated by subtraction of the normalized  $\Delta RBR_i$ -CTCF matrix to WT-CTCF matrix. Red indicates “bleed-through,” that is, loss of TAD insulation (black arrows). Blue indicates weaker TADs in  $\Delta RBR_i$ -CTCF mESCs.

(B) Size distribution of TADs (boundary or insulator-flanked regions). Inset: Venn diagram.  $\Delta RBR_i$ -CTCF mESCs lose 1,474 of 3,666 insulators identified in WT-CTCF mESCs.

(C) Aggregate peak analysis for TADs. TADs in WT-CTCF mESCs were identified through an additional TAD calling algorithm (arrowhead) and rescaled and aggregated ( $n = 4,448$ ) at the center of the plot with ICE normalization (left) or distance normalization (right). WT-CTCF is shown on the top half and  $\Delta RBR_i$ -CTCF is shown on the bottom half.

(D) Genome-wide averaged insulation plotted versus distance around insulation center. Insulation strength is weaker in  $\Delta RBR_i$ -CTCF mESCs when centering at WT insulators, but there is no significant change when centering at  $\Delta RBR_i$ -CTCF insulators.

(E) Browser tracks. Snapshot regions (~200 kb) around the arrows (a, b, and c) indicated in (A) are shown with CTCF and cohesin (Smc1a) ChIP-seq data. (a) and (b) display regions with strong depletion of insulation in  $\Delta RBR_i$ -CTCF mESCs, and (c) shows a region with little effect. The blue arrows indicate examples for loss of Smc1a peaks in  $\Delta RBR_i$ -CTCF mESCs, and the pink arrow indicates an example for gain or shift of Smc1a peak.

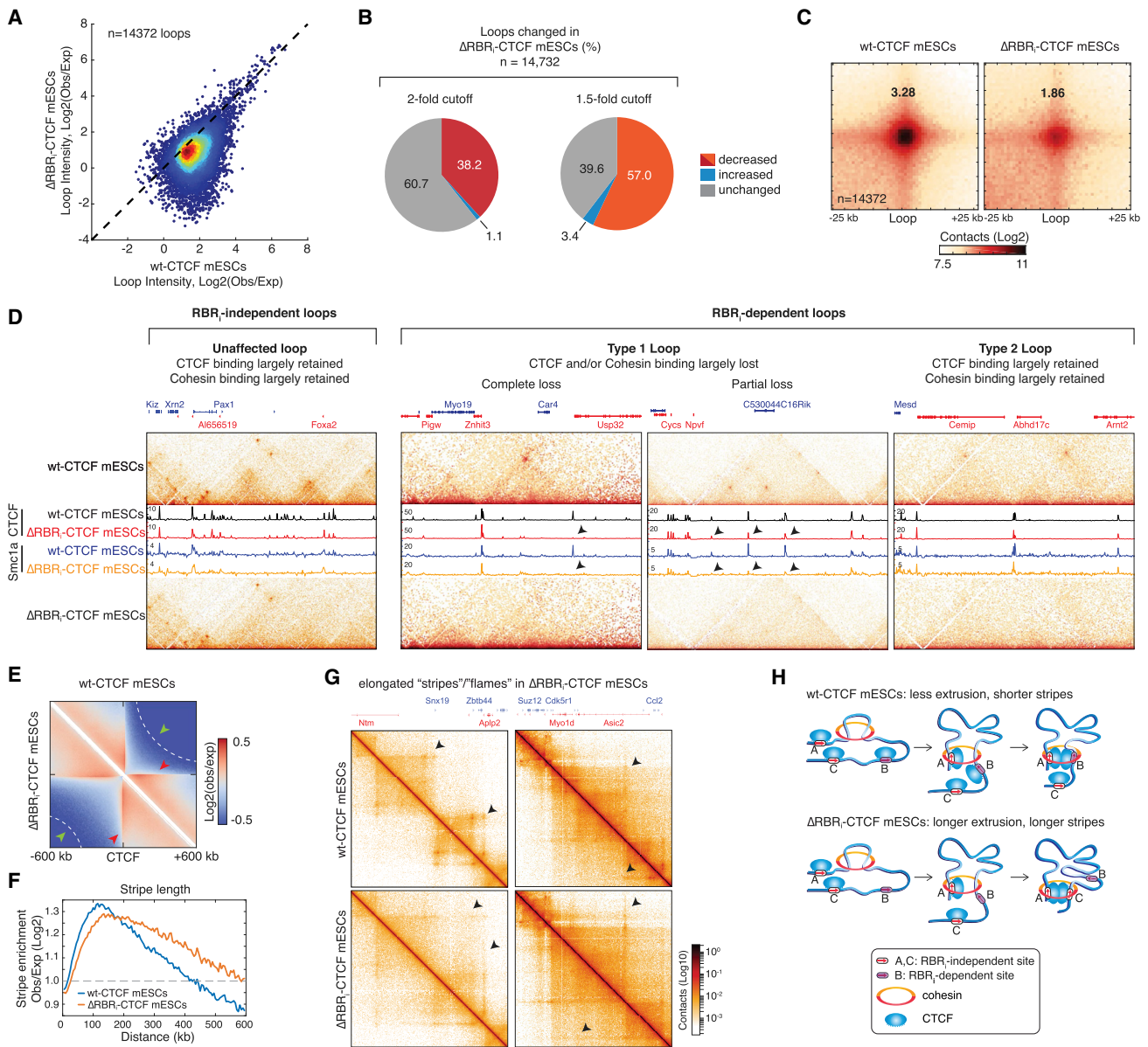
See also [Figures S2–S6](#).

(Forcato et al., 2017). Although the inferred number and size of TADs depends on the algorithm and the resolution of the maps (Forcato et al., 2017), we generally observed fewer and larger TADs in  $\Delta RBR_i$ -CTCF mESCs (Figures 4A and S4A). In brief, our insulation analysis called 3,666 and 2,793 TADs with average TAD sizes of ~715 kb and ~936 kb in WT-CTCF and  $\Delta RBR_i$ -CTCF mESCs, respectively (Figure 4B). We next aggregated over all TADs genome-wide (Figure 4C) and found TADs to be somewhat weaker in  $\Delta RBR_i$ -CTCF mESCs and characterized by weaker insulation strength (Figures 4D and S4B–S4E).

We next inspected local regions that were altered in  $\Delta RBR_i$ -CTCF mESCs, superimposing Micro-C and ChIP-seq results. Of note, when using spike-in normalization for ChIP-seq analysis, the  $\Delta RBR_i$ -CTCF signal appeared globally reduced compared to WT-CTCF, while Smc1a binding was largely unaltered at preserved sites (~60% of WT Smc1a binding sites;

Figures S5B and S5C). Because biochemical experiments showed reduced stability of the  $\Delta RBR_i$ -CTCF protein after cell lysis (Figure 6A), we could not determine whether the dampened ChIP-seq signal resulted from reduced ChIP efficiency, diminished genomic occupancy of  $\Delta RBR_i$ -CTCF, or both. We thus decided to normalize data by sequencing depth instead and avoid direct comparisons between WT-CTCF and  $\Delta RBR_i$ -CTCF ChIP-seq signals to draw conclusions. When inspecting local genomic regions, we noticed that CTCF and cohesin (Smc1a) binding was strongly depleted at some specific loci at  $\Delta RBR_i$ -affected boundaries (Figure 4E, blue arrows in a and b). Conversely, CTCF and cohesin binding was largely retained at unaffected boundaries (Figure 4E, browser track c). We conclude that the RBR<sub>i</sub> contributes to CTCF’s role in forming TADs. This is unlikely to be an indirect effect, because (1) the cell cycle phase distribution was identical between WT-CTCF





**Figure 5. Genome Organization at the Level of Both Loops and Stripes Is Altered in  $\Delta$ RBR<sub>1</sub>-CTCF mESCs**

(A) Scatterplot showing individual loop intensities in WT-CTCF versus  $\Delta$ RBR<sub>1</sub>-CTCF mESCs. A total of 14,372 loops were identified in WT-CTCF mESCs, with a false discovery rate < 0.1. Loop intensity was calculated as log<sub>2</sub> enrichment of center pixel over expected bottom left pixels at 1, 5, or 10 kb resolution.

(B) Pie charts showing affected loops. Approximately 8,189 loops are decreased by at least 1.5-fold, and 5,490 loops are decreased by at least 2-fold in  $\Delta$ RBR<sub>1</sub>-CTCF compared with WT-CTCF mESCs.

(C) Aggregate peak analysis for loops. The called loops were aggregated at the center of a 50 kb window at 1 kb resolution. The genome-wide averaged loop enrichment was calculated by the fold enrichment (center pixel/expected bottom left pixels).

(D) Snapshots of four representative genomic regions of different CTCF loop types. Zoomed-in contact maps were plotted on the top and bottom panels for WT-CTCF and  $\Delta$ RBR<sub>1</sub>-CTCF mESCs, respectively. CTCF and cohesin (Smc1a) ChIP-seq data are overlaid. From left to right, examples are shown of RBR<sub>1</sub>-independent loops and of the two sub-types of RBR<sub>1</sub>-dependent loops (with two examples of partial and complete loss of CTCF and cohesin binding for loop type 1).

(E) Aggregation plot centered at top CTCF peaks. The contact matrices were aggregated around the top 10,000 CTCF ChIP-seq peaks using a  $\pm$ 600 kb window. WT-CTCF mESCs are shown on the top half of the plot, and  $\Delta$ RBR<sub>1</sub>-CTCF mESCs are shown on the bottom half. Red arrows indicate stripes or flames. Green arrows and white dashed lines indicate insulation strength.

(legend continued on next page)



and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs, despite the growth defect of the latter (Figures S4F and S4G); (2) although the  $\Delta$ RBR<sub>i</sub>-CTCF expression level was somewhat lower (reduced by 28%) compared with WT-CTCF (Figures 1C and S1G), Nora et al. (2017) previously demonstrated that TAD organization in mESCs is preserved for the most part even after 85% reduction of CTCF levels; and (3) fluorescence recovery after photobleaching (FRAP) experiments show that the residence time for binding to cognate sites is approximately the same for WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF (Hansen et al., 2018b).

### CTCF Loops Fall into RBR<sub>i</sub>-Dependent and -Independent Sub-classes, and Loss of the CTCF RBR<sub>i</sub> Causes Longer Stripes

Many TADs show corner peaks of “C” signal at their summit, suggesting that they are held together as loop structures (Fudenberg et al., 2017; Rao et al., 2014) (see also Figures 3A and 4C). These loops are thought to be formed when pairs of chromatin-bound CTCF proteins block a loop-extruding cohesin (Fudenberg et al., 2017), yet the protein domain(s) in CTCF required for this are unknown. To test whether the RBR<sub>i</sub> plays any role in loop formation and/or maintenance, we analyzed the contact maps at high resolution (~1–5 kb) and identified ~14,372 loops in WT-CTCF mESCs using the method described by Rao et al. (2014). Overall, out of 14,372 called loops, 57% (8,189 loops) were weakened by at least 1.5-fold in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs and 39% (5,490 loops) by at least 2-fold relative to wild type (WT; Figures 5A and 5B), and loop strength was reproducible between replicates (Figure S4H). We next performed genome-wide loop aggregation analysis. The loop strength in C59 WT-Halo-CTCF mESCs is about as strong as in mESCs with untagged CTCF (Bonev et al., 2017) (Figure S4I), confirming that our endogenously tagged Halo-CTCF mESCs behave as WT mESCs (Hansen et al., 2017). However, the loop strength was greatly reduced in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figures 5C and S4I). As a comparison, we re-analyzed Hi-C data at loops in mESCs with a CTCF degradation tag from Nora et al. (2017) and found that the loss in loop strength upon near complete CTCF degradation is actually comparable with the defect in loop strength we observe for  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figures S4I and S4K). Although technical differences between Micro-C and Hi-C make a direct comparison difficult, these results nevertheless emphasize the loop strength defect in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs.

Surprisingly, the effect of deleting the RBR<sub>i</sub> was highly heterogeneous: some CTCF loops were unaffected or even strengthened, whereas others were significantly weakened or completely disrupted in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figure 5D). Qualitatively, we could distinguish two general categories of loops: an RBR<sub>i</sub>-independent class (Figure 5D, left) and an RBR<sub>i</sub>-dependent class (Figure 5D, right). When we overlaid the ChIP-seq tracks on the

Micro-C contact maps, we noticed that CTCF and cohesin (Smc1a) binding was largely preserved at the anchors of RBR<sub>i</sub>-independent loops, as expected. However, we could distinguish at least two sub-types of loops that were lost in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs: (1) partial or complete loss of  $\Delta$ RBR<sub>i</sub>-CTCF and/or cohesin binding at least at one loop anchor (Figure 5D, type 1 loops) and (2) no significant change in either  $\Delta$ RBR<sub>i</sub>-CTCF or cohesin binding (Figure 5D, type 2 loops). Thus, whereas loop loss for type 1 loops can be explained through loss of CTCF and/or cohesin binding, differential changes in CTCF and cohesin binding cannot readily explain loss of type 2 loops. We discuss the mechanistic implications of these findings in greater detail below.

Finally, we analyzed stripes or flames (Fudenberg et al., 2017). We compiled contact matrices using the top 10,000 WT-CTCF ChIP signals at the center of the plot and found that stripes in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs are less intense at shorter distances (<200 kb from the CTCF peaks) but continue for ~200 kb longer than in WT-CTCF cells (Figures 5E and 5F; red arrow; examples in Figure 5G). Although the mechanistic basis of stripes remains unclear, the loop extrusion model posits that they are formed by cohesin-mediated extrusion ((Fudenberg et al., 2017); Figure 5H). We speculate that longer stripes in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs could be due to ~200 kb larger TADs in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figures 4B and S4D). If cohesin has to extrude longer, on average, to reach a functional CTCF site in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs, this might result in longer stripes, as outlined in Figure 5E. In summary, our Micro-C analysis reveals that the CTCF RBR<sub>i</sub> domain regulates genome organization at the level of TADs, loops, and stripes in mESCs, without affecting A and B compartments.

### Loss of the CTCF RBR<sub>i</sub> Reveals Distinct Sub-classes of TADs and Loops

We next asked why some CTCF boundaries depend on the RBR<sub>i</sub> but others do not (Figure 5D). First, we tested whether the RBR<sub>i</sub> is required for CTCF interaction with cohesin using coIPs. Both WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF immunoprecipitation pulled down cohesin (subunits Rad21 and Smc1a in Figures 6A and S5A). This is especially notable because the protein stability of  $\Delta$ RBR<sub>i</sub>-CTCF during the IP procedure was significantly reduced (compare CTCF inputs in Figures 6A and S5A). Thus, CTCF interacts with cohesin in an RBR<sub>i</sub>-independent manner, implying that loop loss is not due simply to a failure of  $\Delta$ RBR<sub>i</sub>-CTCF to interact with cohesin.

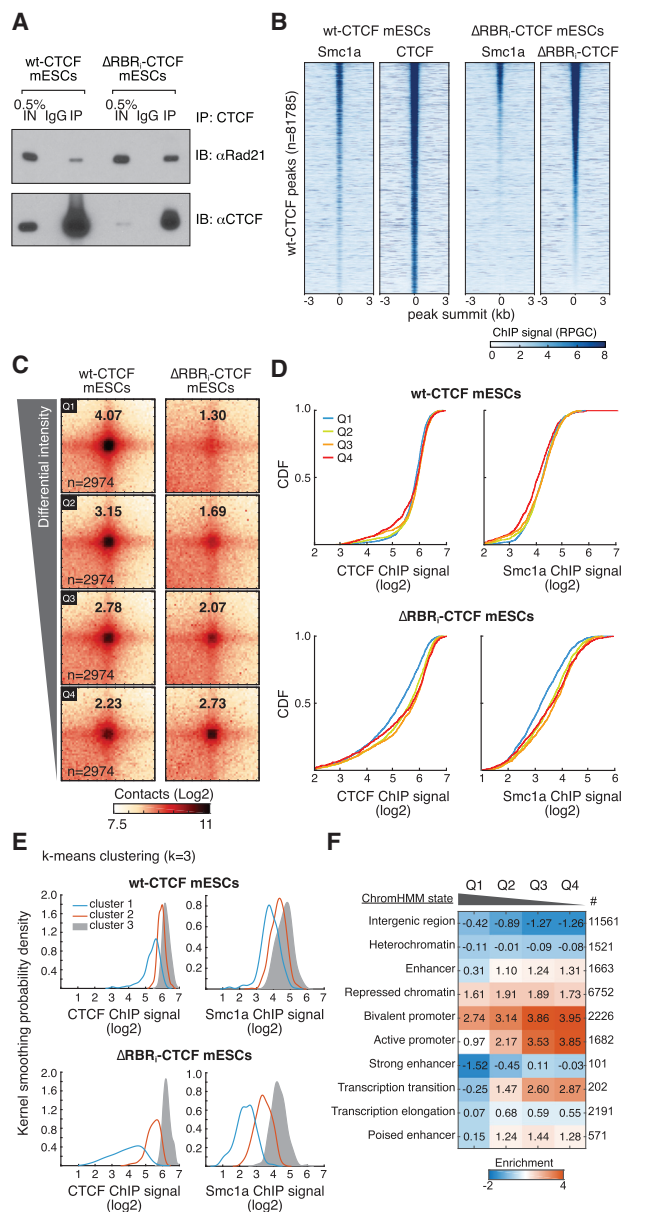
Next, we analyzed our CTCF and cohesin (Smc1a) ChIP-seq data for WT-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs in more detail (Figures 6B, S5B, and S5C). Our ChIP-seq data were both reproducible between replicates and consistent with other studies in mESCs (Figure S6). Consistent with FRAP experiments, which showed no detectable change in residence time

(F) Quantification of stripe length. Stripes enrichments were calculated in log<sub>2</sub> ratio of observed over expected contacts. Significant enrichment was defined as 2-fold enrichment labeled shown as a gray dashed line in the plot.

(G) Representative contact maps at specific regions showing elongated “stripes” or “flames” in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs.

(H) Loop extrusion sketch. Speculative illustration of why loss of a subset of CTCF boundaries might result in longer stripes assuming loop extrusion (Fudenberg et al., 2017).

See also Figures S4 and S6.



**Figure 6. ΔRBR<sub>1</sub>-CTCF Still Interacts with Cohesin, and Loops Lost in ΔRBR<sub>1</sub>-CTCF mESCs Have Less CTCF and Cohesin Bound**

(A) CoIP experiment showing that ΔRBR<sub>1</sub>-CTCF stills interacts with cohesin. CTCF antibodies can pull down Rad21 cohesin subunit in both WT- and ΔRBR<sub>1</sub>-CTCF mESCs.

(B) Heatmaps of CTCF and Smc1a ChIP-Seq signal (deepTools RPGC [reads per genomic content]) around WT-CTCF peaks as called by MACS2, sorted by ΔRBR<sub>1</sub>-CTCF peak intensity.

(C) Aggregate peak analysis for differential loop intensity. Loops were sorted into four quartiles on the basis of differential loop intensities between WT-CTCF and ΔRBR<sub>1</sub>-CTCF mESCs. A total of 2,974 loops in each quartile were aggregated at the center of a 50 kb window and quantified as above.

(D) Cumulative distribution function (CDF) curves of ChIP enrichment at the loop anchors. CTCF and Smc1a ChIP signals were quantified as the log<sub>2</sub> enrichment ± 250 bp around the loop anchor.

(E) *k*-Means clustering analysis of CTCF and cohesin (Smc1a) ChIP-seq data in the Q1 loop anchors. The filtered Q1 loop anchor sites were analyzed using *k*-means clustering (*k* = 3). Clustering analysis output are plotted as kernel

at cognate binding sites for ΔRBR<sub>1</sub>-CTCF (Hansen et al., 2018b), ΔRBR<sub>1</sub>-CTCF still binds the majority of CTCF sites, although the number and occupancy levels were generally reduced (63% of 81,785 WT-CTCF ChIP-seq peaks maintained in ΔRBR<sub>1</sub>-CTCF mESCs, Figure S5C; spike-in normalized ChIP-seq in Figure S5B). Similarly, about 60% of the cohesin binding sites detected in WT-CTCF mESCs were also occupied in ΔRBR<sub>1</sub>-CTCF mESCs (Figure S5C).

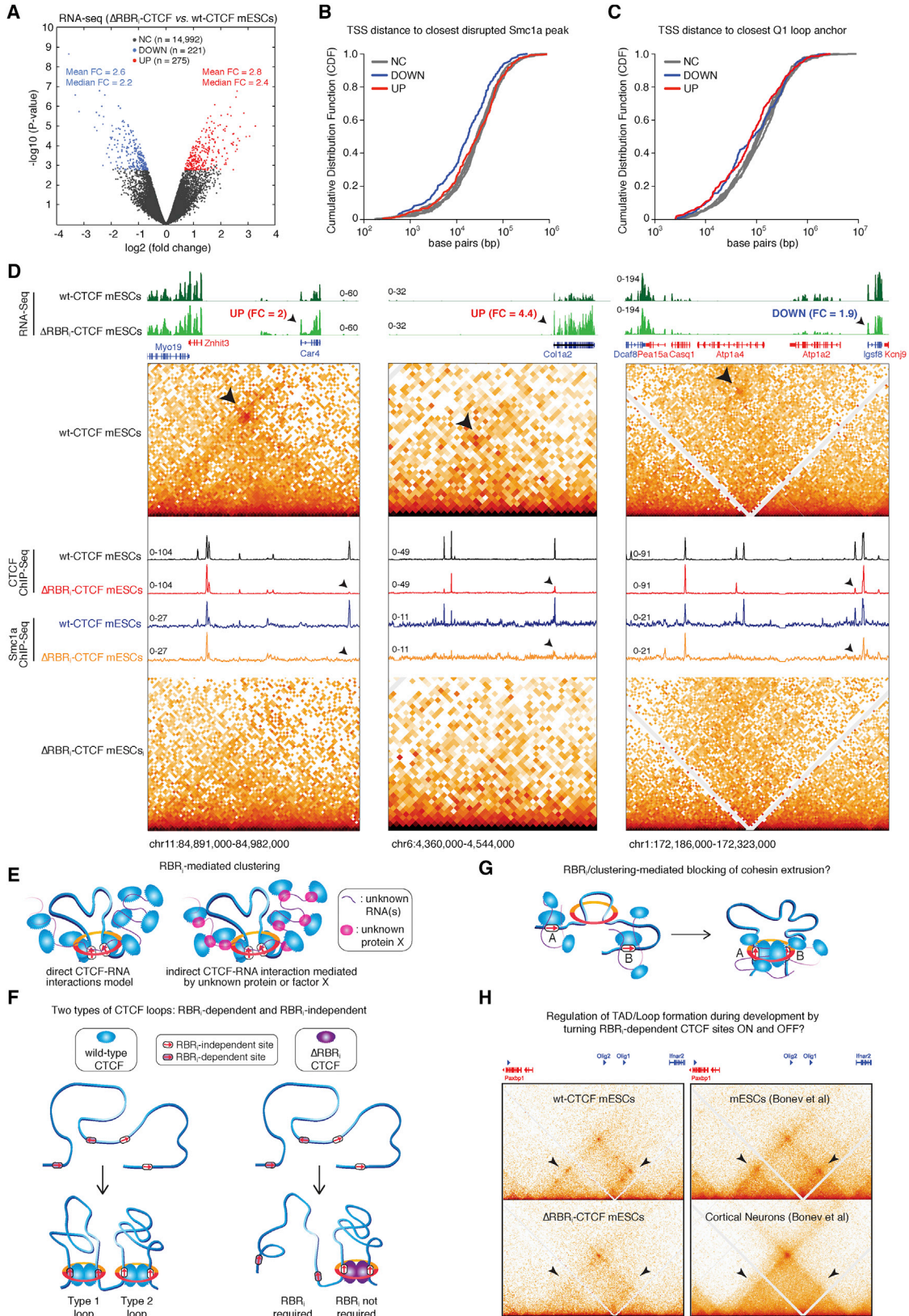
To further dissect the site-specific features from the genome-wide average, we divided loops into four quartiles (Figure 6C), such that Q1 contains loops that are largely lost in ΔRBR<sub>1</sub>-CTCF mESCs and Q4 contains loops that are largely unaffected or even strengthened in ΔRBR<sub>1</sub>-CTCF mESCs. We then characterized the CTCF and Smc1a binding profiles at both anchors of loops and only analyzed loops that satisfy three prerequisites: (1) CTCF shows ChIP-seq signal at both anchors in WT cells, (2) cohesin (Smc1a) shows ChIP-seq signal at both anchors in WT cells, and (3) a pair of convergent CTCF cognate sites are present at both anchors. We then analyzed CTCF and cohesin (Smc1a) ChIP enrichment at the filtered loop anchors for each quartile (Figure 6D). Consistent with a key role for CTCF and cohesin, Q1 loops that were disrupted the most in ΔRBR<sub>1</sub>-CTCF mESCs also had the lowest CTCF and cohesin occupancy in ΔRBR<sub>1</sub>-CTCF mESCs (see also histograms in Figure S5D), while they were just as strongly, if not more, occupied as Q2–Q4 loops in WT-CTCF mESCs.

Our qualitative analysis in Figure 5D suggested that RBR<sub>1</sub>-dependent loops can be subdivided into two types depending on their CTCF and cohesin dependence. If this interpretation is correct and robust, we should be able to recover these types naturally after applying an unsupervised clustering algorithm. To test this, we applied *k*-means clustering (using *k* = 3) on the most affected loops (Q1) and recovered three loop clusters, similar to Figure 5D (Figures 6E and S5E). Cluster 1 and 2 loops (76%) are lost because of partial and near complete loss of CTCF and cohesin binding, respectively (type 1 in Figure 5D); cluster 3 loops (24%) are affected loops without strong CTCF and cohesin loss (type 2 in Figure 5D). Thus, this analysis confirms our qualitative assessment in Figure 5D.

Could the CTCF loop type be encoded in the DNA-binding sequence motif? We performed *de novo* motif discovery on the four loop quartiles and observed distinct CTCF binding sequence preferences and potential co-regulators (Figures S7A and S7B). We conclude that loops can be classified into two classes, RBR<sub>1</sub> dependent and RBR<sub>1</sub> independent, and that the RBR<sub>1</sub>-dependent class can be further sub-classified into two types with distinct CTCF and cohesin binding profiles, and that each class correlates with a distinct CTCF DNA-binding motif preference.

smoothed histograms. Heatmaps with peaks at the center across a ±3 kb region are shown in Figure S5E.

(F) Enrichments of genomic features at loop anchors by ChromHMM analysis (Bogu et al., 2015). Heatmap shows log<sub>2</sub> enrichment of the loop anchors in each chromatin state. Q1 loops are largely depleted in most chromatin states and only slightly enriched in H3K27me3 chromatin. See also Figures S5–S7.



(legend on next page)



Finally, we asked which other genomic features correlate with RBR<sub>i</sub>-dependent versus RBR<sub>i</sub>-independent loops. We performed an extensive bioinformatics comparison using 70 previously published datasets in mESCs (Figure S7C). Notably, Q4 loops that were not disrupted in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs correlated with transcriptionally active genomic regions (enhancers, promoters; Figure 6F) and were more frequently found in the A compartment (Figure S7D), which is generally associated with active genes. In contrast, Q1 loops were relatively larger and more enriched in the B compartment, which is generally associated with transcriptional repression. These results, albeit inherently correlative, argue against a “cis-model” in which nascent RNA transcripts stabilize CTCF boundaries in an RBR<sub>i</sub>-dependent manner. Instead, because active sites of transcription are enriched at TAD boundaries (Dixon et al., 2012; Merckenschlager and Nora, 2016), it seems plausible that active transcription may compensate for CTCF boundary weakening in Q4 loops through a CTCF-independent mechanism.

### Loss of CTCF RBR<sub>i</sub> Affects Gene Expression

To evaluate the functional impact of CTCF RBR<sub>i</sub> deletion on gene expression, we compared RNA-seq of total, ribo-depleted RNA extracted from  $\Delta$ RBR<sub>i</sub>-CTCF mESCs with that obtained from WT-CTCF mESCs (two replicates each). A stringent differential expression analysis between the two cell lines (edgeR false discovery rate  $\leq 0.05$  and DESeq2 adjusted p value  $\leq 0.05$ ; see STAR Methods) revealed 496 deregulated genes upon loss of CTCF RBR<sub>i</sub>, 275 being upregulated and 221 being downregulated, with a mean fold change of  $\sim 2.7$  (Figures 7A and S7; complete gene list in Table S2; Gene Ontology analysis in Table S3).

Do gene expression changes correlate with the partial loss of CTCF and cohesin binding and altered chromatin loops described above in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs? Indeed, genes that were downregulated in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs compared with WT-CTCF mESCs had a higher probability to lie nearby a disrupted Smc1a binding site than any random set of unaltered genes (Figures 7B and S7E). In contrast, upregulated genes were not detectably closer to disrupted Smc1a peaks (Figure 7B). The transcription start site (TSS) of downregulated genes was also significantly closer than that of upregulated genes to CTCF peaks disrupted in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figure S7G). Consistent with these observations, acute depletion of most CTCF protein revealed that early downregulated genes, but not upregulated genes, tended to be close to an affected CTCF site (Nora et al., 2017). Nevertheless, both downregulated and upregulated genes were located closer than the control unchanged gene sets to Q1 loop anchors, the most severely disrupted in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figures 7C and S7F; Q2–4 in Figures S7H–S7J). Inspecting single genomic loci, we found several examples of both upregulated and downregulated genes proximal to the anchors of loops disrupted in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (Figures 7D and S7L). Notably, several—and certainly more than expected by chance—of the deregulated genes in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs changed in the same direction as seen after acute CTCF depletion in mESCs (Nora et al., 2017) (Figure S7K; full overlap analysis in Table S2). Taken together, these results show that the CTCF RBR<sub>i</sub> regulates both chromatin looping and gene expression.

### DISCUSSION

In this study, we have identified unexpected roles for an internal RNA-binding region (RBR<sub>i</sub>) in CTCF. We confirmed that CTCF

#### Figure 7. Altered Gene Expression in $\Delta$ RBR<sub>i</sub>-CTCF mESCs and Speculative Models for the Role of CTCF's RBR<sub>i</sub>

(A) Volcano plot comparing gene expression in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs against WT-CTCF mESCs, measured by RNA-seq followed by differential expression analysis with edgeR and DeSeq2. Plotted are edgeR p values and fold changes. Gray dots, genes not changed (NC) upon CTCF RBR<sub>i</sub> deletion; blue dots, genes called downregulated (DOWN) by both edgeR and DeSeq2 in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs; red dots, genes called upregulated (UP) by both edgeR and DeSeq2 in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (see STAR Methods). edgeR fold change (FC) mean and median values are specified for both downregulated (blue) and upregulated (red) genes. Full analysis in Table S2.

(B) For each differentially regulated gene (DOWN or UP) in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs, we measured the distance in base pairs from its transcription start site (TSS) to the closest disrupted Smc1a called ChIP-seq peak in WT mESCs (and plotted the results as a cumulative distribution function [CDF]). As controls, we randomly selected five groups of  $\sim 500$  unaltered genes each (not changed [NC]). Scatterplots with single data points in Figure S7.

(C) Same as (B), but plotting the distance to the closest Q1 loop anchor. Scatterplots with single data points in Figure S7.

(D) Snapshots of three genomic regions showing two genes (*Car4*, *Col2a1*) upregulated and one gene (*Igsf8*) downregulated in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs compared with WT-CTCF mESCs (more examples in Figure S7). RNA-seq tracks are plotted (top) and deregulated genes marked by black arrowhead. Fold change (FC) in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs versus WT-CTCF mESCs are also specified. Blue genes are transcribed from the “plus” strand, red genes from the “minus” strand. Zoomed-in contact maps at 1 kb (right, left) or 2 kb resolution (middle). Arrowheads highlight disrupted loops. CTCF and cohesin (Smc1a) ChIP-seq data are overlaid, with arrowheads pointing at disrupted right loop anchors in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. ChIP-seq and RNA-seq units: reads per genomic content (deepTools RPGC).

(E) Sketch of a CTCF cluster. We observe that CTCF self-association is sensitive to RNase *in vitro* and that CTCF clustering is partially mediated by its RBR<sub>i</sub> *in vivo*. As such, our results are consistent both with direct CTCF-RNA interactions (left) and indirect CTCF-RNA interactions, perhaps mediated by an unknown factor X.

(F) Two types of CTCF loops. Our analysis of  $\Delta$ RBR<sub>i</sub>-CTCF mESCs uncovers the existence of at least two classes of CTCF loops: RBR<sub>i</sub>-dependent and RBR<sub>i</sub>-independent loops.

(G) Does CTCF clustering help block extruding cohesin? Speculative model that clustering of an otherwise small CTCF protein may contribute to efficiently blocking extruding cohesins.

(H) Regulation of loops and TADs during differentiation. The ability to turn on and off RBR<sub>i</sub>-dependent CTCF boundaries could potentially provide the means for regulating specific TADs and loops during development by regulating RBR<sub>i</sub> interaction partners. As an illustration, we show a side-by-side comparison of 3D genome reorganization in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs and differentiated cells at the region around the *Olig1* and *Olig2* genes (Hi-C data from Bonev et al., 2017). Subdomains and loops (black arrows) are lost in both  $\Delta$ RBR<sub>i</sub> mESCs and cortical neurons.

See also Figures S6 and S7.



self-associates in a largely RNA-mediated manner (Saldaña-Meyer et al., 2014) (Figure 1C) and now demonstrate that the CTCF RBR<sub>i</sub> contributes to RNA binding, CTCF self-association, and clustering *in vivo* (Figure 7E). Moreover, we surprisingly find that almost half of all CTCF loops are lost in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs, suggesting that CTCF-mediated loops can be classified into at least two major classes (Figure 7F): RBR<sub>i</sub>-independent and RBR<sub>i</sub>-dependent CTCF loops. Intriguingly, this may provide a means for differentially engaging or disrupting specific CTCF loops during development and cellular differentiation (Bonev et al., 2017; Pękowska et al., 2018). We discuss some of the implications below.

### How Do CTCF and Cohesin Interact?

Despite their critical role in 3D genome organization, we know surprisingly little mechanistically about CTCF and cohesin. Although the related SMC-complex condensin has been observed to extrude loops *in vitro* (Ganji et al., 2018), *in vitro* single-molecule studies of cohesin failed to detect extrusion (Davidson et al., 2016; Kanke et al., 2016; Stigler et al., 2016). Moreover, whether a hypothetical cohesin-based extrusion complex would exist as a single ring or perhaps as a pair of rings remains unclear and a matter of active debate (Cattoglio et al., 2019; Kim et al., 2019; Nasmyth, 2011; Skibbens, 2016). Finally, how CTCF and cohesin interact *in vivo* remains to be elucidated. Xiao et al. (2011) reported that the 575–611 region in human CTCF interacts directly with the SA2 subunit of cohesin and that interaction with the other cohesin subunits is indirect. This region largely corresponds to the RBR<sub>i</sub> and is entirely deleted in our  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. Nevertheless, we observed robust colP of the cohesin subunits Rad21 and Smc1a with  $\Delta$ RBR<sub>i</sub>-CTCF (Figure 6A; Figure S5A). Similarly, colP between human  $\Delta$ RBR<sub>i</sub>-CTCF with the cohesin subunit SA1 was observed (Saldaña-Meyer et al., 2014). Therefore, both our new studies and that of Saldaña-Meyer et al. (2014) show that  $\Delta$ RBR<sub>i</sub>-CTCF can still interact with cohesin, which contradicts the findings of Xiao et al. (2011). We suggest that fully elucidating how CTCF and cohesin interact should be an important direction for future research.

### What Does the CTCF RBR<sub>i</sub> Bind?

We find that CTCF self-association is strongly reduced upon treatment with RNase A *in vitro* (Figure 1C) and that  $\Delta$ RBR<sub>i</sub>-CTCF shows substantially less clustering in cells (Figures 2G–2I). Consistently, the CTCF RBR<sub>i</sub> was reported on the basis of fractionation studies to be necessary for CTCF multimerization *in vitro* (Saldaña-Meyer et al., 2014). Saldaña-Meyer et al. (2014) also reported that CTCF directly binds the *hWRAP53* RNA and that ZF10–11 contributes to RNA binding. Here, we show that  $\Delta$ RBR<sub>i</sub>-CTCF shows substantially reduced, but not abolished, RNA binding *in vitro* (Figure 2E) and in cells (Figure 2F). After the present work appeared on bioRxiv, Saldaña-Meyer et al. (2019) further identified two additional RNA-binding regions in CTCF ZF1 and ZF10. Loss of ZF1 or ZF10 impairs RNA binding by CTCF as assayed using PAR-CLIP and causes deregulation of gene expression in mESCs (Saldaña-Meyer et al., 2019). Taken together with the results reported here, this suggests that CTCF interacts with RNA(s) through several protein regions,

including ZF1, ZF10, and the RBR<sub>i</sub>. However, although our results clearly show that the CTCF RBR<sub>i</sub> is required for about half of all chromatin loops and mediates CTCF clustering, we do not know the mechanism at this stage. Specifically, our results cannot distinguish a model in which RNA(s) directly bound by the CTCF RBR<sub>i</sub> regulates looping and clustering, from indirect models in which the CTCF RBR<sub>i</sub> binds another factor, which then indirectly contributes to CTCF self-association and clustering in an RNase-sensitive manner and to loop formation (Figure 7E). Moreover, we note that serine residues in the RBR<sub>i</sub> are differentially phosphorylated during stem cell differentiation (El-Kady and Klenova, 2005; Rigbolt et al., 2011).

Nevertheless, it is worth considering other CTCF-RNA interactions that have been reported beyond Wrap53. CTCF has been reported to directly bind the lincRNAs HOTTIP (Wang et al., 2018), CCAT1-L (Xiang et al., 2014), and Firre (Yang et al., 2015); the RNA Jpx has been reported to evict CTCF from the X chromosome (Sun et al., 2013); CTCF has been shown to bind RNAs specifically and with high affinity *in vitro* (Kung et al., 2015); and CTCF was also reported to bind the RNA helicase p68/DDX5 together with the noncoding RNA, SRA (Yao et al., 2010). Finally, CTCF was identified as an RNA-binding protein in three recent independent screens for RNA-binding proteins (Brannan et al., 2016; Caudron-Herger et al., 2019; He et al., 2016), and transcription elongation by RNA Pol II can displace both CTCF and cohesin from chromatin (Heinz et al., 2018). However, there are likely many more CTCF RBR<sub>i</sub> interaction partners, and identifying these will be an important but challenging future endeavor.

### There Are at Least Two Classes of CTCF Binding Sites and Chromatin Loops

The loop extrusion model can elegantly explain most experimental observations through a parsimonious mechanism (Fudenberg et al., 2017). In the model's simplest form, any correctly oriented chromatin-bound CTCF should block cohesin-mediated loop extrusion. Accordingly, all CTCF binding sites should form loops. However, only a minority of CTCF binding sites form loops visible in Hi-C contact maps (Merkenschlager and Nora, 2016; Rao et al., 2014). Why is that? At a minimum, this suggests that not all CTCF sites are equivalent and that only a subset of CTCF sites can stabilize loops. Accordingly, we show here that CTCF sites fall into at least two distinct classes: RBR<sub>i</sub>-dependent and RBR<sub>i</sub>-independent sites.

How is the RBR<sub>i</sub> dependence of a CTCF binding site determined? CTCF binds DNA through 11 ZFs, and which ZFs contribute to DNA binding is somewhat idiosyncratic and binding site dependent (Hashimoto et al., 2017; Nakahashi et al., 2013; Yin et al., 2017). Although the core CTCF DNA motif is bound by the central ZFs, only the upstream motif is bound by ZF9–11 (Nakahashi et al., 2013). Because the RBR<sub>i</sub> is just downstream of ZF9–11 (Figures 1A and 2A), it is tempting to speculate that depending on whether ZF9–11 are engaged in DNA binding, there could be allosteric control over which potential RBR<sub>i</sub> interaction partners would be engaged. Consistent with this interpretation, we observed distinct DNA motifs bound by RBR<sub>i</sub>-dependent and RBR<sub>i</sub>-independent CTCF loops (Figures S7A and S7B).

### Does CTCF Clustering Contribute to Halting Cohesin-Mediated Loop Extrusion?

Within the context of the loop extrusion model, it is unclear how a small  $\sim$ 3- to 5-nm-sized protein, CTCF, would efficiently block a large and rapidly extruding cohesin complex with a lumen of  $\sim$ 40–50 nm—and do so in an orientation-specific manner (Guo et al., 2015; Rao et al., 2014; Vietri Rudan et al., 2015; de Wit et al., 2015). We previously showed that CTCF forms clusters in mESCs and U2OS cells (Hansen et al., 2017), and Zirkel et al. (2018) reported that CTCF forms large foci in senescent cells. Here, we now show that CTCF clustering is partly mediated by the RBR<sub>i</sub> and, simultaneously, that the RBR<sub>i</sub> is required for a large subset of loops. It is thus tempting to speculate that cluster and loop formation are related: in particular, RBR<sub>i</sub>-mediated CTCF clustering could make CTCF a more efficient boundary to cohesin-mediated extrusion in at least two ways (Figure 7G): (1) a cluster containing several CTCF proteins, aided by binding to polymers such as RNA, should be much larger and thus more efficient at arresting cohesin than a single chromatin-bound CTCF protein, and (2) if CTCF binds cohesin through a specific protein region, having more CTCFs present would increase the probability of a correct encounter between this target interaction surface and cohesin.

### Loss of the CTCF RBR<sub>i</sub> Causes Dereglulation of Gene Expression

Here we demonstrate that loss of the CTCF RBR<sub>i</sub> causes deregulation of  $\sim$ 500 genes (Figure 7A) as well as loss of about half of all chromatin loops (Figure 5B). Similarly, disruption of two other RNA-binding regions in CTCF also causes deregulation of  $\sim$ 400–500 genes (Saldaña-Meyer et al., 2019), whereas CTCF depletion for 4 days causes deregulation of 4,996 genes (Nora et al., 2017). Compared with such auxin-induced depletion studies (Nora et al., 2017; Saldaña-Meyer et al., 2019; Wutz et al., 2017), one advantage of the endogenous deletion approach that we use here is that no residual WT-CTCF protein remains to confound interpretation. However, a disadvantage of our approach is that we cannot readily distinguish acute and direct effects of CTCF on transcription from indirect effects (e.g., deregulation of a gene by CTCF, which then causes indirect deregulation of other genes). Nevertheless, we do observe that deregulated genes tend to be closer to a disrupted loop compared with genes whose expression did not change (Figure 7C). This is consistent with chromatin looping directly contributing to the regulation of gene expression, although only for a subset of genes and only modestly (average fold change  $\sim$ 2.7). Taken together with (Nora et al., 2017; Saldaña-Meyer et al., 2019), our work emphasizes that CTCF is a significant regulator of transcription, although the fraction of genes whose expression is directly affected by CTCF and chromatin looping in a given cell type remains unclear.

### Regulation of CTCF Loops during Differentiation and Development

An enduring paradox has been the fact that CTCF and cohesin are present in all cell types. Thus, if they were the only factors forming loops and TADs, how can we explain the observation that some TADs and loops change during differentiation (Bonev et al., 2017; Pękowska et al., 2018)? Here we report that CTCF

loops can be divided into at least two classes: RBR<sub>i</sub> dependent and RBR<sub>i</sub> independent. Moreover, within the RBR<sub>i</sub> dependent CTCF loop class, we identify at least two types (Figures 5D and 6E). Having multiple types of CTCF boundaries provides potential mechanisms through which individual boundaries can be regulated. For example, if CTCF RBR<sub>i</sub>-dependent boundaries function in part by binding other proteins or RNAs, then regulating the abundance or function of these yet to be identified factors would provide a potential mechanism for distinct cell types to regulate specific boundaries and CTCF loops during development and differentiation (Figure 7H). Ultimately, this may enable cells to dissolve and form new CTCF-mediated chromatin loops during development and differentiation to regulate enhancer-promoter contacts and establish proper cell type-specific gene expression programs.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
- METHOD DETAILS
  - CRISPR/Cas9-mediated genome editing
  - Cell Cycle phase analysis
  - CTCF FACS abundance quantification
  - Growth Assay
  - PALM
  - PALM analysis
  - Western Blotting
  - Co-immunoprecipitation (CoIP) assays
  - CoIP DNA and RNA extraction and quantification
  - Chromatin immunoprecipitation (ChIP)
  - Expression and purification of recombinant wt-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF proteins
  - *In vitro* RNA binding assay
  - PAR-CLIP
  - ChIP-Seq library preparation
  - ChIP-Seq analysis
  - RNA-Seq library preparation and analysis
  - Micro-C
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY
  - Data and Code Availability Statement
  - Data availability
  - Computer code

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.molcel.2019.07.039>.

### ACKNOWLEDGMENTS

We thank Luke Lavis for generously providing JF dyes, Hervé Marie-Nelly for help with R code, Gina M. Dailey for assistance with cloning, Carla J. Inouye

for help with the biochemical assays, Assaf Amitai for insightful discussion, Astou Tangara and Ana Robles for microscope assembly and maintenance, Ji Yu and Jean-Yves Tinevez for coding discussions, Daniel J. Lee for help with genotyping, and Dr. Kartoosh Heydari at the Li Ka Shing Facility for flow cytometry assistance. We thank Elphege Nora, Thomas Graham, and other members of the Tjian and Darzacq labs for comments on the manuscript. A preprint describing this work first appeared on bioRxiv on December 13, 2018 under the title “An RNA-Binding Region Regulates CTCF Clustering and Chromatin Looping.” This work was performed in part at the CRL Molecular Imaging Center, supported by the Gordon and Betty Moore Foundation. This work used the Vincent J. Coates Genomics Sequencing Laboratory at the University of California (UC), Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. A.S.H. was a postdoctoral fellow of the Siebel Stem Cell Institute and is supported by a NIH National Institute of General Medical Sciences (NIGMS) K99 Pathway to Independence Award (K99GM130896). This work was supported by NIH 4D Nucleome Common Fund grants UO1-EB021236 and U54-DK107980 (X.D.), California Institute of Regenerative Medicine grant LA1-08013 (X.D.), and the Howard Hughes Medical Institute (grant CC34430 to R.T.).

### AUTHOR CONTRIBUTIONS

A.S.H., T.-H.S.H., C.C., X.D., and R.T. conceived the project. C.C., A.S.H., and I.P. performed genome editing and generated cell lines. A.S.H. characterized the C59D2 line (growth rate, cell cycle, and protein abundance). C.C. and I.P. performed and analyzed biochemistry experiments including western blots and coIPs. C.C. performed *in vitro* RNA-binding assays. R.S.-M. performed PAR-CLIP. D.R. supervised PAR-CLIP. A.S.H. performed and analyzed imaging experiments. T.-H.S.H. performed and analyzed Micro-C experiments. C.C. performed and analyzed ChIP-seq and RNA-seq experiments. T.-H.S.H. led and performed most bioinformatic analyses with input from A.S.H. and C.C. A.S.H., T.S.H., and C.C. drafted the manuscript, and all authors edited the manuscript. X.D. and R.T. supervised the project.

### DECLARATION OF INTERESTS

D.R. is a co-founder of Constellation Pharmaceuticals and Fulcrum Therapeutics. All other authors declare no competing interests.

Received: January 3, 2019

Revised: April 16, 2019

Accepted: July 29, 2019

Published: September 12, 2019

### REFERENCES

Abdennur, N., and Mirny, L. (2019). Cooler: scalable storage for Hi-C data and other genomically-labeled arrays. *Bioinformatics*. Published online July 10, 2019. <https://doi.org/10.1093/bioinformatics/btz540>.

Alipour, E., and Marko, J.F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* *40*, 11202–11212.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* *18*, 285–298.

Besag, J.E. (1977). Contribution to the discussion of Dr. Ripley’s paper. *J. R. Stat. Soc. B* *39*, 193–195.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol. Chapter 19*, 1–21.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kocik, G., Dailey, G.M., Cramer, P., et al. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* *25*, 833–840.

Bogu, G.K., Vizán, P., Stanton, L.W., Beato, M., Di Croce, L., and Marti-Renom, M.A. (2015). Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Mol. Cell. Biol.* *36*, 809–819.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., and Cavalli, G. (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell* *171*, 557–572.e24.

Brannan, K.W., Jin, W., Huelga, S.C., Banks, C.A.S., Gilmore, J.M., Florens, L., Washburn, M.P., Van Nostrand, E.L., Pratt, G.A., Schwinn, M.K., et al. (2016). SONAR discovers RNA-binding proteins from analysis of large-scale protein-protein interactomes. *Mol. Cell* *64*, 282–293.

Cattoglio, C., Pustova, I., Walther, N., Ho, J.J., Hantsche-Grininger, M., Inouye, C.J., Hossain, M.J., Dailey, G.M., Ellenberg, J., Darzacq, X., et al. (2019). Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife* *8*, 40164.

Caudron-Herger, M., Rusin, S.F., Adamo, M.E., Seiler, J., Schmid, V.K., Barreau, E., Kettenbach, A.N., and Diederichs, S. (2019). R-DeeP: proteome-wide and quantitative identification of RNA-dependent proteins by density gradient ultracentrifugation. *Mol. Cell* *75*, 184–199.e10.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* *133*, 1106–1117.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* *367*, 412–415.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., and Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* *367*, eaar2555.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* *523*, 240–244.

Davidson, I.F., Goetz, D., Zaczek, M.P., Molodtsov, M.I., Huis In ’t Veld, P.J., Weissmann, F., Litos, G., Cisneros, D.A., Ocampo-Hafalla, M., Ladurner, R., et al. (2016). Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* *35*, 2671–2685.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Versteegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krjger, P.H.L., and de Laat, W. (2015). CTCF binding polarity determines chromatin looping. *Mol. Cell* *60*, 676–684.

Dekker, J., and Mirny, L. (2016). The 3D genome as moderator of chromosomal communication. *Cell* *164*, 1110–1121.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* *485*, 376–380.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* *3*, 95–98.

El Beheiry, M., and Dahan, M. (2013). ViSP: representing single-particle localizations in three dimensions. *Nat. Methods* *10*, 689–690.

El-Kady, A., and Klenova, E. (2005). Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Lett.* *579*, 1424–1434.

Eimokadem, A., and Yu, J. (2015). Optimal drift correction for superresolution localization microscopy with Bayesian inference. *Biophys. J.* *109*, 1772–1780.

Forcato, M., Nicoletti, C., Pal, K., Livi, C.M., Ferrari, F., and Bicciato, S. (2017). Comparison of computational methods for Hi-C data analysis. *Nat. Methods* *14*, 679–685.

- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* *15*, 2038–2049.
- Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L.A. (2017). Emerging evidence of chromosome folding by loop extrusion. *Cold Spring Harb. Symp. Quant. Biol.* *82*, 45–55.
- Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Nat. Science* *360*, 102–105.
- Gassler, J., Brandão, H.B., Imakaev, M., Flyamer, I.M., Ladstätter, S., Bickmore, W.A., Peters, J.-M., Mirny, L.A., and Tachibana, K. (2017). A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* *36*, 3600–3618.
- Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* *15*, 1451–1455.
- Goecks, J., Nekrutenko, A., and Taylor, J.; Galaxy Team (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* *11*, R86.
- Grimm, J.B., English, B.P., Choi, H., Muthusamy, A.K., Mehl, B.P., Dong, P., Brown, T.A., Lippincott-Schwartz, J., Liu, Z., Lionnet, T., and Lavis, L.D. (2016). Bright photoactivatable fluorophores for single-molecule imaging. *Nat. Methods* *13*, 985–988.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* *162*, 900–910.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* *141*, 129–141.
- Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* *6*, e25776.
- Hansen, A.S., Cattoglio, C., Darzacq, X., and Tjian, R. (2018a). Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* *9*, 20–32.
- Hansen, A.S., Amitai, A., Cattoglio, C., Tjian, R., and Darzacq, X. (2018b). Guided nuclear exploration increases CTCF target search efficiency. *bioRxiv*. <https://doi.org/10.1101/495457>.
- Hashimoto, H., Wang, D., Horton, J.R., Zhang, X., Corces, V.G., and Cheng, X. (2017). Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell* *66*, 711–720.e3.
- He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* *64*, 416–430.
- Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription elongation can affect genome 3D structure. *Cell* *174*, 1522–1536.e22.
- Hnisz, D., Schuijers, J., Li, C.H., and Young, R.A. (2017). Regulation and dysregulation of chromosome structure in cancer. *Annu. Rev. Cancer Biol.* *2*, 21–40.
- Hsieh, T.H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* *162*, 108–119.
- Hsieh, T.S., Fudenberg, G., Goloborodko, A., and Rando, O.J. (2016). Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods* *13*, 1009–1011.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* *37*, 1–13.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* *4*, 44–57.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* *467*, 430–435.
- Kanke, M., Tahara, E., Huis In't Veld, P.J., and Nishiyama, T. (2016). Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *EMBO J.* *35*, 2686–2698.
- Kentepozidou, E., Aitken, S.J., Feig, C., Stefflova, K., Ibarra-Soria, X., Odom, D.T., Roller, M., and Fliece, P. (2019). Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *bioRxiv*. <https://doi.org/10.1101/668855>.
- Kerpedjiev, P., Abdennur, N., Lekschas, F., McCallum, C., Dinkla, K., Strobelt, H., Luber, J.M., Ouellette, S.B., Azhir, A., Kumar, N., et al. (2018). HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* *19*, 125.
- Kim, E., Kerssemakers, J., Shaltiel, I.A., Haering, C.H., and Dekker, C. (2019). DNA-loop extruding condensin complexes can traverse one another. *bioRxiv*. <https://doi.org/10.1101/682864>.
- Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., León, J., et al. (2001). Functional phosphorylation sites in the C-terminal region of the multi-valent multifunctional transcriptional factor CTCF. *Mol. Cell. Biol.* *21*, 2221–2234.
- Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Mol. Cell* *57*, 361–375.
- 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.
- Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* *505*, 495–501.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Lieberman-aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* *326*, 289–293.
- Liu, R., Holik, A.Z., Su, S., Jansz, N., Chen, K., Leong, H.S., Blewitt, M.E., Asselin-Labat, M.L., Smyth, G.K., and Ritchie, M.E. (2015). Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res.* *43*, e97–e97.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Lupiañez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* *161*, 1012–1025.
- Martinez, S.R., and Miranda, J.L. (2010). CTCF terminal segments are unstructured. *Protein Sci.* *19*, 1110–1116.
- McSwiggen, D.T., Hansen, A.S., Teves, S.S., Marie-Nelly, H., Hao, Y., Heckert, A.B., Umemoto, K.K., Dugast-Darzacq, C., Tjian, R., and Darzacq, X. (2019). Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *eLife* *8*, e47098.



- Merkenschlager, M., and Nora, E.P. (2016). CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.* *17*, 17–43.
- Nakahashi, H., Kieffer Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* *3*, 1678–1689.
- Nasmyth, K. (2011). Cohesin: a catenase with separate entry and exit gates? *Nat. Cell Biol.* *13*, 1170–1177.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* *485*, 381–385.
- Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Ueberohrn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from higher-order genomic compartmentalization. *Cell* *169*, 930–944.e22.
- Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., Klenova, E., Lobanekov, V., and Ohlsson, R. (2004). Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol. Cell. Biol.* *24*, 3497–3504.
- Pękowska, A., Klaus, B., Xiang, W., Severino, J., Daigle, N., Klein, F.A., Oleś, M., Casellas, R., Ellenberg, J., Steinmetz, L.M., et al. (2018). Gain of CTCF-anchored chromatin loops marks the exit from naive pluripotency. *Cell Syst.* *7*, 482–495.e10.
- Pélissier, R., and Goreaud, F. (2015). ads package for R: a fast unbiased implementation of the K-function family for studying spatial point patterns in irregular-shaped sampling windows. *J. Stat. Softw.* *63* (6).
- Pettitt, S.J., Liang, Q., Rairdan, X.Y., Moran, J.L., Prosser, H.M., Beier, D.R., Lloyd, K.C., Bradley, A., and Skarnes, W.C. (2009). Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat. Methods* *6*, 493–495.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* *44* (W1), W160–W165.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* *8*, 2281–2308.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* *159*, 1665–1680.
- Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin loss eliminates all loop domains. *Cell* *171*, 305–320.e24.
- Rasko, J.E.J., Klenova, E.M., Leon, J., Filippova, G.N., Loukinov, D.I., Vatolin, S., Robinson, A.F., Hu, Y.J., Ulmer, J., Ward, M.D., et al. (2001). Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Res.* *61*, 6002–6007.
- Rigbolt, K.T.G., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.* *4*, rs3.
- Ripley, B.D. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab.* *13*, 255–266.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* *26*, 139–140.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.
- Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* *19*, 789–800.
- Rubin-Delanchy, P., Burn, G.L., Griffié, J., Williamson, D.J., Heard, N.A., Cope, A.P., and Owen, D.M. (2015). Bayesian cluster identification in single-molecule localization microscopy data. *Nat. Methods* *12*, 1072–1076.
- Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* *28*, 723–734.
- Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsigros, A., Furlan-Magaril, M., Skok, J., et al. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol. Cell* *76*. Published online September 12, 2019. <https://doi.org/10.1016/j.molcel.2019.08.015>.
- Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U S A* *112*, E6456–E6465.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* *9*, 676–682.
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., and Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* *551*, 51–56.
- Sergé, A., Bertaux, N., Rigneault, H., and Marguet, D. (2008). Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* *5*, 687–694.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* *16*, 259.
- Sheff, M.A., and Thorn, K.S. (2004). Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* *21*, 661–670.
- Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. *Science* *357*, eaaf4382.
- Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* *6*, e21856.
- Skibbens, R.V. (2016). Of rings and rods: regulating cohesin entrapment of DNA to generate intra- and intermolecular tethers. *PLoS Genet.* *12*, e1006337.
- Stigler, J., Çamdere, G.Ö., Koshland, D.E., and Greene, E.C. (2016). Single-molecule imaging reveals a collapsed conformational state for DNA-bound cohesin. *Cell Rep.* *15*, 988–998.
- Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA activates Xist by evicting CTCF. *Cell* *153*, 1537–1551.
- Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* *24*, 390–400.
- Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* *14*, 178–192.
- van Steensel, B., and Belmont, A.S. (2017). Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* *169*, 780–791.
- Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* *10*, 1297–1309.
- Wang, F., Tang, Z., Shao, H., Guo, J., Tan, T., Dong, Y., and Lin, L. (2018). Long noncoding RNA HOTTIP cooperates with CCCTC-binding factor to coordinate HOXA gene expression. *Biochem. Biophys. Res. Commun.* *500*, 852–859.
- Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017).

- Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* 36, 3573–3599.
- Xiang, J.F., Yin, Q.F., Chen, T., Zhang, Y., Zhang, X.O., Wu, Z., Zhang, S., Wang, H.B., Ge, J., Lu, X., et al. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* 24, 513–531.
- Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.* 31, 2174–2183.
- Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y., et al. (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* 16, 52.
- Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., and Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 24, 2543–2555.
- Yin, M., Wang, J., Wang, M., Li, X., Zhang, M., Wu, Q., and Wang, Y. (2017). Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.* 27, 1365–1377.
- Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* 13, 291–298.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137.
- Zirke, A., Nikolic, M., Sofiadis, K., Mallm, J.-P., Brackley, C.A., Gothe, H., Drechsel, O., Becker, C., Altmüller, J., Jospovic, N., et al. (2018). HMGB2 loss upon senescence entry disrupts genomic organization and induces CTCF clustering across cell types. *Mol. Cell* 70, 730–744.e6.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Anti-ACTB (WB)	Sigma-Aldrich # A2228	RRID:AB_476697
Anti-CTCF (IP, ChIP)	Abcam # ab128873	RRID:AB_11144295
Anti-CTCF (WB)	Millipore # 07-729	RRID:AB_441965
Anti-FLAG (IP)	Sigma-Aldrich # F7425	RRID:AB_439687
Anti-FLAG (WB, IP)	Sigma-Aldrich # F3165	RRID:AB_259529
Anti-H3 (WB)	Abcam # ab1791	RRID:AB_302613
Anti-HA tag (WB)	Abcam # ab9110	RRID:AB_307019
Anti-HaloTag (WB)	Promega # G9211	RRID:AB_2688011
Anti-Smc1a (ChIP)	Bethyl # A300-055A	RRID:AB_2192467
Anti-V5 (IP)	Abcam # ab9166	RRID:AB_307024
Anti-V5 (WB)	Thermo Fisher Scientific # R960-25	RRID:AB_2556564
Mouse IgG (IP, ChIP)	Jackson ImmunoResearch Labs # 015-000-003	RRID:AB_2337188
Rabbit IgG (IP, ChIP)	Jackson ImmunoResearch Labs # 011-000-003	RRID:AB_2337118
ANTI-FLAG M2 Affinity Gel ( <i>in vitro</i> RNA binding assay)	Sigma-Aldrich #A2220	RRID:AB_10063035
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
DAPI 4',6-Diamidine-2'-phenylindole dihydrochloride	Sigma-Aldrich	Cat. # 10236276001
HaloTag TMR ligand	Promega	Cat. # G8251
HaloTag PA-JF549 ligand	<a href="#">Grimm et al., 2016</a>	N/A
Benzonase	Millipore	Cat. # 71205
RNase A	Thermo Scientific	Cat. # EN0531
DNase I	Ambion	Cat. # AM2222
Formaldehyde	Polysciences	Cat. # 1881420
10 mM dNTPs	KAPA Biosystems	Cat. # KK1017
10 mM ATP	New England Biolabs	Cat. # P0756S
5 U/μl T4 DNA polymerase	Invitrogen	Cat. # 18005025
5 U/μl Taq DNA polymerase	ThermoFisher Scientific	Cat. # EP0401
KAPA HS HIFI polymerase	KAPA Biosystems	Cat. # KK2502
Rapid DNA ligase	Enzymatics	Cat. # L6030-HC-L
AMPure XP beads	Agencourt	Cat. # A63880
DSG (disuccinimidyl glutarate)	ThermoFisher Scientific	Cat. # 20593
Micrococcal Nuclease	Worthington Biochem	Cat. # LS004798
100mM ATP	ThermoFisher Scientific	Cat. # R1441
DNA Polymerase I, Large (Klenow) Fragment	New England Biolabs	Cat. # M0210
T4 Polynucleotide Kinase	New England Biolabs	Cat. # M0201
T4 DNA Ligase	New England Biolabs	Cat. # M0202
Exonuclease III ( <i>E. coli</i> )	New England Biolabs	Cat. # M0206
Biotin-14-dATP	Jena Bioscience	Cat. # NU-835-BIO14
Biotin-11-dCTP	Jena Bioscience	Cat. # NU-809-BIOX
20X Proteinase K solution	Sigma Aldrich	Cat. # 3115879001

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dynabeads MyOne Streptavidin C1	ThermoFisher Scientific	Cat. # 65001
SuperScript III Reverse Transcriptase	Invitrogen	Cat. # 18080044
SYBR Gold Nucleic Acid Gel Stain	Invitrogen	Cat. # S11494
PageBlue Protein Staining Solution	ThermoFisher Scientific	Cat. # 24620
TRIzol Reagent	ThermoFisher Scientific	Cat. # 15596026
Ni-NTA agarose	QIAGEN	Cat. # 30210
3X FLAG Peptide	Sigma Aldrich	Cat. # F4799
recombinant 3xFLAG-Halo-wt-CTCF-6xHis protein	This manuscript	r-wt-CTCF
recombinant 3xFLAG-Halo- $\Delta$ RBR <sub>i</sub> -CTCF-6xHis protein	This manuscript	r- $\Delta$ RBR <sub>i</sub> -CTCF
EMPIGEN BB detergent	Sigma-Aldrich	Cat. # 30326
<b>Critical Commercial Assays</b>		
Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit	ThermoFisher Scientific	Cat. # C10425
NEBNext Ultra II	New England Biolabs	Cat. # E7645
End-It DNA End-Repair	Lucigen	Cat. # ER81050
RNeasy Mini Kit	QIAGEN	Cat. # 74104
DNA-free DNA Removal Kit	Invitrogen	Cat. # AM1906
Ribo-Zero rRNA Removal Kit	Illumina	Cat. # MRZH116
TruSeq RNA Sample Preparation v2 Kit	Illumina	Cat. # RS-122-2001
HiScribe T7 Quick High Yield RNA Synthesis Kit	New England Biolabs	Cat. # E2050S
Bac-to-Bac Baculovirus Expression System	ThermoFisher Scientific	Cat. # 10359016
<b>Deposited Data</b>		
Raw imaging data (all raw data for This manuscript)	This manuscript and Hansen et al., 2018b	<a href="https://zenodo.org/record/2208323">https://zenodo.org/record/2208323</a>
Micro-C, ChIP-Seq and RNA-Seq data	This manuscript	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123636">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123636</a>
Raw images and uncropped gels and blots deposited in Mendeley Data	This manuscript	DOI: <a href="https://doi.org/10.17632/5zdrpcsbt9.2">https://doi.org/10.17632/5zdrpcsbt9.2</a>
CTCF ChIP-Seq in mESCs	Chen et al., 2008	GSE11431
Smc1a ChIP-Seq in mESCs	Kagey et al., 2010	GSE22562
RNA-seq in mESCs after CTCF degradation	Nora et al., 2017	GSE98671
<b>Experimental Models: Cell Lines</b>		
Mouse: JM8.N4 mouse embryonic stem cells	Pettitt et al., 2009 and UC Davis KOMP Repository	<a href="https://www.komp.org/pdf.php?clonelD=8669">https://www.komp.org/pdf.php?clonelD=8669</a>
mESC C59 FLAG-Halo-CTCF; Rad21-SNAP <sub>T</sub> -V5 (knock-in)	Hansen et al., 2017	C59
mESC C59D2 $\Delta$ RBR <sub>i</sub> -FLAG-Halo-CTCF (e10::3xHA); Rad21-SNAP <sub>T</sub> -V5 (knock-in)	This manuscript	C59D2 or $\Delta$ RBR <sub>i</sub>
mESC C62 3xFLAG-Halo-CTCF (allele 1); V5-SNAP <sub>T</sub> -Halo-CTCF (allele 2) (both knock-in)	This manuscript	C62
<b>Oligonucleotides</b>		
See Table S1		N/A
<b>Recombinant DNA</b>		
pBlueScript SK II (+)	Addgene # 212205	GenBank: X52328.1
pBSII HR mCtcf. delRBR <sub>i</sub> (link-3XHA) Repair Vector	This manuscript	N/A
pUC57 V5 Snap(f) mCTCF Repair Vector	This manuscript	N/A

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pUC57 3xFLAG Halo TEV mCTCF Repair Vector	This manuscript	N/A
pFastBac Dual Expression Vector	ThermoFisher Scientific	Cat. # 10712024
<b>Software and Algorithms</b>		
MATLAB 2014b	The Mathworks	2014b
PALM analysis pipeline (MATLAB)	This manuscript	<a href="https://gitlab.com/anders.sejr.hansen/palm_pipeline">https://gitlab.com/anders.sejr.hansen/palm_pipeline</a>
FCSREAD (MATLAB)	Mathworks File Exchange	<a href="https://www.mathworks.com/matlabcentral/fileexchange/8430-flow-cytometry-data-reader-and-visualization">https://www.mathworks.com/matlabcentral/fileexchange/8430-flow-cytometry-data-reader-and-visualization</a>
MTT-Algorithm (MATLAB implementation)	<a href="#">Sergé et al., 2008</a>	<a href="https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack">https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack</a>
SimpleTracker (MATLAB)	Jean-Yves Tinevez	<a href="https://www.mathworks.com/matlabcentral/fileexchange/34040-simple-tracker">https://www.mathworks.com/matlabcentral/fileexchange/34040-simple-tracker</a>
Spatial Point Patterns Analysis (ads package on CRAN)	<a href="#">Pélissier and Goreaud, 2015</a>	<a href="https://cran.r-project.org/web/packages/ads/index.html">https://cran.r-project.org/web/packages/ads/index.html</a>
RStudio	RStudio	<a href="https://www.rstudio.com">https://www.rstudio.com</a>
BaSDI (MATLAB)	<a href="#">Elmokadem and Yu, 2015</a>	<a href="https://github.com/jiyuuchc/BaSDI">https://github.com/jiyuuchc/BaSDI</a>
ImageJ ( <a href="https://imagej.net/">https://imagej.net/</a> )	<a href="#">Schindelin et al., 2012</a>	RRID:SCR_003070
Samtools	<a href="#">Li et al., 2009</a>	RRID:SCR_002105
Integrative Genomics Viewer	<a href="#">Robinson et al., 2011</a> ; <a href="#">Thorvaldsdóttir et al., 2013</a>	RRID:SCR_011793
Bowtie	<a href="#">Langmead et al., 2009</a>	RRID:SCR_005476
deepTools	<a href="#">Ramírez et al., 2016</a>	RRID:SCR_016366
FastQC	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc">http://www.bioinformatics.babraham.ac.uk/projects/fastqc</a>	RRID:SCR_014583
MACS2	<a href="#">Zhang et al., 2008</a>	<a href="https://github.com/taoliu/MACS/">https://github.com/taoliu/MACS/</a>
Python 3.7	Python	<a href="https://www.python.org/">https://www.python.org/</a>
Anaconda 3.7	Anaconda	<a href="https://www.anaconda.com/">https://www.anaconda.com/</a>
HIC-Pro	<a href="#">Servant et al., 2015</a>	<a href="https://github.com/nservant/HiC-Pro">https://github.com/nservant/HiC-Pro</a>
Cooler	<a href="#">Abdennur and Mirny, 2019</a>	<a href="https://github.com/mirnylab/cooler">https://github.com/mirnylab/cooler</a>
Juicer tools	<a href="#">Durand et al., 2016</a>	<a href="https://github.com/aidenlab/juicer">https://github.com/aidenlab/juicer</a>
Higlass	<a href="#">Kerpedjiev et al., 2018</a>	<a href="http://higlass.io/">http://higlass.io/</a>
STAR	<a href="#">Dobin et al., 2013</a>	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
HTSeq	<a href="#">Anders et al., 2015</a>	<a href="https://pypi.org/project/HTSeq/">https://pypi.org/project/HTSeq/</a>
edgeR	<a href="#">Robinson et al., 2010</a> <a href="#">Liu et al., 2015</a>	<a href="https://bioconductor.org/packages/release/bioc/html/edgeR.html">https://bioconductor.org/packages/release/bioc/html/edgeR.html</a>
Galaxy	<a href="#">Blankenberg et al., 2010</a> ; <a href="#">Giardine et al., 2005</a> ; <a href="#">Goecks et al., 2010</a>	<a href="https://usegalaxy.org/">https://usegalaxy.org/</a>
DESeq2	<a href="#">Love et al., 2014</a>	<a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
DAVID Bioinformatics Resources 6.8	<a href="#">Huang et al., 2009a, 2009b</a>	<a href="https://david.ncifcrf.gov/home.jsp">https://david.ncifcrf.gov/home.jsp</a>
VISP	<a href="#">El Beheiry and Dahan, 2013</a>	<a href="https://www.nature.com/articles/nmeth.2566">https://www.nature.com/articles/nmeth.2566</a>

**LEAD CONTACT AND MATERIALS AVAILABILITY**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Robert Tjian ([jmlim@berkeley.edu](mailto:jmlim@berkeley.edu)).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cell culture

JM8.N4 mouse embryonic stem cells (Pettitt et al., 2009) (male mESCs; Research Resource Identifier: RRID:CVCL\_J962; obtained from the KOMP Repository at UC Davis) were grown and handled as described previously (Hansen et al., 2017). Briefly, mES cells were grown on plates pre-coated with a 0.1% autoclaved gelatin solution (Sigma-Aldrich, St. Louis, MO, G9391) under feeder free conditions in knock-out DMEM with 15% FBS and LIF (full recipe: 500 mL knockout DMEM (ThermoFisher, Waltham, MA, #10829018), 6 mL MEM NEAA (ThermoFisher #11140050), 6 mL GlutaMax (ThermoFisher #35050061), 5 mL Penicillin-streptomycin (ThermoFisher #15140122), 4.6  $\mu$ L 2-mercapoethanol (Sigma-Aldrich M3148), 90 mL fetal bovine serum (HyClone Logan, UT, FBS SH30910.03 lot #AXJ47554)) and LIF. mES cells were fed by replacing half the medium with fresh medium daily and passaged every two days by trypsinization. Cell lines were pathogen tested (IMPACT II test for mESC C59) as described previously (Hansen et al., 2017). All cell lines will be provided upon request.

## METHOD DETAILS

### CRISPR/Cas9-mediated genome editing

Genome-editing was performed as previously described (Hansen et al., 2017). Briefly, we co-transfected cells with a repair plasmid and a plasmid encoding Cas9 and the sgRNA (using 2  $\mu$ g and 1  $\mu$ g, respectively, per well in a 6-well plate). The Cas9 plasmid was slightly modified from that distributed from the Zhang lab (Ran et al., 2013): 3xFLAG-SV40NLS-pSpCas9 was expressed from a CBh promoter; the sgRNA was expressed from a U6 promoter; and mVenus was expressed from a PGK promoter. We generally designed 2–4 sgRNAs per knock-in and transfected each (or two of them when necessary) in a separate well. The day of transfection, we pooled all transfected cells and FACS-sorted for transfected cells using the mVenus encoding by the Cas9 plasmid. For edits where there was no tag added (e.g., to replace the RBR<sub>i</sub> with 3xHA), we immediately plated single clones after the FACS. But for knock-ins with tags, e.g., 3xFLAG-Halo-CTCF or V5-SNAP<sub>F</sub>-CTCF, we first grew up cells and then labeled cells with dye (Halo-TMR for 3xFLAG-Halo-CTCF; SNAP-JF646 for V5-SNAP<sub>F</sub>-CTCF) and then did a second round of FACS-sorting to increase the efficiency. Selected cells were plated at very low density ( $\sim$ 0.1 cells per mm<sup>2</sup>), and single colonies were then picked, expanded and genotyped by PCR. Successfully edited clones were further verified by PCR followed by Sanger sequencing and western blotting.

For knock-ins with 2 different tags, we generated them using the above protocol in 2 steps. We first isolated a heterozygous knock-in clone for one tag and we then re-edited that clone to introduce the second tag. This was the case for C62, where one of the diploid CTCF alleles is V5-SNAP<sub>F</sub>-tagged and the other 3xFLAG-Halo-tagged. We first isolated a clone with a correct V5-SNAP<sub>F</sub>-tagged CTCF allele and a null CTCF allele, where non-homologous end joining event following Cas9 cleavage introduced a 4-nucleotide deletion (81\_84delACGC), leading to a premature stop codon. We then designed sgRNAs specific for the null CTCF allele and re-targeted this clone with a 3xFLAG-Halo-CTCF repair vector. To build the repair vectors, we modified a pUC57 plasmid to contain the tag of interest flanked by  $\sim$ 500 bp of genomic homology sequence on either side (IDT gBLOCKs). To prevent the Cas9-sgRNA complex from cutting the repair vector, we introduced synonymous mutations in the first nine codons after the ATG. To link the SNAP and Halo proteins to CTCF, we used the Sheff and Thorn linker (GDGAGLIN) (Sheff and Thorn, 2004) and a TEV linker sequence (EDLYFQS), respectively. mESC clones were screened using a three-primer PCR (two genomic primers external to the left and right homology sequences, and one internal to the tag).

To endogenously and homozygously delete the RBR<sub>i</sub> region in the previously published C59 mESC line (Hansen et al., 2017), we generated by Gibson Assembly a repair vector modifying a pBlueScript II SK (+) plasmid to contain the Sheff and Thorn linker followed by a 3xHA tag (Figure S1F), and flanked by  $\sim$ 500 bp of genomic homology sequence on either side. mESC clones were screened using a three-primer PCR (one genomic primer external to the left homology sequence, one internal to the right homology region, and an internal HA primer). Notably, we failed to generate clones with a simple deletion of the RBR<sub>i</sub>, possibly because shortening of the already small exon 10 (only 135 bp-long, 27 bp upon RBR<sub>i</sub> deletion) causes exon skipping and aberrant splicing.

All plasmids used in the editing are available upon request as are any of the cell lines. See Table S1 for sgRNA and primer sequences.

### Cell Cycle phase analysis

Cell cycle phase analysis was performed using the Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (ThermoFisher Scientific Cat. # C10425) according to manufacturer's instructions, but with minor modifications. C59 mESCs (Halo-CTCF; Rad21-SNAP<sub>F</sub>) and C59D2 mESCs ( $\Delta$ RBR<sub>i</sub>-Halo-CTCF; Rad21-SNAP<sub>F</sub>) were grown overnight in a 6-well plate and labeled with 10  $\mu$ M EdU for 30 min at 37°C/5.5% CO<sub>2</sub> in a TC incubator (one well was unlabeled, as a negative control). Cells were harvested, washed with 1% BSA in PBS, permeabilized (using 100  $\mu$ L 1x Click-iT saponin-based permeabilization and wash reagent (Component D; see kit manual), mixed well and then incubated for 15 min. 0.5 mL Click-iT reaction was added to each tube and incubated for 30 min in the dark. Cells were washed with 1x Click-iT saponin-based permeabilization and wash reagent and resuspended in 1x Click-iT saponin-based permeabilization and wash reagent with DAPI (5 ng/mL) and incubated for 10 min. Cells were then spun down and re-suspended in 1% BSA in PBS and FACS performed on a LSR Fortessa Cytometer. DAPI fluorescence was excited using a 405 nm laser and collected using a 450/50 bandpass emission filter. Alexa Fluor 488 fluorescence was excited using a 488 nm laser and collected using

a 525/50 bandpass emission filter. Cells were gated based on forward and side scattering using identical settings for C59 and C59D2 mESCs. Cell cycle analysis was performed using custom-written MATLAB code using identical settings for C59 and C59D2 mESCs as illustrated in [Figures S4F and S4G](#). Three independent biological replicates were performed.

### CTCF FACS abundance quantification

FACS was performed as previously described ([Hansen et al., 2017](#)). We grew C59 mESCs (Halo-CTCF; Rad21-SNAP<sub>f</sub>) and C59D2 mESCs ( $\Delta$ RBR<sub>f</sub>-Halo-CTCF; Rad21-SNAP<sub>f</sub>) overnight in a 6-well plate and labeled 1 well with 500 nM Halo-TMR (Promega Cat. # G8521) and left 1 well unlabeled (negative control for baseline fluorescence). Cells were labeled for 30 min at 37°C/5.5% CO<sub>2</sub> in a TC incubator, washed with PBS and incubated with medium for 5 min in a TC incubator. Cells were then washed again with PBS, harvested, filtered and fluorescence quantified in live cells on a LSR Fortessa Cytometer, exciting fluorescence with a 561 nm laser and collecting fluorescence through a 610/20 bandpass emission filter. Live cells were gated based on forward and side scattering (using identical settings for C59 and C59D2 mESCs) using custom-written MATLAB code and the relative abundance quantified as the relative background-subtracted mean fluorescence as illustrated in [Figures 2C and S1G](#).

### Growth Assay

When passaging cells, two processes contribute to the apparent growth rate: 1) the fraction of cells that survive passaging and 2) the growth rate. To compare exclusively the growth rate of mESC C59 Halo-CTCF and mESC C59D2  $\Delta$ RBR<sub>f</sub>-Halo-CTCF, we therefore took the following approach. On day 0, we plated 250,000 cells in 2 wells in a 6-well plate. On day 1, we collected and counted the number of cells from 1 well. This gave us the number of cells that survived plating. Let this number be  $N_1$ . On day 2, we then collected and counted the number of cells from the second well. Let this number be  $N_2$  and the time between the measurements be  $\Delta\tau$ . The doubling time is then given by:

$$\tau_{\text{DOUBLING}} = \frac{\Delta\tau \ln(2)}{\ln\left(\frac{N_2}{N_1}\right)}$$

We performed 4 biological replicates and grew C59 and C59D2 side-by-side at the same time and handled them identically. The bar graph in [Figure 2D](#) shows the mean and standard error of the mean from the 4 replicates.

### PALM

PALM was performed as previously described ([Hansen et al., 2017](#)) but with minor modifications. C59 mESCs (Halo-CTCF; Rad21-SNAP<sub>f</sub>) and C59D2 mESCs ( $\Delta$ RBR<sub>f</sub>-Halo-CTCF; Rad21-SNAP<sub>f</sub>) were grown overnight on Matrigel coated plasma-cleaned 25 mm circular no 1.5H cover glasses (Marienfeld, Germany, High-Precision 0117650), labeled with 500 nM PA-JF549 ([Grimm et al., 2016](#)) for 30 min at 37°C/5.5% CO<sub>2</sub> in a TC incubator, washed twice (medium removed; PBS wash; fresh medium for 5 min), and then fixed in 4% Formaldehyde / 0.2% Glutaraldehyde in PBS for 20 min at 37°C, washed with PBS and then imaged in PBS with 0.01% (w/v) NaNO<sub>3</sub> on the same day. All PALM movies were acquired at room temperature using continuous HiLo illumination on the same microscope as previously described ([Hansen et al., 2017](#)). We used the following laser lines: main excitation laser (561 nm for PA-JF549) and photo-activation laser (405 nm). However, the intensity of the 405 nm laser was gradually increased over the course of the illumination sequence to image all molecules and at the same time avoid too many molecules being activated at any given frame. The following camera settings were used: 25 ms exposure time; frame transfer mode; vertical shift speed: 0.9  $\mu$ s; ROI: variable. In total, 40,000 frames were recorded for each cell (~20 min), which was sufficient to image and bleach all labeled molecules at an effective pixel size of 106.67 nm, which resulted in a mean localization error (defined as the standard deviation) of ~13-14 nm ([Figure S1H](#)). We recorded 6-10 movies per cell line per day (and always imaged both C59 and C59D2 on the same day) and performed 3 biological replicates. Each movie contained several nuclei (generally 3-6), which improved the robustness of the algorithmic drift-correction ([Elmokadem and Yu, 2015](#)). We obtained and analyzed a total of 52 cells for C59 and 46 cells for C59D2.

Molecules in PALM data were localized using a custom-written MATLAB implementation of the MTT-algorithm ([Sergé et al., 2008](#)); code is available on GitLab: [https://gitlab.com/tjian-darzaccq-lab/SPT\\_LocAndTrack](https://gitlab.com/tjian-darzaccq-lab/SPT_LocAndTrack)) and the following settings: Localization error: 10<sup>-6</sup>; deflation loops: 0. After localization, the data was analyzed as described below using code available on GitLab: [https://gitlab.com/anders.sejr.hansen/palm\\_pipeline](https://gitlab.com/anders.sejr.hansen/palm_pipeline)

### PALM analysis

Full details on PALM analysis as well as code to reproduce our results are available on GitLab: [https://gitlab.com/anders.sejr.hansen/palm\\_pipeline](https://gitlab.com/anders.sejr.hansen/palm_pipeline). Here we summarize the major steps. First, drift-correction and merging of blinks is achieved through the main script “DriftCorrectMergeBlinks.m,” which calls a number of functions and runs in parallel as default, so the parallel processing toolbox in MATLAB is necessary. Drift-correction is first performed using a custom-modified implementation of BaSDI ([Elmokadem and Yu, 2015](#)) (“BaSDI\_ASH”). This is achieved through the function “IterativeBaSDI\_DriftCorrect.m” using FramesBin = 2000; PixelBin = 10; Iterations = 5. Compared with BaSDI, the main difference is that we found multiple iterations to be necessary to reach

convergence and we have therefore custom-written the wrapper “IterativeBaSDI\_DriftCorrect.m” to achieve this. Since the inferred drift is binned according to “FramesBin,” we use linear interpolation to drift-correct each frame. Once drift-correction has been achieved, we merge photo-blinking using a custom implementation of SimpleTracker (<https://www.mathworks.com/matlabcentral/fileexchange/34040-simple-tracker>), which was modified to be substantially more memory-efficient for large PALM movies (SimpleTracker\_ASH). An important aspect of PALM, especially with very photo-stable dyes such as PA-JF549 (Grimm et al., 2016), is that the same molecule can appear in multiple adjacent frames and also blink such that there are gaps. It is therefore essential to link these appearances, which we accomplish using SimpleTracker’s implementation of nearest neighbor tracking and we allow a maximal linking distance of 75 nm and maximally 2 gaps. We note that 75 nm is quite lenient since the localization error is less than 15 nm, but we chose it so to ensure we fully correct for multiple appearances. For each molecule with multiple appearances, we collapse all the localizations to a single localization and take the x,y coordinates to be the means.

After drift-correcting and merging, individual nuclei are segmented after Gaussian smoothing of reconstructed images using a 60 nm pixel size. Since the movies contain several nuclei, each nucleus is manually segmented using polygon-segmentation. For each nucleus, a series of summary statistics are then displayed and saved (e.g., localization error, number of localization per frame, nuclear reconstructions) and each nucleus is saved to a separate directory together with code for running K-Ripley analysis (Besag, 1977; Boehning et al., 2018; Ripley, 1976) using the ads package in R (Pélissier and Goreaud, 2015) as well as code for running a Bayesian cluster identification algorithm (Rubin-Delanchy et al., 2015).

The R-code for running K-Ripley analysis was written by Herve Marie-Nelly and is described elsewhere (Boehning et al., 2018). The version included here is a slightly modified version and we refer the reader to the tutorial on GitLab for how to run it (requires both Python and R). Finally, the results of the K-Ripley analysis were plotted with “PLOT\_K\_L\_g\_Ripley.m” and Figure 2I show the mean and standard error of the mean across the population. More generally, Ripley’s *K* function analyzes pointillist data. PALM generates pointillist data. Specifically, we have in 2 dimensions the X,Y-coordinates for each CTCF protein inside the nucleus. Ripley’s *K* function is defined as:

$$\widehat{K}(r) = \lambda^{-1} \sum_{i \neq j} \frac{I(d_{ij} < r)}{n}$$

where  $d_{ij}$  is the Euclidian distance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  points,  $\lambda$  is the average density of points,  $r$  is the search radius, where the total number of data points (i.e., CTCF protein X,Y-coordinates) is  $n$ .  $I$  is the indicator function (equal to 1 only if the distance  $d_{ij}$  is smaller than  $r$ ; otherwise, 0).  $K(r)$  scales as  $\pi r^2$  in 2 dimensions, if CTCF is randomly distributed. For this reason, typically, Ripley’s *L* function is used instead (this formulation was introduced by Besag in 1977):

$$\widehat{L}(r) - r = \sqrt{\frac{\widehat{K}(r)}{\pi}} - r$$

Interpreting plots of  $\widehat{L}(r) - r$  such as shown in Figure 2I is straightforward: If the data are randomly distribution,  $\widehat{L}(r) - r = 0$ . If below 0, there is dispersion (“repulsion”). And if above 0, there is clustering (“attraction” between the CTCF proteins).

Ripley’s *K* and *L* function are normalized for the abundance. In other words, clustering does not depend on protein abundance and the 27.7% lower expression level of  $\Delta$ RBR<sub>1</sub>-CTCF cannot explain the lower clustering that we observe. For full details, we refer to the original papers by Ripley and Besag (Besag, 1977; Ripley, 1976).

The example reconstructions of CTCF nuclear localization in Figures 2G and 2H were plotted using ViSP (El Beheiry and Dahan, 2013). Each molecule was plotted using 25 nm (FWHM) and colored according to the neighbor density (0-200 Neighbors (min/max); Neighborhood Radius: 100 nm; Jet colormap (cMin-cMax: 0-0.35) with identical settings for C59 Halo-CTCF and C59D2  $\Delta$ RBR<sub>1</sub>-Halo-CTCF.

### Western Blotting

Cells were grown in 6-well plates to confluency, washed twice with ice-cold PBS with protease inhibitors and scraped in 300  $\mu$ L of high salt lysis buffer (0.5 M NaCl, 25 mM HEPES, 1 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5% NP-40 and protease inhibitors). Lysates were immediately transferred to 1.5 mL tubes containing 100  $\mu$ L of 4X protein loading buffer (16% 2-Mercaptoethanol, 200 mM Tris-HCl pH 6.8, 8% SDS, 40% glycerol, 400 mM DTT, 0.4% bromophenol blue), boiled for 20’ and loaded to 8% Bis-Tris protein gels (10  $\mu$ L per lane). Proteins were transferred onto nitrocellulose membranes (Amershan Protran 0.45  $\mu$ m NC, GE Healthcare) for 2 hr at 100V. Membranes were blocked in TBS-Tween with 10% milk for at least 1 hr at room temperature and blotted with the specified antibodies in TBS-T with 5% milk at 4°C overnight. HRP-conjugated secondary antibodies were diluted 1:5000 in TBS-T with 5% milk and incubated at room temperature for an hour prior to the chemiluminescence reaction. Band intensities were measured with the ImageJ “Analyze Gels” function (Schindelin et al., 2012) and used to calculate IP and CoIP efficiencies.

### Co-immunoprecipitation (CoIP) assays

For CoIP experiments, cells were scraped from plates in ice-cold phosphate-buffered saline (PBS) with PMSF and aprotinin, pelleted, and flash-frozen in liquid nitrogen. Cell pellets were thawed on ice, resuspended to 1 ml/10 cm plate of cell lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40 and protease inhibitors), and incubated on ice for 10’. Nuclei were pelleted in a tabletop centrifuge at



4°C, at 4000 rpm for 10', and resuspended to 0.5 ml/10 cm plate of low salt lysis buffer either with or without benzonase (600U/ml) and rocked 4 hr at 4°C. After the incubation the salt concentration was adjusted to 0.2 M NaCl final and the lysates were incubated for another 30' at 4°C. 50  $\mu$ L of each lysate were used for DNA and RNA extraction (see below), while the rest was cleared by centrifugation at maximum speed at 4°C and the supernatants quantified by Bradford. In a typical CoIP experiment, 1 mg of proteins was diluted in 1 mL CoIP buffer (0.2 M NaCl, 25 mM HEPES, 1 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5% NP-40 and protease inhibitors) and pre-cleared for 2 hr at 4°C with protein-A/G Sepharose beads (GE Healthcare Life Sciences) before overnight immunoprecipitation with 4 mg of either normal serum IgGs or specific antibodies. Some pre-cleared lysate was kept at 4°C overnight as input. Protein-A/G-Sepharose beads pre-cleared overnight in CoIP buffer with 0.5% BSA were then added to the samples and incubated at 4°C for 2 hr. Beads were then washed extensively with CoIP buffer, and proteins were eluted by boiling the beads for 5' in 2X SDS-loading buffer. The immunoprecipitated material was split to two SDS-PAGE gels followed by Western Blotting: 90% of the IP was loaded to probe CoIP efficiencies, while 10% of the IP was loaded to probe IP efficiencies.

### CoIP DNA and RNA extraction and quantification

For DNA extraction, 50  $\mu$ L of lysates were added to 150  $\mu$ L of CoIP buffer and extracted twice with 200  $\mu$ L of phenol-chloroform (UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v)). After centrifugation at room temperature and maximum speed for 5', the aqueous phase containing DNA was added of 2 volumes of 100% ethanol and precipitated 30' at -80°C. After centrifugation at 4°C for 20' at maximum speed, DNA was re-dissolved in 25  $\mu$ L water and quantified by nanodrop. About 100 ng of the untreated sample DNA, or an equal volume from the nuclease treated samples, were used for relative quantification by quantitative PCR (qPCR) with SYBR Select Master Mix for CFX (Applied Biosystems, ThermoFisher) on a BIO-RAD CFX Real-time PCR system (primer sequences in [Table S1](#)).

RNA was extracted from 50  $\mu$ L of lysates with 500  $\mu$ L of TRIzol reagent, following manufacturer's instructions. The RNA pellet was re-dissolved in 25  $\mu$ L of water and quantified by nanodrop. About 1  $\mu$ g of the untreated sample RNA, or an equal volume from the nuclease treated samples, was retrotranscribed with SuperScript III Reverse Transcriptase and random examers. cDNA was diluted 1:20 and 2  $\mu$ L quantified by qPCR as above.

### Chromatin immunoprecipitation (ChIP)

Smc1a, CTCF and control IgG ChIP assays were performed in the parental C59 ES cell line (wt-CTCF) and in its derivative clone C59D2 ( $\Delta$ RBR<sub>1</sub>-CTCF). Cells were cross-linked for 5' at room temperature with 1% formaldehyde-containing Knockout D-MEM; cross-linking was stopped by PBS-glycine (0.125 M final). Cells were washed twice with ice-cold PBS, scraped, centrifuged for 10' at 4000 rpm and flash-frozen in liquid nitrogen. Cell pellets were thawed in ice, resuspended in cell lysis buffer (5 mM PIPES, pH 8.0, 85 mM KCl, and 0.5% NP-40, 1 ml/15 cm plate) and incubated for 10' on ice. During the incubation, the lysates were repeatedly pipetted up and down every 5 minutes. Lysates were then centrifuged for 10' at 4000 rpm. Nuclear pellets were measured and resuspended in 6 volumes of sonication buffer (50 mM Tris-HCl, pH 8.1, 10 mM EDTA pH 8.0, 0.1% SDS), incubated on ice for 10', and sonicated to obtain DNA fragments below 2000 bp in length (Covaris S220 sonicator, 20% Duty factor, 200 cycles/burst, 150 peak incident power, 30-40 cycles of 20" on and 40" off). Sonicated lysates were cleared by centrifugation (20' at 13200 rpm) and 625-800  $\mu$ g of chromatin were diluted in RIPA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA pH 8.0, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, 140 mM NaCl) to a final concentration of 0.8  $\mu$ g/ $\mu$ L, pre-cleared with Protein A Sepharose (GE Healthcare) for 2 hr at 4°C and immunoprecipitated overnight with 6.25-8  $\mu$ g of normal mouse IgGs (ChromPure rabbit normal IgG; Jackson ImmunoResearch), anti-Smc1a (Abcam ab154769) or anti-CTCF antibodies (Abcam ab128873), which we have extensively validated for ChIP in a previous paper ([Hansen et al., 2017](#)). 4% of the pre-cleared chromatin was saved as input. After the overnight incubation, samples were added to 20  $\mu$ L of Protein A Sepharose beads pre-cleared overnight in RIPA buffer with 0.5% (w/v) BSA and incubated for 2 hr at 4°C. Immunoprecipitated samples were washed 5 times with RIPA buffer, once with LiCl buffer (0.5% NP-40, 0.5% Na-deoxycholate, 250 mM LiCl, 1 mM EDTA pH 8.0), and once with TE. After the last wash, immunoprecipitated complexes were eluted from the beads twice with 150  $\mu$ L of TE with 1% SDS, each time incubating 30' in a thermomixer set at 37°C and 900 rpm. To the 300  $\mu$ L eluted material was added of 1  $\mu$ L of RNaseA (10 mg/ml) and 18  $\mu$ L 5M NaCl, and incubated at 67°C for 4-5 hr to reverse formaldehyde cross-linking. To inputs were added elution buffer to 300  $\mu$ L total volume, and subject to the same treatment. To reverse cross-linked samples were added 2.5 volumes of ice-cold ethanol and precipitated overnight at -20°C. DNA was pelleted by centrifugation (20' at 13,200 rpm and 4°C), and pellets resuspended in 100  $\mu$ L TE, 25  $\mu$ L 5X PK buffer (50 mM Tris-HCl, pH 7.5, 25 mM EDTA pH 8.0, 1.25% SDS), and 1.5  $\mu$ L of proteinase K (20 mg/ml), and incubated 2 hr at 45°C. After proteinase K digestion, DNA was purified with the QIAGEN QIAquick PCR Purification Kit, eluted in 60  $\mu$ L of water and used for ChIP-Seq library preparation as described below.

### Expression and purification of recombinant wt-CTCF and $\Delta$ RBR<sub>1</sub>-CTCF proteins

Recombinant Bacmid DNAs for the fusion mouse proteins 3xFLAG-Halo-wt-CTCF-6xHis (1086 amino acids; 123.5 kDa) and 3xFLAG-Halo- $\Delta$ RBR<sub>1</sub>-CTCF-6xHis (1086 amino acids; 123.7 kDa) were generated from pFastBAC constructs according to manufacturer's instructions (Invitrogen). Recombinant baculovirus for the infection of Sf9 cells was generated using the

Bac-to-Bac Baculovirus Expression System (Invitrogen). Sf9 cells ( $\sim 2 \times 10^6$  / ml) were infected with amplified baculoviruses expressing recombinant wt- or  $\Delta$ RBR<sub>i</sub>-CTCF. Infected Sf9 suspension cultures were collected at 48 hr post infection, washed extensively with cold PBS, lysed in 5 packed cell volumes of high salt lysis buffer (HSLB; 1.0 M NaCl, 50 mM HEPES pH 7.9, 0.05% NP-40, 10% glycerol, 10 mM 2-mercaptoethanol, and protease inhibitors), and sonicated. Lysates were cleared by ultracentrifugation, supplemented with 10 mM imidazole, and incubated at 4°C with Ni-NTA resin (QIAGEN) for 90 minutes. Bound proteins were washed extensively with HSLB with 20 mM imidazole, equilibrated with 0.5 M NaCl HGN (50 mM HEPES pH 7.9, 10% glycerol, 0.01% NP-40) with 20 mM imidazole, and eluted with 0.5 M NaCl HGN supplemented with 0.25 M imidazole. Eluted fractions were analyzed by SDS-PAGE followed by staining with PageBlue Protein Staining Solution. Peak fractions were pooled and incubated with anti-FLAG® M2 Affinity Gel (Sigma) for 3 hr at 4°C. Bound proteins were washed extensively with HSLB, equilibrated to 0.2M NaCl HGN, and eluted with 3xFLAG peptide (Sigma) at 0.4 mg/ml. Protein concentrations were determined by PageBlue staining compared to a BSA standard.

### **In vitro RNA binding assay**

Binding of CTCF recombinant proteins to RNA was assessed *in vitro* as described by (Saldaña-Meyer et al., 2014) with some modifications. The first exon of human *WRAP53* (nucleotides 1-167) was PCR amplified from HEK293T genomic DNA using PrimeSTAR HS DNA Polymerase (Takara R010B) and a forward primer that included a T7 promoter sequence (see Table S1). The gel-purified PCR product (200 ng) served as a template for T7 *in vitro* transcription (NEB), carried out in a total volume of 30  $\mu$ L and incubated at 37°C for 4 hours and 30 minutes. The transcribed RNA was added to 30  $\mu$ L of water and 2  $\mu$ L RNase-free DNase I and incubated at 37°C for 15 minutes. The total volume was then adjusted to 360  $\mu$ L with water. 40  $\mu$ L of 3.3 M sodium acetate pH 5.2 was then added before two sequential extractions with 1 volume of phenol/chloroform followed by ethanol precipitation over night. The RNA pellet was washed with 500  $\mu$ L of 70% ethanol, resuspended in water and quantified by nanodrop. 4 pmol RNA were incubated with 18.5 pmol of wt- or  $\Delta$ RBR<sub>i</sub>-CTCF recombinant proteins in a 40  $\mu$ L reaction containing 20  $\mu$ L of 2X low-salt RNA binding buffer (100 mM Tris-HCl pH 7.9, 200 mM KCl, 0.2% NP-40, 1.5 mM MgSO<sub>4</sub>) at 4°C for 15 minutes. Reactions were added to 20  $\mu$ L anti-FLAG® M2 Affinity Gel (Sigma) and rocked at 4°C for at least 1 hour and 30 minutes. FLAG beads were washed twice with 1X high-salt RNA binding buffer (50 mM Tris-HCl pH 7.9, 500 mM KCl, 0.1% NP-40, 0.75 mM MgSO<sub>4</sub>), resuspended in 500  $\mu$ L of 1X low-salt RNA binding buffer, and split in half for either RNA extraction or protein analysis. “No protein” reactions containing RNA only were run in parallel to control for pulldown specificity. RNA was extracted with 500  $\mu$ L of TRIzol reagent, following manufacturer’s recommendation but performing an additional extraction with chloroform prior to the isopropanol precipitation. The RNA pellet was added to 20  $\mu$ L of 2X RNA loading dye (95% formamide, 0.02% SDS, 0.00625% bromophenol blue, 0.00625% xylene cyanol, 1mM EDTA), dissolved at 55°C for 10 minutes, denatured at 95°C for 5 minutes and placed on ice immediately prior to loading 10  $\mu$ L to a 5% polyacrylamide urea gel in 1X TBE. RNA was stained with SYBR Gold Nucleic Acid Gel Stain (Invitrogen) in 1X TBE for 10-40 minutes and visualized on a Bio-Rad ChemiDoc imaging system. Proteins were extracted from the FLAG resin by adding 10  $\mu$ L of 2X protein loading buffer (8% 2-Mercaptoethanol, 100 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 200 mM DTT, 0.2% bromophenol blue), boiling for 5 minutes and adjusting the final volume to 20  $\mu$ L with low-salt 1X RNA binding buffer. Half of the recovered proteins (10  $\mu$ L) were loaded to 8% Bis-Tris protein gels and stained with PageBlue. Band intensities were measured with the ImageJ “Analyze Gels” function (Schindelin et al., 2012) and used to calculate RNA pulldown efficiencies, normalizing each RNA sample by total recovered protein.

### **PAR-CLIP**

PAR-CLIP was performed as in (Saldaña-Meyer et al., 2014) with some modifications. Briefly, mESC C59 Halo-wt-CTCF and mESC C59D2 Halo- $\Delta$ RBR<sub>i</sub>-CTCF cells were grown under standard conditions and pulsed with 400 mM 4-SU (Sigma) for 2 h. After washing the plates with PBS, cells were cross-linked with 400 mJ/cm<sup>2</sup> UVA (312 nm) using a Stratalinker UV cross-linker (Stratagene). Whole nuclear lysates (WNLs) were obtained by fractionation and nuclei were then incubated for 10 min at 37°C in an appropriate volume of CLIP buffer (20 mM HEPES at pH 7.4, 5 mM EDTA, 150 mM NaCl, 2% EMPIGEN) supplemented with protease inhibitors, 20 U/mL Turbo DNase (Life technologies), and 200 U/mL murine RNase inhibitor (New England Biolabs). After clearing the lysate by centrifugation, immunoprecipitations were carried out using 200  $\mu$ g of WNLs in the same CLIP buffer for 4 h at 4°C and then added protein G-coupled Dynabeads (Life Technologies) for an additional hour. Contaminating DNA was removed by treating the beads with Turbo DNase (2 U in 20 mL). Cross-linked RNA was labeled by successive incubation with 5 U of Antarctic phosphatase (New England Biolabs) and 5 U of T4 PNK (New England Biolabs) in the presence of 10 mCi [g-32P] ATP (PerkinElmer). Labeled material was resolved on 8% Bis-Tris gels, transferred to nitrocellulose membranes, and visualized by autoradiography.

### **ChIP-Seq library preparation**

ChIP-Seq libraries were prepared independently from two ChIP biological replicates using the Solexa rapid library protocol. Briefly, immunoprecipitated DNA or 50 ng of input DNA was end-repaired, phosphorylated and adenylated in a single 50  $\mu$ L reaction containing 31.5  $\mu$ L of DNA, 5  $\mu$ L of spike-in yeast DNA from MNase treated nucleosomes (10 ng/ml) (Skene and Henikoff, 2017) and 13.5  $\mu$ L of end-repair/3' A mix. Reactions were incubated in a thermal cycler for 15' at 12°C, 15' at 37°C, 20' at 72°C, and held at 4°C.

End-repair/3' A mix component	Final concentration	Cat #
10X T4 DNA ligase buffer	1X	NEB #B0202S
10 mM dNTPs	0.5 mM each	KAPA #KK1017
10 mM ATP	0.25 mM	NEB #P0756S
40% PEG 4000	2.5%	
10 U/ $\mu$ L T4 PNK	0.0025 U/ $\mu$ L	NEB #M0201S
5U/ $\mu$ L T4 DNA polymerase*	0.0025 U/ $\mu$ L	Invitrogen #18005025
5U/ $\mu$ L Taq DNA polymerase**	0.0025 U/ $\mu$ L	Thermo #EP0401

\*diluted 1:20 in 1x T4 DNA ligase buffer  
\*\*diluted 1:20 in 1X standard Taq buffer (NEB #B9014S)

To reactions were added 4  $\mu$ L of water, 1  $\mu$ L of Illumina TruSeq adapters, 55  $\mu$ L of 2x Rapid DNA ligase buffer (Enzymatics #B101L) and 5  $\mu$ L of DNA ligase (Enzymatics #L6030-HC-L), and incubated for 15' at 20°C. Ligations were cleaned up twice with AMPure XP beads (Agencourt #A63880) diluted 1:2 with 20% PEG, 1.25M NaCl (first cleanup: 38  $\mu$ L; beads eluted with 53  $\mu$ L of 10 mM Tris-HCl pH 8.0, 50  $\mu$ L transferred to a new tube and added of 55  $\mu$ L of beads:PEG solution). Final elution volume was in 22  $\mu$ L of 10 mM Tris-HCl pH 8.0, 20  $\mu$ L of which were transferred to a new tube and amplified by PCR (45" at 98°C; 14 cycles of 15" at 98°C and 10" at 60°C; 1' at 72°C; hold at 4°C).

PCR mix component	Final concentration	Cat #
5X KAPA buffer	1X	KAPA #KK2502
10 mM dNTPs	0.3 mM each	KAPA #KK1017
5 $\mu$ M TruSeq PCR primers	0.5 $\mu$ M	Primer sequence in <a href="#">Table S1</a>
KAPA HS HIFI polymerase	1 U	KAPA #KK2502
Nuclease-free water to 30 $\mu$ L		

PCR reactions were cleaned up once with 38  $\mu$ L of AMPure XP beads diluted 1:2 with 20% PEG, 1.25M NaCl and eluted with 33  $\mu$ L of 10 mM Tris-HCl pH 8.0, 30  $\mu$ L of which were transferred to a new tube. We assessed library quality and fragment size by qPCR and Fragment analyzer, and sequenced 8-12 multiplexed libraries per lane on the Illumina HiSeq4000 sequencing platform (single end-reads, 50 bp long) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (supported by NIH S10 OD018174 Instrumentation Grant).

### ChIP-Seq analysis

Input, IgG, Smc1a and CTCF ChIP-Seq raw reads from wt-CTCF (C59) and  $\Delta$ RBR<sub>1</sub>-CTCF (C59D2) ESCs (16 libraries total) were quality-checked with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and aligned onto the mouse and the yeast genome (mm10 and sacCer3 assembly, respectively) using Bowtie (Langmead et al., 2009), allowing for two mismatches (-n 2) and no multiple alignments (-m 1). We used Samtools (Li et al., 2009) version 1.9) to sort and index bowtie output .bam files, remove duplicates from mapped reads (rmdup -s) and merge ChIP-Seq replicates, after assessing a good reproducibility between them (Figure S6). Peaks were called with MACS2 (--nomodel --extsize 300) (Zhang et al., 2008) using input DNA as a control. Overlap between ChIP-Seq peaks across samples were computed through Galaxy (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010), requiring a minimum 1-bp overlap between peak intervals.

For spike-in control normalization, we performed pairwise comparisons (e.g., C59 input versus C59D2 input) and selected the sample with the lowest number of unique yeast alignments (C59D2 input: 94,642 reads versus C59 input: 119,846 reads). We then used this value to compute a scale factor (*sf*) for the other sample (C59 *sf*: 94,642 / 119,846 = 0.79), to be used in the downstream analyses (see below).

To create heatmaps we used deepTools (version 2.4.1) (Ramírez et al., 2016). We first ran bamCoverage (--binSize 50 --extendReads 300 -of bigwig) and normalized read numbers to either 1x sequencing depth (--normalizeTo1x 2150570000) or to the spike-in yeast DNA (--scaleFactor *sf*), obtaining read coverage per 50-bp bins across the whole genome (bigWig files). We then used the bigWig files to compute read numbers across 6 Kb centered on C59 CTCF or Smc1a peak summits as called by MACS2 (computeMatrix reference-point --referencePoint TSS --upstream 3000 --downstream 3000 --missingDataAsZero --sortRegions no). We sorted the output matrices by decreasing C59D2 enrichment, calculated as the total number of reads within a MACS2 called ChIP-Seq peak. Finally, heatmaps were created with the plotHeatmap tool (--averageTypeSummaryPlot mean --colorMap 'Blues' --sortRegions no).

To generate the scatterplots in [Figure S6A](#) we used deepTools multiBigwigSummary (BED-file mode) on the bigWig output files generated by deepTools bamCoverage, and computed the average scores for each of the files in every CTCF or Smc1a peak called by MACS2 in wt-CTCF mESCs on the merged replicates.

Enriched regions were visualized on the mm10 genome with the Integrative Genomics Viewer (IGV) ([Robinson et al., 2011](#); [Thorvaldsdóttir et al., 2013](#)) using the bigWig output files from deepTools bamCoverage.

Previously published data describing CTCF ([Chen et al., 2008](#)) and Smc1a ([Kagey et al., 2010](#)) binding profiles in mESCs were downloaded from GEO, analyzed with the very same pipeline described above and compared to the data generated in this study ([Figures S6C and S6D](#)).

### RNA-Seq library preparation and analysis

RNA-Seq was performed in the parental C59 ES cell line (wt-CTCF) and in its derivative clone C59D2 ( $\Delta$ RBR<sub>1</sub>-CTCF), with two biological replicates each. RNA was extracted with the QIAGEN RNeasy Mini Kit according to manufacturer's instructions, lysing cells directly into 6-well plates with buffer RTL plus. 5  $\mu$ g of the eluted RNA were treated with DNase I in a 25- $\mu$ l reaction at 37°C for 30' (Invitrogen DNA-free DNA Removal Kit). DNA-free RNA was quantified by nanodrop and quality checked by Bioanalyzer and 2.5  $\mu$ g were subjected to ribosomal RNA depletion following Illumina Ribo-zero rRNA Removal Kit's instructions. Precipitated RNA was re-suspended to 17  $\mu$ l End Repair Mix (ERP) from the TruSeq RNA Sample Preparation v2 Kit (Illumina RS-122-2001) and stored overnight at  $-80^{\circ}\text{C}$  until library preparation. RNA fragmentation, first and second strand cDNA synthesis were performed according to the TruSeq RNA Sample Preparation v2 Kit but using Superscript III for reverse transcription instead of Superscript II (50°C for 50' incubation time). cDNA was purified with AMPure XP beads diluted 1:2 with 20% PEG, 1.25M NaCl, and eluted in 38.5  $\mu$ l 10 mM Tris-HCl pH 8.0, 36.5  $\mu$ l of which were transferred to a new tube and subjected to the ChIP-Seq Solexa rapid library protocol described above.

RNA-Seq raw reads from wt-CTCF (C59) and  $\Delta$ RBR<sub>1</sub>-CTCF (C59D2) ESCs (4 libraries total) were quality-checked with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and aligned onto the mouse genome (mm10) using STAR RNA-Seq aligner ([Dobin et al., 2013](#)) with the following options:--outSJfilterReads Unique --outFilterMultimapNmax 1 --outFilterIntronMotifs RemoveNoncanonical --outSAMstrandField intronMotif. We used Samtools ([Li et al., 2009](#)) version 1.9) to convert STAR output .sam files into .bam files, and to sort and index them. We then counted how many reads overlapped an annotated gene (GENECODE vM19 annotations) using HTSeq ([Anders et al., 2015](#)) (htseq-count --stranded no -f bam --additional-attr gene\_name -m union), and used the output counts files to find differentially expressed genes with edgeR ([Liu et al., 2015](#); [Robinson et al., 2010](#)) and DESeq2 ([Love et al., 2014](#)), both run within the Galaxy platform. We used the following edgeR parameters: genes with  $\leq 0.5$  counts per million (CPM) in at least 3 samples were filtered out (38956 out of 54445); TMM was the method used to normalize library sizes; the edgeR quasi-likelihood test was used with robust settings (robust = TRUE with estimateDisp and glmQLFit). DESeq2 was run with Galaxy default parameters. 496 genes were called differentially expressed in  $\Delta$ RBR<sub>1</sub>-CTCF ESCs compared to wt-CTCF cells by both edgeR (false discovery rate  $\leq 0.05$ ) and DESeq2 (adjusted P value  $\leq 0.05$ ), 275 of which being upregulated and 221 being downregulated. 5 groups of  $\sim 500$  genes each were randomly sampled from the unchanged genes with  $> 0.5$  CPM in at least 3 samples as controls for downstream analyses. Gene ontology analysis was performed with DAVID 6.8 Functional Annotation Tool ([Huang et al., 2009a, 2009b](#)). Gene transcript levels were visualized on the mm10 genome with the Integrative Genomics Viewer (IGV) ([Robinson et al., 2011](#); [Thorvaldsdóttir et al., 2013](#)) using the bigWig output files from deepTools bamCoverage (--binSize 50 --extendReads 250 --normalizeTo1x 2150570000 -of bigwig).

Previously published RNA-Seq data measuring transcription changes in mESCs 1, 2 and 4 days after CTCF degradation ([Nora et al., 2017](#)) were downloaded from GEO, analyzed with the very same pipeline described above and compared to the data generated in this study ([Figure S7K](#); [Table S2](#)). Of note, when using our stringent differential expression analysis we found 76, 262 and 3039 deregulated genes at day 1, 2 and 4 after CTCF degradation, respectively, which are significantly fewer than those reported by Nora and coworkers (370 differentially expressed genes 1 day after CTCF depletion, 1353 after 2 days and 4996 after 4 days). We might thus be underestimating the overlap between the RNA-seq data generated in this study and the one reported by Nora et al.

### Micro-C

Mammalian Micro-C protocol and analysis were modified from ([Hsieh et al., 2016](#)). Here, we briefly summarize the key concepts of Micro-C experiment and data analysis. The detailed step-by-step protocol can be found in Supplemental Protocol.

#### I. Prepare crosslinked chromatin from cell culture

One to five million of trypsinized mouse embryonic stem cells were directly resuspended and crosslinked with freshly made 1% formaldehyde at room temperature for 10 minutes. Crosslinking reaction was quenched by adding Tris buffer (pH = 7.5) to final 0.75M at room temperature. Crosslinked cells were washed twice by 1x PBS and subjected to the second crosslinking with 3mM DSG crosslinking solution for 45 minutes at room temperature. Cells were snap-frozen and can be stored at  $-80^{\circ}\text{C}$  up to a year.

#### II. Digest crosslinked chromatin by micrococcal nuclease



Crosslinked cells were permeabilized in ice-cold Micro-C Buffer #1 (50mM NaCl, 10mM Tris-HCl pH = 7.5, 5mM MgCl<sub>2</sub>, 1M CaCl<sub>2</sub>, 0.2% NP-40, 1x Protease Inhibitor Cocktail) for 20 minutes. Chromatin from permeabilized cells was digested by pre-titrated concentration of micrococcal nuclease to about 90% of mononucleosomes and 10% dinucleosome at 37°C for 10 minutes. Digestion reaction was stopped by adding EGTA to a final concentration at 4mM and incubated at 65°C for 10 minutes to completely deactivate enzyme activity. MNase-digested chromatin was washed twice with ice-cold Micro-C Buffer #2 (50mM NaCl, 10mM Tris-HCl pH = 7.5, 10mM MgCl<sub>2</sub>).

### III. Repair fragment ends

Digested chromatin fragments were then subjected to dephosphorylation, phosphorylation, end-chewing reactions by T4 Polynucleotide Kinase and DNA Polymerase I Klenow Fragment in Micro-C end-repair buffers (50mM NaCl, 10mM Tris-HCl pH = 7.5, 10mM MgCl<sub>2</sub>, 100ug/mL BSA, 2mM ATP, 5mM DTT, no dNTPs) at 37°C for 30 minutes. Blunt-end reaction was triggered by adding biotin-dATP, biotin-dCTP, dGTP, and dTTP to a final concentration at 66mM and incubated at 25°C for 45 minutes. Fill-in reaction was stopped by 30mM EDTA in 65°C for 20 minutes. Chromatin was washed once with ice-cold Micro-C Buffer #3 (50mM Tris-HCl pH = 7.5, 10mM MgCl<sub>2</sub>).

### IV. Proximity ligation and purge biotin-dNTP from unligated ends

Chromatin fragments with biotin-dNTPs were then ligated by T4 DNA ligase at room temperature for at least 2 hours. Unligated ends containing biotin-dNTPs were then removed by 5' to 3' exonuclease III in 37°C for at least 15 minutes. Chromatin was subjected to reverse crosslinking and protein digestion in proteinase K buffer (2mg/mL Proteinase K, 1% SDS, 1x TE buffer) in 65°C overnight.

### V. Purify dinucleosomal DNA

DNA from Micro-C sample was extracted by Phenol:Chloroform:Isoamyl Alcohol (25:24:1) solution and ethanol precipitation method. DNA was cleaned-up by Zymo DNA clean and concentration column prior to gel size-selection for dinucleosomal DNA at 200 to 400 bp. DNA was then extracted from agarose gel by Zymo Gel purification column.

### VI. Micro-C library preparation for deep sequencing

Purified DNA with biotin-dNTPs was captured by Dynabeads® MyOne Streptavidin C1. Standard library preparation protocol including end-repair, A-tailing, and adaptor ligation was performed on beads with individual enzymes purchased from Lucigen or NEBnext Ultral kit. Sequencing library was amplified by Kapa HiFi PCR enzyme with lowest possible cycles to reduce PCR duplicates. Sequencing library was sequenced paired-end 50x50 or 100x100 in Illumina HiSeq 4000 sequencer.

### VII. Micro-C data analysis

#### a. Mapping and pairing Micro-C contacts.

Micro-C data were preprocessed by a streamlined pipeline HiC-pro (Servant et al., 2015). Briefly, sequencing reads were mapped to mouse mm10 genome by Bowtie2 in -very-sensitive-local mode. Mapped reads were paired and pairs with multiple hits, low MAPQ, self-circle, and PCR duplicates were removed. Output file containing all valid pairs were used for following analysis.

#### b. Visualize Micro-C data.

Micro-C data was converted to standard 4DN formats (e.g., .cool or .hic file) with multiple resolutions, typically ranging from 500bp to 10Mb. Cool file can be visualized on Hiclass browser (<http://hiclass.io/>) (Kerpedjiev et al., 2018) and hic file can be visualized on Juicebox (<https://github.com/aidenlab/Juicebox>). All files can be found at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123636>. In this study, all snapshots of contact matrices were generated by Hiclass browser.

#### c. Binning and balancing of Micro-C data.

Valid pairs were binned to 500bp resolution by using cooler packages (<https://github.com/mirnylab/cooler>) for cool file or juicer packages (<https://github.com/aidenlab/juicer>) for hic files (Durand et al., 2016). Low mappability or noisy regions were precluded prior to matrix balancing. Matrix was balanced by using iterative correction (IC) for cool file or Knight-Ruiz (KR) for hic file. Visually, both normalization methods generate equal quality of contact maps. Multiple resolutions of contact maps can be generated by using matrix coarsen or zoomify functions in cooler package.

#### d. Contact probability analysis

Only intra-chromosomal contacts were used to calculate contact density in bins with exponentially increasing widths from 100bp to 10Mb. Contacts shorter than 100bp were removed to minimize noises introduced by self-ligation or unligated products. The numbers of intra-chromosomal contacts within the range of distance were calculated and normalized by the all contacts within this range. Plot shown in Figure 3D only included pairs in "UNI" direction to minimize noise from unligated products.

#### e. Compartment analysis

Chromosomal compartments were identified using principal component analysis (PCA) on contact maps in 100kb resolution. The first component typically represents the compartment profile in mammalian genome – positive eigenvector value enriches with A compartment (gene-rich regions) and negative eigenvalue enriches with B compartment (gene-poor regions). The rank of compartment strength shown in saddle plot was analyzed by rearrangement and aggregation of genome-wide distance-normalized contact matrix in the order of increasing eigenvector values.

#### f. Domain/boundary analysis

We used two approaches to identify TADs and TAD boundaries/insulators. The detailed methods were described in (Crane et al., 2015) for the insulation score analysis or (Rao et al., 2014) for the arrowhead transformation analysis. Briefly, the optimal condition for calling insulation score was determined by testing multiple sizes of sliding window on 10kb - 100kb resolutions of contact maps. A sliding window at 200kb on 20kb-binned contact maps was used to analyze insulation score in this study. The signal within the sliding window was assigned to the corresponding bin across the entire genome. The insulation score was normalized by calculating the log<sub>2</sub> ratio of individual score and the mean of genome-wide averaged insulation score. TAD boundaries/insulators can be identified by calling the local minima along the normalized insulation score. The arrowhead analysis defined as  $A_{i,i+d} = (M_{i,i-d}^* - M_{i,i+d}^*) / (M_{i,i-d}^* + M_{i,i+d}^*)$ .  $A_{i,i+d}$  can be thought of as the measurement of the directionality preference of locus  $i$ , restricted to contacts at a linear distance of  $d$ .  $A_{i,i+d}$  will be strong positive / negative if either one of  $i,i-d$  or  $i,i+d$  is inside the domain and one another is not, but  $A_{i,i+d}$  will be close to zero if both loci are inside or outside the domain. Assigning this query across the genome, the edges of domain will be sharpened and TADs can be detected. For aggregate domain analysis, the individual TADs were rescaled to the same size as  $TAD_{i,j} = ((C_j - TAD_{start}) / (TAD_{end} - TAD_{start}), (C_j - TAD_{end}) / (TAD_{end} - TAD_{start}))$ , where  $C_{i,j}$  is a pair of contact loci within a TAD. The rescale matrices were then aggregated at the center of plot with either ICE or distance normalization.

#### g. Loop analysis

Loops in mouse ES cells were discovered by HiCCUPS as described in (Rao et al., 2014) HiCCUPS uses a modified Benjamini-Hochberg FDR control procedure to reduce the rate of false positive and identify highly reliable loop annotations on contact map. Loops were called on multiple resolutions (1, 5, and 10kb) of KR-normalized Micro-C contact matrices with a false discovery rate smaller than 0.1. Peak widths and windows of peak-to-merging were set as 5kb for 1kb contact maps and 20kb for 5kb and 10kb contact maps. Genome-wide loop comparison/quantification was assessed by using aggregate peak analysis. All called loops were compiled on a center of 25kb x 25kb matrix with 1kb resolution of KR-normalized data. Loops within 55kb of diagonal were excluded to avoid distance decay effects. The ratio of loop enrichment was calculated by dividing observed contact in a searching window by the expected bottom-left submatrix.

#### h. Pile-up analysis

The concept of pile-up analysis is similar to aggregate domain analysis described above. Briefly, we used a set of ChIP-Seq peaks of interest (e.g., CTCF ChIP-Seq in this study) as baits to extract 600kb x 600kb snippets of contact map from 5kb resolution of Micro-C data, in which the coordination of ChIP peak was centered at the center point of each snippet. The snippets were then piled-up on the center of plot and normalized by the expected matrix.

#### i. Motif analysis

Sequences of loop anchors were extracted for CTCF cognate binding motifs scanning by MEME suit (<http://meme-suite.org/>). We also investigated the sequence enrichment for 20 bp upstream and downstream of CTCF motif.

#### j. Reproducibility analysis

Reproducibility of Micro-C data was measured by four algorithms with different aspects of principles (packages are available at [https://github.com/kundajelab/3DChromatin\\_ReplicateQC](https://github.com/kundajelab/3DChromatin_ReplicateQC)). 1). QuASAR (<https://github.com/bxlab/hifive>): contact matrix was transformed based on correlation matrix of distance-based enrichment. The reproducibility score was calculated by correlation of values in two transformed matrices. 2). GenomeDISCO (<https://github.com/kundajelab/genomedisco>): contact matrix was smoothed by graph diffusion. The matrix smoothing considers the input matrix as a network, in which nodes equal to the genomic bins and edges are weighted by contact maps. The reproducibility score was calculated by the difference in two smoothed matrices. 3). HiC-Rep (<https://github.com/qunhualilab/hicrep>): contact matrix was transformed by 2D mean filter. The reproducibility score was measured by the weighted sum of correlation coefficients. 4). HiC-Spector (<https://github.com/gersteinlab/HiC-spector>): contact map was transformed by eigenvalues of Laplace operator. The reproducibility score was analyzed by the difference of weighted eigenvectors. The detailed principle of algorithm can be found in the provided links.

## QUANTIFICATION AND STATISTICAL ANALYSIS

For information about the number of replicates, the meaning of error bars (e.g., standard error of the mean) and other relevant statistical considerations, please see the figure legend associated with the figure showing the data. For information about how data was analyzed and/or quantified, please see the relevant section in METHOD DETAILS and/or the figure legend. And for the code and algorithms used in the analysis, please see the “Software and Algorithms” section of the Key Resources Table.

## DATA AND CODE AVAILABILITY

### Data and Code Availability Statement

The datasets and code generated during this study are available as detailed in the Key Resources Table.

### Data availability

All the Micro-C, ChIP-Seq and RNA-Seq data is available at GEO under accession number GSE123636: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123636>. The imaging data supporting this manuscript and (Hansen et al., 2018b) is available at <https://zenodo.org/record/2208323>. The raw uncropped gels, blots and confocal micrographs can be found at Mendeley Data: DOI: <https://doi.org/10.17632/5zdrpcsbt9.2>

### Computer code

The code used for analyzing and processing the PALM data can be found at [https://gitlab.com/anders.sejr.hansen/palm\\_pipeline](https://gitlab.com/anders.sejr.hansen/palm_pipeline). Please see Key Resources Table for a full list of all the codes and softwares and where to find them.

**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

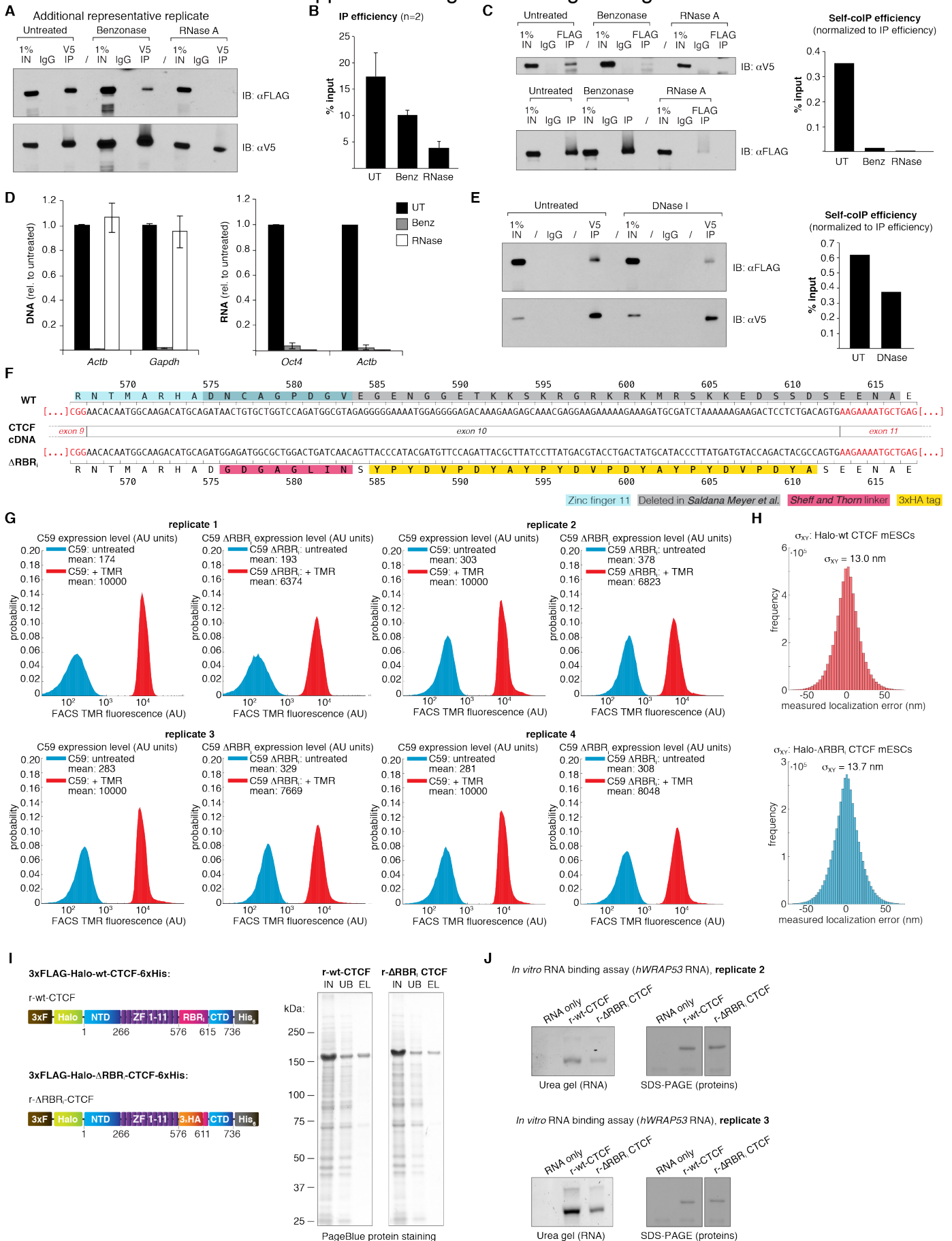
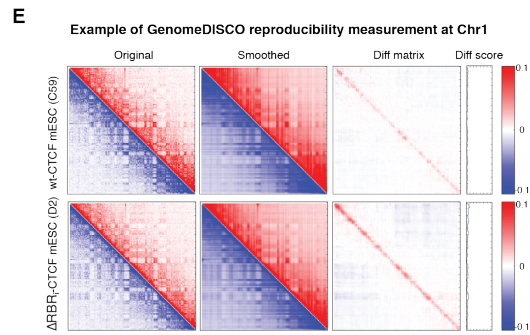
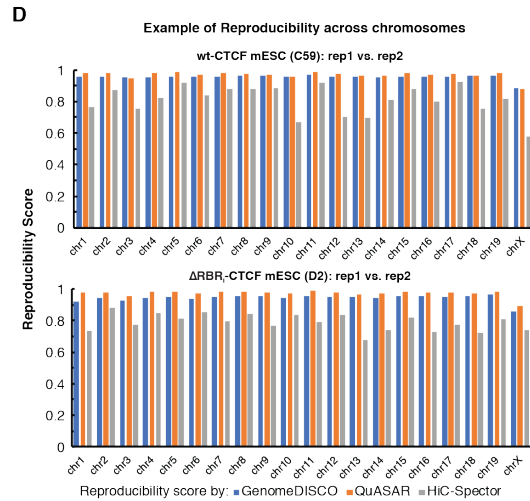
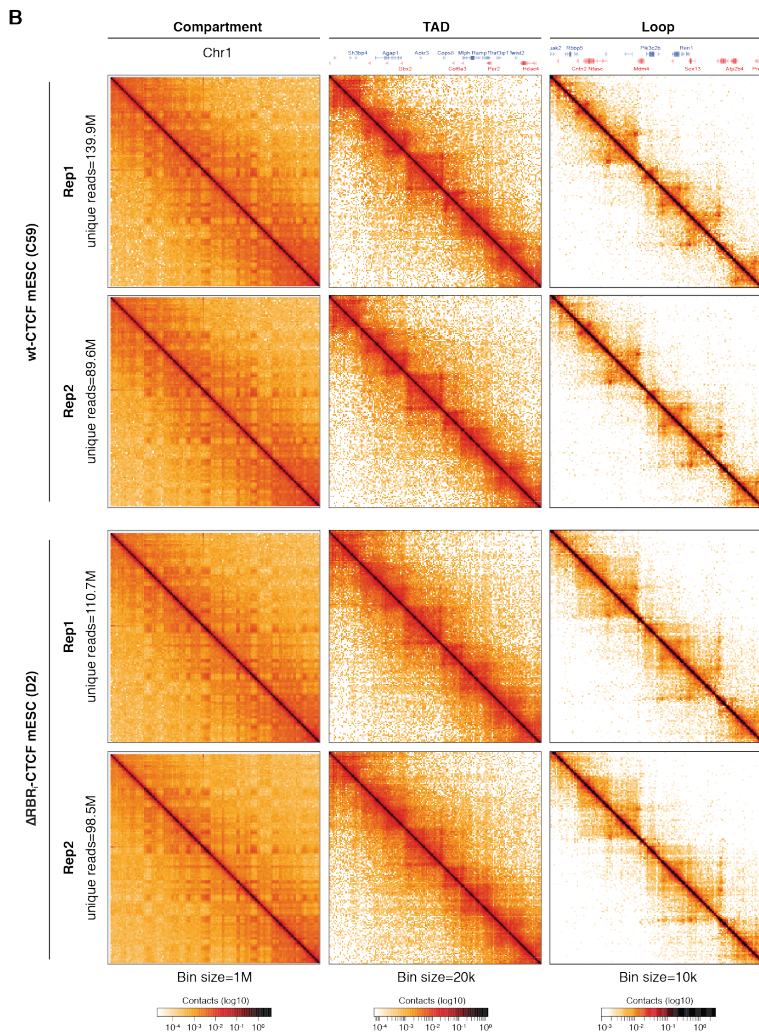
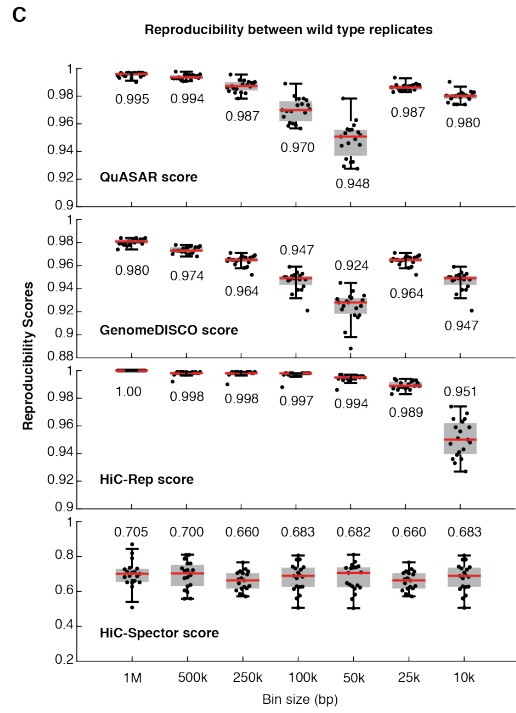
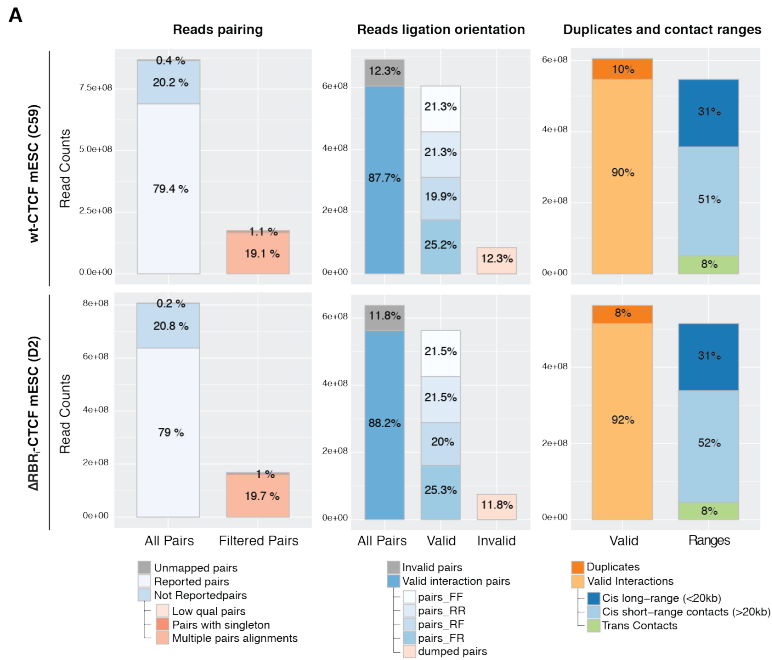


Figure S1. Related to Figure 1 and 2.

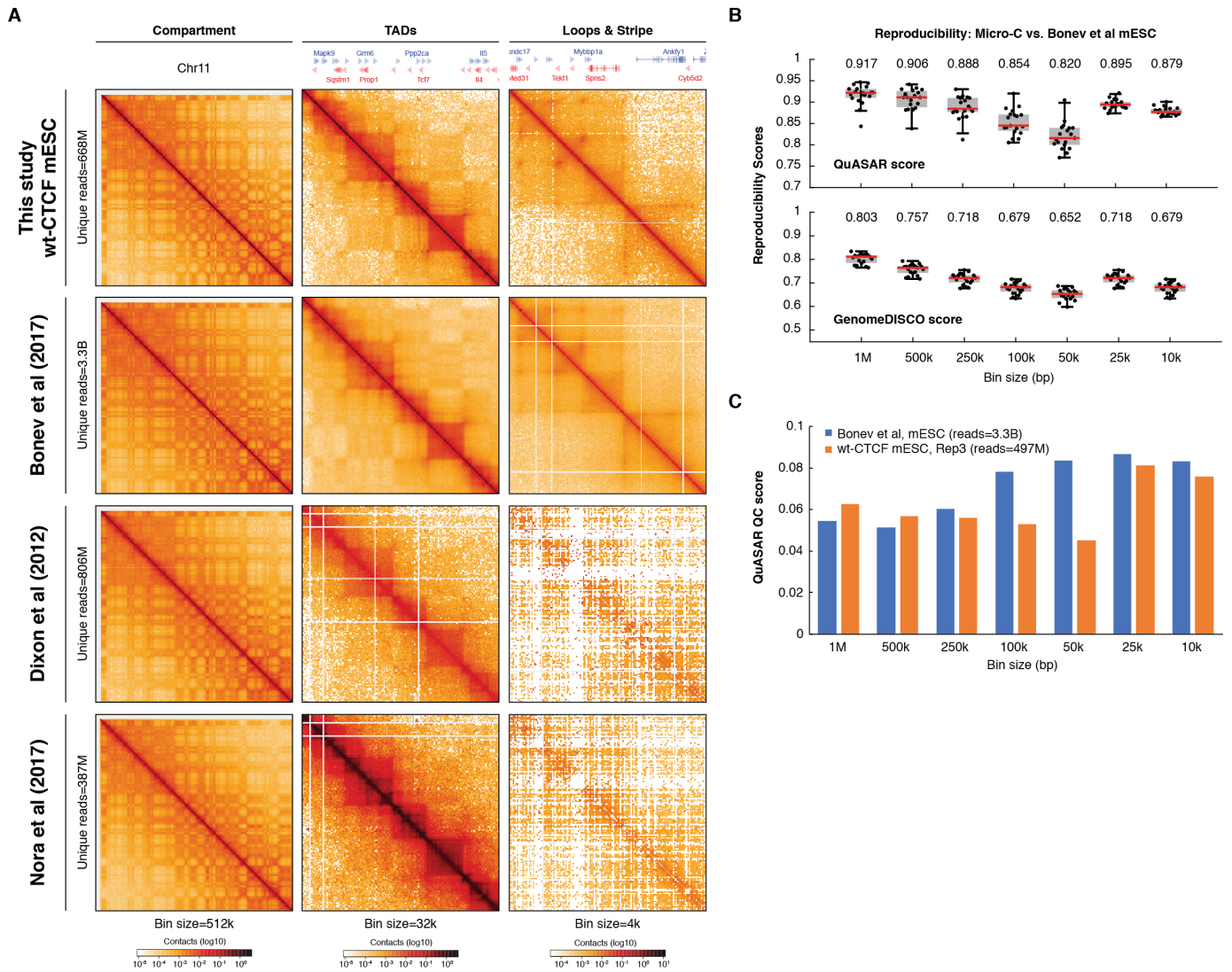
**Additional colP experiments and controls.** (A) Representative colP experiment (V5 IP) indicating RNA-dependent CTCF self-interaction (additional replicate). Top: V5 IP followed by FLAG immunoblotting measures self-colP efficiency (90% of total IP material loaded); bottom: V5 IP followed by V5 immunoblotting controls for IP efficiency (remaining 10% of IP sample loaded). (B) IP efficiency for the V5 colP experiments of Figure 1C and S1A, used to calculate the self-colP efficiency of Figure 1D. Error bars are SD, n=2. (C) Reciprocal colP experiment (left) and quantification (right). Top left: FLAG IP followed by V5 immunoblotting measures self-colP 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 colP experiments. DNA (left) and RNA (right) were extracted from colP 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) ColP experiment with DNase I treatment (left) and quantification (right). Top left: V5 IP followed by FLAG immunoblotting measures self-colP 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  $\Delta$ RBR<sub>i</sub>-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- $\Delta$ RBR<sub>i</sub>-CTCF expression levels in the mESC clones C59 and C59  $\Delta$ RBR<sub>i</sub>, 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- $\Delta$ RBR<sub>i</sub>-CTCF-6xHis (r- $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-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).



## Figure S2. Related to Figure 3

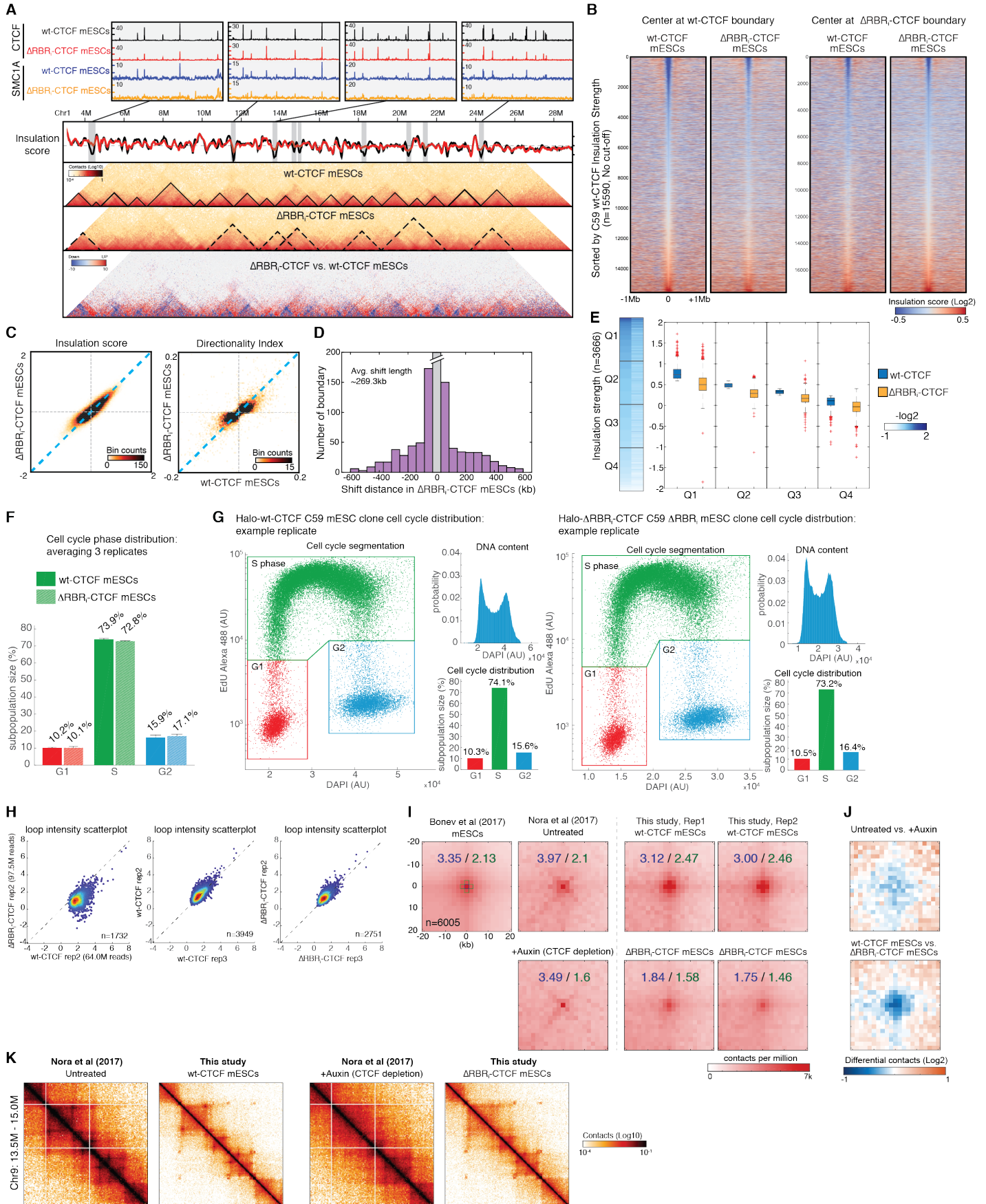
**Micro-C mapping summary and reproducibility.** (A) Statistics of Micro-C assays in wt-CTCF and  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. About 80% of reads were successfully aligned and paired. Interaction orientations are equally distributed in four directionalities: forward-forward, reverse-reverse, 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**

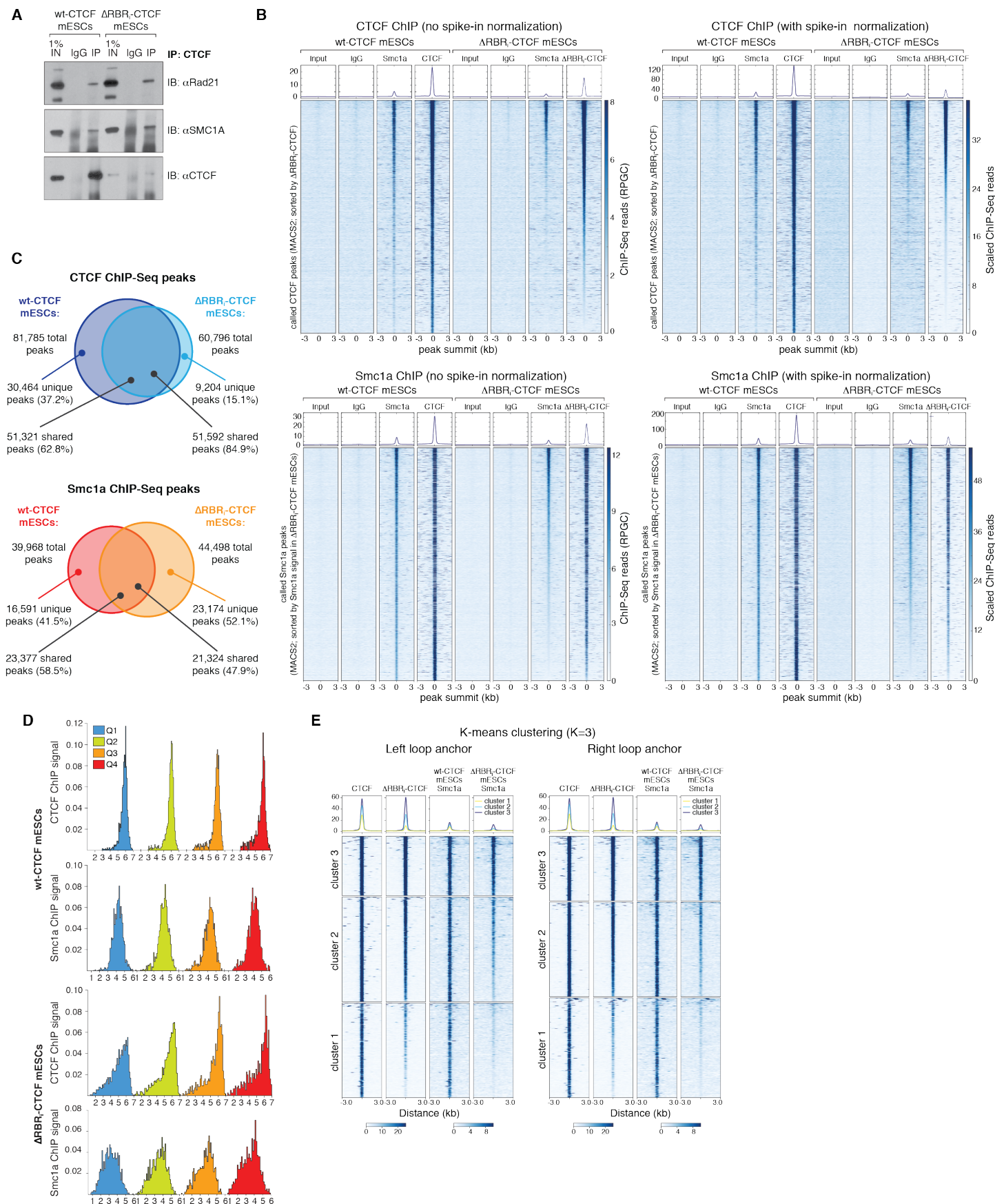
**Comparison of Micro-C and published Hi-C datasets.** (A) Snapshots of chromatin organization in wt-CTCF mESC mapped by Micro-C, Bonev et al (2017), Dixon et al (2012), and Nora et al (2017). Micro-C and Hi-C successfully identify large-scale chromatin structures like compartments and TADs with similar 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.



## 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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs in all the cases shown here. Bottom panel shows a snapshot for insulation score, contact maps for wt-CTCF and  $\Delta$ RBR<sub>i</sub>-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.  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. Gain of contacts between domains in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs also supports this observation. **(B)** Heatmaps of insulation strength centered at the called insulators (n=15590) in wt-CTCF or  $\Delta$ RBR<sub>i</sub>-CTCF mESCs flanked by  $\pm$ 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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs when centering at  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. **(D)** Histogram of shift distance of  $\Delta$ RBR<sub>i</sub>-CTCF insulators. The insulators unique to  $\Delta$ RBR<sub>i</sub>-CTCF mESCs were taken to calculate one's distance to the closest insulator in wt-CTCF mESCs. The majority of insulators unique to  $\Delta$ RBR<sub>i</sub>-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 (-log<sub>2</sub>) 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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. **(F)** Distribution of cell cycle phases in wt-CTCF and  $\Delta$ RBR<sub>i</sub>-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 G<sub>1</sub>, S and G<sub>2</sub> 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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESC nevertheless robustly recapitulate the finding in Figure 5A. Also, loop intensities within the same cell types are highly correlated (Spearman's  $\rho$  ~0.7), while the correlation coefficient value drops to ~0.3 between wt-CTCF and  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-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 log<sub>2</sub> change between untreated vs. +auxin (top) or wt-CTCF vs.  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs (right panel). This further confirms our genome-wide analysis that both CTCF-depletion and  $\Delta$ RBR<sub>i</sub>-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<sub>i</sub>. CTCF antibodies can pull down Rad21 and Smc1a cohesin subunits in both wt- and  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF (top) and Smc1a in  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-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  $\pm$  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  $\Delta$ RBR<sub>i</sub>-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<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs.

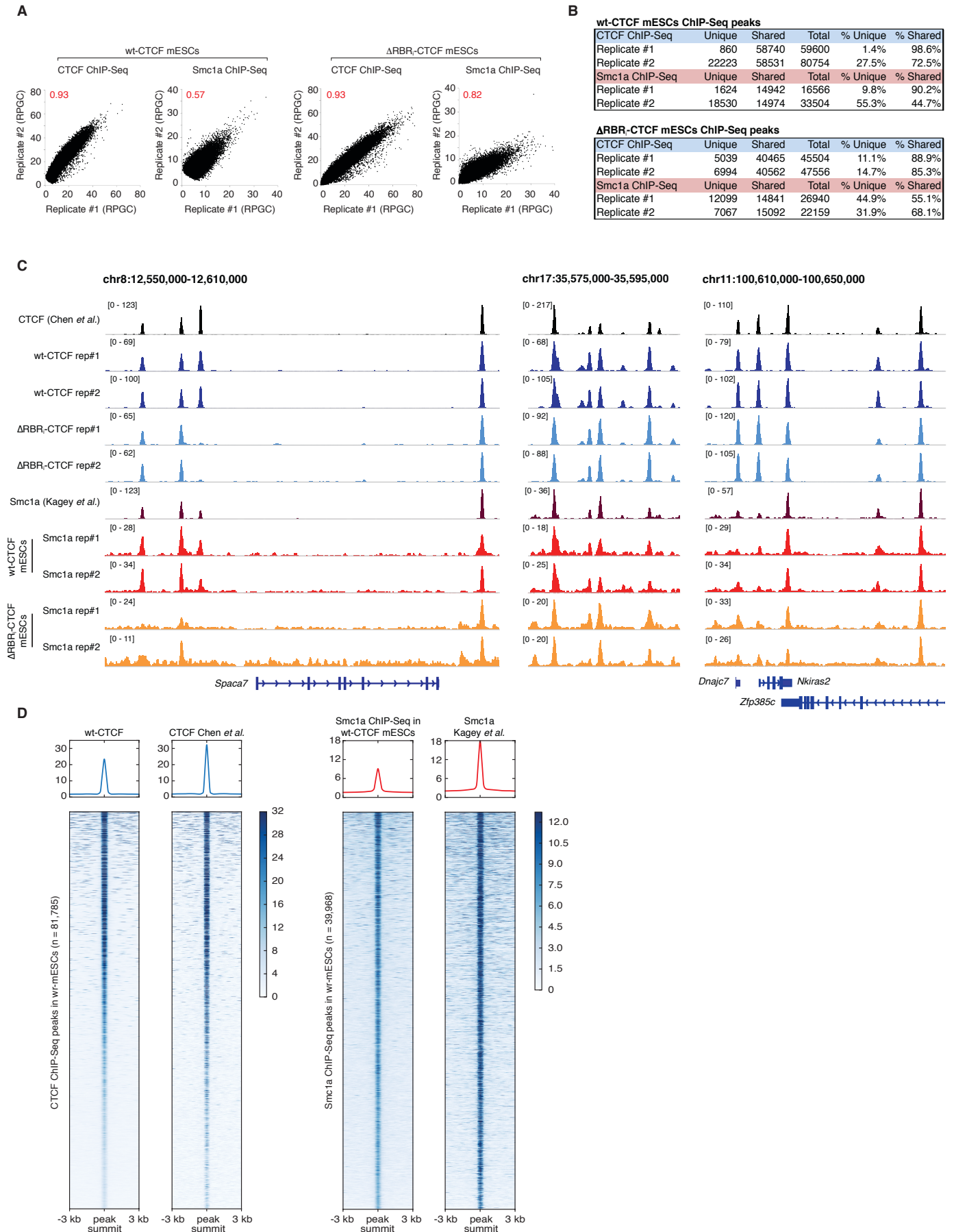
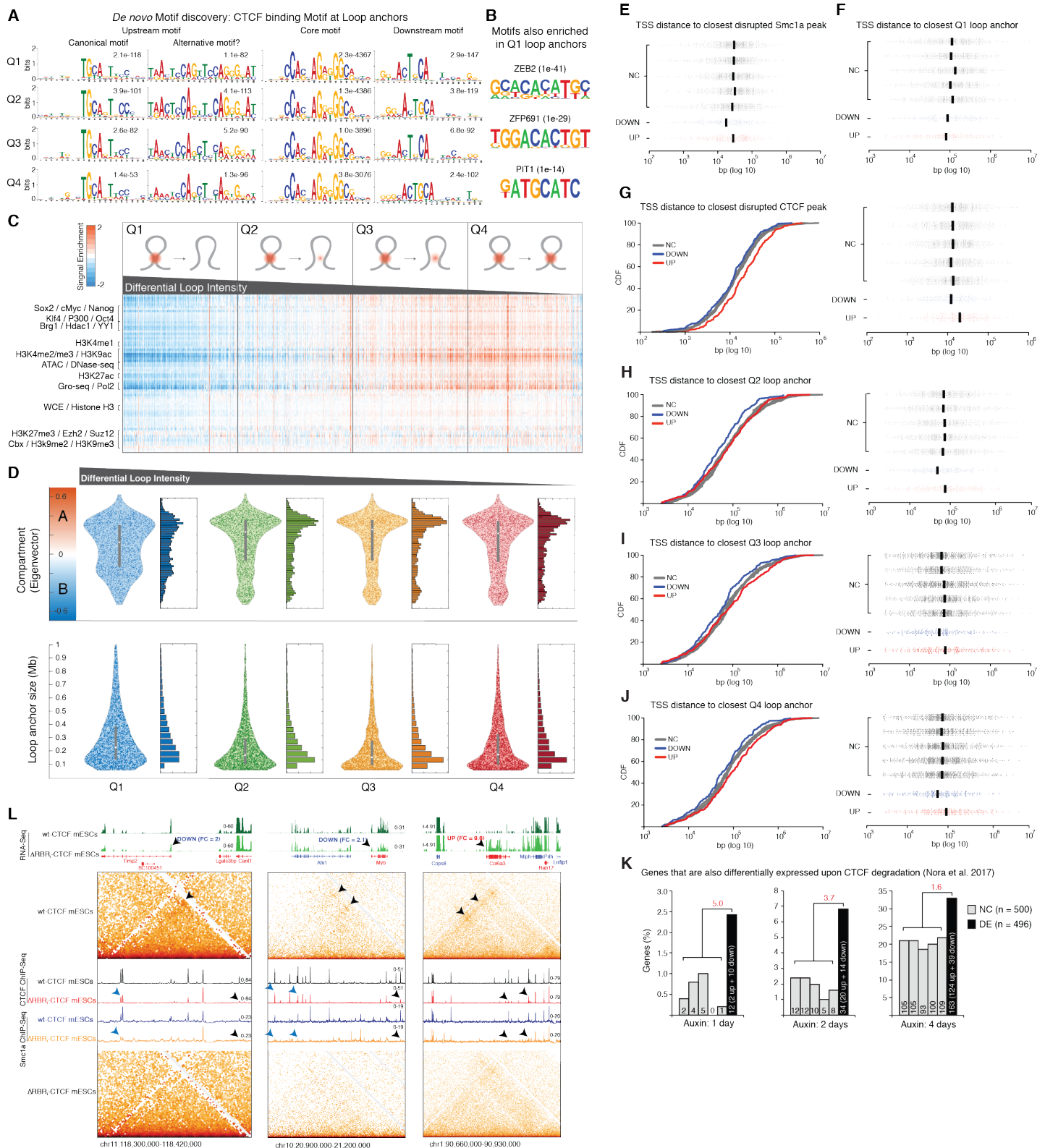


Figure S6. Related to Figure 4-6.

**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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. As can be 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  $\pm$  3 kb around peak summits.



**Figure S7. Related to Figure 6 and 7.**

**DNA determinants of RBR<sub>1</sub>-dependent and RBR<sub>1</sub>-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  $\pm 20$  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<sub>i</sub> domain. On the contrary, Q1 loop anchors, the least affected by RBR<sub>i</sub> 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  $\Delta$ RBR<sub>i</sub>-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<sub>1</sub> represents compartment A (active chromatin) and negative EV<sub>2</sub> 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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs compared to wt-CTCF mESCs. **(G)** Left: for each gene deregulated (DOWN or UP) in  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs compared to wt-CTCF mESCs. Snapshots of three genomic regions showing two genes (*Timp2*, *Myb*) downregulated and one gene (*Col6a3*) upregulated in  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-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  $\Delta$ RBR<sub>i</sub>-CTCF mESCs, respectively. Arrowheads highlight loops disrupted following CTCF RBR<sub>i</sub> deletion. CTCF and cohesin (Smc1a) ChIP-Seq data is overlaid, with blue/black arrowheads pointing at the disrupted left/right loop anchors in  $\Delta$ RBR<sub>i</sub>-CTCF mESCs. ChIP-Seq and RNA-Seq scale is reads per genomic content (deepTools RPGC).