Context-dependent 3D genome regulation by cohesin and related factors

Ryuichiro Nakato[1,§,*], Toyonori Sakata[2,3,§], Jiankang Wang[1], Luis Augusto Eijy Nagai[1], Gina Miku Oba[1], Masashige Bando[2] and Katsuhiko Shirahige[2,3,*]

[1]Laboratory of Computational Genomics, Institute for Quantitative Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-Ku, Tokyo 113-0032, Japan

[2]Laboratory of Genome Structure and Function, Institute for Quantitative Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-Ku, Tokyo 113-0032, Japan

[3]Karolinska Institutet, Department of Biosciences and Nutrition, Biomedicum, Quarter A6, 171 77, Stockholm, Sweden

[§]These authors contributed equally to this work.

*To whom correspondence should be addressed: rnakato@iqb.u-tokyo.ac.jp, kshirahi@iqb.u-tokyo.ac.jp

ABSTRACT

Cohesin plays vital roles in chromatin folding and gene expression regulation, cooperating with such factors as cohesin loaders, unloaders, acetyltransferase, and the insulation factor CTCF. Although various models of regulation have been proposed (e.g., loop extrusion), how cohesin and related factors collectively or individually regulate the hierarchical chromatin structure and gene expression remains unclear. In this study, we have depleted cohesin and related factors and then conducted a comprehensive evaluation of the resulting 3D genome, transcriptome and epigenome data. We observed substantial variation in depletion effects among factors at topologically associating domain (TAD) boundaries and on interTAD interactions, which were partly related to epigenomic status. Gene expression changes were highly correlated with direct cohesin binding and gain of TAD

boundaries than with the loss of boundaries. Our results suggested that cohesin positively regulates gene expression, whereas other mechanisms (e.g., cohesin turnover and acetylation) add to the diversity of this pattern of dysregulation. Moreover, cohesin was broadly enriched in active compartment A, but not in compartment B, which were retained even after CTCF depletion. Our rich dataset and the subsequent data-driven analysis support the context-specific regulation of chromatin folding by cohesin and related factors.

Keywords: Computational genomics, 3D genome, transcriptome, epigenome, cohesin, CTCF, NIPBL

INTRODUCTION

The cohesin complex is crucial for gene transcription and chromatin folding in mammalian cells (Merkenschlager and Nora 2016; van Ruiten and Rowland 2018). Cohesin colocalizes with the CCCTC-binding factor CTCF to function as an insulator (Wendt et al. 2008), whereas a small proportion of cohesin binds the genome independently of CTCF, regulating gene expression with tissue-specific transcription factors (Schmidt et al. 2010; Faure et al. 2012). At least a subset of CTCF-independent cohesin mediates chromatin interactions between enhancer and promoter sites of active genes with mediator complexes (Kagey et al. 2010). Cohesin also participates in transcription elongation machinery that interacts with RNA polymerase II (Pol2) (Izumi et al. 2015). Mutations in the cohesin loader NIPBL (~60%) and in cohesin subunits (~10%) have been found in the human developmental disorder Cornelia de Lange syndrome (CdLS) (Kline et al. 2018).

Recent studies using whole-genome chromatin-conformation capture (Hi-C) uncovered a hierarchical three-dimensional (3D) genome structure regulated by cohesin and its related factors. Chromosomes are spatially segregated into active “compartment A” and inactive “compartment B” (Lieberman-Aiden et al. 2009). At a finer scale, chromosomes are folded into topologically associating domains (TADs), the boundaries of which are strongly enriched for cohesin and CTCF (Dixon et al. 2012). TADs can be nested, and interactions between TADs (interTAD interactions) are more rare than those within TADs (intraTAD interactions) (Bonev and Cavalli 2016). The depletion

of cohesin or NIPBL causes a dramatic loss of TADs and chromatin loops (Rao et al. 2017; Schwarzer et al. 2017), whereas CTCF depletion affects TAD boundaries and loops more locally (Nora et al. 2017). These observations can be explained by the "loop extrusion" model, in which cohesin extrudes chromatin until it encounters CTCF, resulting in the formation of TADs (Fudenberg et al. 2016; Davidson et al. 2019; Kim et al. 2019). This model can also explain depletion effects of cohesin unloading factors (WAPL, PDS5A and PDS5B), which prevent the release of cohesin from DNA and cause loop extension, resulting in the appearance of longer loops than usual (Haarhuis et al. 2017; Wutz et al. 2017). Importantly, such extended loops are rare but also occur in wild-type cells (Allahyar et al. 2018), suggesting that CTCF boundaries are not absolute and the dynamics of TADs/loop formation may depend on the amount of cohesin on chromatin, which is balanced by continuous loading and unloading (turnover).

Despite these extensive efforts, the detailed mechanism of the hierarchical chromosome organization and the functional relationships involved in regulating transcription are still unclear (Sikorska and Sexton 2020). Extensive loss of TADs/loops and loop extension have a limited impact on gene expression and does not cause the spread of histone modifications (Haarhuis et al. 2017; Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017; Ghavi-Helm et al. 2019). A dCas9-mediated insertion of boundary sequence was insufficient for creating TAD boundaries *de novo* (Bonev et al. 2017). Moreover, cohesin and CTCF also localize within TADs without forming boundaries. These results suggest a more complicated set of rules for chromatin structure formation and gene expression regulation by cohesin and related factors than the current models. Although each cohesin-related factor has been studied using different cell lines, a study to explore how cohesin and its related factors collectively or individually regulate chromatin folding, gene expression and the epigenome is needed.

Here we conducted a large-scale *in situ* Hi-C analysis after depletion of a variety of cohesin-related factors, with multiple replicates (31 samples, 14 billion paired-end reads in total). Combined with transcriptome and epigenome marks data, we comprehensively evaluated the similarities and differences in the resulting effects after depletion of individual factors. The resulting extensive dataset

and subsequent analysis provide new insights into the context-specific roles of cohesin-related factors on gene expression and chromatin folding.

RESULTS

Datasets

Here we used human retinal pigment epithelial (RPE) cells to avoid the effect of aneuploidy or other genomic rearrangements (Figure 1A). We depleted cohesin (Rad21), cohesin loaders (NIPBL and Mau2), cohesin unloaders (WAPL, PDS5A and PDS5B), boundary element (CTCF) and cohesin acetyltransferase (ESCO1) and also carried out two sets of co-depletions (Rad21 and NIPBL, PDS5A and B). We confirmed that the depletion efficiencies were sufficient for all samples (Figure 1A) and that the majority of asynchronous cells were in G1 phase (Figure S1A). We used a 72-h treatment with short interfering RNA (siRNA) for most samples, but we also explored the effect of different treatment times (24, 48 and 120 h, Figure 1B). We also generated a sample that had been treated with the BET bromodomain inhibitor JQ1 because the bromodomain protein BRD4 is reported to interact with NIPBL and to be mutated in CdLS (Olley et al. 2018). Using these samples, we prepared *in situ* Hi-C, RNA-seq and spike-in ChIP-seq data (Tables S1–3). In the spike-in ChIP-seq, we observed that 60–80% of the peaks in control cells were lost after siRNA (Figure S1B).

We evaluate the overall similarity among our Hi-C samples and found that the depletion effects can be categorized into four groups that correspond to the siRNA target (Figure 1C): cohesin and loaders, CTCF, cohesin unloaders and acetyltransferase, and control and JQ1. The exception was PDS5A and B co-depletion, which showed less correlation with NIPBL depletion (siNIPBL). Cohesin unloader and acetyltransferase depletion showed a milder effect on chromosome structure as compared with the cohesin loading and localization at CTCF sites. Having confirmed the sufficient similarity among replicates, we merged all replicates into a single deep Hi-C dataset for control,

102 siRad21, siNIPBL (except for the 24-h treatment), siCTCF and siESCO1, resulting in at most 3 billion
103 reads, for further analysis of these depletions.

104

105 **Comparative Hi-C analysis reveals diverse depletion effects on chromatin folding**

106 We first analyzed Hi-C data and compared the depletion effects on TADs and loops. We
107 observed a dramatic loss of TADs and loops after siRad21 and siNIPBL (Figures 1D and 1E),
108 consistent with the previous studies (Rao et al. 2017; Schwarzer et al. 2017). Co-depletion of Rad21
109 and NIPBL showed a more severe effect. Mau2 depletion showed a milder effect than siNIPBL,
110 possibly because some amount of cohesin can be loaded without Mau2 (Haarhuis et al. 2017).
111 Although CTCF depletion strongly affected loops, it had a limited effect on TAD numbers and
112 intraTAD interactions as compared with cohesin depletion (Figures 1D and 1E), indicating the
113 function of CTCF as a boundary element (Nora et al. 2017; Hansen 2020). Most loops in the control
114 samples anchored convergent CTCF motif sites as reported (Vietri Rudan et al. 2015), which was
115 slightly violated after siWAPL, siPDS5AB and siCTCF (Figure S1C).

116 Compartmentalization can be uncoupled from TAD formation, which is strengthened by the
117 depletion of cohesin and loaders (Haarhuis et al. 2017; Rao et al. 2017; Schwarzer et al. 2017) but not
118 of CTCF (Nora et al. 2017). We observed a similar tendency in our data, as indicated by the “plaid
119 pattern” (Figure S2A) and quantitative compartment strength estimated by a saddle plot (Figures 1E
120 and S2B). siMau2 showed stronger compartmentalization than siRad21, in contrast to its milder effect
121 on TADs and loops, suggesting the importance of cohesin loaders for compartmentalization. This
122 tendency was also indicated by the relative contact frequency of mapped reads (Figure 1F). The
123 depletion of cohesin and loaders diminished interactions at a length consistent with TADs (~1 Mb),
124 whereas the long-range interactions corresponding to the compartment (~10 Mb) drastically increased.
125 In contrast, depletion of CTCF, WAPL, PDS5B, PDS5AB or ESCO1 decreased long-range
126 interactions, suggesting weakened compartmentalization. The depletion of PDS5A alone did not show

a clear tendency. Lastly, we did not observe prominent compartment switching among any of the samples (Figure S2C).

We next explored the loop length distribution that showed a distinct tendency from the relative contact probability (Figure 1G). After siRad21, most short loops were depleted, and the distribution peaked at a longer length (~10 Mbp) than did the control (~500 kbp). siCTCF showed a similar but less drastic effect. After siNIPBL and after NIPBL and Rad21 co-depletion, a dramatic loss of short loops was observed, whereas a small number of long-range interactions appeared (> 5 Mbp, possibly due to the cohesin-independent long-range loops (Rao et al. 2017)). In contrast, siMau2 resulted in highly depleted long loops (~1 Mbp), and the distribution then peaked at a shorter length than the control (~400 kbp), which was similar to the effect of siPDS5A and JQ1. Based on the loop extrusion model, it was likely that shorter loops were retained under the mild loss of cohesin after siMau2. After depletion of PDS5B, PDS5A and B or WAPL, the peak distribution increased slightly relative to control samples (~500 kbp), consistent with their function as cohesin unloaders. siESCO1 also caused the appearance of longer loops, similar to the effect of siPDS5B.

Additionally, we investigated the allele-specific depletion effect on chromosome X. Whereas the active chromosome X (Xa) forms the typical chromosome structure, inactive chromosome X (Xi) is partitioned into two megadomains, the boundary between which was affected by depletion of cohesin (Wang et al. 2018; Kriz et al. 2021). Although our data did not show an explicit disruption of the megadomain boundary in Xi, possibly due to incomplete siRNA depletion, we did observe a difference in depletion effects between Xi and Xa (Figure S3). Xi showed a “coarser” plaid pattern than Xa, which was strengthened by siRad21 and siNIPBL. siCTCF showed an asymmetric tendency of interaction frequency between the megadomain boundary and the two megadomains (black arrows), whereas there was no similarly clear chromosome-wide pattern in Xa. In addition, the interaction within the smaller megadomain was less affected in all samples. In summary, our Hi-C analysis showed consistent tendencies with previous studies, confirming its reliability, and provided multiple new findings of diverse depletion effects on chromatin folding.

Gene expression changes were correlated with direct cohesin binding

Next, we explored the depletion effect on gene expression. We detected 2,000–7,000 differentially expressed genes (DEGs) for each sample (false discovery rate [FDR] < 0.01; Figures 2A and S4A). We selected the top-ranked 1,000 DEGs from all samples and merged them into a single DEG list (4,240 genes in total). Pairwise comparisons showed extensive overlap of DEGs between cohesin and loaders and between individual unloaders (Figure 2A). Interestingly, siNIPBL was more similar to siPDS5B than siRad21 and siMau2, suggesting DEGs from dysregulation of cohesin turnover. siCTCF and JQ1 showed less correlation with the others, suggesting their distinct roles for gene expression regulation.

To identify the pattern of expression dysregulation, we applied k-means clustering (k = 20) based on the overlap of up- and downregulated genes among siRNA treatments (Figure 2B and Table S4). For example, clusters 6 and 10 represent down- and upregulated genes after cohesin and loader depletion, respectively. Gene ontology (GO) analysis suggested that cluster 6 was mainly enriched in "growth factor activity," consistent with slower growth under such depletions (Waizenegger et al. 2000). Clusters 9 and 18 contained down- and upregulated genes after NIPBL and unloader depletions, respectively. These DEGs were not observed after siRad21 and therefore would be correlated with cohesin turnover. Their GO terms were correlated with fundamental functions related to the cytoskeleton and extracellular matrix. These diverse expression patterns suggested multiple roles for cohesin-related factors in gene expression regulation.

We next examined the enrichment of ChIP-seq peaks and Hi-C loops at transcription start sites (TSSs) of the DEGs (Figure 2C). Most clusters were enriched for Rad21 and Mau2 peaks, suggesting that expression dysregulation of these clusters was caused by loss of Rad21 and Mau2 binding to each gene, rather than by region-wide effects caused by TAD disruption. We also found that the siRad21 DEGs were less likely to be located around disrupted TADs compared with non-

differential boundaries (Figure S4B), suggesting little correlation between TAD disruption and gene expression dysregulation after siRad21. The exception was downregulated genes after siWAPL (cluster 3 and 9), which were independent of cohesin binding, implying the indirect or unrelated regulation relative to cohesin. In addition, loops mediated by acetylated cohesin (Smc3ac) were enriched in upregulated DEGs associated with siNIPBL and siPDS5B (clusters 1, 7, 18), whereas clusters not enriched for Smc3ac (12, 13, 15) were downregulated. Considering that acetylated cohesin sites are more stable and thus were more persistent even under siNIPBL and siRad21 as compared with non-acetylated sites (Figure 2C), this result suggests the necessity of cohesin at TSSs for gene expression.

Figure 2D shows the ChIP-seq distribution around several top-ranked DEG loci, each of which have cohesin peaks around their TSSs. Remarkably, Rad21 peaks around TSSs were lost after siNIPBL, whereas they remained after siCTCF (red arrows), suggesting that cohesin binds at TSSs in a more CTCF-independent manner. In contrast, Rad21 peaks in other regions were lost after siCTCF (black arrows). We confirmed that this tendency was genome-wide (Figure 2E). siNIPBL significantly depleted cohesin peaks at upstream and exon regions, whereas siCTCF affected intron and intergenic regions. In summary, our result suggested that cohesin positively regulates gene expression via direct binding at TSSs, whereas other mechanisms (e.g., turnover and acetylation) add to the diversity of this pattern of dysregulation.

Quantitative classification of insulation levels reveals diversity among boundaries

To further study the depletion effects on chromatin folding, we calculated a multi-scale insulation score (Crane et al. 2015) (Figures 3A and 3B). We found various patterns of insulation perturbation at TAD boundaries: (i) boundaries weakened by siRad21 and siNIPBL but not by siCTCF (cohesin-dependent), (ii) boundaries strengthened by siNIPBL and siRad21 (cohesin-separated), (iii) boundaries depleted by siRad21, siNIPBL and siCTCF (all-dependent) and (iv)

boundaries that were barely affected by any siRNA (robust). To quantify the observed patterns across the genome, we classified all boundaries into six types based on insulation score (Figure 3C and Table S5). In this classification, over half of the boundaries were annotated "robust," indicating their stability in the presence of reduced amounts of a targeted protein. Most of the rest were classified as weakened boundaries after siRNA. Depletion of unloader proteins did not show an explicit perturbation, which is consistent with their minimal influence on the number of TADs (Figure 1E). Regarding the comparison with compartments, CTCF-dependent and cohesin- separated boundaries occurred more frequently between compartments A and B, whereas cohesin-dependent ones occurred less frequently (Figure 3D). This result suggests that cohesin has a role in connecting neighboring TADs (Schwarzer et al. 2017), especially those from compartments A and B, whereas CTCF is involved in partitions within compartment A and, to a lesser extent, compartment B.

We also investigated the overlap of boundaries and ChIP-seq peaks and DEGs (Figures 3E and S6). While cohesin-dependent and CTCF-dependent boundaries were enriched for loops and CTCF peaks, there were few DEGs there. In contrast, cohesin-separated boundaries significantly overlapped with upregulated DEGs after depletion of cohesin and loaders. Upregulated DEGs associated with unloader siRNA were also enriched, although not significantly. At the boundaries, Mau2 was strikingly enriched, but Rad21, CTCF and loops were not. This result suggested that loss of cohesin loading enhances insulation at the Mau2 peak regions (indicating cohesin loading points), which often occur at A-A and A-B boundaries (Figure 3D), resulting in the dysregulation of expression of some genes. As CTCF-separated boundaries also overlapped with DEGs (although not significantly), a gain of boundaries would be more highly correlated with DEGs than a loss of boundaries. Interestingly, DEGs were not enriched at cohesin-dependent boundaries, even though Pol2 was enriched there. Whereas CTCF-independent boundaries were preferred by active genes (Bonev et al. 2017), the loss of such boundaries may not necessarily cause gene expression dysregulation. These observations highlight the necessity of considering the boundary type when investigating the correlation between chromatin folding and gene expression as regulated by cohesin.

229

230 **Context-specific depletion effects on interTAD interactions**

231 In addition to the six boundary types (Figure 3C), we also found long-range insulation
232 boundaries that appeared after siNIPBL and siRad21 (500 kbp~; Figure 3B, red rectangle). The
233 insulation likely reflected the strong depletion of the interaction between an active TAD (A4, enriched
234 by active markers and Pol2) and an inactive compartment B TAD (B2; Figure 4A, black arrows).
235 Although a decreased interaction between active and inactive regions is compatible with finer
236 compartmentalization (Schwarzer et al. 2017), this depletion effect was more region specific and was
237 not symmetric (e.g., there was a milder effect between B2 and A5; Figure 4A). Moreover, we also
238 observed a difference even between siRad21 and siNIPBL on interTAD interactions (e.g., A3-B2;
239 Figure 4A, black rectangles), despite their closely similar effects on TAD and loop structures. We
240 were therefore interested in the variation in perturbations of interTAD interactions among different
241 siRNA targets.

242 To identify such a strong effect of depletions on interTAD interactions, we calculated the
243 directional relative frequency (DRF) (Wang and Nakato 2021), which evaluates the directional bias of
244 long-range depletion effects (Figure 4B). We scanned the whole genome and identified 241 regions in
245 which DRF values significantly changed after cohesin or loader depletion (Figure 4C and Table S6).
246 Some of them corresponded to a decrease across broad regions (Figure 4C, C1 and C2), while other
247 regions showed a decreased interaction at one side of TADs (Figure 4C, C3 and C4), reminiscent of
248 the “stripe” structure (Vian et al. 2018). While stripes were reported to be located near super-enhancer
249 regions (Vian et al. 2018), the differential DRF regions in our data were often located at the changing
250 points of compartment PC1 value (Figures 4C and 4D). The strong depletion in interTAD interactions
251 is likely to be distinct from the strengthened compartmentalization because PC1 values around the
252 changing points were not perturbed by siNIPBL (Figure 4D). Moreover, these interactions often
253 increased after siRNA of unloaders; therefore, the effect was inversely correlated. The interaction-
254 increased regions were often larger than the detected TADs in the control (e.g., C1 and C2),

suggesting loop extension (Haarhuis et al. 2017; Wutz et al. 2017). Interestingly, siESCO1 showed a smaller but similar effect relative to unloaders in some cases, in which the edge interactions were strengthened (Figure 4C, also supported by the loop length distribution; Figure 1G). We examined this tendency across all 241 regions and confirmed the contrasting depletion effects between cohesin/loaders and unloaders, as well as the positive correlation between ESCO1 and unloaders (except for siPDS5A, Figures 4E and S7B). Considering that ESCO1 facilitates loop stabilization and boundary formation together with CTCF (Wutz et al. 2020), our results suggested that inhibition of cohesin acetylation causes the more frequent pass-through of cohesin at CTCF roadblocks, resulting in loop extension.

Depletion effects of cohesin and loaders were uncoupled for long-range interactions

We next evaluated the correlation between epigenomic features and depletion effects on long-range interactions using a structured interaction matrix analysis (SIMA) (Lin et al. 2012; Seitan et al. 2013) for interactions at a distance of 500 kbp–5 Mbp. Strikingly, there was a context-dependent difference among depletions (Figures 5A and S8A). Whereas interactions between active markers (H3K4me2, H3K4me3, H3K27ac, Med1 and Pol2) increased after both siRad21 and siNIPBL, interactions between the suppressive marker H3K27me3 and the active markers decreased only after siNIPBL. CTCF depletion increased interactions, especially between promoter marks (Pol2, H3K4me2 and H3K4me3). Again, the depletion of cohesin unloaders showed the opposite tendency relative to that of siNIPBL, whereas SIMA also indicated a difference between PDS5A and B. siPDS5A mainly affected interactions across active markers and H3K27me3, whereas siPDS5B affected cohesin (Rad21 and Smc3ac) and CTCF binding sites, similar to the effect of siWAPL. The effect of siPDS5AB was equivalent to the combined effect of siPDS5A and siPDS5B. siESCO1 affected cohesin, CTCF and enhancer markers.

We found that the difference between siRad21 and siNIPBL was mainly derived from global increase of long-range interactions (>2 Mbp) in siRad21, which correspond to interTAD interactions (Figures 5B and S8B). The increase could be involved in strengthened compartmentalization (Rao et al. 2017; Schwarzer et al. 2017), but it cannot explain the difference between siRad21 and siNIPBL. We also found sporadic interactions increased only in siRad21, which might correspond to H3K27ac peaks (Figure 6A). In contrast to our findings in Figure 4, depletion of cohesin unloaders did not show the opposite tendency. Rad21 and NIPBL co-depletion showed a similar effect to siNIPBL alone. This suggested that the amount of cohesin on chromatin has a different effect on long-range interactions relative to the frequency of cohesin loading.

Finally, we tested whether there is a region-wide depletion effect between TADs and the relationship to the epigenome, as was shown in Figure 4A. For this, we annotated all TADs with epigenomic marks and then classified all TAD pairs based on the relative change in interactions between them (Figures 6B–D). The interactions increased after siRad21 but decreased after siNIPBL in cluster 4. The cluster was enriched for interactions between H3K9me3 and active marks (Figure 6D), which included the difference shown in Figure 4A (A3-B2, black rectangle). The interactions decreased after both siRad21 and siNIPBL in clusters 1 and 3, which included interactions among H3K27me3, H3K9me3 and active markers. This tendency was consistent with the depleted long-range interactions (A4-B2; Figure 4A). In summary, depletion of cohesin and a cohesin loader differently affected the long-range interTAD interactions in an epigenomic-dependent manner.

Cohesin is broadly distributed but is more important for TAD formation in the active compartment

To directly analyze the depletion effect on the epigenome, we detected the genomic regions significantly changed after siRNA compared with control (Figures 7A and S9A). H3K27ac marks were largely perturbed after NIPBL and Rad21 depletion, whereas H3K36me3 and H3K9me3 were

not substantially affected. Though H3K27me3 enrichment also increased after siNIPBL (Figure S9A), we did not observe a spread of it caused by the loss of TAD boundaries. Therefore, the H3K27ac perturbation was likely derived from changes in intraTAD or interTAD interactions.

Remarkably, we observed a broad decrease in cohesin after siRad21 and siNIPBL, whereas siCTCF depleted only Rad21 binding at CTCF binding sites (Figure 7A). This indicated that cohesin is located not only in Rad21 peak regions but also in background regions, as assumed in the loop extrusion model (Fudenberg et al. 2016). CTCF acts as an obstacle for cohesin translocation (resulting in sharp cohesin peaks at CTCF sites) but does not control the amount of cohesin on chromatin. This also explains why CTCF depletion decreased loops but had less of an effect on intraTAD interactions (Figures 1D and 1E). However, in compartment B regions, such a substantial depletion of cohesin in the background was not observed (Figure 7A, blue bars). We further investigated this tendency across the genome by dividing compartments A and B into “strong” and “weak” ones based on their PC1 values and confirmed the larger amount of cohesin particularly in “strongA” regions (Figures 7B and 7C). We further investigated the cohesin density using extended ChromHMM (Wang and Nakato 2021) and found that cohesin accumulated to the highest levels at highly active sites (enriched by H3K27ac and H3K4me, Figure S9B). Heterochromatin regions enriched for H3K9me3 in compartment B showed subtle cohesin enrichment. siCTCF did not show such a context-specific tendency (Figures 7C and S9B), and therefore the imbalance in the amount of cohesin between compartments A and B was retained even after loss of CTCF-dependent boundaries, which are often located between A and B (Figure 3D). We also found a milder loss of intraTAD interactions in StrongB than in StrongA, although siCTCF affected both (Figures 7D and 7E). Taken together, our data showed that cohesin also accumulated in non-peak regions, mainly in compartment A, possibly to regulate genes and enhancer activity. In heterochromatic regions, a very small amount of cohesin may be sufficient to maintain TADs, or other systems may maintain TADs such as phase separation (Strom et al. 2017) instead of loop extrusion.

DISCUSSION

Despite multiple promising models (Wendt et al. 2008; Kagey et al. 2010; Schmidt et al. 2010; Izumi et al. 2015; Fudenberg et al. 2016), the cooperative or distinct roles of cohesin in combination with related factors with respect to chromatin folding and gene expression are not fully understood, especially in a context-specific manner. In this study, we found a variety of TAD boundaries and interTAD interactions that should be considered when investigating the functional and mechanistic relationships of cohesin. Whereas several studies have reported genome clustering based on a single Hi-C data—e.g., a third compartment (Yaffe and Tanay 2011) and six subcompartments (Rao et al. 2014)—our analysis focused on the variation of depletion effects and classified genomic regions using multiple Hi-C data. Although it should be noted that cohesin- and CTCF-independent boundaries may be lost after extreme depletion (e.g., by an auxin-inducible degradation system; (Rao et al. 2017)), our data have delineated the dominant factors for boundaries. The perturbation of long-range interTAD interactions observed in this study cannot be captured by a typical analysis that evaluates only the number/strength of TADs and loops.

Most of the cohesin-related DEGs were related to the direct binding of cohesin around TSSs, which was more CTCF-independent. This result is reminiscent of a report using mouse embryonic fibroblast (MEF) cells (Busslinger et al. 2017), in which cohesin accumulated near TSSs of active genes after CTCF knockout. Some DEGs were also enriched near cohesin-separated boundaries, in which Mau2 specifically accumulated. In contrast, the disruption of TAD boundaries was less correlated with DEGs and histone modifications. We also found DEGs that were similarly dysregulated among siNIPBL and depletion of unloaders, possibly due to dysregulation of cohesin turnover. Because the deficiency of cohesin turnover is one of the causes for CdLS (Deardorff et al. 2012), the DEGs could be candidates for CdLS studies. Whereas the BRD4 mutation also causes a CdLS-like syndrome (Olley et al. 2018), there was little effect of JQ1 on chromatin folding despite the many isolated JQ1-related DEGs. Considering the report that BET inhibition does not disrupt enhancer–promoter contact (Crump et al. 2021), the CdLS phenotype might not involve the

356 perturbation of chromatin structure but could be caused by direct transcription regulation by cohesin,
357 e.g., transcription machinery interacting with Pol2 (Izumi et al. 2015; Busslinger et al. 2017).

358 For interTAD interactions, although we also found several genes whose changes in expression
359 were consistent with the depletion effect on the interTAD interaction, the effect was often not region-
360 wide (e.g., RUNX1; Figures S9C and S9D). Region-wide gene expression dysregulation may occur
361 with gene clusters: for example, most *PCDH* genes located on chromosome 5 were detected as DEGs
362 and showed cooperative dysregulation (upregulated with unloader depletion, downregulated with
363 loader depletion; Figure S10). Whether these DEGs are the result of TAD disruption or of changes in
364 interTAD interactions, or if their expression is regulated independently via cohesin binding at TSSs or
365 other gene-specific factors, are important questions for future studies.

366 Despite their closely similar effects on TADs and loops, we also observed a difference in
367 siRad21 and siNIPBL with respect to H3K27me3 and H3K9me3 enrichment. Together with the
368 strengthened compartmentalization under siMau2, the frequency of cohesin loading should be
369 considered separately from the amount of cohesin on chromatin. Meanwhile, the epigenomic state
370 alone cannot fully explain the trend for interTAD interactions, given that the same phenomenon does
371 not occur in all regions that show the same epigenomic pattern.

372 Compared with the cohesin loaders, the depletion of unloader proteins had different effects on
373 transcriptome and chromatin folding, suggesting their nonredundant roles. WAPL and PDS5B showed
374 a more unloader-like effect, whereas siPDS5A was more related to effects on the epigenome (e.g.,
375 Figure 5A). siESCO1 showed a weaker but similar effect relative to that of unloader depletion on
376 chromatin folding, suggesting the involvement of cohesin acetylation in CTCF roadblocks.
377 Considering that DEGs of siESCO1 overlapped to a greater extent with cohesin loader and unloaders
378 than did those associated with siCTCF, the role of cohesin acetylation in gene expression regulation
379 would be mainly related to the cohesin pass-through at CTCF roadblocks in loop extrusion, rather
380 than the formation of stable chromatin loops mediated by acetylated cohesin (Wutz et al. 2020).

381 We demonstrated that cohesin is broadly distributed within compartment A, not only at peak
382 sites of Rad21 or CTCF localization. Because cohesin should simultaneously function to regulate

gene expression and form TADs, a large amount of cohesin could be required to maintain dynamic structures in compartment A. Cohesin did not accumulate substantially in compartment B, in which there was a smaller effect of cohesin depletion. Under the loop-extrusion model, cohesin distribution should have become more uniform after CTCF depletion because of the loss of cohesin stalling at CTCF sites. However, the unequal amounts of cohesin between compartments A and B were retained after siCTCF, whereas we did observe increased interactions between neighboring TADs with the depletion of CTCF-dependent boundaries (e.g., Figure 4A). Therefore, the amount of cohesin on the genome would not be merely derived from loop extrusion but be affected by the genomic context. Moreover, because the interTAD pattern and DEGs of siCTCF were distinct from siRNA of unloaders, the 3D genome segmentation that is retained after siCTCF, and therefore which is likely to be CTCF independent, may be an independent mechanism from the extended TADs observed previously with depletion of unloaders (Haarhuis et al. 2017). How the genomic segmentation within the whole genome is regulated by (extended) loop extrusion and other mechanisms and whether it is conserved among cell types and during cell differentiation remain essential questions for future studies.

METHODS

Cell culture and siRNA

We used the siRNA system for depletion, as the auxin-inducible degradation system reduces protein levels even in the absence of auxin, which is not suitable as a control (Wutz et al. 2020). RPE cells (Wendt et al. 2008) were cultured in DMEM (Wako) supplemented with Penicillin-Streptomycin-L-Glutamine Solution (Wako), 10% fetal bovine serum (Biosera) and 20 mM HEPES-KOH (pH 7.4). All siRNA transfections were performed using Lipofectamine RNAiMAX (Thermo Fisher Scientific) in accordance with the manufacturer's protocol 2 or 3 days before sample preparation, using a final RNA duplex concentration of 50 nM. The siRNA sequences are shown in Supplementary Table S7 and are the same as those described previously (Deardorff et al. 2012; Minamino et al. 2015). For inhibition of BET family proteins, cells were treated with JQ1 for 6 h at a

1 μM final concentration. We labeled JQ1-treated and the corresponding control samples as JQ1(+) and JQ1(-), respectively.

Antibodies

Antibodies used for ChIP and immunoblotting were as follows. Antibodies against histone H3 lysine-27 acetylation (H3K27ac) (Stasevich et al. 2014), H3 lysine-4 trimethylation (H3K4me3), H3 lysine-9 trimethylation (H3K9me3), H3 lysine-36 trimethylation (H3K36me3) and Pol2ser2 were provided by Dr. Kimura (Tokyo Institute of Technology, Tokyo, Japan). We also used antibodies against Rad21, Smc3ac and ESCO1, which were described previously (Minamino et al. 2015). Antibodies against NIPBL (A301-779A, BETHYL), Mau2 (ab46906, abcam), CTCF (07-729, Merck), BRD4 (A301-985A50, BETHYL), AFF4 (A302-538A, BETHYL), Pol2 (14958, Cell signaling technology) and H3 lysine-27 trimethylation (H3K27me3, ab192985, abcam) were used for ChIP. Antibodies against NIPBL (sc-374625, Santa Cruz Biotechnology), a-tubulin (T6074, Merck), WAPL (16370-1-AP, Proteintech), PDS5A (A300-088A, BETHYL), PDS5B (ab70299, abcam) and CTCF (3417, Cell signaling technology) were used for immunoblotting. The mouse monoclonal antibody against Mau2 was generated using a synthetic peptide corresponding to residues 596-613 (PVQFQAQNGPNTSLASLL) of human Mau2 and used for immunoblotting. Antibody dilutions in immunoblotting were 1:500 (NIPBL and ESCO1) and 1:1,000 (other antibodies).

Protein analysis

Cells were lysed with lysis buffer (20 mM HEPES-KOH, pH 7.5; 100 mM NaCl; 10 mM KCl; 10% glycerol; 340 mM sucrose; 1.5 mM $MgCl_2$; 10 mM sodium butyrate; 0.25% Triton X-100; 1 mM dithiothreitol; 1× cOmplete proteinase inhibitor cocktail) as described (Deardorff et al. 2012). The resulting lysate was mixed with SDS-PAGE sample buffer (50 mM Tris-HCl, pH 6.8; 2% SDS; 0.005% BPB; 7% glycerol; 5% 2-mercaptoethanol) and boiled for 5 min. The proteins were analyzed

by Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad) in accordance with the manufacturer's protocol.

In situ Hi-C

We used the *in situ* Hi-C protocol as described in Rao *et al.* (Rao et al. 2014). In brief, ~3 × 10^6 RPE cells were crosslinked with 1% formaldehyde for 10 min at room temperature, followed by an additional 5 min with 200 mM glycine in phosphate-buffered saline (PBS). Fixed cells were permeabilized in Hi-C lysis buffer (10 mM Tris-HCl, pH 8.0; 10 mM NaCl; 0.2% Igepal CA630; 1× protease inhibitor cocktail [Sigma]) on ice. The cells were treated with 100 U of MboI (New England Biolabs) for chromatin digestion, and the ends of digested fragments were labeled with biotinylated nucleotides followed by ligation. After DNA reverse crosslinking and purification, ligated DNA was sheared to a size of 300–500 bp using a Covaris S2 focused-ultrasonicator (settings: Duty Cycle, 10%; Intensity, 4; Cycles per Burst, 200; Duration, 55 sec). The ligated junctions were then pulled down with Dynabeads MyOne Streptavidin T1 beads (Thermo Fisher Scientific). The pulled-down DNA was end-repaired, ligated to sequencing adaptors, amplified on beads and purified using Nextera Mate Pair Sample Preparation Kit (Illumina) and Agencourt AMPure XP (Beckman Coulter). DNA was then sequenced to generate paired-end 150-bp reads using the Illumina HiSeq-2500 or X Ten system.

Hi-C data processing with Juicer

Sequenced reads were processed using Juicer version 1.5.7 and Juicer tools version 1.9.9 (Durand et al. 2016), and the definition of TADs and loops follows Rao *et al.* (Rao et al. 2014). The detailed steps are as follows: Sequenced paired-end reads were mapped by BWA version 0.7.17 (Li and Durbin 2009). We generated contact map files with square root vanilla coverage (VC_SQRT) normalization. We used 25-kbp resolution maps unless otherwise described. We called TADs using Arrowhead. Because the obtained TADs can be nested, we also generated a list of non-overlapping

TADs by segmenting the genome based on all TAD boundaries. TAD boundaries were defined as edges for all annotated TADs. Loops were called at 5-kbp, 10-kbp and 25-kbp resolution by HiCCUPS. To obtain peak-overlapping loops (Figure 2C), we used BEDTools v2.28.0 (https://bedtools.readthedocs.io/en/latest/) and extracted loops for which both anchor sites overlapped with the peaks. CTCF motif analysis was implemented using MotifFinder. High-resolution data that combined all replicates were generated by *mega.sh* script provided by Juicer. Eigenvector (PC1) values for compartment analysis were generated with the *eigenvector* command in Juicertools. For allele-specific Hi-C analysis of chromosome X, we obtained single-nucleotide polymorphism data for RPE cells from Darrow *et al. (Darrow et al. 2016)*, which was then converted to genome build hg38 by the liftOver tool (https://genome-store.ucsc.edu/). We modified the *diploid.sh* script provided by Juicer and made interaction map files for active and inactive chromosome X.

The samples and mapping statistics are summarized in Tables S1. Since we generated six replicates as control samples, we merged them into a single high-resolution Hi-C data and used it to obtain reference data for the TADs, loops and compartment data. For the comparative analysis, we normalized Hi-C matrices based on the number of mapped reads on each chromosome. Therefore, the tendency for increases and decreases is relative; that is, increased long-range interactions might be compensated for by increased short-range interactions (Nora et al. 2020). siRad21 and siNIPBL were most affected by this fact, because almost all TADs and loops were depleted after these treatments. Therefore, in our analysis, we focused on the variation of depletion effects across samples to capture the context-specific tendency, rather than translating the biological meaning of increased/decreased interaction itself.

Hi-C data processing with other tools

To evaluate the quality and reproducibility of our Hi-C data, we used 3DChromatin_ReplicateQC (Yardimci et al. 2019), which internally implements QuASAR (Sauria and Taylor 2017) and HiCRep (Yang et al. 2017). Because of the large computational complexity

485 involved, we used only chromosomes 21 and 22 with a 50-kbp bin for the quality evaluation. We
486 confirmed that all of our Hi-C data had sufficient quality (QuASAR-QC scores > 0.05, Table S1).
487 HiCRep was used to evaluate the overall similarity of the depletion effects among our Hi-C samples
488 by calculating a stratum-adjusted correlation coefficient that captures the similarity of chromatin
489 features including TADs and loops. We used Cooler (Abdennur and Mirny 2020) and cooltools
490 (https://cooltools.readthedocs.io/) for APA plots, averaged TAD plots and saddle plots. The
491 compartment strength was defined as (AA + BB) –2(AB), where AA, BB and AB indicate the
492 interaction frequency between compartments A and A, compartments B and B and compartments A
493 and B, respectively, of the saddle plot. The visualization of Hi-C matrices with ChIP-seq distributions
494 were visualized using Python.

495

496 **Structured interaction matrix analysis**

497 To explore interactions between specific chromatin features (e.g., ChIP-seq peaks), we used
498 structured interaction matrix analysis (SIMA) (Lin et al. 2012) implemented in HOMER
499 (http://homer.ucsd.edu/homer/). SIMA assembles information for multiple occurrences of each
500 feature, providing an overview of Hi-C interactions associated with a genomic feature between each
501 pair of specified domains. In this study, the genomic features included the ChIP-seq peak list (AFF4,
502 CTCF, H3K27ac, H3K27me3, H3K36me3, H3K4me3, H3K4me2, Smc3ac, Mau2, Med1, Pol2,
503 Pol2ser2 and Rad21), differentially expressed genes (DEGs) (siCTCF, siNIPBL, siRad21) and
504 transcriptional start sites (TSSs). Domains of interest were defined as TAD lists, and the distances
505 between two TADs were specified to be <5 Mbp, <2 Mbp or 2–<5 Mbp with '-max -min' parameters.
506 By making comparisons with the background model, we obtained an enrichment score, representing
507 the degree to which a genomic feature pair was enriched in the Hi-C interactions between two TADs.
508 To compare differences in enrichment scores between cohesin-knockdown and control Hi-C samples,
509 we used the paired Wilcoxon signed-rank test to calculate the p-value and effect size for each

genomic feature pair, as described (Seitan et al. 2013). We used Cytoscape (Shannon et al. 2003) to visualize the results.

Multi-scale insulation score

A multi-scale insulation score was generated as described (Crane et al. 2015). In brief, the insulation score was calculated at a resolution of 25 kbp as the log-scaled relative contact frequency across pairs of genomic loci located around the genomic positions from 100 kbp to 1 Mbp. The 500-kbp distance was used in the insulation score analysis. For the classification of insulation boundaries into six types, we used the following criteria based on this 500-kbp insulation score:

1. if(siNIPBL – control) > T_{ins} and if(siCTCF – control) > T_{ins} : "all-dependent"

2. else if(siNIPBL – control) > T_{ins} or if(siRad21 – control) > T_{ins} : "cohesin-dependent"

3. else if(siCTCF – control) > T_{ins} : "CTCF-dependent"

4. else if(control – siCTCF) > T_{ins} : "CTCF-separated"

5. else if(control – siNIPBL) > T_{ins} or if(control – siRad21) > T_{ins} : "cohesin-separated"

6. else if(siNIPBL – control) < T_{ins} and if(siCTCF – control) < T_{ins} and if(siCTCF – control) < T_{ins} : "robust"

where we set T_{ins}, the threshold of insulation score, as 0.13. We excluded siMau2 as a criterion because it had a smaller effect than siNIPBL and siRad21 on the insulation score. We excluded chromosomes X and Y and the mitochondrial genome from this boundary analysis. The obtained six types of boundaries are summarized in Table S5.

Directional relative frequency (DRF)

DRF measures the bias in the relative interaction frequency ($T = log\ (C_{siRNA}) - log\ (C_{control})$) between regions up- and downstream of each genomic region, where C is a normalized contact matrix. Therefore, the DRF can be calculated by

$$DRF_i = \sum_{j=l_{min}}^{l_{max}} T_{i,i+j} - \sum_{j=l_{min}}^{l_{max}} T_{i,i-j},$$

where l_{min} and l_{max} indicate the range of interaction. In this study, we set $l_{min} = 500\ kbp$ and $l_{max} = 2\ Mbp$.

To obtain differential DRF regions, we classified Hi-C samples into “cohesin and loaders,” “cohesin unloaders” and “others (including control)” and calculated the averaged DRF values and a 99% confidence interval (CI). Then we identified the regions that satisfied the following criteria: the 99% CI ranges of “cohesin and loaders” and “others” did not overlap, and the averaged DRF value of “cohesin and loaders” was > T_{DRF} or < -T_{DRF}, where T_{DRF} refers to the threshold of DRF. We set T_{DRF} = 0.7 in this study. The obtained differential DRF regions are summarized in Table S6.

RNA-seq

Total RNA was isolated using Trizol (Thermo Fisher Scientific) and a Nucleospin RNA kit (Macherey-Nagel). rRNA was removed with the Ribo-Zero Gold rRNA Removal Kit (Illumina), followed by sequencing library preparation with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). Single-end 65-bp reads were sequenced by Illumina HiSeq-2500 system. Sequenced reads were mapped to the human reference sequence (GRCh38) by STAR version 2.7.3a (Dobin et al. 2013) with the following options "SortedByCoordinate --quantMode TranscriptomeSAM --outSAMattributes All". The samples and mapping statistics are summarized in Tables S2. The gene expression levels were estimated by RSEM version 1.3.1 (Li and Dewey 2011) with the option "--estimate-rspd --strandedness reverse". We used DESeq2 (Love et al. 2014) to

identify DEGs (protein-coding genes, false discovery rate [FDR] < 0.01). We focused on protein-coding genes to avoid the effects of repetitive non-coding RNAs.

To mitigate the indirect effect and the technical variances, we generated the list of DEGs by merging the top-ranked 1,000 DEGs from each pairwise comparison between each siRNA and the controls. We used clusterProfiler (Yu et al. 2012) for the GO enrichment analysis.

Spike-in ChIP-seq

Spike-in ChIP-seq enables us to explore the absolute-level difference in read enrichment among samples (Bonhoure et al. 2014; Nakato and Shirahige 2017). Chromatin preparation for ChIP was performed as described (Izumi et al. 2015). In brief, ~8×10^6 RPE cells were crosslinked with 1% formaldehyde for 10 min at room temperature, followed by an additional 5 min with glycine in PBS added at a final concentration of 125 mM. Fixed cells were lysed in LB1 (50 mM HEPES-KOH, pH 7.4; 140 mM NaCl; 1 mM EDTA; 10% glycerol; 0.5% NP-40; 0.25% Triton X-100; 10 mM dithiothreitol; 1 mM PMSF) on ice. The lysate was washed sequentially with LB2 (20 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 1 mM PMSF) and LB3 (20 mM Tris-HCl, pH 7.5; 150 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% SDS; 1× cOmplete protease inhibitor cocktail [Roche]) on ice. The lysate was resuspended in LB3 and sonicated using Branson Sonifier 250D (Branson) for chromatin shearing (12 sec with amplitude setting at 17% of the maximum amplitude, six times). In addition, lysate containing fragmented chromatin was also prepared from ~2×10^6 mouse cells (C2C12) with the same procedures. Human cell lysate and mouse cell lysate (as a spike-in internal control) were combined (~4:1 ratio) and incubated with protein A or G Dynabeads (Thermo Fisher Scientific) conjugated with the relevant antibodies for 14 h at 4°C. The beads were then washed five times with cold RIPA wash buffer (50 mM HEPES-KOH, pH 7.4; 500 mM LiCl; 1 mM EDTA; 0.5% sodium deoxycholate; 1% NP-40) and once with cold TE50 (50 mM Tris-HCl, pH 8.0; 10 mM EDTA). Material captured on the beads was

eluted with TE50 containing 1% SDS. The eluted material and input were incubated for 6 h at 65°C to reverse crosslinks and were treated with 100 ng RNaseA (Roche) for 1 h at 50°C, followed by treatment with 100 ng Proteinase K (Merck) overnight at 50°C. The input and ChIP DNA were then purified with a PCR purification kit (Qiagen). DNA from the ChIP and input fractions was end-repaired, ligated to sequencing adaptors, amplified and size-selected using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) and Agencourt AMPure XP (Beckman Coulter). DNA was then sequenced to generate single-end 65-bp reads using the Illumina HiSeq-2500 and NextSeq 2000 systems.

Reads were aligned to the human genome build hg38 and mouse genome build mm10 using Bowtie2 version 2.4.1 (Langmead and Salzberg 2012) with default parameters. Quality assessment was performed with SSP version 1.2.2 (Nakato and Shirahige 2018) and DROMPAplus version 1.12.1 (Nakato and Sakata 2020). Spike-in read normalization, peak calling and visualization were performed with DROMPAplus. The mapping statistics, quality values and the scaling factors for spike-in normalization are summarized in Table S3. The default parameter set was used for peak calling (100-bp bin, --pthre_internal 5, --pthre_enrich 4) except for H3K9me3 (--pthre_internal 1 --pthre_enrich 2) due to the lower signal-to-noise ratio. For read visualization (Figures 2–4 and 6), we displayed $-\log_{10}(p)$ scores of ChIP/input enrichment (--showpenrich 1 option), which is recommended for distinguishing the signal from the noise (Roadmap Epigenomics et al. 2015). For Figure 7, to look for significant changes in ChIP-seq data, we similarly compared ChIP (control) and ChIP (siRNA) and visualized $-\log_{10}(p)$.

Permutation test for overlapping analysis

To compare the overlapping ratio of TSSs of DEGs and ChIP-seq peaks, Hi-C loops and insulation boundaries, we implemented the permutation test (n = 1,000) that compared the relative overlap frequency against the background distribution. As a background, we used all DEGs obtained

by DESeq2 (11,345 genes, FDR < 0.01) for DEGs, and all boundaries (7,421) for the six types of boundaries. We randomly picked up the same number of genes or boundaries from the background in each permutation and generated the frequency distribution. For the boundary analysis (Figure 3E), we counted DEGs and the peaks that overlapped within 50 kbp of them.

Correlation of interactions with epigenomes

For interTAD interaction comparisons (Figures 6B–D), we extracted all TAD regions with widths of >100 kbp and annotated them using the epigenomic marks (H3K36me3, H3K27me3, H3K9me3 and Pol2) with the following criterion: whether the marks covered > 40.0% of the TAD length. To avoid a low read coverage at long-range distances and the technical effect derived from the different resolution of Hi-C matrices, we used the log-fold change $log_2(N_{siRNA}/N_{control})$, where N_{siRNA} and $N_{control}$ indicate the total number of fragments mapped within the interTAD regions between a TAD pair (≤2 Mbp in distance) annotated with the epigenomic marks. We calculated the score for all TAD pairs and applied k-means clustering (k = 5). Then we calculated the z-score–normalized fraction of epigenomic status for the TAD pairs included in each cluster to estimate the epigenomic-dependent depletion effect of interTAD interactions.

Extended ChromHMM

Our previous study showed that several 1D metrics for Hi-C data are effective for annotating chromatin states in detail (Wang and Nakato 2021). In this study, we added CTCF, boundary and compartment information in addition to five core histone marks (H3K4me3, K3K27ac, H3K27me3, H3K36me3 and H3K9me3) to ChromHMM (Ernst and Kellis 2012) and annotated 15 chromatin states.

DATA ACCESS

The raw sequencing data and processed files of Hi-C, RNA-seq and ChIP-seq data from this study have been submitted to the Gene Expression Omnibus (GEO) under the accession number GSE196450. Custom code used for the principal analysis is available at Docker Hub (https://hub.docker.com/r/rnakato/juicer) and at GitHub (https://github.com/rnakato/RPE_Hi-C_Analysis).

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

ACKNOWLEDGEMENTS

We thank all members of the Nakato and Shirahige Laboratories for discussions and comments on the manuscript. This work was supported by a Grant-in-Aid for Scientific Research (17H06331 to R.N. and 15H05970, 20H05686 and 20H05940 to K.S), the Japan Agency for Medical Research and Development under grant number JP22gm6310012h0003 and the Japan Science and Technology Agency under grant number JPMJCR18S5.

AUTHOR CONTRIBUTIONS

R.N. conceived this project and wrote the manuscript. R.N., J.W., L.A.E.N. and G.M.O. implemented the computational analysis. T.S. prepared Hi-C, ChIP-seq and RNA-seq samples. M.B. prepared ChIP-seq and RNA-seq samples. K.S. supervised the sample preparation and sequencing and suggested ways to improve the analysis and the manuscript.

REFERENCES

Abdennur N, Mirny LA. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**: 311-316.

Allahyar A, Vermeulen C, Bouwman BAM, Krijger PHL, Verstegen M, Geeven G, van Kranenburg M, Pieterse M, Straver R, Haarhuis JHI et al. 2018. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet* **50**: 1151-1160.

Bonev B, Cavalli G. 2016. Organization and function of the 3D genome. *Nat Rev Genet* **17**: 772.

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A et al. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**: 557-572 e524.

Bonhoure N, Bounova G, Bernasconi D, Praz V, Lammers F, Canella D, Willis IM, Herr W, Hernandez N, Delorenzi M et al. 2014. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome research* **24**: 1157-1168.

Busslinger GA, Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters JM. 2017. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**: 503-507.

Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ. 2015. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**: 240-244.

Crump NT, Ballabio E, Godfrey L, Thorne R, Repapi E, Kerry J, Tapia M, Hua P, Lagerholm C, Filippakopoulos P et al. 2021. BET inhibition disrupts transcription but retains enhancer-promoter contact. *Nat Commun* **12**: 223.

Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang SC, Sanborn AL, Machol I, Shamim M et al. 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A* **113**: E4504-4512.

679 Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters JM. 2019. DNA loop extrusion by human
680 cohesin. *Science* **366**: 1338-1345.
681 Deardorff MA, Bando M, Nakato R, Watrin E, Itoh T, Minamino M, Saitoh K, Komata M, Katou Y,
682 Clark D et al. 2012. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin
683 acetylation cycle. *Nature* **489**: 313-317.
684 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains
685 in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376-380.
686 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR.
687 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.
688 Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer
689 Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**:
690 95-98.
691 Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat*
692 *Methods* **9**: 215-216.
693 Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, Xu H, Ramsay RG, Odom DT, Flicek P.
694 2012. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-
695 regulatory modules. *Genome research* **22**: 2163-2175.
696 Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of
697 Chromosomal Domains by Loop Extrusion. *Cell reports* **15**: 2038-2049.
698 Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. 2019. Highly
699 rearranged chromosomes reveal uncoupling between genome topology and gene expression.
700 *Nat Genet* **51**: 1272-1282.
701 Haarhuis JHI, van der Weide RH, Blomen VA, Yanez-Cuna JO, Amendola M, van Ruiten MS,
702 Krijger PHL, Teunissen H, Medema RH, van Steensel B et al. 2017. The Cohesin Release
703 Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**: 693-707 e614.
704 Hansen AS. 2020. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a
705 multi-step mechanism. *Nucleus* **11**: 132-148.

706 Izumi K, Nakato R, Zhang Z, Edmondson AC, Noon S, Dulik MC, Rajagopalan R, Venditti CP,
707 Gripp K, Samanich J et al. 2015. Germline gain-of-function mutations in AFF4 cause a
708 developmental syndrome functionally linking the super elongation complex and cohesin. *Nat*
709 *Genet* **47**: 338-344.
710 Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J,
711 Rahl PB, Levine SS et al. 2010. Mediator and cohesin connect gene expression and chromatin
712 architecture. *Nature* **467**: 430-435.
713 Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H. 2019. Human cohesin compacts DNA by loop extrusion.
714 *Science* **366**: 1345-1349.
715 Kline AD, Moss JF, Selicorni A, Bisgaard AM, Deardorff MA, Gillett PM, Ishman SL, Kerr LM,
716 Levin AV, Mulder PA et al. 2018. Diagnosis and management of Cornelia de Lange
717 syndrome: first international consensus statement. *Nat Rev Genet* **19**: 649-666.
718 Kriz AJ, Colognori D, Sunwoo H, Nabet B, Lee JT. 2021. Balancing cohesin eviction and retention
719 prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-
720 inactivation. *Mol Cell* **81**: 1970-1987 e1979.
721 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.
722 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without
723 a reference genome. *BMC bioinformatics* **12**: 323.
724 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
725 *Bioinformatics* **25**: 1754-1760.
726 Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie
727 BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions
728 reveals folding principles of the human genome. *Science* **326**: 289-293.
729 Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK,
730 Murre C. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain
731 genomic interactions that orchestrate B cell fate. *Nat Immunol* **13**: 1196-1204.

732 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-
733 seq data with DESeq2. *Genome Biol* **15**: 550.
734 Merkenschlager M, Nora EP. 2016. CTCF and Cohesin in Genome Folding and Transcriptional Gene
735 Regulation. *Annu Rev Genomics Hum Genet* **17**: 17-43.
736 Minamino M, Ishibashi M, Nakato R, Akiyama K, Tanaka H, Kato Y, Negishi L, Hirota T, Sutani T,
737 Bando M et al. 2015. Esco1 Acetylates Cohesin via a Mechanism Different from That of
738 Esco2. *Curr Biol* **25**: 1694-1706.
739 Nakato R, Sakata T. 2020. Methods for ChIP-seq analysis: A practical workflow and advanced
740 applications. *Methods* doi:10.1016/j.ymeth.2020.03.005.
741 Nakato R, Shirahige K. 2017. Recent advances in ChIP-seq analysis: from quality management to
742 whole-genome annotation. *Brief Bioinform* **18**: 279-290.
743 Nakato R, Shirahige K. 2018. Sensitive and robust assessment of ChIP-seq read distribution using a
744 strand-shift profile. *Bioinformatics* **34**: 2356-2363.
745 Nora EP, Caccianini L, Fudenberg G, So K, Kameswaran V, Nagle A, Uebersohn A, Hajj B, Saux AL,
746 Coulon A et al. 2020. Molecular basis of CTCF binding polarity in genome folding. *Nat*
747 *Commun* **11**: 5612.
748 Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA,
749 Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of
750 Chromosome Domains from Genomic Compartmentalization. *Cell* **169**: 930-944 e922.
751 Olley G, Ansari M, Bengani H, Grimes GR, Rhodes J, von Kriegsheim A, Blatnik A, Stewart FJ,
752 Wakeling E, Carroll N et al. 2018. BRD4 interacts with NIPBL and BRD4 is mutated in a
753 Cornelia de Lange-like syndrome. *Nat Genet* **50**: 329-332.
754 Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I,
755 Omer AD, Lander ES et al. 2014. A 3D map of the human genome at kilobase resolution
756 reveals principles of chromatin looping. *Cell* **159**: 1665-1680.

Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID et al. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**: 305-320 e324.

Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317-330.

Sauria MEG, Taylor J. 2017. QuASAR: Quality Assessment of Spatial Arrangement Reproducibility in Hi-C Data. *bioRxiv (preprint)* doi:10.1101/204438.

Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome research* **20**: 578-588.

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L et al. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**: 51-56.

Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, Lenhard B, Giorgetti L, Heard E, Fisher AG et al. 2013. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome research* **23**: 2066-2077.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**: 2498-2504.

Sikorska N, Sexton T. 2020. Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated. *J Mol Biol* **432**: 653-664.

Stasevich TJ, Hayashi-Takanaka Y, Sato Y, Maehara K, Ohkawa Y, Sakata-Sogawa K, Tokunaga M, Nagase T, Nozaki N, McNally JG et al. 2014. Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**: 272-275.

Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. 2017. Phase separation drives heterochromatin domain formation. *Nature* **547**: 241-245.

van Ruiten MS, Rowland BD. 2018. SMC Complexes: Universal DNA Looping Machines with Distinct Regulators. *Trends Genet* **34**: 477-487.

Vian L, Pekowska A, Rao SSP, Kieffer-Kwon KR, Jung S, Baranello L, Huang SC, El Khattabi L, Dose M, Pruett N et al. 2018. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **173**: 1165-1178 e1120.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. 2015. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell reports* **10**: 1297-1309.

Waizenegger IC, Hauf S, Meinke A, Peters JM. 2000. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* **103**: 399-410.

Wang CY, Jegu T, Chu HP, Oh HJ, Lee JT. 2018. SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell* **174**: 406-421 e425.

Wang J, Nakato R. 2021. HiC1Dmetrics: framework to extract various one-dimensional features from chromosome structure data. *Brief Bioinform* doi:10.1093/bib/bbab509.

Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**: 796-801.

Wutz G, Ladurner R, St Hilaire BG, Stocsits RR, Nagasaka K, Pignard B, Sanborn A, Tang W, Varnai C, Ivanov MP et al. 2020. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin(STAG1) from WAPL. *Elife* **9**.

Wutz G, Varnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ 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.

Yaffe E, Tanay A. 2011. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet* **43**: 1059-1065.

811 Yang T, Zhang F, Yardimci GG, Song F, Hardison RC, Noble WS, Yue F, Li Q. 2017. HiCRep:
812 assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient.
813 *Genome research* **27**: 1939-1949.
814 Yardimci GG, Ozadam H, Sauria MEG, Ursu O, Yan KK, Yang T, Chakraborty A, Kaul A, Lajoie
815 BR, Song F et al. 2019. Measuring the reproducibility and quality of Hi-C data. *Genome Biol*
816 **20**: 57.
817 Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological
818 themes among gene clusters. *OMICS* **16**: 284-287.

819

FIGURE LEGENDS

Figure 1. Comparative Hi-C analysis to explore the variation of depletion effects. **(A)** Summary of protein targets (left) and immunoblots after 72 h of siRNA (right). **(B)** Aggregate peak analysis (APA) of loops. **(C)** Correlation heatmap among samples based on stratum-adjusted correlation coefficients. **(D)** Normalized Hi-C matrix of a representative chromosomal region (chromosome 21, 26.0–29.0 Mb). **(E)** The number of TADs, loops, APA strength (observed/expected ratio of the center) and compartment strength (from saddle plots, see also Figure S2B). The dots indicate the Hi-C samples (including different treatment times). **(F)** Relative contact probability for the likelihood of a contact at increasing length scales (averaged by replicates). **(G)** Loop length distribution.

Figure 2. Comparative analysis of siRNA effects on the transcriptome and epigenome. **(A)** Left: Number of DEGs for each siRNA treatment (FDR < 0.01). Right: Correlation heatmap based on the Simpson index (showing overlap among top-ranked 1,000 DEGs) for all sample pairs. **(B)** Clustering of all top-ranked DEGs (rows) and samples (columns) based on the overlap of up- and downregulated genes. The significant GO terms are also shown (left). See Figure S5 for the full list of GO terms. **(C)** Log-scale relative enrichment of ChIP-seq peaks (left) and Hi-C loops (right) at the TSSs of DEGs in the clusters corresponding to **(B)** against all DEGs. *p < 0.01, the permutation test (n = 1,000). **(D)** ChIP-seq distribution ($-\log_{10}$(p), 100-bp bin) among top-ranked DEGs (*BMP6*, *SSBP2* and *TNFRSF19*). The significantly enriched regions ($p < 10^{-4}$) are highlighted in red. Red and black arrows (bottom) indicate CTCF-independent and -dependent Rad21 peaks, respectively. **(E)** The relative enrichment of Rad21 peaks after depletions and Smc3ac around genes, as compared with Rad21 (control). *: p < 0.001; **: p < 0.0001, Fisher exact test against control.

Figure 3. Multi-scale insulation scores reveal the diversity of insulation perturbation at TAD boundaries. **(A-B)** A representative chromosomal region (chromosome 21, 24.0–32.0 Mb). **(A)** Normalized Hi-C matrices. Black dashed lines and blue circles indicate TADs and loops, respectively. **(B)** Top: ChIP-seq distribution ($-\log_{10}(p)$, 5-kbp bin). Red regions: $p < 10^{-3}$. Middle: Multi-scale insulation scores. Red regions indicate insulated regions (boundaries). The numbers along the bottom (i–iv) indicate four of the six boundary types (see text). The red rectangle indicates the long-range insulation boundaries present after siNIPBL. Bottom: Compartment PC1 and labeled TADs. **(C)** Insulation score distribution for six types of boundaries. The number in parentheses below each boundary type indicates the number of boundaries. **(D)** The proportion of boundaries located between compartments A and B (A-B) or within compartments A (A-A) and B (B-B). **(E)** Relative enrichment of DEGs (top), loops (middle), ChIP-seq peaks (lower middle) and broad histone enrichment (bottom) that overlap the six boundary types and intraTAD regions for compartments A and B against all boundaries. For broad histone marks (bottom), we calculated the fraction of regions covered by the obtained peaks. *$p < 0.01$, the permutation test (n = 1,000).

Figure 4. InterTAD-level insulation is increased by cohesin loss. **(A)** Relative enrichment of the interaction frequency (log scale) relative to control. The region and TAD labels (A1–A5, B1 and B2) are as in Figures 3A and 3B. See Figure S7A for all samples. **(B)** Top: Schematic illustration of DRF. Middle: Heatmap of DRF for depletions (row) at the same region with **(A)**. Bottom: Averaged DRF of cohesin and loaders (red), unloaders (blue) and others (black). The shaded regions indicate the 99% confidence interval. The purple bars indicate the identified differential DRF regions. **(C)** Examples of differential DRF regions. Top: Normalized Hi-C matrix (control) and relative interaction frequency after depletions. The dashed black triangles indicate TADs in the control sample. Bottom: Compartment PC1 and ChIP-seq distribution ($-\log_{10}(p)$, 5-kbp bin). **(D)** Compartment PC1 of the control and siNIPBL sample centered on 241 differential DRF regions. Rows were ordered using hierarchical clustering based on the control sample. **(E)** The cosine similarity distribution of the

relative frequency in all 241 differential DRF regions (~2 Mbp from the center of each region). See Figure S7B for all comparisons.

Figure 5. The depletion effect on long-range interactions was different between siRad21 and siNIPBL. **(A)** SIMA analysis after siRNA. See Figure S8A for all samples. **(B)** Relative interaction frequency and compartment PC1 (chromosome 15, 38–101 Mb).

Figure 6. InterTAD interaction variation as compared with epigenomic marks. **(A)** Three example chromosomal regions of increased interactions only in siRad21. Top: Relative frequency and compartment PC1. Bottom: ChIP-seq distribution ($-\log_{10}$(p), 5-kbp bin). **(B-D)** K-means clustering (k = 5) of all TAD pairs based on the depletion effect on interactions between them **(B)**, distribution of depletion effects on the clusters **(C)**, and fraction of the epigenomic state of TAD pairs included in the clusters **(D)**.

Figure 7. Cohesin was broadly enriched in compartment A but was not enriched in compartment B. **(A)** The depletion effect distribution from ChIP-seq data ($-\log_{10}$(p), control/siRNA, 5-kbp bin) on chromosome 21, 24–32 Mbp. The colored bars (bottom) indicate the four compartment types (red, StrongA; yellow, WeakA; green, WeakB; blue, StrongB). **(B)** A graphical representation of the four compartments. **(C)** The $-\log_{10}$(p) distribution for the depletion effect of siNIPBL and siCTCF on Rad21 ChIP-seq in the four compartments. **(D)** siRad21 had less of an effect on StrongB TADs (black rectangles). The region is the same with **(A)**. **(E)** Averaged interactions (observed/expected) in StrongA and StrongB TADs for representative samples.

Figure 1

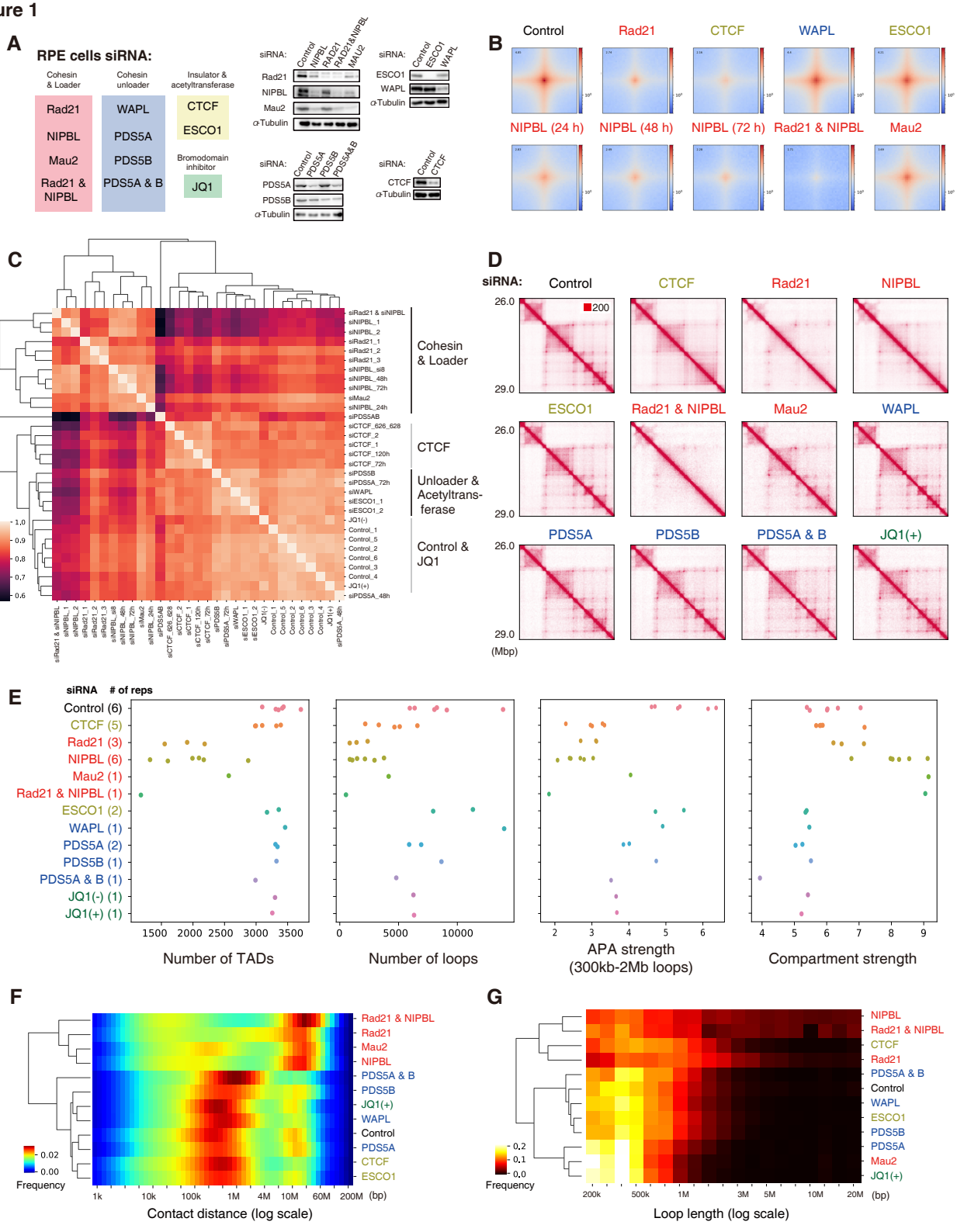

A
RPE cells siRNA:
Cohesin & Loader
Rad21
NIPBL
Mau2
Rad21 & NIPBL
Cohesin unloader
WAPL
PDS5A
PDS5B
PDS5A & B
Insulator & acetyltransferase
CTCF
ESCO1
Bromodomain inhibitor
JQ1
siRNA:
Control
NIPBL
Rad21
Rad21&NIPBL
Mau2
Rad21
NIPBL
Mau2
α-Tubulin
siRNA:
Control
ESCO1
WAPL
ESCO1
WAPL
α-Tubulin
siRNA:
Control
PDS5A
PDS5B
PDS5A&B
PDS5A
PDS5B
α-Tubulin
siRNA:
Control
CTCF
CTCF
α-Tubulin
B
Control
Rad21
CTCF
WAPL
ESCO1
NIPBL (24 h)
NIPBL (48 h)
NIPBL (72 h)
Rad21 & NIPBL
Mau2
C
siRad21 & siNIPBL
siNIPBL_1
siNIPBL_2
siRad21_1
siRad21_2
siRad21_3
siNIPBL_si8
siNIPBL_48h
siNIPBL_72h
siMau2
siNIPBL_24h
siPDS5AB
siCTCF_626_628
siCTCF_2
siCTCF_1
siCTCF_120h
siCTCF_72h
siPDS5B
siPDS5A_72h
siWAPL
siESCO1_1
siESCO1_2
JQ1(-)
Control_1
Control_5
Control_2
Control_6
Control_3
Control_4
JQ1(+)
siPDS5A_48h
Cohesin & Loader
CTCF
Unloader & Acetyltrans-ferase
Control & JQ1
1.0
0.9
0.8
0.7
0.6
D
siRNA:
Control
CTCF
Rad21
NIPBL
ESCO1
Rad21 & NIPBL
Mau2
WAPL
PDS5A
PDS5B
PDS5A & B
JQ1(+)
26.0
29.0
200
(Mbp)
E
siRNA # of reps
Control (6)
CTCF (5)
Rad21 (3)
NIPBL (6)
Mau2 (1)
Rad21 & NIPBL (1)
ESCO1 (2)
WAPL (1)
PDS5A (2)
PDS5B (1)
PDS5A & B (1)
JQ1(-) (1)
JQ1(+) (1)
1500 2000 2500 3000 3500
Number of TADs
0 5000 10000
Number of loops
2 3 4 5 6
APA strength
(300kb-2Mb loops)
4 5 6 7 8 9
Compartment strength
F
Rad21 & NIPBL
Rad21
Mau2
NIPBL
PDS5A & B
PDS5B
JQ1(+)
WAPL
Control
PDS5A
CTCF
ESCO1
0.02
0.00
Frequency
1k 10k 100k 1M 4M 10M 60M 200M (bp)
Contact distance (log scale)
G
NIPBL
Rad21 & NIPBL
CTCF
Rad21
PDS5A & B
Control
WAPL
ESCO1
PDS5B
PDS5A
Mau2
JQ1(+)
0.2
0.1
0.0
Frequency
200k 500k 1M 3M 5M 10M 20M (bp)
Loop length (log scale)

Figure 2

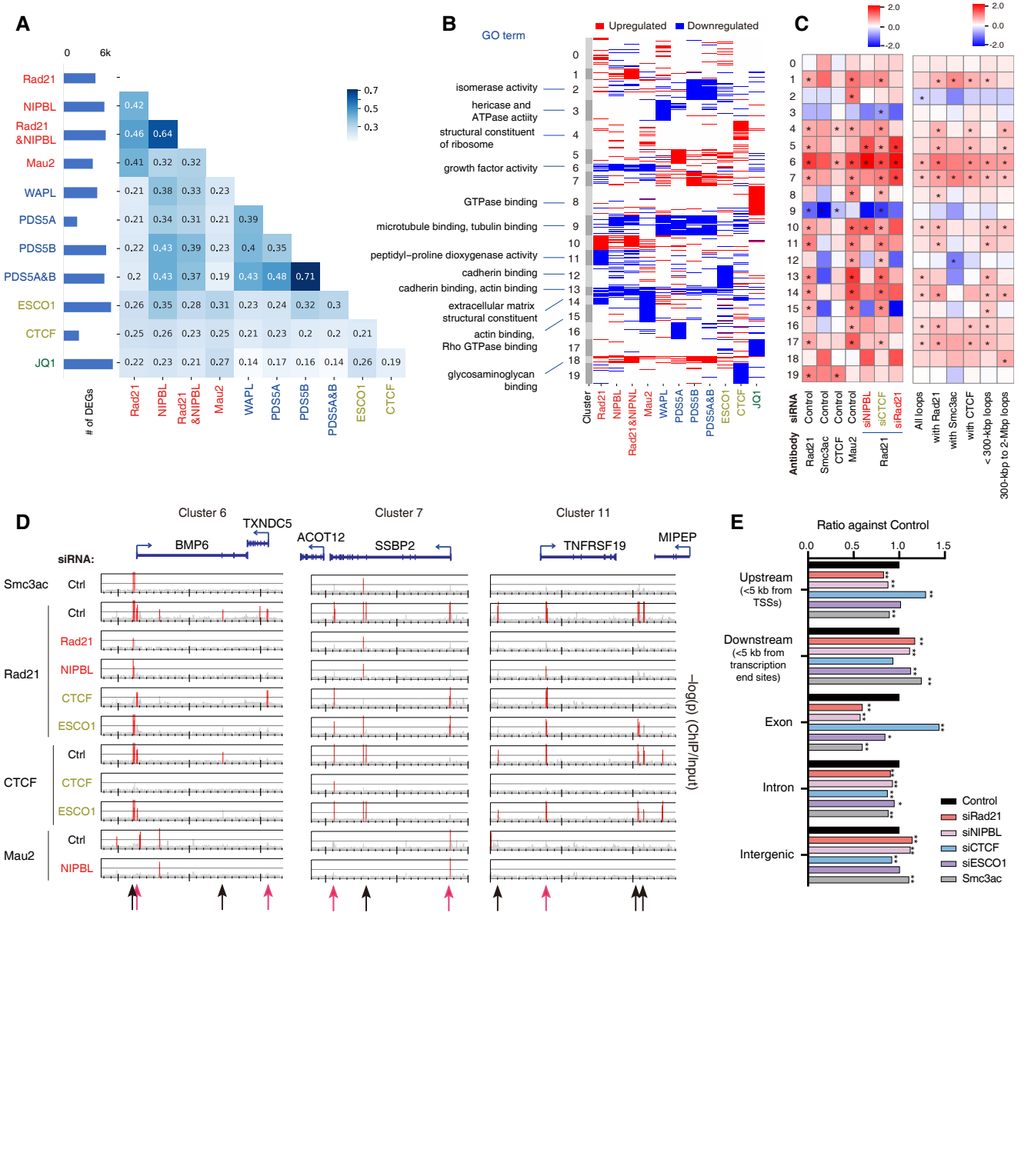

A
0
6k
Rad21
NIPBL
Rad21 &NIPBL
Mau2
WAPL
PDS5A
PDS5B
PDS5A&B
ESCO1
CTCF
JQ1
0.42
0.46 0.64
0.41 0.32 0.32
0.21 0.38 0.33 0.23
0.21 0.34 0.31 0.21 0.39
0.22 0.43 0.39 0.23 0.4 0.35
0.2 0.43 0.37 0.19 0.43 0.48 0.71
0.26 0.35 0.28 0.31 0.23 0.24 0.32 0.3
0.25 0.26 0.23 0.25 0.21 0.23 0.2 0.2 0.21
0.22 0.23 0.21 0.27 0.14 0.17 0.16 0.14 0.26 0.19
0.7
0.5
0.3
# of DEGs
B
GO term
Upregulated
Downregulated
isomerase activity
hericase and ATPase actiity
structural constituent of ribosome
growth factor activity
GTPase binding
microtubule binding, tubulin binding
peptidyl−proline dioxygenase activity
cadherin binding
cadherin binding, actin binding
extracellular matrix structural constituent
actin binding, Rho GTPase binding
glycosaminoglycan binding
Cluster
Rad21
NIPBL
Rad21&NIPNL
Mau2
WAPL
PDS5A
PDS5B
PDS5A&B
ESCO1
CTCF
JQ1
C
2.0
0.0
-2.0
siRNA
Control
Control
Control
Control
siNIPBL
siCTCF
siRad21
Antibody
Rad21
Smc3ac
CTCF
Mau2
Rad21
All loops
with Rad21
with Smc3ac
with CTCF
< 300-kbp loops
300-kbp to 2-Mbp loops
D
Cluster 6
Cluster 7
Cluster 11
TXNDC5
BMP6
ACOT12
SSBP2
TNFRSF19
MIPEP
siRNA:
Smc3ac
Ctrl
Rad21
Ctrl
Rad21
NIPBL
CTCF
ESCO1
CTCF
Ctrl
CTCF
ESCO1
Mau2
Ctrl
NIPBL
−log(p) (Ctrl/Untputt)
E
Ratio against Control
0.0
0.5
1.0
1.5
Upstream (<5 kb from TSSs)
Downstream (<5 kb from transcription end sites)
Exon
Intron
Intergenic
Control
siRad21
siNIPBL
siCTCF
siESCO1
Smc3ac

Figure 3

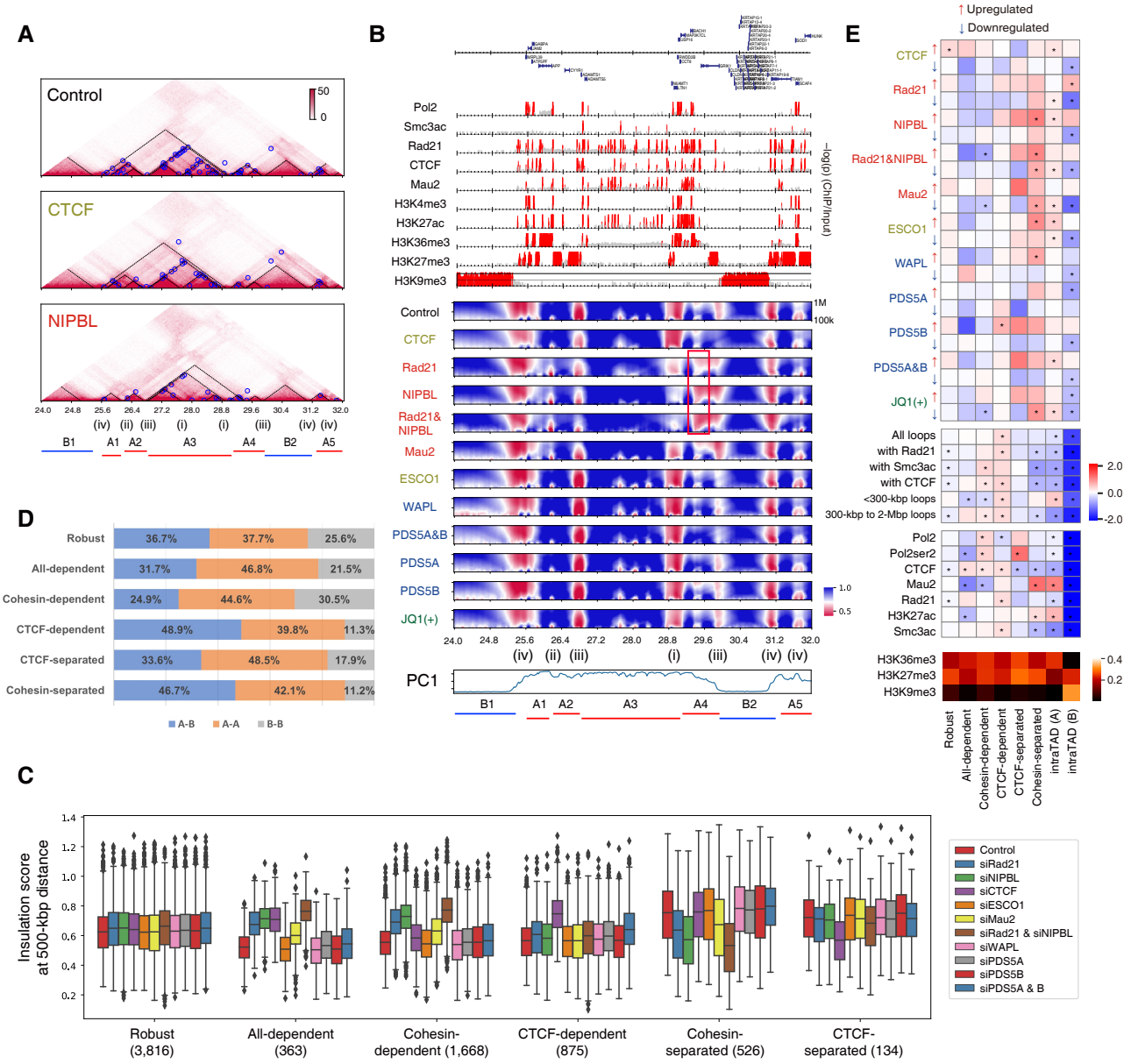

A
Control
CTCF
NIPBL
50
0
24.0 24.8 25.6 26.4 27.2 28.0 28.8 29.6 30.4 31.2 32.0
(iv) (ii) (iii) (i) (i) (iii) (iv) (iv)
B1 A1 A2 A3 A4 B2 A5
B
Pol2
Smc3ac
Rad21
CTCF
Mau2
H3K4me3
H3K27ac
H3K36me3
H3K27me3
H3K9me3
−log(p) (ChIP/Input)
Control
CTCF
Rad21
NIPBL
Rad21& NIPBL
Mau2
ESCO1
WAPL
PDS5A&B
PDS5A
PDS5B
JQ1(+)
1M
100k
1.0
0.5
24.0 24.8 25.6 26.4 27.2 28.0 28.8 29.6 30.4 31.2 32.0
(iv) (ii) (iii) (i) (iii) (iv) (iv)
PC1
B1 A1 A2 A3 A4 B2 A5
E
↑ Upregulated
↓ Downregulated
CTCF
Rad21
NIPBL
Rad21&NIPBL
Mau2
ESCO1
WAPL
PDS5A
PDS5B
PDS5A&B
JQ1(+)
All loops
with Rad21
with Smc3ac
with CTCF
<300-kbp loops
300-kbp to 2-Mbp loops
2.0
0.0
-2.0
Pol2
Pol2ser2
CTCF
Mau2
Rad21
H3K27ac
Smc3ac
H3K36me3
H3K27me3
H3K9me3
0.4
0.2
Robust
All-dependent
Cohesin-dependent
CTCF-dependent
CTCF-separated
Cohesin-separated
intraTAD (A)
intraTAD (B)
D
Robust 36.7% 37.7% 25.6%
All-dependent 31.7% 46.8% 21.5%
Cohesin-dependent 24.9% 44.6% 30.5%
CTCF-dependent 48.9% 39.8% 11.3%
CTCF-separated 33.6% 48.5% 17.9%
Cohesin-separated 46.7% 42.1% 11.2%
A-B
A-A
B-B
C
Insulation score at 500-kbp distance
1.4
1.2
1.0
0.8
0.6
0.4
0.2
Robust (3,816)
All-dependent (363)
Cohesin-dependent (1,668)
CTCF-dependent (875)
Cohesin-separated (526)
CTCF-separated (134)
Control
siRad21
siNIPBL
siCTCF
siESCO1
siMau2
siRad21 & siNIPBL
siWAPL
siPDS5A
siPDS5B
siPDS5A & B

Figure 4

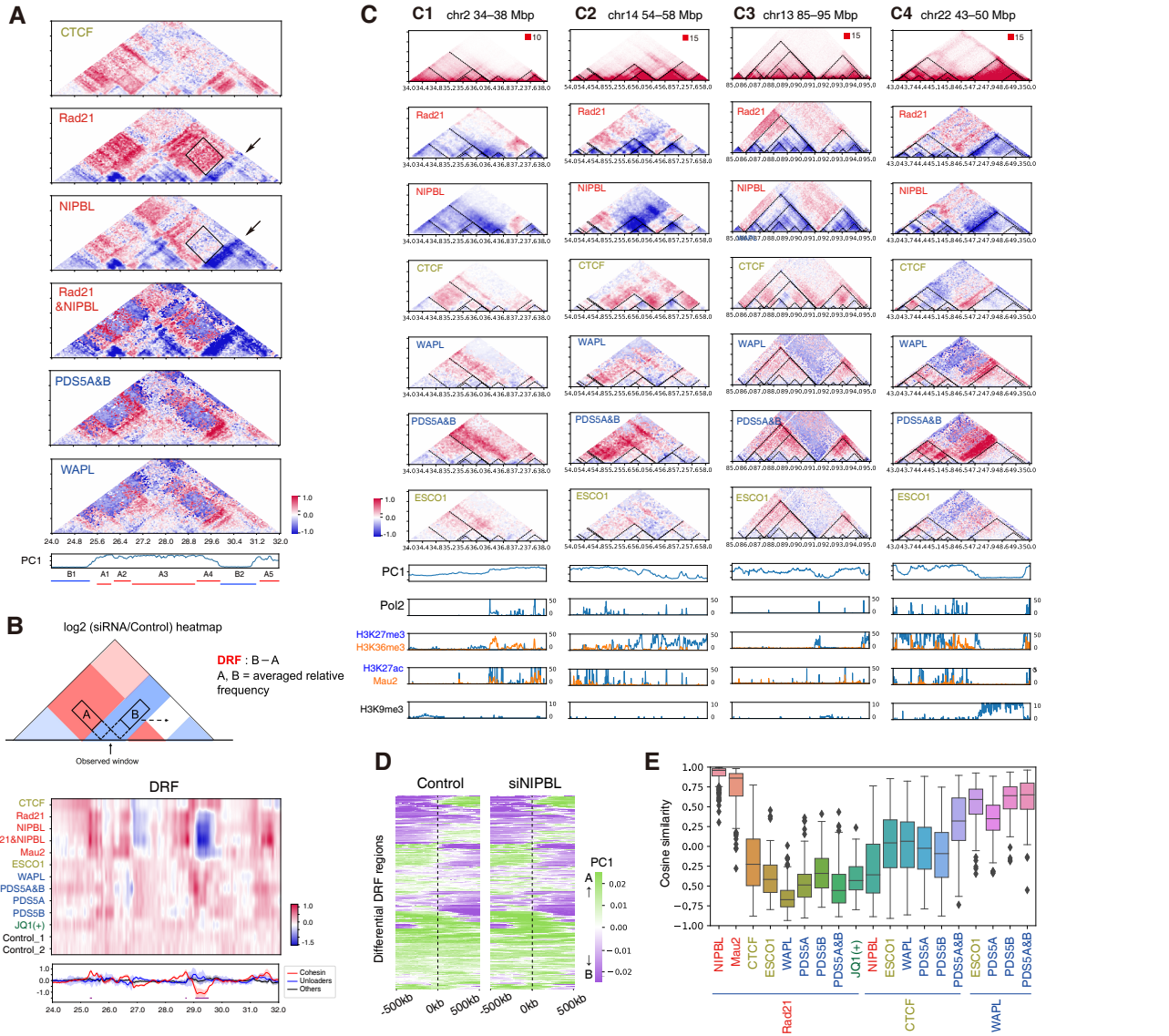
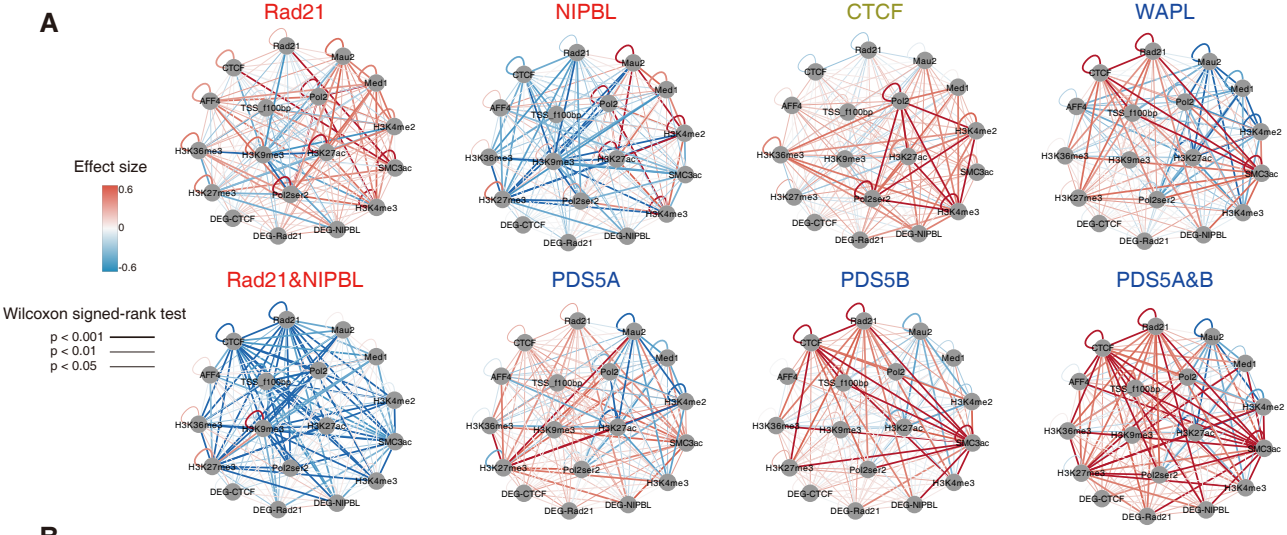

A
CTCF
Rad21
NIPBL
Rad21 &NIPBL
PDS5A&B
WAPL
PC1
B1
A1
A2
A3
A4
B2
A5
B
log2 (siRNA/Control) heatmap
DRF : B – A
A, B = averaged relative frequency
Observed window
DRF
Control_1
Control_2
Cohesin
Unloaders
Others
C
C1 chr2 34–38 Mbp
C2 chr14 54–58 Mbp
C3 chr13 85–95 Mbp
C4 chr22 43–50 Mbp
ESCO1
Pol2
H3K27me3
H3K36me3
H3K27ac
Mau2
H3K9me3
D
Control
siNIPBL
Differential DRF regions
PC1
-500kb
0kb
500kb
E
Cosine similarity
Rad21
CTCF
WAPL

Figure 5


A
Rad21
NIPBL
CTCF
WAPL
Rad21&NIPBL
PDS5A
PDS5B
PDS5A&B
Effect size
0.6
0
-0.6
Wilcoxon signed-rank test
p < 0.001
p < 0.01
p < 0.05
Rad21
Mau2
CTCF
Med1
AFF4
TSS_1100bp
Pol2
H3K4me2
H3K36me3
H3K9me3
H3K27ac
SMC3ac
H3K27me3
Pol2ser2
H3K4me3
DEG-CTCF
DEG-Rad21
DEG-NIPBL

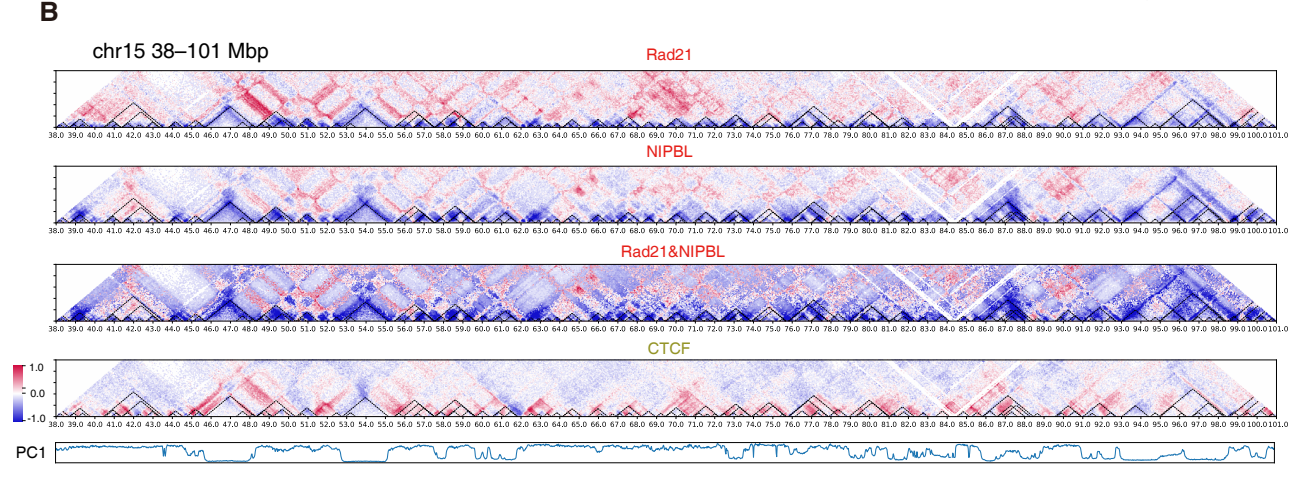

B
chr15 38–101 Mbp
Rad21
NIPBL
Rad21&NIPBL
CTCF
1.0
0.0
-1.0
PC1
38.0 39.0 40.0 41.0 42.0 43.0 44.0 45.0 46.0 47.0 48.0 49.0 50.0 51.0 52.0 53.0 54.0 55.0 56.0 57.0 58.0 59.0 60.0 61.0 62.0 63.0 64.0 65.0 66.0 67.0 68.0 69.0 70.0 71.0 72.0 73.0 74.0 75.0 76.0 77.0 78.0 79.0 80.0 81.0 82.0 83.0 84.0 85.0 86.0 87.0 88.0 89.0 90.0 91.0 92.0 93.0 94.0 95.0 96.0 97.0 98.0 99.0 100.0 101.0

Figure 6

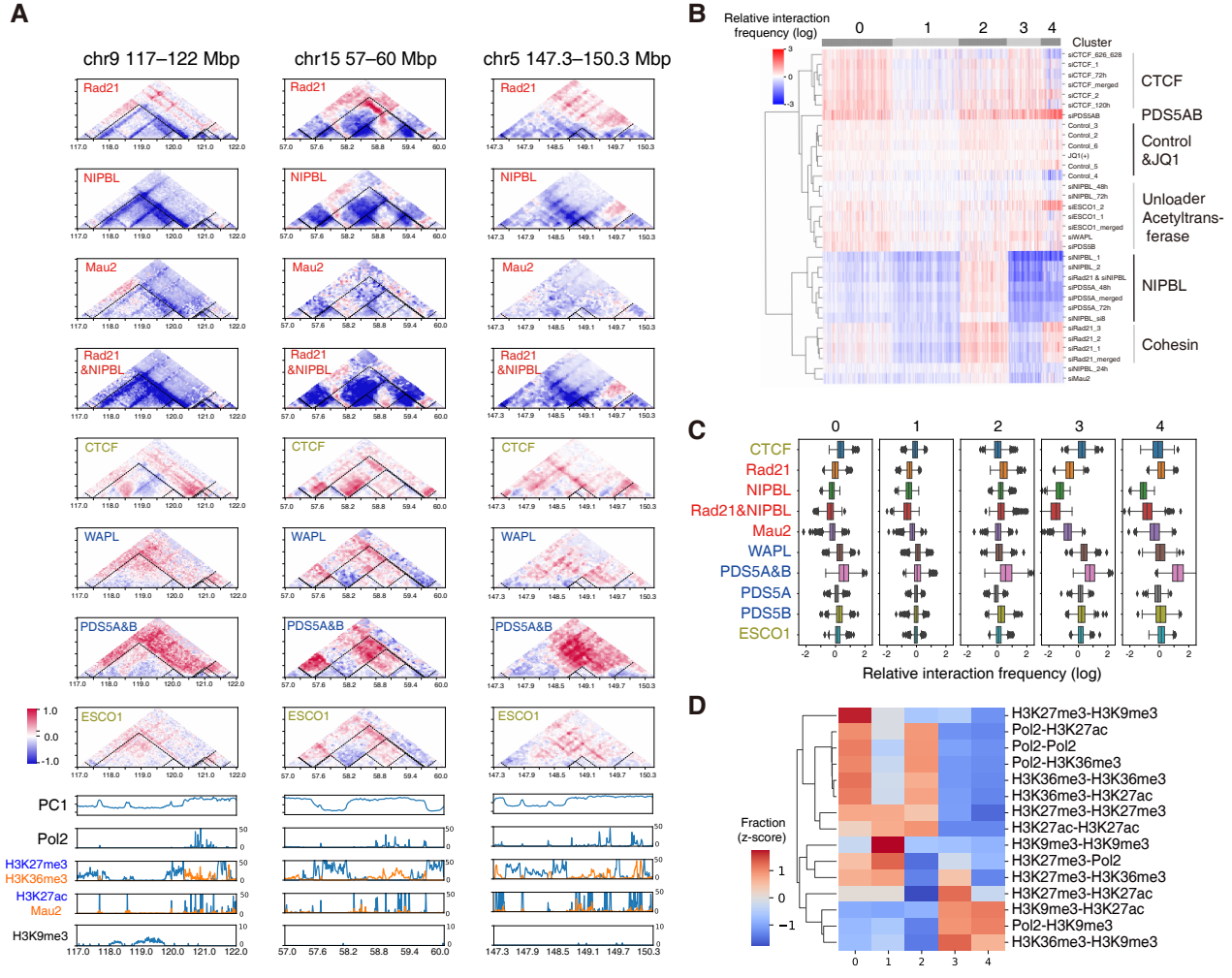
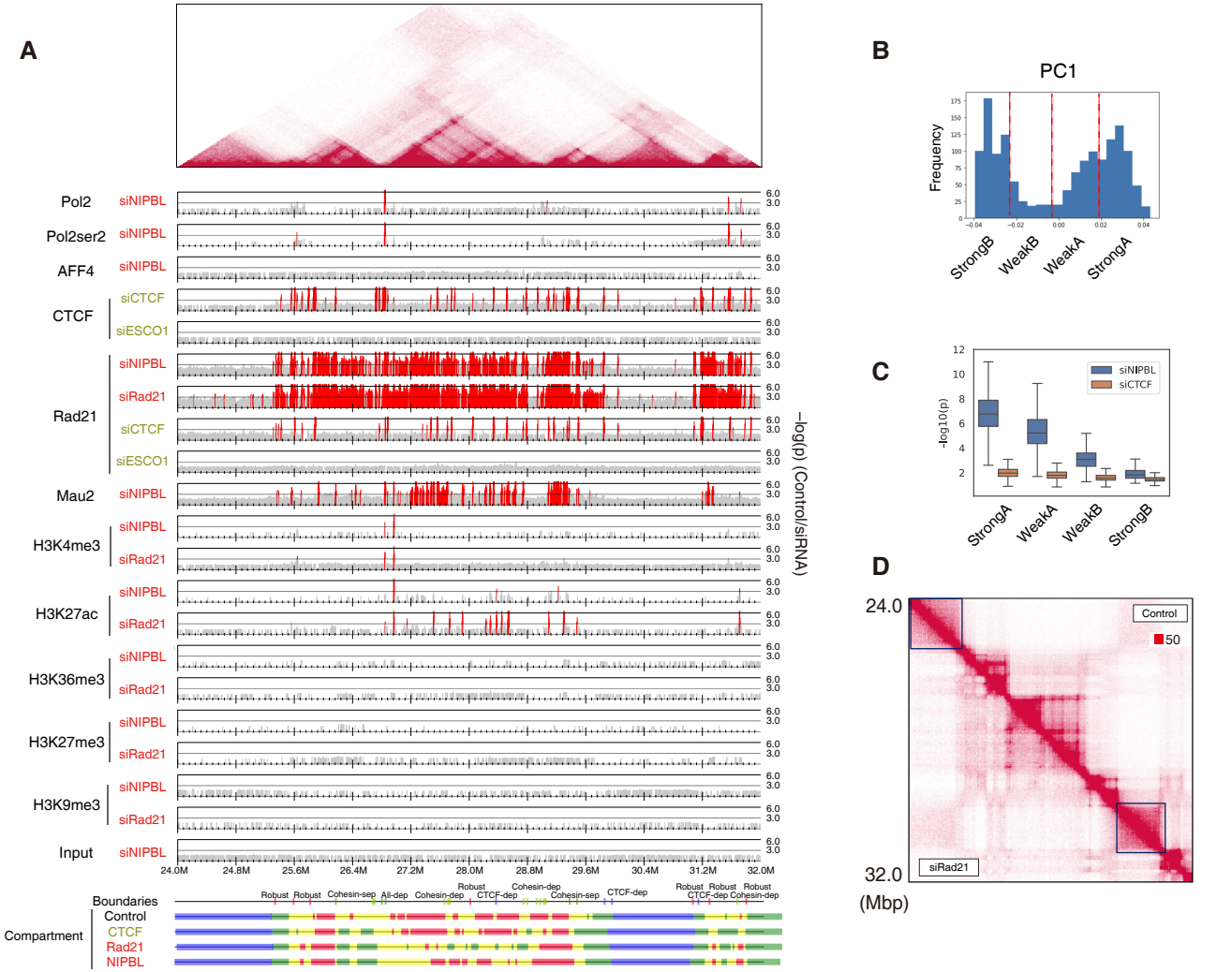
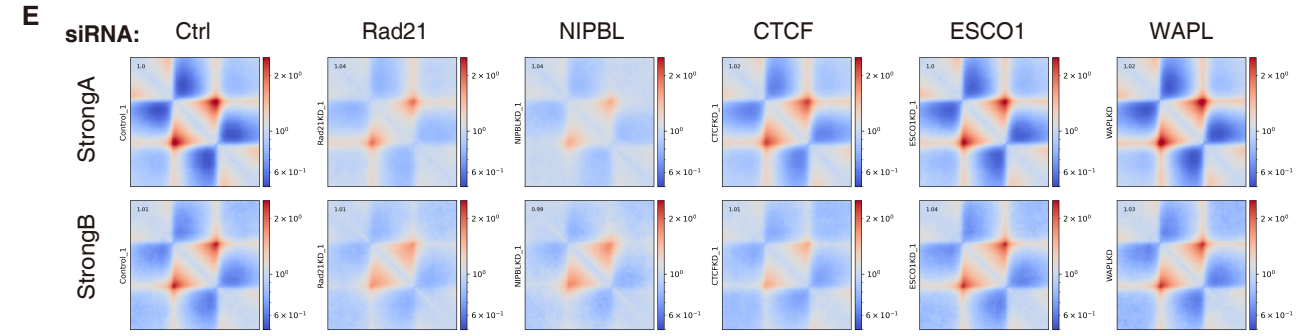

A
chr9 117–122 Mbp
chr15 57–60 Mbp
chr5 147.3–150.3 Mbp
Rad21
NIPBL
Mau2
Rad21 &NIPBL
CTCF
WAPL
PDS5A&B
ESCO1
PC1
Pol2
H3K27me3
H3K36me3
H3K27ac
Mau2
H3K9me3
B
Relative interaction frequency (log)
Cluster
CTCF
PDS5AB
Control &JQ1
Unloader Acetyltrans- ferase
NIPBL
Cohesin
C
CTCF
Rad21
NIPBL
Rad21&NIPBL
Mau2
WAPL
PDS5A&B
PDS5A
PDS5B
ESCO1
Relative interaction frequency (log)
D
Fraction (z-score)
H3K27me3-H3K9me3
Pol2-H3K27ac
Pol2-Pol2
Pol2-H3K36me3
H3K36me3-H3K36me3
H3K36me3-H3K27ac
H3K27me3-H3K27me3
H3K27ac-H3K27ac
H3K9me3-H3K9me3
H3K27me3-Pol2
H3K27me3-H3K36me3
H3K27me3-H3K27ac
H3K9me3-H3K27ac
Pol2-H3K9me3
H3K36me3-H3K9me3

Figure 7


A
Pol2 siNIPBL
Pol2ser2 siNIPBL
AFF4 siNIPBL
CTCF siCTCF
siESCO1
Rad21 siNIPBL
siRad21
siCTCF
siESCO1
Mau2 siNIPBL
H3K4me3 siNIPBL
siRad21
H3K27ac siNIPBL
siRad21
H3K36me3 siNIPBL
siRad21
H3K27me3 siNIPBL
siRad21
H3K9me3 siNIPBL
siRad21
Input siNIPBL
6.0
3.0
24.0M 24.8M 25.6M 26.4M 27.2M 28.0M 28.8M 29.6M 30.4M 31.2M 32.0M
−log(p) (Control/siRNA)
Boundaries
Robust Robust Cohesin-sep All-dep Cohesin-dep Robust CTCF-dep Cohesin-dep Cohesin-sep CTCF-dep Robust CTCF-dep Robust Cohesin-dep
Compartment
Control
CTCF
Rad21
NIPBL
B
PC1
Frequency
175
150
125
100
75
50
25
0
−0.04 −0.02 0.00 0.02 0.04
StrongB WeakB WeakA StrongA
C
-log10(p)
12
10
8
6
4
2
siNIPBL
siCTCF
StrongA WeakA WeakB StrongB
D
24.0
32.0
(Mbp)
Control
50
siRad21


E
siRNA: Ctrl Rad21 NIPBL CTCF ESCO1 WAPL
StrongA
StrongB
Control_1
Rad21KD_1
NIPBLKD_1
CTCFKD_1
ESCO1KD_1
WAPLKD
2 × 10^0
10^0
6 × 10^−1