On the origin of the integrity of a subset of Topologically Associating Domains upon Cohesin loss in Interphase Chromosomes

Davin Jeong[1], Guang Shi[1], Xin Li[1] and D. Thirumalai[1,2]

Department of Chemistry[1] *and Department of Physics*[2],

University of Texas at Austin, Austin, TX 78712, USA

Abstract

Experiments have established that in interphase chromosomes, active and inactive loci are spatially segregated on a few mega base pairs (Mbp) scale, resulting in compartment formation. On the sub-Mbp scales, Topologically Associating Domains (TADs) appear as squares along the diagonal in the contact matrix (CM). Hi-C experiments showed that most of the TADs vanish upon deleting cohesin, while the compartment structure is maintained or even modestly enhanced. However, closer inspection of the data reveals that a non-negligible fraction of TADs is preserved (P-TADs), even after cohesin loss. Super resolution tracing, on the other hand, shows that, at the single cell level, TAD-like structures are visible after depleting cohesin. To rationalize these disparate findings, we used polymer simulations, which reveal that TADs with epigenetic mismatches across their boundaries survive after deleting the loops. Analyses of 3D structures reveal that many of the P-TADs have sharp physical boundaries. Informed by the simulations, we analyzed the Hi-C maps (with and without cohesin) in mouse liver and HCT-116, which affirmed that epigenetic mismatches and physical boundaries (calculated using the 3D structures from Hi-C contact maps) explain the origin of P-TADs. The calculated single-cell structures, using cohesin depleted Hi-C as input, have TAD-like features in single cells that are remarkably similar to the findings in imaging experiments. Cumulatively our study, utilizing a variety of methods, shows that preservation of a subset of TADs upon deleting cohesin is a robust phenomenon.

I. INTRODUCTION

Advances in experimental techniques have provided details of the three-dimensional (3D) organization of chromosomes in diverse species[1–6]. The contact map (CM)[4,6], inferred using chromosome conformation capture technique and related variants (referred to as Hi-C from now on), is a two-dimensional (2D) matrix, whose elements indicate the ensemble average probability that two loci separated by a certain genomic distance are spatially adjacent. The Hi-C experiments on different mammalian cells suggest that there are two major length scales in the organization of interphase chromosomes. On the scale, $L_c \sim$ (2-5) Mbp (mega base pairs or Mbp), one observes checkerboard patterns in the CMs[4], which are thought to be associated with micro-phase separation between the two major epigenetic states, active (A) or euchromatin and inactive (B) or heterochromatin. On the shorter length scale, L_{TAD}, from tens of kb up to a few Mb, chromatin domains appear as squares along the diagonal of the CMs[7,8]. These are referred to as Topologically Associating Domains (TADs). Contacts are enriched within the TADs, and are reduced across the boundaries between the TADs. The use of the two length scales, L_c and L_{TAD}, in characterizing the organization of interphase chromosomes is now entrenched in the field.

TADs are thought to regulate gene expression by constraining the contacts between target gene and regulatory regions[9,10]. As a consequence, perturbation or disruption of their integrity such as deletions, duplications or inversions of DNA segments within the TADs, could lead to aberrant gene expression[8,11–17]. A class of chromatin loops, mediated by cohesin and CTCF protein ("cohesin-associated CTCF loop"), organizes a subset of the TADs[18]. An ATP-dependent molecular motor, cohesin[19] extrudes DNA loops of varying length, and it is suspected that the loop is terminated when cohesin encounters the transcriptional insulator CCCTC-binding factor (CTCF)[20]. This implies that cohesin and CTCF, a zinc finger DNA binding protein, are often colocalized at the TAD boundary[4–6,18,21,22].

Several experiments have shown that depletion of architectural proteins (Nipbl, RAD21, and CTCF) disrupts the organization of interphase chromosomes[15,23–30]. For instance, Schwarzer et al.[23] showed that the removal of the cohesin loading factor, *Nipbl* in the mouse liver cell results in loss of TADs. They concluded that cohesin-independent compartment formation is a consequence of the underlying epigenetic landscape, while TAD formation is driven by the motor protein cohesin. Similarly, it was suggested that if the cohesin subunit, *RAD21* is deleted, then cohesin-associated CTCF loops and TADs are abolished[15,28,29]. Deletion of *RAD21* results in the complete loss of the so-called loop domains[15], which are formed

when CTCF colocalizes with cohesin. Remarkably, imaging experiments[29] show that TAD-like structures, with sharp boundaries are present, at the single cell level even after deleting cohesin. However, the preferential positioning observed in the Hi-C data are lost. Three headlines emerge from these studies are: (i) They reinforce the two-length scale description of genome organization, described above. (ii) Factors that prevent the association of cohesin with chromosomes globally abolish the TADs and the Hi-C peaks, but preserve (or *even enhance*) compartmentalization. Experimental studies[4,15,23,28,29] and polymer simulations[31?,32] have shown that the global epigenetic state determines compartment formation, while the more dynamic TADs, with finite lifetimes[33], require ATP-dependent cohesin motor. (iii) TAD-like features persist at the single cell before and after auxin treatment, albeit with changes in the locations of sharp domain boundaries.

Similarly, majority of the TADs are abolished upon depleting CTCF in mouse embryonic stem cells (mESCs). Comparison of the CMs with and without CTCF shows that a subset of TADs remains even after CTCF loss. Analysis of the CMs (Figure 3B in[27]) using insulation score shows that $\sim 18\%$ of the TADs are retained after CTCF loss. Inspired by this observation, we wondered if similar conclusions could be found by analyzing the experimental CMs from mouse liver and HCT-116 cells in the presence and absence of cohesin. We found that, on an average, $\approx 16\%$ and $\approx 26\%$ of the TADs, identified using the TopDom method[34], are preserved in the mouse liver and HCT116 cell lines, respectively (Fig.1). First, these findings reinforce the conclusion reached for mESCs[27], although mechanistically depleting CTCF and cohesin could be different. Second, preservation of certain TADs raises the question: What is the underlying mechanism that a small but not insignificant fraction of TADs are retained after cohesin loss? Third, is it possible to explain the results in super resolution imaging experiments that show TAD-like structures at the single level, with and without cohesin, solely using CMs inferred from Hi-C experiments?

We attempt to answer the questions posed above by using the following strategy: (1) We first performed polymer simulations using a generalization of the Chromosome Copolymer Model (CCM)[35] with and without loop anchors, which roughly mimics the wild type (WT) and the absence of cohesin-associated CTCF loops, respectively. Although the CCM simulations generate three dimensional (3D) structures directly, we calculated the CMs in order to compare with Hi-C experiments. The TADs were identified by the TopDom method[34] using the calculated CMs as inputs. The simulations of two full length chromosomes (Chr13 and Chr10) from the GM12878 cell line showed that one of the factors that accounts for the preservation of certain TADs is a mismatch in the epigenetic states across the TAD

boundary. However, some TADs without epigenetic mismatches are also retained. We used the simulated structures to demonstrate the presence of physical boundaries in preserved TADs (P-TADs) after loop loss. (2) Informed by the results from the CCM simulations, we analyzed the experimental data from the two cell lines[15,23]. Just as in the CCM simulations, we discovered that epigenetic mismatch does account for a reasonable fraction ($\approx$ 0.4 in mouse liver and $\approx$ 0.3 in HCT-116 cell lines) of P-TADs. We also generated the 3D structural ensemble of the loci for all the chromosomes for both the cell lines using the HIPPS method[36]. The analyses using the 3D coordinates accounted for about 53% of the P-TADs. (3) Analyses of 3D structures, calculated using HIPPS with experimental CMs as inputs, revealed TAD-like structures at the single cell level both in the presence and absence of cohesin, which is in accord with the super resolution imaging data[29]. Our work shows that the effects of cohesin removal on chromatin structures are nuanced, requiring analyses of both the epigenetic landscape as well as 3D structures in order to obtain a comprehensive picture of how distinct factors determine interphase chromosome organization in the nucleus. Despite the limitations of different methods the conclusion that a subset of P-TADs is intact after deleting cohesin is robust.

II. RESULTS

A non-negligible fraction of TADs are preserved upon deletion of CTCF or cohesin: It has been asserted, through careful experimentation[15,23–26,28,29], that deletion of cohesin loaders (*Nipbl in mouse liver, SCC2 in yeast*), cohesin subunit (*RAD21*) abolishes a substantial fraction of both cohesin-associated CTCF loops and the TADs. The observation that[27] a subset of TAD boundaries (18% of the initial boundaries identified before CTCF loss) remains after depleting CTCF raises an important question: Do all the TADs completely lose their contact patterns after removal of cohesin? To answer this question, we first analyzed 50kb-resolution CMs from the two cell lines (mouse liver[23] and HCT-116[15]) before and after degradation of *Nipbl* and *RAD21*, respectively (see SI Analyses of the experimental data for details). Using TopDom[34] to identify the TADs in the CMs, we discovered that roughly 16% (26%) of TADs in mouse liver (HCT116) cell are preserved upon *Nipbl* (*RAD21*) removal (Fig.1). In particular, about 659 TADs out of 4176 are preserved (Fig.1a) after removing *Nipbl* in the mouse liver cells. In the HCT-116 cells, 1226 TADs out of 4733 are preserved (Fig.1b) upon *RAD21* loss. Fig.1c and Fig.1d show that the number of preserved TADs (P-TADs) depends on the chromosome type. The fraction of P-TADs is

greater in all the chromosomes in HCT-116 cells than in the mouse liver cells.

CCM simulations reproduce Hi-C maps: To explore the mechanism that results in compartmental domains and CTCF loop domains, we first created the chromosome copolymer model, CCM (SI Fig.1a). We showed previously that simulations based on the CCM[35] reproduced many features of interphase chromosome organization accurately. To ensure that there is no bias in formulating a hypothesis that would account for the P-TADs (Fig.1) in experiments, we calculated the CMs for Chr13, shown in SI Fig.1 (Chr10 in SI Fig.2) from GM12878 cell line. We find that the CCM simulations (SI Fig.1b) reproduce the ubiquitous checker board patterns well. The rectangle in SI Fig.1b represents the border of one such compartment formed primarily by interactions between the B-type loci.

In order to quantitatively compare the CMs from Hi-C and simulations, we transformed the contact maps to Pearson correlation maps, which is used to assess if two loci have correlated interaction profiles (SI Fig.1c). The Kullback-Leibler (KL) divergence between the two probability distributions for the Pearson correlation coefficients, PCC (ρ_{ij}) from simulations and experiments is 0.04 (see SI Fig.1d). We also performed Principal Component Analysis (PCA) on the Pearson correlation matrix to identify the compartment structure. Comparison of the PCA-derived first principal components (PC1) across the Chr13 reveals that A/B compartments observed in the CCM correspond to those found in the experiments (SI Fig.1e).

We then compared the three-dimensional spatial organization between the simulations and experiments using the Ward Linkage Matrix (WLM), which is based on an agglomerative clustering algorithm. The simulated WLM is calculated from a spatial distance map of the organized chromosome, as described in SI Fig.1f-h. We constructed the experimental WLM by converting the Hi-C contact map to a distance map using the approximate relationship[2,35,36], $P_{ij} \propto R_{ij}^{-4.1}$. Here, P_{ij} is the contact probability and R_{ij} is the 3D spatial distance between loci i and j (see SI Data Analyses). The Pearson correlation coefficient ($\rho_{WLM_{exp},WLM_{CCM}}$) between experimental and simulated WLMs is 0.86. The results in SI Fig.1b-h show that the agreement between the CCM simulations and Hi-C data is reasonable, especially considering that (i) error estimates in the Hi-C experiments are essentially unknown, and (ii) more importantly, we did not adjust any parameter to fit the experimental CMs.

Topologically associating domains (TADs): Identification of the TADs depends on the resolution of the Hi-C data, and the method used to annotate them[4,7,8,37,38]. We used the Multi-Chromatin Domain (Multi-CD) method[39] to visualize the TADs and the sub-

TAD structures. SI Fig.1j and m illustrate that a subset of TADs is defined by the CTCF loops, whereas others are not associated with loops. These could arise from segregation between the chromatin states of the neighboring domains, referred to as "compartmental domains" in certain experimental studies[15,40,41]. Snapshots of TADs in SI Fig.1k and n show that they are relatively compact but structurally diverse. The average sizes of the TADs detected using Multi-CD method from Hi-C and simulated contact maps are $\sim 750kbs$ and $\sim 700kbs$, respectively. TopDom[34], with the parameter, w=5, predicts, $\sim 615kbs$ and $\sim 565kbs$, respectively. Overall the emerging picture of the compartment and TAD structures using different methods are qualitatively consistent with each other. From the quantitative comparisons made in SI Fig.1, we surmise that the CCM faithfully reproduces the important features of the Hi-C maps for Chr13 (see SI Fig.2 for Chr10 results).

Loss of CTCF loops leads to an increase in the degree of compartmentalization: The first experimental finding is that cohesin knockdown results in preservation or even enhancement of the compartment structure[15,23,25,28,29]. To ensure that CCM reproduces this finding, we first explored how CTCF loop deletion affects compartmentalization by performing simulations without loop anchors, which is implemented by setting $P_L = 0$ (see Methods). We find that the plaid patterns persist after the deletion of the loops (Fig.2a). Visual comparison between the two CMs in Fig.2a shows that this is indeed the case. Moreover, Fig.2a also shows that finer features, absent when P_L is unity, emerge when P_L=0, which is a qualitative indication of increase in the compartment strength. Pearson correlation maps (lower panel in Fig.2a) illustrate the reduction in the contact profiles between A and B loci (dark-blue color) after loop loss. There is an enhancement in interactions between same type loci (A-A or B-B) (dark-red color). The results in Fig.2a confirm the experimental finding that disruption of the loops not only creates finer features in the compartment but also results in increased number of contacts between loci of the same type, which is simultaneously accompanied by a decrease in the number of A-B interactions.

We then investigated if the enhancement of compartments observed in CMs has a structural basis. An imaging study[42] showed that active loci form larger spatial clusters in cohesin depleted cells. To probe whether this finding is reproduced in our CCM simulations, we first calculated the spatial distance matrix (DM) using 10,000 simulated 3D structures. For individual matrices, with and without CTCF loops, we identified A (B)-dense regions as A (B) clusters, respectively, using the DBSCAN (Density-based spatial clustering of applications with noise) method[43] (see SI Methods for details). Fig.2b shows the number of clusters obtained using DBSCAN. The number of A clusters varied between individual

structures, which reflects the heterogeneity in the chromosome organization. On an average, there are about five A clusters for $P_L = 1$, which decreases to four with $P_L = 0$, which is one indication of the increase in the compartment strength. We also calculated the size of the clusters, which is the average fraction of loci in each cluster in individual structures (right panel in Fig.2c). On an average, the size of A clusters is greater after loop loss. This implies that active clusters merge in space to form larger and more connected clusters upon deleting the loops. In contrast, there is no change in the number and size of B clusters (see SI Fig.3a). This could potentially be due to the larger proportion of B loci compared to the A loci (five-fold) in Chr13 (see SI Fig.3d). Taken together, our observations for A and B clusters show that the A loci cluster is enhanced upon loop loss, which naturally leads to stronger segregation between A and B loci after the loops are deleted.

Relation to experiments: Besides being consistent with imaging experiments[42], we wondered whether the structural changes inferred from the Hi-C maps of both the cell lines could be used to demonstrate the strengthening of compartments upon deleting cohesin. We generated an ensemble of 10,000 3D structures for chromosomes from the two cell lines (mouse liver and HCT-116)[15,23] using the HIPPS method[36]. Consistent with the simulation results, we also observed variations in the spatial organization of A and B loci in individual structures in WT and cohesin depleted cells. DBSCAN analysis identified, on an average, smaller number A/B clusters with larger size in $\Delta Nipbl$ ($\Delta RAD21$) cells compared to WT cells, as shown in Fig.2d for Chr11 and Chr19, and SI Fig.3b (see SI Fig.3c for results of HCT-116).

To validate these structural changes, we constructed Pearson correlation matrices using contact maps calculated from the 3D structures. Figs.2e and f show that cohesin loss induces stronger plaid patterns, which is consistent with simulation results (lower panel of Fig.2a) for Chr13 from GM12878 cell line. Overall, the 3D structural changes demonstrate that the micro-phase separation between active and inactive loci is strengthened with larger A and B physical clusters upon depletion of cohesin, which accords well with Hi-C experiments[42].

Epigenetic mismatch accounts for a large fraction of P-TADs: Most of the TADs are invisible after loop loss, as evidenced by the blurred edges in the contact maps (Fig.3c). Just as in experiments, we find in the CCM simulations that a subset of TAD structures remains even after CTCF loop loss. Fig.3a shows that $\approx 44\%$ (Chr10) and 28% (Chr13) of initial TADs identified in the $P_L = 1$ simulations using TopDom[34] are preserved after loop loss ($P_L = 0$). These values are in line with TAD preservation in chromosomes in HCT-116[15] cells (see Fig.1d). It is likely that the percentages of P-TADs depends on both the resolution

of the Hi-C experiments as well as the algorithm used to identify the TADs. By using the same method to analyze both the simulation results and experimental data, it is hoped that the conclusions would be qualitatively robust.

We used the simulation results to determine the plausible reasons for preservation of some of the TADs by comparing the results for P_L=1 and P_L=0. The first observation is that some TADs, even with CTCF loops, consist mostly of sequences in the same epigenetic state (A or B). Fig.3d compares the fate of one TAD in the region (19.3 -21.8 Mb) in Chr13 between P_L=1 and P_L=0. The highlighted TAD is preserved upon loop loss although the probabilities of contact within this TAD is reduced when $P_L = 0$ (bottom) compared to $P_L = 1$ (top). Their boundaries correspond to a switch in the epigenetic state (the sequence location where the change in the epigenetic states occurs). In contrast, a TAD in Fig.3e, that is present in the WT, is abolished when P_L=0. The disruption of this particular TAD, lacking in at least one epigenetic mismatch, occurs because it can interact more frequently with neighboring TADs composed of similar epigenetic states, which in this case is B-type loci.

The results in Figs.3d and e show that a mismatch in the epigenetic state across a TAD boundary in the WT is likely to result in its preservation after cohesin-associated CTCF loop loss. To test this assertion, we calculated the number of P-TADs that are associated with mismatches in the epigenetic states (see SI Data Analyses and SI Fig.5 for details). We considered the TAD boundary and epigenetic mismatch as overlapping if they are less than 100kb apart, which is reasonable given that Hi-C resolution adopted here is 50kb. By using 100kb as the cut off, we only consider mismatch occurrences that exceed two loci. With this criterion, out of 216 (169) TADs calculated using TopDom 96 (47) are P-TADs for Chr13 (Chr10) (vertical blue bars in Fig.3b in which there are epigenetic mismatches in the WT).

The P-TADs due to epigenetic mismatches are illustrated in Fig.3f, which shows TADs in the 2.5Mbs region in Chr13. Among the three P-TADs, two of them, whose boundaries are marked by dashed blue lines, have epigenetic mismatch across the TAD boundary. These two TADs survive after removal of CTCF loop.

P-TADs have prominent spatial domain boundaries: Because there are a number of P-TADs that are preserved even without epigenetic mismatches across their boundaries, we wondered if the distance matrix, which requires 3D structures, would offer additional insights about P-TADs upon CTCF loop loss. Recent imaging experiments[29,32,44] revealed TAD-like domain structures with spatially segregated heterogeneous conformations in single cells. The boundaries of TAD-like domains, identified from individual 3D structures, only

vary from cell to cell. They do exhibit a preference to reside at specific genomic positions only upon an averaging, as found in the Hi-C experiments. The boundary probability at each locus is the proportion of chromosome structures in which the locus is identified as a domain boundary in 3D space. The locations of prominent peaks in the boundary probability overlap with TADs detected by the population level Hi-C maps.

To explore the relation between P-TADs in ensemble averaged CMs and preferential boundaries in individual 3D structures of chromosomes, we first constructed an individual spatial distance matrix (DM) using 10,000 simulated 3D structures, which were used to identify the single-cell domain boundaries (see SI Data Analyses for details and SI Fig.6). We find preferential domains with high peaks in the boundary probability along the genomic region as well as variations in single-cell domains both in $P_L = 1$ and $P_L = 0$ (see SI Fig.8).

The simulations show that TADs with epigenetic mismatches across the boundary are likely to be preserved after CTCF loop loss. Furthermore, Fig.3f (blue dashed lines) shows that single-cell domain boundaries preferentially reside at the TADs boundaries with epigenetic mismatches, leading to a prominent signature for the ensemble boundary after averaging over a population of cells. Interestingly, the P-TAD has prominent peaks in boundary probability, even without epigenetic mismatch, at the same genomic position as in the CM (green lines in Fig.3f and g). These observations imply that the presence of physical boundaries in the 3D structures can be used to identify P-TADs.

As noted in the experimental study[29], not all preferential boundaries identified in the distance matrices of the WT cells coincide with the TADs detected using the CM. We also find discordance in TAD boundaries identified using TopDom and high peaks in boundary probability. The top panel in Fig.3h (magenta lines) shows that in the (102 - 104.5) Mb range, TopDom predicts that there are three TADs in the WT that are preserved after loop loss (see the top panel with $P_L = 0$). There are two prominent peaks in the WT boundary probability whose boundaries coincide with the TADs predicted by TopDom (see bottom panel with P_L=1 in Fig.3h). But the peak height for the third TAD is very small. At best, one can deduce from the boundary probabilities (compare the results in Fig.3h for P_L=1 and P_L=0) that the middle TAD is preserved, which would be consistent with the TopDom prediction.

We calculated a standardized Z-score for the boundary probability in the genomic region to determine the preferred boundaries in single-cell domains. The number of P-TADs that are accounted for by prominent boundary peaks increases when a smaller Z-score is used. This implies that some P-TADs detected in the CMs using TopDom have weak physical

boundaries in the 3D structures. We considered the maxima, with Z-score values are larger than 0.7, as preferred boundaries in order to determine if P-TADs arise due to the presence of strong physical boundary. With this criterion, we obtained good agreement for the mean length of the TADs detected in the CM using the TopDom method. The averaged sizes of the TADs in Chr13 using TopDom and boundary probability are $\sim 565kbs$ and $\sim 535kbs$, respectively. Quantitative analysis of the boundary probabilities along the genomic region revealed $\approx 66\%$ of the P-TADs in Chr10 and Chr13 have preferential positioning in single-cell domains (green bars in Fig. 3b). We find that most P-TADs with epigenetic mismatches display prominent peaks in the boundary probability ($\approx 85\%$).

Let us summarize the primary lessons from the simulations, which form the basis for analyzing the experiments on chromosomes from mouse-liver and HCT-116 cell lines. (i) Mismatch in the epigenetic state across the TAD boundary is a predominant factor in determining the P-TADs after CTCF loop deletion. (ii) In addition, the presence of physical boundaries in the 3D structures accounts for certain fraction of P-TADs, although in some instances TopDom predictions (used in (i)) are not compatible with boundaries deduced from 3D structures.

P-TADs upon cohesin deletion: In order to assess the extent to which the conclusions from the simulations hold in accounting for the experimental data, we first calculated the number of P-TADs whose boundaries have switches in the A/B epigenetic states in two cell lines (mouse-liver and HCT-116). We assigned chromatin state A (active, red) or B (repressive, blue) by analyzing the combinatorial patterns of histone marks using ChromHMM[45] (see SI analyses of the experimental data and SI Fig.7). An average over the 20 chromosomes shows that 280 P-TADs were associated with a mismatch between A and B epigenetic states upon $\Delta Nipbl$ (blue bar in Fig.4a). The corresponding number of P-TADs, averaged over 23 chromosomes, with epigenetic mismatches is 396 after deleting *RAD21* (blue bar in Fig.4b). Thus, TADs with epigenetic mismatches across their boundaries are preserved with high probability.

In order to make direct comparison with imaging experiments[29], which have examined the consequences of $\Delta RAD21$, we first investigated the heterogeneity in the distance map (DM). The results of our calculations for a 2.5Mbp in Chr21 (34.6-37.1Mb) region from HCT-116 cell line for the WT and $\Delta RAD21$ are presented in Fig.5. The DMs were calculated from the 3D structures that were generated using the HIPPS method with the CMs from Hi-C as inputs. (Fig.5a). The mean DMs for the WT and $\Delta RAD21$ are shown on the left and right panels in Fig.5b. Similar results for Chr 4 in mouse liver cell (Chr2 in HCT-116)

are displayed in SI Fig.9 (SI Fig.10). Several inferences may be drawn from the results in Fig.5. (1)There are large variations in the distance matrices and boundary strengths from cell to cell (Fig.5c and d). This finding is in excellent agreement with imaging data[29]. (2) In both experiments and our calculations, there are TAD-like structures at the single level after removal of *RAD21*, (see the right panel in Fig.5c for the theoretical predictions). (3) The calculated boundary strength distribution (blue histogram in the left panel in Fig. 5d) for the WT is in reasonable agreement with the measured distribution (purple histogram from[29]). Similarly, the calculated and measured boundary strength distributions for $\Delta RAD21$ cells are also in fair agreement (right panel in Fig.5d). Just as shown in experiments[29], we find that the distributions of boundary strength are the same in the WT and in cells without *RAD21*. (4) We also find that the average locus-dependent boundary probability obtained using theory is in very good agreement with the reported experimental data (compare the curves in the left panel in Fig.5e for the WT and the ones on the right for $\Delta RAD21$ cells).

Having demonstrated that the results from the calculated 3D structures are consistent with imaging data[29], we computed the boundary probability from the 10,000 DMs for all the chromosomes in the two cell lines. Comparison of the peak positions in the averaged boundary probabilities and TAD boundaries identified using TopDom shows that they often coincide but there are discrepancies as well (Fig.6 and SI Figs.12 and13). In addition, when there is a mismatch in the epigenetic states, a substantial fraction of P-TADs have high peaks in the boundary probabilities (see Fig.6 and SI Fig.12). As in the simulations, a substantial fraction of P-TADs ($\approx 67\%$) have high peaks in the boundary probabilities. Taken together, the results show that there is an overall qualitative consistency between the predictions using boundary probabilities and the TopDom method. The overall consistency between the conclusions using the calculated and measured distance maps shows that the location of physical boundaries may be computed using only the Hi-C experiments. Finally, analyses of the experiments suggest that the epigenetic state as well as the presence of physical boundary in the 3D structures could be combined to determine the P-TADs in cohesin depleted cells.

Corner dots in P-TADs: Our analyses of the experimental data show that TADs whose borders correspond to both epigenetic mismatch and CTCF loop anchors in the WT cells are preserved after removal of cohesin. However, not all P-TADs have loop anchors at their boundaries (corner dots) in the WT cells (see SI Fig.11). Out of the 280 (396) P-TADs with epigenetic mismatches in mouse liver (HCT-116) chromosomes, 117 (170) did not have corner dots at their boundaries before deleting *Nipbl* ($RAD21$) (Fig.4c and d).

This suggests that some TAD boundaries are formed in the absence of corner dots. It is the underlying epigenetic landscape that is the predominant factor in the formation of such domain boundaries. A similar term "compartment domain" was coined in Rao et al[15].

TADs and P-TADs upon depletion of *Nipbl*: With the near quantitative agreement with experiments, we performed detailed analyses, based on the epigenetic mismatches and boundary probabilities for chromosomes from the mouse liver (Fig.6). The SI contains analyses of the experimental data, and the results for HCT-116 cell are given in SI Fig.12. To illustrate the different scenarios, we consider the 2.5Mbs regions from Chr6 (Fig.6a), Chr7 (Fig.6b), and Chr15 (Fig.6c). (i) For Chr 6, there are three TADs according to TopDom (Fig.6a) in the WT. Upon Δ*Nipbl*, these TADs are abolished (compare the top and bottom panels in Fig 6a). The epigenetic track indicates that the region is mostly in the repressive (B) state. Quantification of the boundary probabilities along the $2.5Mb$ region of Chr6 shows that the TADs also lack physical boundaries upon Δ*Nipbl*. (ii) Examples of P-TADs that satisfy the epigenetic mismatch criterion are given in Fig.6b. Using TopDom, we identified several TADs (top panel in Fig.6b) in this region of Chr7. It is interesting that the boundary probabilities obtained from the HIPPS-generated distance matrices are also large when there is a mismatch in the epigenetic states. In the Δ*Nipbl* case, there are P-TADs that are accounted for by both the epigenetic mismatch and boundary probabilities (see the dashed blue lines in Fig.6b). Two TADs in the WT (the ones on the right in the upper panel in Fig.6b) merge to form a single TAD in the Δ*Nipbl*. This observation is in accord with expectation based on epigenetic mismatch, whose corollary is that if there is a TAD within a region that contains predominantly A or B type loci they ought to merge upon Δ*Nipbl*. (iii) In the $2.5Mb$ region of Chr15, there are three TADs in the WT (top panel in Fig.6c). The first and the third TADs have an epigenetic mismatch at only one boundary (blue dashed line), and the expectation is that they would not be preserved upon *Nipbl* removal. However, the boundary probabilities show that the TADs have physical boundaries in both, and thus they are preserved. Taken together, the results in Fig.6 show that by combining the epigenetic mismatches and the boundary probabilities one can account for a number of P-TADs.

III. DISCUSSION

By analyzing the experimental Hi-C data[15,23], we showed even after cohesin loss a non-negligible fraction of TADs is preserved, just is the case when CTCF is deleted in mESCs[27].

To examine the factors that control the P-TADs, we first performed polymer simulations of two chromosomes in the presence and absence of CTCF loops from the GM12878 cell line. The simulations for chromosomes from a cell line, not used in the experiments[15,23], reproduced the salient findings noted in the experiments. The key findings in the simulations are: (1) Deletion of loops results in the emergence of finer compartment details. In addition, the compartments are strengthened. (2) Mismatches in the epigenetic states across the TAD boundary account for a large fraction of P-TADs. Moreover, some TADs, that lack epigenetic mismatches, are preserved as revealed by the presence of physical boundaries in the 3D structures.

Informed by the simulation results, we analyzed the experimental Hi-C data on chromosomes from two cell lines (mouse liver and HCT-116)[15,23]. Analyses of the structures determined from the Hi-C map using HIPPS[36], with and without *Nipbl* or *RAD*21, showed that A-type loci form larger spatial clusters after cohesin removal, consistent with enhancement of compartments inferred from Hi-C CMs. In addition, we found that a substantial fraction of P-TADs can be accounted for by the epigenetic mismatches and the preferential boundaries in 3D structures of chromosomes. Thus, a combination of the two perspectives accounts for the experimental observations pertaining to P-TADs upon cohesin loss[15,23]. Most of the P-TADs with epigenetic mismatches in the CMs have a prominent peak in the boundary probability, which is an indicator of a strong physical boundary in 3D structures. If this interpretation is correct then we are forced to conclude that not only micro-phase separation on the larger scale, L_C but also some special TADs on a shorter scale, L_{TAD} are encoded in the epigenetic sequence. We can only speculate that P-TADs must be functionally significant.

Remarkably, the conclusion that there are cell-to-cell variations in the distance maps, noted in imaging experiments[29], are also fully affirmed in the calculated 3D structures using the Hi-C contact maps as input. In addition, the calculations also show that in both mouse liver and HCT-116 cell lines TAD-like structures are present after removing *Nipbl* and *RAD21*. We established that the Hi-C data sets can be be utilized to generate 3D structures using HIPPS, which allows us to examine the impact of perturbations on three dimensional organization of interphase chromosomes at the single cell level. These results are significant because only a limited number of loci can be directly imaged whereas Hi-C data can be routinely generated at higher resolution.

Comments on the Methods: In order to explore the factors that control the P-TADs, there are two assumptions. (1) The results of the Hi-C experiments are taken at face value

in the sense the reported CMs are assumed to be accurate. We ought to view this as an assumption because errors in the Hi-C readouts may be difficult to evaluate even though such experiments are invaluable[46]. (2) The TADs were identified using TopDom, one of many TAD callers. A recent survey[38] shows that, although the finding that TADs are hierarchically organized domains is robust, there are substantial variations in the identification of these domains predicted by different methods. Although TopDom fairs reasonably well in comparison to other methods, there is no guarantee that it identifies the TAD location or the number of TADs accurately. It is only for convenience that we used TopDom as the reference to which the results using the boundary probabilities are compared.

In a previous study[27], which was the basis of the current investigation, the fraction of P-TADs in mESCs was calculated using insulation score (IS). We used TopDom to reanalyze the Hi-C maps in presence of and absence of CTCF protein (see SI Fig.14). Although the precise number of predicted P-TADs depends on the method, the percentage of P-TADs (roughly 19%) does not. Thus, despite inherent errors in the data and limitations in the method of analyses, it is noteworthy that the conclusion that (a) compartment strength is preserved (or enhanced), and (b) a non-negligible number of TADs are retained when architectural proteins or the motor proteins are deleted is robust.

IV. METHODS

Simulations: We first used simulations of the Chromatin Copolymer Model (CCM) in order to deduce the reasons for preservation of certain TADs when the loop anchors are deleted. In order to formulate a meaningful hypothesis, the simulations must reproduce the two major findings[15,23]: (a) propensity of the A and B loci to segregate is enhanced when cohesin loading factors are depleted; (b) a fraction of TADs should be preserved upon cohesin or CTCF loop loss. To this end, we adopted the CCM that accounts for the epigenetic landscape to simulate interphase chromosomes[35]. Each locus is either A type (active locus is in red) or B type (repressive locus is in blue) (SI Fig.1a). The locus type is determined using the Broad ChromHMM track[47–49]. Following our previous study[35], we adjusted the interaction energies between the A-B pairs, based on the Flory-Huggins theory, which gives the criterion for micro-phase separation between A and B. As formulated here and elsewhere[35], the CCM has only one energy scale, thus making it the simplest polymer model for interphase chromosomes. The energy scale, describing the A-B interaction, is chosen to ensure that phase segregation occurs on large length ($>$ 1Mbp) genomic length

scale.

In the previous study[35], the locus size was 1.2kb, which allowed us to simulate about 10Mbp of chromosomes. In order to simulate the entire length of the chromosomes, we developed a reduced resolution CCM in which the size of each locus is 50,000 bps (50kb). The locations of the loop anchors are obtained from the Hi-C data[4]. In the Wild type (WT) simulations, the probability of loop anchor, P_L being present is unity. To model cohesin deletion, we set $P_L = 0$ to assess the impact of deleting the loops on compartments and TADs. Using the two values of P_L (0 and 1), we performed Langevin dynamics simulations (see SI Methods for details) to construct the ensemble of chromosome structures.

To avoid biases in the formulation of the hypothesis to explain TAD preservation, we simulated chromosomes from the Human Cell line (GM12878), which is different from the cell lines used in experiments[15,23] to assess the consequences of cohesin loading factors. In the main text, we report results for Chr13 (19-115.10 Mbp). The total number of $50kb$ loci is $N = 1923$, and the total number of loop anchor pairs is 72. To ensure that the results are robust, we also simulated (see the SI Information) Chr10. The details of the simulations and data analyses are described in the SI.

Acknowledgements: We are grateful to Elphege P. Nora for providing the Hi-C maps for mESCs. This work was supported by the National Science Foundation (CHE 19-00033) and the Welch Foundation (F-0019) through the Collie-Welch Regents Chair.

---

[1] Fraser, J.; Williamson, I.; Bickmore, W. A.; Dostie, J. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* **2015**, *79*, 347–372.

[2] Wang, S.; Su, J.-H.; Beliveau, B. J.; Bintu, B.; Moffitt, J. R.; Wu, C.-t.; Zhuang, X. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **2016**, *353*, 598–602.

[3] Boettiger, A. N.; Bintu, B.; Moffitt, J. R.; Wang, S.; Beliveau, B. J.; Fudenberg, G.; Imakaev, M.; Mirny, L. A.; Wu, C.-T.; Zhuang, X. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **2016**, *529*, 418–422.

[4] Rao, S.; Huntley, M.; Durand, N.; Stamenova, E.; Bochkov, I.; Robinson, J.; Sanborn, A.; Machol, I.; Omer, A.; Lander, E.; Aiden, E. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **2014**, *159*, 1665–1680.

[5] Dekker, J.; Marti-Renom, M. A.; Mirny, L. A. Exploring the three-dimensional organization of

genomes: interpreting chromatin interaction data. *Nat. Rev. Genet* **2013**, *14*, 390–403.

[6] Lieberman-Aiden, E. et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **2009**, *326*, 289–293.

[7] Jesse R. Dixon, S. S.; Feng Yue, A. K.; Yan Li, Y. S.; Ming Hu, J. S. L.; Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012**, *485*, 376–380.

[8] Szabo, Q.; Bantignies, F.; Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **2019**, *5*, eaaw1668.

[9] Dowen, J.; Fan, Z.; Hnisz, D.; Ren, G.; Abraham, B.; Zhang, L.; Weintraub, A.; Schuijers, J.; Lee, T.; Zhao, K.; Richard A, *Cell* **2014**, *159*, 374–387.

[10] Özdemir, I.; Gambetta, M. C. The role of insulation in patterning gene expression. *Genes* **2019**, *10*, 767.

[11] Hnisz, D.; Day, D. S.; Young, R. A. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **2016**, *167*, 1188 – 1200.

[12] Ji, X.; Dadon, D.; Powell, B.; Fan, Z.; Borges-Rivera, D.; Shachar, S.; Weintraub, A.; Hnisz, D.; Pegoraro, G.; Lee, T.; Misteli, T.; Jaenisch, R.; Young, R. *Cell Stem Cell* **2016**, *18*, 262–275.

[13] Lupiáñez, D. G.; Kraft, K.; Heinrich, V.; Krawitz, P.; Brancati, F.; Klopocki, E.; Horn, D.; Kayserili, H.; Opitz, J. M.; Laxova, R., et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **2015**, *161*, 1012–1025.

[14] Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **2016**, *351*, 1454–1458.

[15] Rao, S. S.; Huang, S.-C.; St Hilaire, B. G.; Engreitz, J. M.; Perez, E. M.; Kieffer-Kwon, K.-R.; Sanborn, A. L.; Johnstone, S. E.; Bascom, G. D.; Bochkov, I. D., et al. Cohesin loss eliminates all loop domains. *Cell* **2017**, *171*, 305–320.

[16] Ren, B.; Dixon, J. A CRISPR Connection between Chromatin Topology and Genetic Disorders. *Cell* **2015**, *161*, 955 – 957.

[17] Bianco, S.; Lupiáñez, D. G.; Chiariello, A. M.; Annunziatella, C.; Kraft, K.; Schöpflin, R.; Wittler, L.; Andrey, G.; Vingron, M.; Pombo, A., et al. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet* **2018**, *50*, 662–667.

[18] Fudenberg, G.; Imakaev, M.; Lu, C.; Goloborodko, A.; Abdennur, N.; Mirny, L. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **2016**, *15*, 2038 – 2049.

[19] Kim, Y.; Yu, H. Shaping of the 3D genome by the ATPase machine cohesin. *Exp. Mol. Med.* **2020**, *52*, 1891–1897.

[20] Phillips, J. E.; Corces, V. G. CTCF: master weaver of the genome. *Cell* **2009**, *137*, 1194–1211.

[21] Guo, Y.; Xu, Q.; Canzio, D.; Shou, J.; Li, J.; Gorkin, D. U.; Jung, I.; Wu, H.; Zhai, Y.; Tang, Y., et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **2015**, *162*, 900–910.

[22] Vietri Rudan, M.; Barrington, C.; Henderson, S.; Ernst, C.; Odom, D.; Tanay, A.; Hadjur, S. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep.* **2015**, *10*, 1297 –1309.

[23] Schwarzer, W.; Abdennur, N.; Goloborodko, A.; Pekowska, A.; Fudenberg, G.; Loe-Mie, Y.; Fonseca, N. A.; Huber, W.; Haering, C. H.; Mirny, L., et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **2017**, *551*, 51–56.

[24] Nuebler, J.; Fudenberg, G.; Imakaev, M.; Abdennur, N.; Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E6697–E6706.

[25] Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **2017**, *36*, 3573–3599.

[26] Haarhuis, J. H.; van der Weide, R. H.; Blomen, V. A.; Yáñez-Cuna, J. O.; Amendola, M.; van Ruiten, M. S.; Krijger, P. H.; Teunissen, H.; Medema, R. H.; van Steensel, B.; Brummelkamp, T. R.; de Wit, E.; Rowland, B. D. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension". *Cell* **2017**, *169*, 693–707.

[27] Nora, E. P.; Goloborodko, A.; Valton, A.-L.; Gibcus, J. H.; Uebersohn, A.; Abdennur, N.; Dekker, J.; Mirny, L. A.; Bruneau, B. G. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **2017**, *169*, 930–944.

[28] Zuin, J.; Dixon, J. R.; van der Reijden, M. I. J. A.; Ye, Z.; Kolovos, P.; Brouwer, R. W. W.; van de Corput, M. P. C.; van de Werken, H. J. G.; Knoch, T. A.; van IJcken, W. F. J.; Grosveld, F. G.; Ren, B.; Wendt, K. S. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 996–1001.

[29] Bintu, B.; Mateo, L. J.; Su, J.-H.; Sinnott-Armstrong, N. A.; Parker, M.; Kinrot, S.; Yamaya, K.; Boettiger, A. N.; Zhuang, X. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **2018**, *362*, eaau1783.

[30] de Wit, E.; Nora, E. P. New insights into genome folding by loop extrusion from inducible degron technologies. *Nat. Rev. Genet* **2022**, 1–13.

[31] Conte, M.; Fiorillo, L.; Bianco, S.; Chiariello, A. M.; Esposito, A.; Nicodemi, M. Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nat. Commun* **2020**, *11*, 3289–3289.

[32] Xie, W. J.; Qi, Y.; Zhang, B. Characterizing chromatin folding coordinate and landscape with deep learning. *PLoS Comput. Biol.* **2020**, *16*, 1–19.

[33] Hansen, A. S.; Pustova, I.; Cattoglio, C.; Tjian, R.; Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *ELife* **2017**, *6*, e25776.

[34] Shin, H.; Shi, Y.; Dai, C.; Tjong, H.; Gong, K.; Alber, F.; Zhou, X. J. TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* **2015**, *44*, e70–e70.

[35] Shi, G.; Liu, L.; Hyeon, C.; Thirumalai, D. Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nat. Commun* **2018**, *9*, 3161.

[36] Shi, G.; Thirumalai, D. From Hi-C Contact Map to Three-Dimensional Organization of Interphase Human Chromosomes. *Phys. Rev. X.* **2021**, *11*, 011051.

[37] Rowley, M. J.; Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet* **2018**, *19*, 789–800.

[38] Zufferey, M.; Tavernari, D.; Oricchio, E.; Ciriello, G. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol.* **2018**, *19*, 1–18.

[39] Bak, J. H.; Kim, M. H.; Liu, L.; Hyeon, C. A unified framework for inferring the multi-scale organization of chromatin domains from Hi-C. *PLoS Comput. Biol.* **2021**, *17*, 1–27.

[40] Rowley, M. J.; Nichols, M. H.; Lyu, X.; Ando-Kuri, M.; Rivera, I. S. M.; Hermetz, K.; Wang, P.; Ruan, Y.; Corces, V. G. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* **2017**, *67*, 837–852.

[41] Beagan, J. A.; Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat. Genet* **2020**, *52*, 8–16.

[42] Xie, L.; Dong, P.; Qi, Y.; Hsieh, T.-H. S.; English, B. P.; Jung, S.; Chen, X.; De Marzio, M.; Casellas, R.; Chang, H. Y.; Zhang, B.; Tjian, R.; Liu, Z. BRD2 compartmentalizes the accessible genome. *Nat. Genet* **2022**, *54*, 481–491.

[43] Sander, J.; Ester, M.; Kriegel, H.; Xu, X. Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Min Knowl Discov* **1998**, *2*, 169–194.

[44] Cheng, Y.; Liu, M.; Hu, M.; Wang, S. TAD-like single-cell domain structures exist on both active and inactive X chromosomes and persist under epigenetic perturbations. *Genome Biol.* **2021**, *22*, 1–26.

[45] Ernst, J.; Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **2012**, *9*, 215–216.

[46] Akgol Oksuz, B.; Yang, L.; Abraham, S.; Venev, S. V.; Krietenstein, N.; Parsi, K. M.; Ozadam, H.; Oomen, M. E.; Nand, A.; Mao, H., et al. Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **2021**, *18*, 1046–1055.

[47] Rosenbloom, K. R.; Sloan, C. A.; Malladi, V. S.; Dreszer, T. R.; Learned, K.; Kirkup, V. M.; Wong, M. C.; Maddren, M.; Fang, R.; Heitner, S. G., et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* **2012**, *41*, D56–D63.

[48] Ernst, J.; Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol* **2010**, *28*, 817–825.

[49] Ernst, J.; Pouya Kheradpour, T. S. M.; Noam Shoresh, L. D. W.; Charles B. Epstein, X. Z.; Li Wang, R. l.; Coyne, M.; Manching Ku, T. D.; Kellis, M.; Bernstein, B. E. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **2011**, *473*, 43–49.

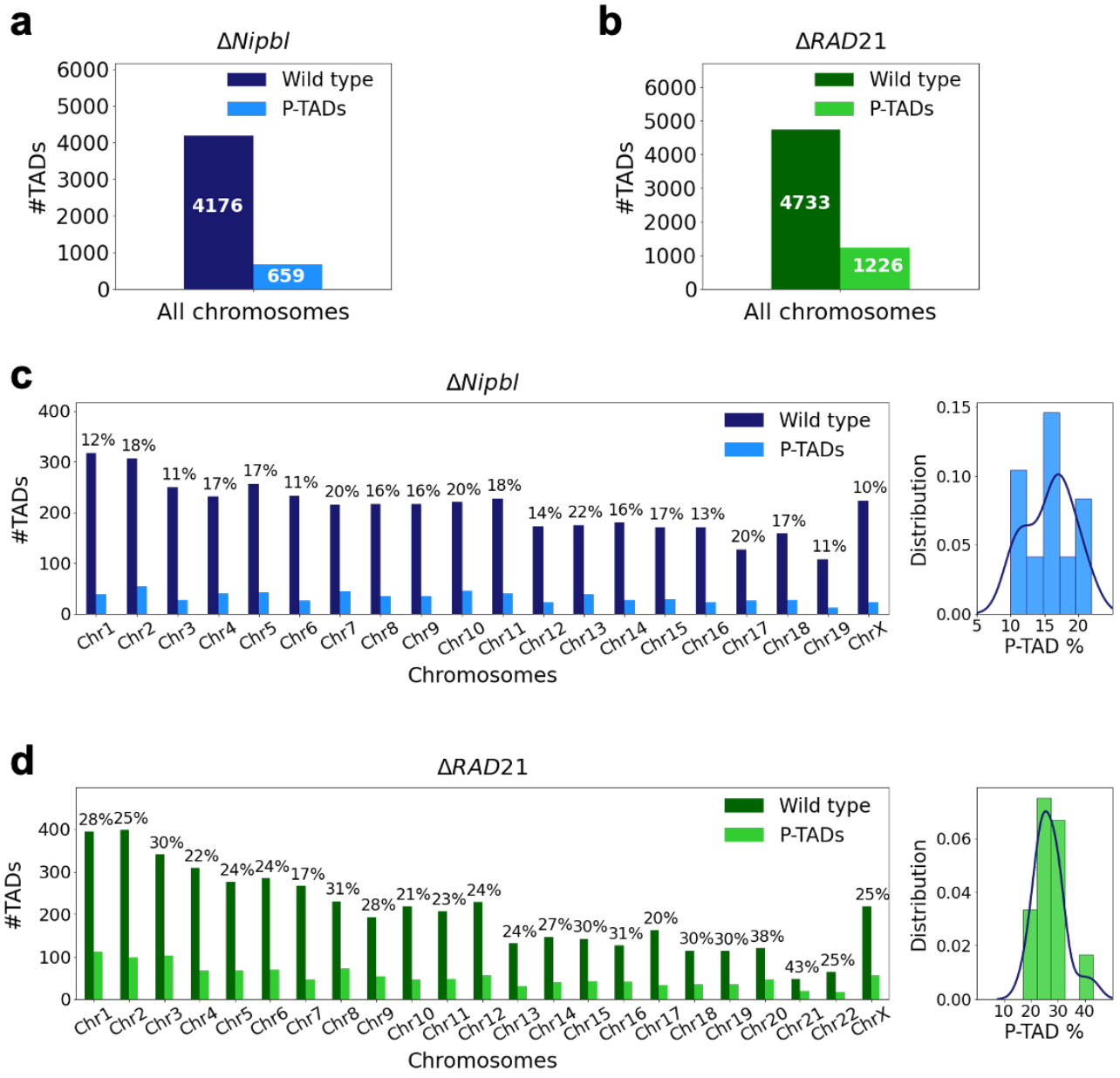

FIG. 1. **Fate of the TADs in chromosomes upon cohesin deletion.**

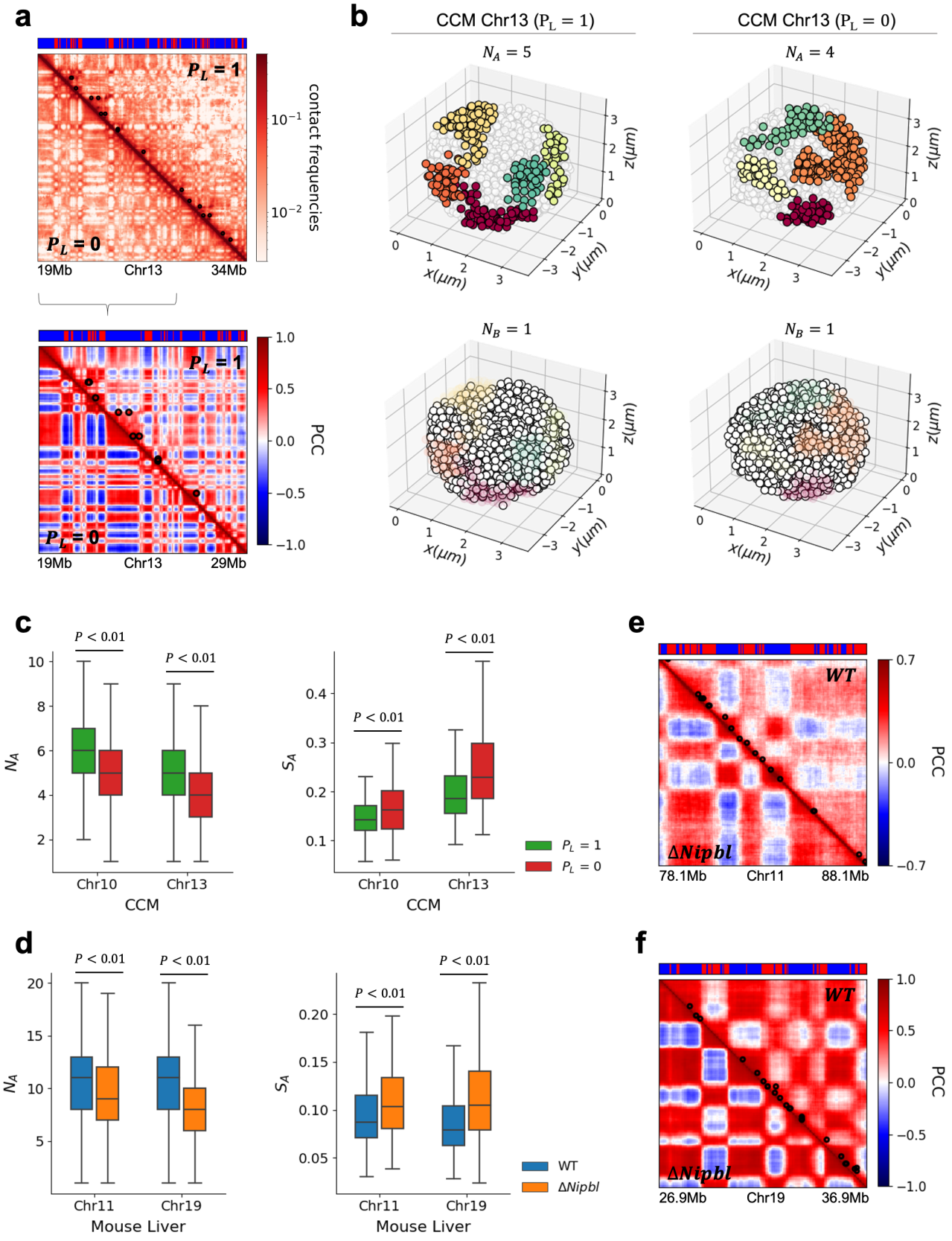

FIG. 2. **Impact of CTCF loop anchor loss on chromosome organization.**

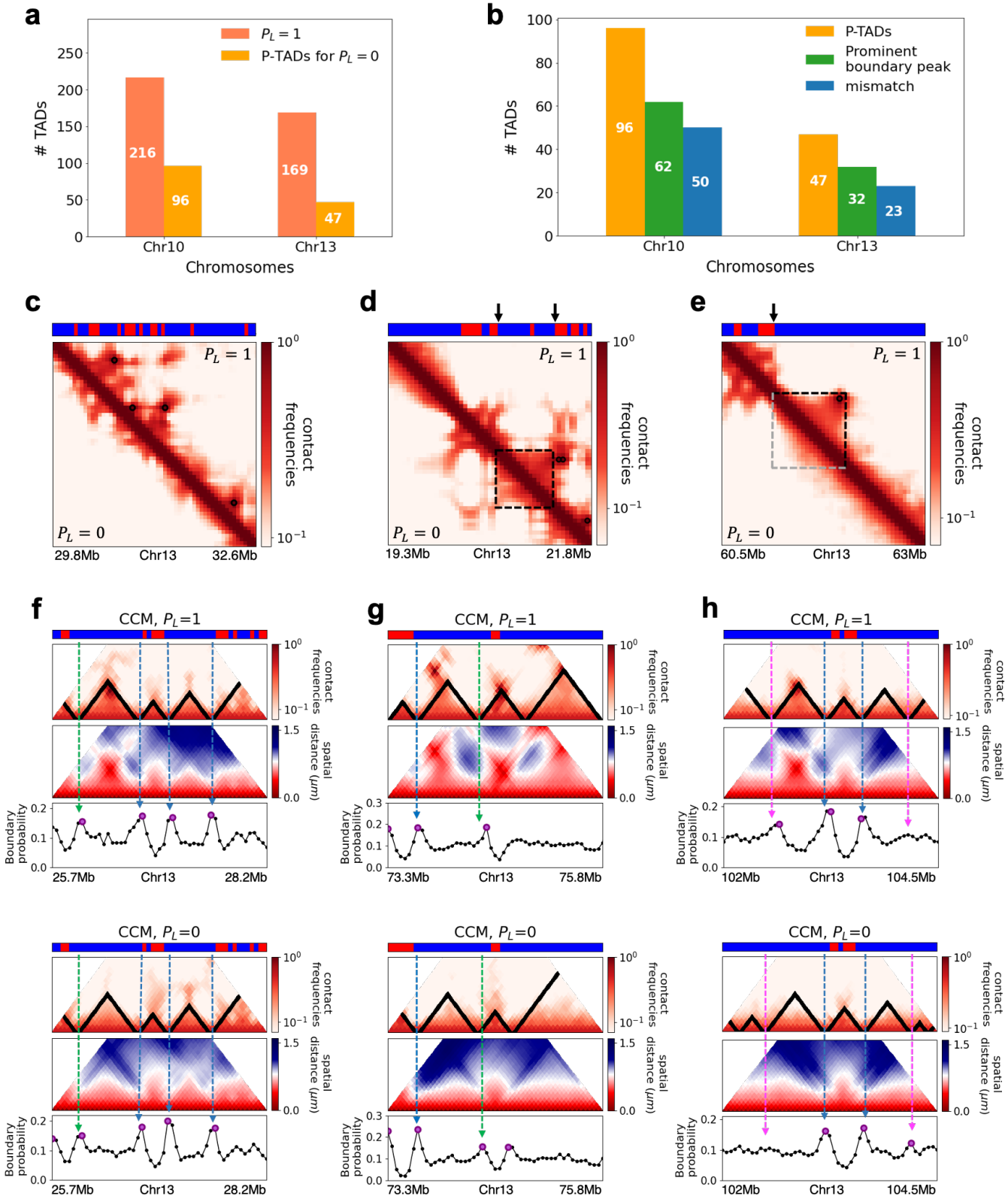

FIG. 3. **CCM simulations reveal characteristics of P-TADs.**

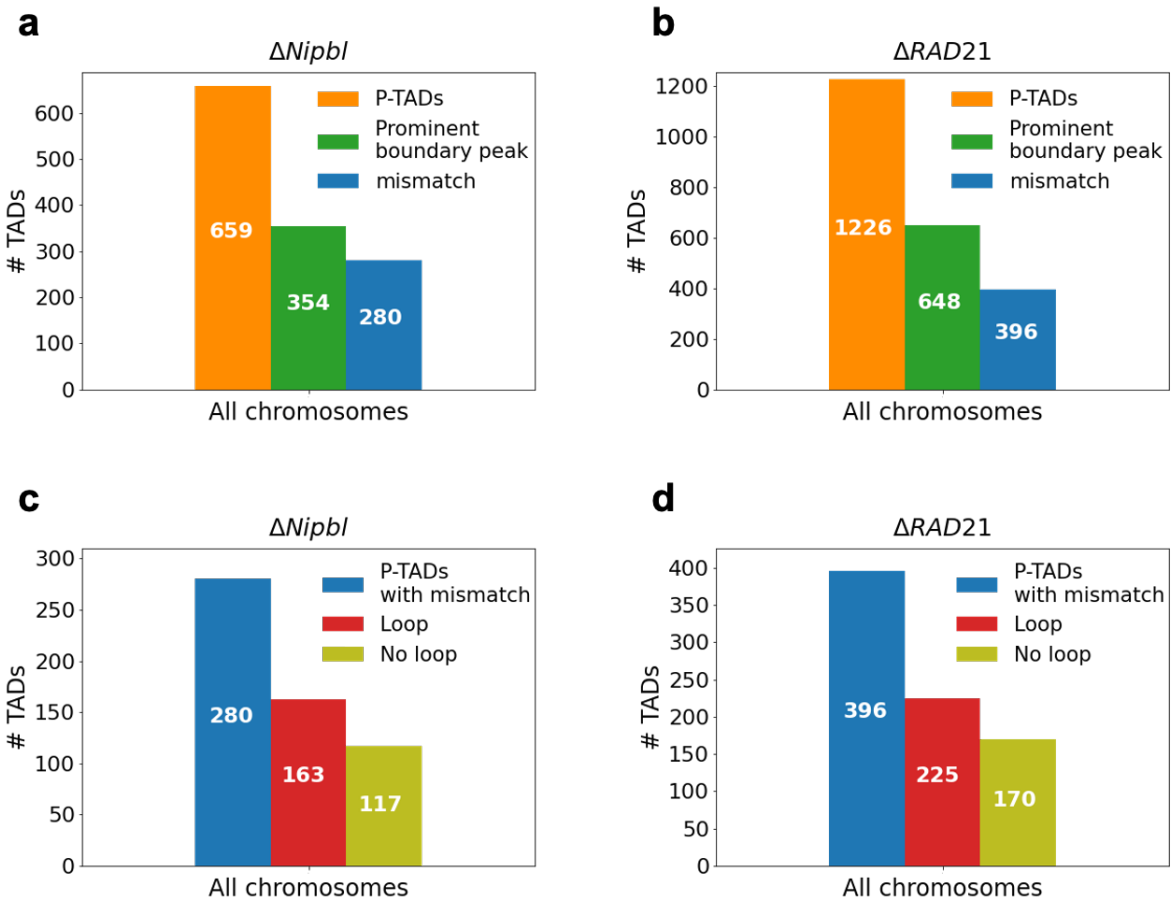

FIG. 4. **Classification of P-TADs from Hi-C maps from two cell lines.**

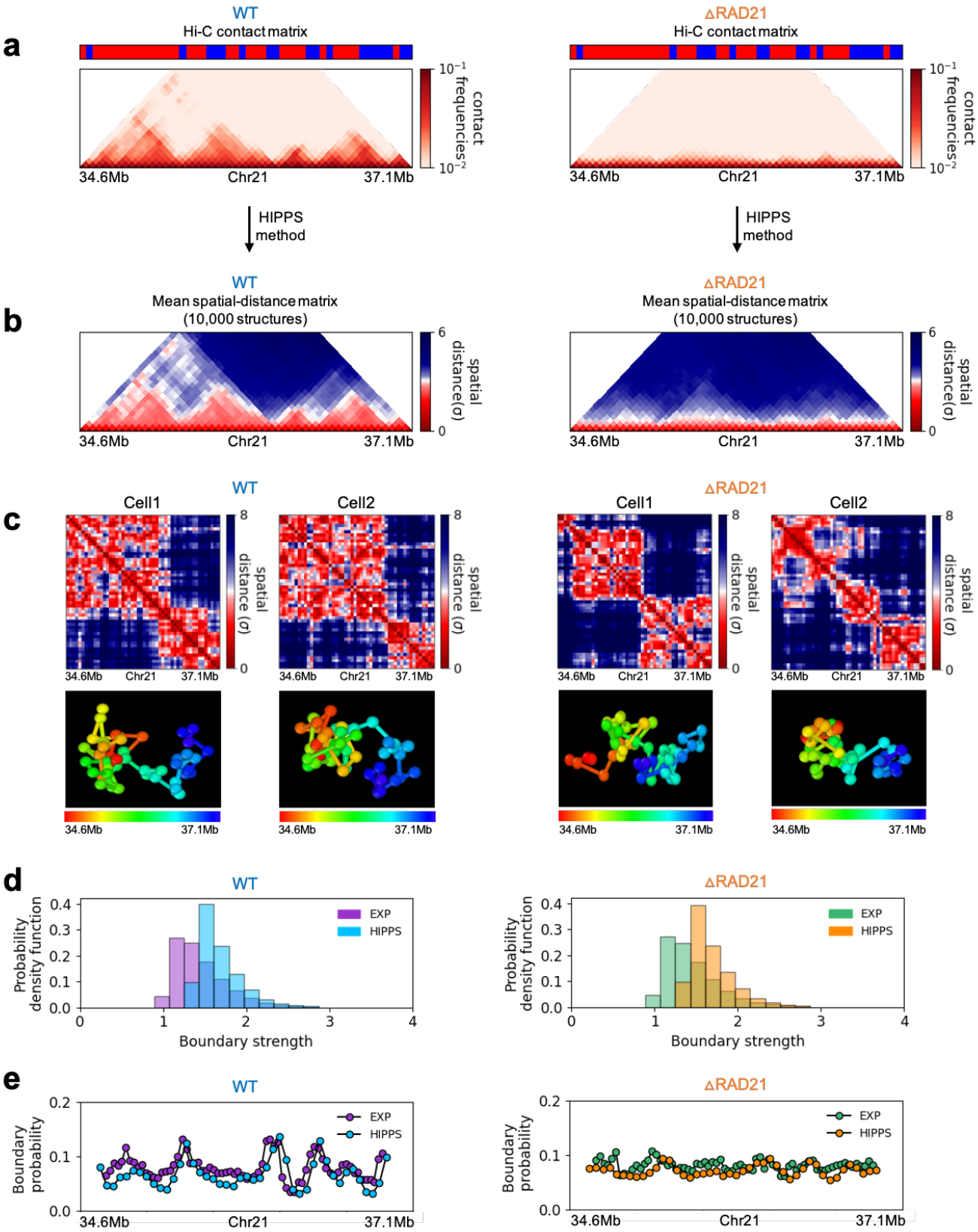

FIG. 5. **Calculated 3D structures from Hi-C produce TAD-like structures found in imaging experiments**

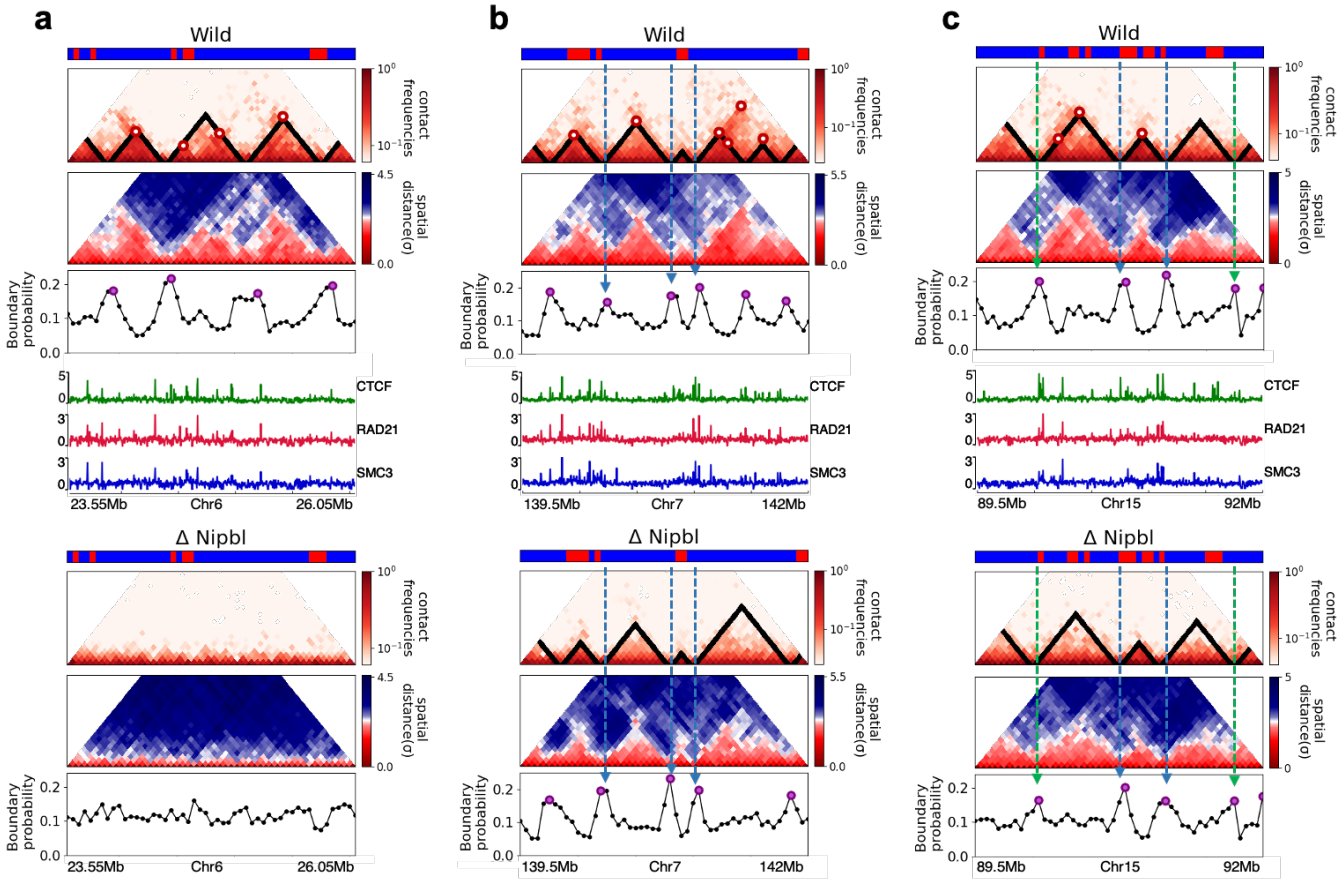

FIG. 6. **Fate of TADs after Δ*Nipbl* in mouse liver cells.**

Fig1. Fate of the TADs in chromosomes upon cohesin deletion. (a) The number of TADs in all the chromosomes, identified by the TopDom method[34], in the wild type (WT) cells and the number of preserved TADs (P-TADs) after deleting cohesin loading factor (ΔNipbl) in mouse liver. The percentage of P-TADs is $\approx$ 16%. (b) Same as (a) except the experimental data are analyzed for HCT-116 cell before (WT) and after RAD21 deletion. The percentage of P-TADs is 26%. (c) The total number of TADs and the number of P-TADs for each chromosome calculated using the mouse-liver Hi-C data. The number above each bar is the percentage of P-TADs in each chromosome. (d) Same as (c) except the results are for chromosomes from the HCT-116 cell line. The percentage of P-TADs is greater in the HCT-116 cell line than in mouse-liver for almost all the chromosomes, a feature that is more prominent in the distribution of P-TADs proportions (Right).

Fig2. Impact of cohesin/CTCF loops loss on the organization of chromosomes. (a) Comparison between simulated contact maps using CCM (19-34Mb, upper panel) and Pearson correlation maps (19-29Mb, lower panel) for Chr13 (GM12878 cell line). Upper triangle (lower triangle) was calculated with (without) CTCF loops. The black circles in the upper triangle are the positions of the CTCF loop anchors detected in the Hi-C experiment[4]. The bar on top marks the epigenetic states with red (blue) representing active (repressive) loci. Upon CTCF loop loss, the plaid patterns are more prominent, and finer details of the compartment organization emerge. (b) 3D snapshots of A and B clusters identified using the DBSCAN algorithm with $P_L = 1$ (left panel) and $P_L = 0$ (right panel) computed from simulations of Chr13 with and without loops, respectively. Five A clusters (Upper panel; Red, orange, yellow, dark-green, light-green) and one B cluster (Lower panel; white) are detected in this 3D structure with P_L =1. Four A clusters and one B cluster are detected for P_L =0. The size of a locus $\sigma_{50K} \approx$ 243nm[35]. (c) Box plot of the number (upper) and average size (lower) of A clusters determined using 10,000 individual 3D structures for P_L =1 and P_L =0 for simulated Chr10 and Chr13. The size of the A cluster, S_A is defined as, (the number of A loci within the cluster)/(the total number of A loci within the chromosome). Boxes depict median and quartiles. The black line with caps describes the range of values in the number and size. Loop loss creates a smaller number (enhancement in compartment strength) of A-type clusters whose sizes are larger. Two-sided Mann-Whitney U test was performed for the statistical analysis. (d) Same as (c) except the results were determined using 10,000 3D structures generated with the HIPPS method from the experimental Chr11 and Chr19 CMs from mouse liver for the WT and Δ*Nipbl*[23], respectively. The number of A clusters decreases by 18% and 27% after *Nipbl* loss in Chr11 and Chr19,

respectively. (e)-(f) Pearson correlation matrix derived from 3D structures for Chr11 and Chr19 of mouse liver, respectively. Two loci, separated by a distance smaller than 1.75σ are in contact (σ is the mean distance between i and $i+1$ loci for WT and Δ*Nipbl*, respectively). The black circles in the upper triangle are loop anchors detected in Hi-C map[23] using HiCCUPS[4].

Fig3. CCM simulations reveal characteristics of P-TADs. (a) The number of TADs in the simulated Chr10 and Chr13 chromosomes identified by the TopDom method[34] for P_L=1. The number of P-TADs after CTCF loop depletion (P_L=0) is also shown. (b) The number of P-TAD with epigenetic mismatches (blue) and those identified by the peaks in the boundary probability. (c)-(e) Comparison between CMs for the region of Chr13 with upper (lower) triangle with $P_L = 1$ ($P_L = 0$). The black circles at the corner of the TADs are the CTCF loop anchors. The bar above the contact map are the epigenetic states with red (blue) representing A (B) loci. Arrows above the bar show the epigenetic switch. (c) After loop deletion, TAD structures disappear. (d) TAD whose boundaries are marked by epigenetic mismatches are preserved. (e) TAD lacking at least one epigenetic mismatch is disrupted after CTCF loop loss. (f)-(h) Comparison of the CM and mean spatial-distance matrices for the 2.5Mb genomic regions (25.7-28.2Mbp,73.3-75.8Mbp and 102-104.5Mbp, respectively) with (upper) and without (lower) CTCF loop. Bottom graph shows the boundary probability, with the high values indicating population averaged TAD boundary. Purple circles in the boundary probability graph represent the preferred boundaries. A subset of P-TADs boundaries match epigenetic mismatches (blue lines). P-TADs with high boundary probability is in green line. The magenta line describes P-TADs which is not accounted by epigenetic mismatch or physical boundary in 3D space but are found using the TopDom method.

Fig4. Classification of P-TADs from Hi-C maps from two cell lines. (a) The number of P-TADs in all the chromosomes (orange bar taken from Fig. 1a) that are accounted for by epigenetic mismatches (blue bar) and peaks in the boundary probability (green bar) after *Nipbl* loss in mouse liver. (b) Same as (a) except the analyses is done using experimental data are for HCT-116 cell after Δ*RAD21*. (c) The number of P-TADs (after Δ*Nipbl*) whose boundaries coincide with both epigenetic mismatches and corner dots (CTCF loop anchors) (red color) and only epigenetic mismatches (olive color) in the WT chromosomes from mouse liver. (d) Same as (c) except the results are obtained using experimental data from HCT-116 cell.

Fig5. Calculated 3D structures from Hi-C produce TAD-like structures found in imaging experiments: On the left (right) panels are results for WT (Δ*RAD21*)

for (Chr21:34.6-37.1Mb). For visualization purposes we adopted the color scheme used in the imaging study[29]. (a) Hi-C contact maps with (left) and without Δ*RAD21*[15]. (b) Mean distance matrices calculated from HIPPS-generated 3D strictures. (c) Examples of calculated single cell distance matrices with (left) and without (right) *RAD21*. Schematic of structures for the two cells under the two conditions are given below. (d) Distribution of the boundary strengths, describing the steepness in the changes in the spatial distance across the boundaries. The left (right) is for the WT (Δ*RAD21*) cells. The blue (purple) histogram is calculated using HIPPS (experiments). (e) Position dependent boundary probability for the WT (left) *RAD21* deleted cells (right). The curve in blue (purple) is the calculated (measured) boundary probability for the WT cells. The orange (green) curve is from the calculations (experiments). The plots show that the location of prominent peaks in the calculated boundary probability are in excellent agreement with experiments for the WT cells (left panel). Without RAD21, high peaks are absent in both the cases (right panel). The calculated results were obtained *without* adjusting any parameter in the theory.

Fig6. Fate of TADs after Δ*Nipbl* in mouse liver cells. (a) Complete Loss (Chr6:23.55Mb-26.05Mb) of TADs. (b)-(c) P-TADs (Chr7:139.5Mb-142Mb and Chr15:89.5Mb-92Mb). In all cases, the plots below the scale on top, identifying the epigenetic states[45], compare 50kb-resolution Hi-C CMs for the genomic regions of interest with *Nipbl* (upper) and without *Nipbl* (lower). Mean spatial-distance matrices, obtained from the Hi-C contact matrices using the HIPPS method[36], are below the CMs. The dark-red circles at the boundaries of the TADs in the CMs are loop anchors detected using HiCCUPS[4]. ChIP-seq tracks for CTCF, RAD21 and SMC3 in the WT cells[23] illustrate the correspondence between the locations of the most detected loop anchors and the ChIP-seq signals. Bottom plots give the probabilities that each genomic position is at a single-cell domain boundary in the specified regions. Purple circles in the boundary probability graph represent the physical boundaries. A subset of physical boundaries in P-TADs coincide with epigenetic mismatches (blue lines), indicating that the probabilities of contact at these boundaries are small. P-TADs in (c), demarcated by green lines, have high peaks in the boundary probability without epigenetic mismatch.