MIT Open Access Articles

Cohesin Loss Eliminates All Loop Domains

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation	Rao, Suhas S.P. et al. "Cohesin Loss Eliminates All Loop Domains." Cell 171, 2 (October 2017): 305–320 © 2017 Elsevier Inc
As Published	http://dx.doi.org/10.1016/J.CELL.2017.09.026
Publisher	Elsevier
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/118942
Terms of Use	Creative Commons Attribution-NonCommercial-NoDerivs License
Detailed Terms	http://creativecommons.org/licenses/by-nc-nd/4.0/

Massachusetts Institute of Technology

DSpace@MIT

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2018 October 05.

Published in final edited form as:

Cell. 2017 October 05; 171(2): 305–320.e24. doi:10.1016/j.cell.2017.09.026.

Cohesin loss eliminates all loop domains

Suhas S. P. Rao[1,2,3], **Su-Chen Huang**[1,2], **Brian Glenn St Hilaire**[1,2,4], **Jesse M. Engreitz**[5], **Elizabeth M. Perez**[5], **Kyong-Rim Kieffer-Kwon**[6], **Adrian L. Sanborn**[1,4,7], **Sarah E. Johnstone**[5,8], **Gavin D. Bascom**[9], **Ivan D. Bochkov**[1,2], **Xingfan Huang**[1,10], **Muhammad S. Shamim**[1,2,10], **Jaeweon Shin**[1,10], **Douglass Turner**[1,11], **Ziyi Ye**[1,10], **Arina D. Omer**[1,2], **James T. Robinson**[1,5,11], **Tamar Schlick**[9,12,13], **Bradley E. Bernstein**[5,8], **Rafael Casellas**[6,14], **Eric S. Lander**[5,15,16], and **Erez Lieberman Aiden**[1,2,4,5,10,*]

[1]The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA.

[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

[3]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.

[4]Center for Theoretical Biological Physics, Rice University, Houston, TX 77030, USA.

[5]Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA.

[6]Genomics & Immunity, NIAMS, NIH, Bethesda, MD 20892, USA.

[7]Department of Computer Science, Stanford University, Stanford, CA 94305, USA.

[8]Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA.

[9]Department of Chemistry, New York University, New York, NY 10003, USA.

[10]Departments of Computer Science and Computational and Applied Mathematics, Rice University, Houston, TX 77030, USA.

[11]Department of Medicine, University of California San Diego, La Jolla, CA 92037, USA.

[12]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA.

[13]NYU-ECNU Center for Computational Chemistry, NYU Shanghai, Shanghai 200062, China.

[14]Center of Cancer Research, NCI, NIH, Bethesda, MD 20892, USA.

[*]Lead Contact, Correspondence to: erez@erez.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Contributions: S.S.P.R. and E.L.A. conceived this project and designed all experiments. S-C.H. and B.G.S.H., and A.D.O. performed Hi-C experiments. J.M.E. aided with the design of the PRO-Seq experiments and J.M.E. and E.M.P. performed PRO-Seq experiments. K-R.K-K., S.E.J., and S-C.H. performed ChIP-Seq experiments. A.L.S. performed extrusion and compartmentalization simulations. G.B. and T.S. designed and analyzed nucleosome simulations, which were performed by G.B. I.D.B. performed microscopy experiments. D.T, Z.Y., and J.T.R. designed a system for creating interactive figures. S.S.P.R., J.M.E., X.H., M.S.S., J.S., O.D., B.E.B., R.C., E.S.L, and E.L.A. analyzed data. S.S.P.R., E.S.L., and E.L.A. prepared the manuscript with input from all authors.

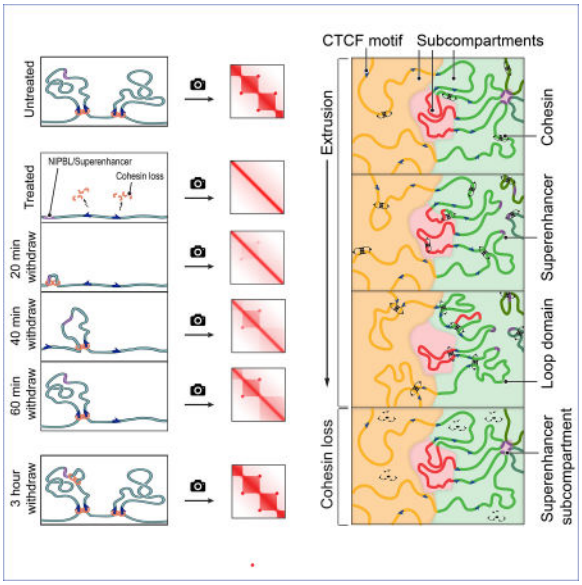
[15]Department of Biology, MIT, Cambridge, MA 02139, USA.

[16]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

SUMMARY

The human genome folds to create thousands of intervals, called “contact domains,” that exhibit enhanced contact frequency within themselves. “Loop domains” form because of tethering between two loci – almost always bound by CTCF and cohesin – lying on the same chromosome. “Compartment domains” form when genomic intervals with similar histone marks co-segregate. Here, we explore the effects of degrading cohesin. All loop domains are eliminated, but neither compartment domains nor histone marks are affected. Loss of loop domains does not lead to widespread ectopic gene activation, but does affect a significant minority of active genes. In particular, cohesin loss causes superenhancers to co-localize, forming hundreds of links within and across chromosomes, and affecting the regulation of nearby genes. We then restore cohesin and monitor the re-formation of each loop. Although reformation rates vary greatly, many megabase-sized loops recovered in under an hour, consistent with a model where loop extrusion is rapid.

eTOC



Cohesin loss results in elimination of loop domains, but neither compartment domains nor histone marks are affected.

INTRODUCTION

Many studies have shown that the insulator protein CTCF and the ring-shaped cohesin complex colocalize on chromatin (Wendt et al., 2008) and lie at the anchors of loops (Rao et al., 2014; Splinter et al., 2006) and the boundaries of contact domains (also called “topologically constrained domains”, “topologically associated domains”, or “physical domains”) (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Nora et al., 2012; Rao et al.,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2014). This suggests that these proteins help regulate genome folding (Merkenschlager and Nora, 2016). Consistent with this, deletion of CTCF sites interferes with loop and contact domain formation (Guo et al., 2015; Sanborn et al., 2015; de Wit et al., 2015). However, initial, low-resolution experiments examining genome-wide depletion of CTCF and cohesin observed only limited effects, reporting that compartments and contact domains still appear to be present (Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014). These results have made it difficult to ascertain the role of CTCF and cohesin in regulating genome architecture.

Here, we examine the effects of cohesin loss on nuclear architecture, epigenetic state, and transcription.

RESULTS

Rapid degradation of RAD21 using an auxin-inducible degron system

We employed an auxin-inducible degron (AID) (Natsume et al., 2016) to destroy RAD21, a core component of the cohesin complex. In this system, constitutive expression of the auxin-activated ubiquitin ligase TIR1 leads, in the presence of auxin, to rapid degradation of proteins tagged with an AID domain. We used this system in HCT-116, a human colorectal carcinoma cell line. This cell line had been previously modified by (Natsume et al., 2016) so that both RAD21 alleles were tagged with an AID domain and a fluorescent mClover ("RAD21-mAC") (Fig. 1A). We confirmed that RAD21-mAC was efficiently degraded after 6 hours of auxin treatment using fluorescence microscopy and ChIP (Fig. 1B, S1, see Methods), and that cohesin could no longer associate with DNA using ChIP-Seq with antibodies for SMC1, a different cohesin subunit (Fig. 1C, D).

Histone modification patterns are unaffected by cohesin loss

We first examined the effects of cohesin degradation on key epigenomic features associated with genome folding, using ChIP-Seq. We examined the distribution of CTCF (associated with loop anchors) and the histone modifications H3K27me3, H3K36me3, H3K27Ac, H3K4me1, H3K4me3, H3K9me3, H4K16Ac, H4K20me3, H3K79me2, and H2.AZ (associated with compartment intervals). Cohesin loss had little effect on these features (Fig. 1C,D; Fig. S1C–E, see Methods).

Loop domains are rapidly lost after degradation of cohesin

We then turned to study genome folding itself, beginning with loop domains. Loops arise when two loci on the same chromosome are tethered together. (For clarity, the loci will be referred to as "loop anchors", the tethered pair as a "link", and the interval between them as a "loop".) Loop anchors are typically a pair of DNA motifs in the convergent orientation (i.e., the motifs face each other) that bind CTCF and cohesin (Rao et al., 2014). Loops frequently form a contact domain—that is, an interval in which all loci exhibit higher contact frequency with one another (than random loci at similar distance along the genome sequence); this structure is called a "loop domain" (Rao et al., 2014).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To examine loop domains, we used *in situ* Hi-C (Rao et al., 2014), which combines DNA-DNA proximity ligation and high-throughput sequencing to create maps showing the frequency of physical contact between all pairs of loci across the genome. Loop domains are manifest in Hi-C maps as a bright "peak" pixel (indicating the link between the two loop anchors) at the corner of a bright square (indicating the presence of a contact domain).

We generated roughly 6 billion Hi-C contacts from HCT-116 cells before (3.0B) and immediately after (2.9B) auxin treatment. In the untreated cells, our algorithms annotated 3,170 loops, of which 2,140 were loop domains. Strikingly, the loop domains disappeared upon cohesin loss. The result was evident by visual examination (Fig. 2A, Fig. S2C, Data S1, I; (Durand et al., 2016a)). Moreover, the algorithms found only 9 loop domains after auxin treatment. Upon close inspection, all were found to be false positives (see Methods). (We return below to examine loops *not* associated with contact domains.)

To see if these changes were reversible, we performed a time-course analysis in which untreated cells were exposed to auxin for six hours, after which auxin was withdrawn (Fig. 2B, Fig. S2B). Low resolution Hi-C was performed immediately before treatment, as well as at a series of time points during treatment (at 20, 40, 60, 240, and 360 minutes) and after withdrawal (at 20, 40, 60, 180, 360, 1080, and 1440 minutes). To assess whether the anchors of the loop domains seen in the pre-treatment data continued to be linked (that is, co-located in space) subsequently, we used a method called Aggregate Peak Analysis (APA) (Durand et al., 2016b), which superimposes the signals from a set of peak pixels, thus allowing us to observe an aggregate signal even in sparse datasets where individual signals cannot be resolved (Fig. 2B). The APA signal was initially strong, but was gone by 40 minutes after treatment, and remained absent throughout the rest of the treatment period (Fig. 2B, S2B). The disappearance of the loop-domain links closely mirrored the depletion of cohesin levels in the samples during the treatment period, as ascertained by measuring mClover fluorescence (Fig. S1). After auxin was withdrawn, the APA signal quickly increased, largely recovering by 1 hour (Fig. 2B, S2B).

These results indicate that the formation of loop domains requires cohesin; that loop domains rapidly disappear after the cohesin tethering the link has been degraded; and that the restoration of cohesin rescues the loop domains.

Loop domains spanning NIPBL and superenhancers recover more rapidly

To explore the formation process for individual loop domains, we generated 2.6 billion additional contacts from our Hi-C experiments after 20, 40, 60 and 180 minutes of auxin withdrawal (678M, 665M, 618M, and 675M contacts, respectively). Using these improved maps, we were able to calculate individual recovery curves for 2038 of the 2140 loop domains observed in the untreated cells (Fig. 2C, see Methods). (The time-resolved maps were not sufficiently deep to assess the 102 loop domains shorter than 100kb.) Recovery rates for individual loop domains varied dramatically. Faster recovery was strongly associated with high levels of NIPBL binding between the loop anchors, as well as enrichment of promoter and enhancer elements, and of activating histone marks like H3K36me3 and H4K16Ac. By contrast, loop domains that recovered slowly were typically depleted for these features, but enriched for repressive marks like H3K27me3 and H3K9me3

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Fig. 2D–E, Fig. S2D–F, Data S1, II). The most extreme difference we found was in the presence of superenhancers (also called stretch enhancers, or SEs), which are regions of the mammalian genome containing a very high density of enhancer elements, and which are marked by extremely high levels of H3K27 acetylation (Hnisz et al., 2013; Parker et al., 2013). Fast loop domains (recovery rate > 95th percentile) were 159-fold more likely than slow loop domains (<5th percentile) to span an SE (2.94 SEs/Mb vs. 0.02 SEs/Mb), and 26-fold more likely to contain strong NIPBL binding sites (4.23/Mb vs. 0.17/Mb) (Fig. S2E–F).

Taken together, these results indicate that the rate of loop domain re-formation varies greatly across the genome, and is associated with factors ranging from cohesin loading to local regulatory activity.

Loss of cohesin is associated with stronger genome compartmentalization

Next, we examined the effects of cohesin loss on compartmentalization. Compartmentalization refers to the fact that the genome is partitioned into intervals (which can range from 14 kb to more than 5 Mb) belonging to a small number of types, such that intervals of the same type exhibit an enhanced contact frequency with one another (Lieberman-Aiden et al., 2009; Rao et al., 2014). Intervals are thereby assigned to two compartments (A or B, closely associated with open and closed chromatin, respectively) and, more finely, into six subcompartments (A1, A2, B1, B2, B3, B4). The “compartment intervals” that lie in a particular compartment are associated with distinctive patterns of chromatin marks (Rao et al., 2014). Because loci within a compartment interval are of the same type, they exhibit an increased contact frequency with one another and frequently form contact domains. In this case, we call the contact domain a “compartment domain.” The enhanced contact frequency between compartment intervals in the same subcompartment also gives rise to a plaid pattern in Hi-C maps (Lieberman-Aiden et al., 2009).

Whereas loop domains disappear entirely after cohesin loss, compartmentalization is preserved (Fig. 3A). Following auxin treatment, there is no significant change in either the compartment domains, as defined by the presence of the corresponding squares along the diagonal in the Hi-C contact map (Fig. 3B; see Methods), or in the plaid pattern, as defined by the eigenvectors of the Hi-C correlation map (Fig. 3A; mean Pearson’s r = 0.968 across all chromosomes). Our data is consistent with a previous report that genome compartmentalization is preserved after depletion of cohesin (Seitan et al., 2013) or CTCF (Nora et al., 2017).

We then examined the interaction between compartments and loop domains. Specifically, we examined the compartment boundaries (transition points between compartment intervals) that either (i) lay in the interior of a loop domain in untreated cells or (ii) coincided with a loop-domain anchor in untreated cells (Fig. 3C–E). In the former case, the correlation in the genome-wide contact pattern on opposite sides of compartment boundaries showed a much greater decrease in treated vs. untreated cells—that is, the plaid pattern across the genome became much stronger in the absence of cohesin (Fig. 3C–E, Data S1,III). The results were similar when we examined boundaries between intervals that were enriched vs. depleted for H3K27Ac (which marks intervals in the “A” compartment (Rao et al., 2014)) or intervals that were enriched vs. depleted for H3K27me3 (which marks intervals in the “B1”

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

subcompartment (Rao et al., 2014)) (Fig. 3F, S3A–D, see Methods). These data indicate that the compartmentalization process that brings together loci with similar histone marks does not rely on cohesin. On the contrary, the strengthening of the plaid pattern after cohesin loss suggests that the formation of cohesin-dependent loop domains interferes with compartmentalization by promoting the co-localization of locus pairs with different histone modification patterns. Our data is consistent with the observation that genome compartmentalization is weakened in cells where the cohesin-antagonist *WAPL* is knocked out, leading to larger loop domains (Haarhuis et al., 2017).

Links between superenhancers strengthened after loss of cohesin

Next, we examined loops not associated with contact domains. Whereas 1,030 such loops were annotated in untreated cells, only 72 were annotated after cohesin loss. Upon close examination, 57 were false positives (see Methods). (The loop-detection algorithms have a higher false-discovery rate after cohesin loss, since true positives are so rare.) The remaining 15 loops were much larger than those seen in untreated cells (median: 1.75 Mb, vs. 0.275 Mb). Given their large size, we found that loops could be more reliably identified in treated cells by running our peak detection algorithm at coarser resolution (50–100 kb vs. 5–10 kb) (see Methods). This identified an additional 46 loops that were confirmed by manual inspection (Fig. 4A, S4A, see Methods). Afterwards, the size difference between the 61 “cohesin-independent loops” and the cohesin-associated loops was even more dramatic (Fig. 4B, median size: 23.15 Mb).

We sought to understand the basis of these “cohesin-independent loops.” We found that they do not demarcate the boundaries of contact domains (0 of 61 [0%]; vs. 2140 of 3170 [68%] for cohesin-associated loops). Remarkably, many cohesin-independent loop anchors form links with each another – manifest as focal peaks in the Hi-C heatmap – even when the anchors reside on different chromosomes (Fig. 4A, Data S1,IV). In total, we identified 203 such interchromosomal links. Interchromosomal links are not seen between the anchors of cohesin-associated loops.

The anchors of cohesin-independent loops also exhibit very different patterns of protein binding. The proportion that binds CTCF is much lower (20% vs. 90% for cohesin-associated loops; Fig. 4C, Fig. S4A). Moreover, there is no tendency for the CTCF motifs at loop anchors to point into the loop (5 of 9 (56%) point inward vs. 2770 of 2919 (95%) for cohesin-associated loops.

Notably, the cohesin-independent loop anchors are highly enriched for superenhancers. We found that 41 of the 64 cohesin-independent loop anchors overlapped with the 387 superenhancers in HCT116 cells—a 37.5-fold enrichment ($p<10^{-15}$). For the 100 strongest superenhancers, the enrichment was 76- fold (30 of 64, $p<10^{-15}$; Fig. 4D, S4B). Interestingly, cohesin-independent loops and the associated loops and links between superenhancers could be seen in the untreated cells as well, but were much weaker (Fig. 4A,E, Fig. 5A, Fig. S4C–E, Data S1,IV, see Methods).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Strikingly, we observed large cliques forming between the anchors of the cohesin-independent loops (Fig. 5A, Data S1,V). Large cliques are not seen for cohesin-associated loops (Fig. S2A).

In many respects, the cohesin-independent loops resemble the superloops we previously observed on the inactive X chromosome (Darrow et al., 2016; Rao et al., 2014): they are very large (up to 77Mb), the intervals they span do not form contact domains, and their anchors tend to form cliques and are marked by H3K27 acetylation (Fig. S4F). We also found that the superloops tend to occur simultaneously, forming hubs containing 3 or more loci. We therefore wondered if cohesin-independent loop anchors would exhibit the same behavior.

To probe this question, we examined concatemers—that is, Hi-C reads which bring together 3 or more loci, and indicate that the loci in question were simultaneously co-located in a particular cell during the Hi-C experiment. We identified 57 million unique “triples,” which bring together 3 loci; 32 million unique quadruples, and 130 thousand quintuples (Fig. 5B, see Methods). Instead of a 2-dimensional heatmap, concatemers are naturally represented as an n-dimensional matrix, or tensor, showing the collision frequency (i.e. the frequency of simultaneous physical contact) between any set of n loci in the genome (Fig. 5C). Hubs involving n loci manifest in the n-dimensional contact tensor as peaks in collision frequency with respect to the local n-dimensional neighborhood.

Because the number of entries in an n-dimensional contact tensor scales as the genome size to the nth power, contact tensors can be exceedingly sparse. We therefore did not expect to see triples corresponding to any particular set of three cohesin-independent loop anchors. Instead, we developed a variant of APA for contact tensors, superimposing the signal from all possible sets of three cohesin-independent loop anchors that lie on a single chromosome (see Methods). This analysis revealed 11 collisions involving three cohesin-independent loop anchors in the auxin-treated data, as compared to an expected value of 0.41 collisions based on the density of collisions in the local neighborhood (Fig. 5D–F, Fig S5A–B). These findings indicate that, like superloops, cohesin-independent loop anchors tend to form hubs involving three or more loci. By contrast, no collisions were found in the untreated data. This is consistent with our finding that cohesin-independent loops are much weaker in the presence of cohesin.

Molecular dynamics simulations integrating extrusion and compartmentalization can recapitulate Hi-C experimental results

To test the hypothesis that the Hi-C contact maps we observed are consistent with the presence of two distinct folding mechanisms, we modeled a 2.1 Mb region on chromosome 3 (Sanborn et al., 2015). Our model treated the region as a block copolymer consisting of two types of chromatin, A or B, determined by classifying loci based on ChIP-Seq binding data; and containing CTCF binding sites, whose position and strength were derived from CTCF and SMC1 ChIP-Seq tracks, and whose orientation was determined by examining the human genome reference (Fig. 6A, see Methods). We used molecular dynamics simulations to examine the behavior of this polymer in a solvent containing extrusion complexes (thus modeling loop extrusion (Fudenberg et al., 2016; Sanborn et al., 2015)), and in the presence

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of attractive forces between like monomers (thus modeling compartmentalization (Di Pierro et al., 2016)). The resulting ensemble was used to create an *in silico* contact map for the region.

We found that the resulting contact maps accurately recapitulated the experimental results in both untreated and treated cells (Figure 6A,B), and illustrate the change in long-range contact pattern that is seen when a loop spans a compartment boundary (Data S1,VI). These findings suggest that the most prominent features observed in Hi-C data sets (loops, domains, and compartments) can be recapitulated by simulations that use only ChIP-Seq data as input, in the context of a model combining extrusion and compartmentalization.

The above simulations assume that phase separation leads to the compartmentalization of chromatin intervals bearing similar patterns of histone marks. There are multiple models that could account for such a tendency. In one, phase separation is facilitated by protein chaperones that recognize histone tails. Alternatively, similar nucleosomes might directly attract one another through histone tail interactions. To see whether tail interactions can guide folding at the oligonucleosome scale given the mechanical, electrostatic, and entropic constraints on chromatin fibers we simulated short fibers using a mesoscale approach (Bascom and Schlick, 2017; Bascom et al., 2017; Grigoryev et al., 2016). These simulations have three components: linker DNA beads (each representing 9 bp), nucleosome core particles (rigid bodies with charged surfaces), and histone tail beads (each representing 5 aa) (Fig. S6A–B). We found that histone tail interactions overcame constraints on the chromatin chain, leading either to focal contacts between short chromatin intervals or to global condensation, depending on the initial configuration and epigenetic state of the fiber (Fig S6C–E). Of course, these findings do not rule out a central role for protein chaperones *in vivo*.

Cohesin loss results in strong down-regulation of genes near superenhancers, but does not bring about widespread ectopic activation

Finally, we sought to investigate the role of cohesin in regulating gene expression. Cohesin has been proposed to facilitate interactions between enhancers and promoters (Kagey et al., 2010; Merkenschlager and Nora, 2016). Loop domains are thought to regulate this process by preventing enhancers from forming ectopic interactions with targets that lie in a different loop domain (Lupiáñez et al., 2015; Flavahan et al., 2016). We therefore characterized the effects of cohesin loss on nascent transcription by performing precision nuclear run-on sequencing (PRO-Seq) in treated and untreated HCT116 cells (Engreitz et al., 2016; Jonkers and Lis, 2015) (Fig. 7A). We chose an early timepoint – 6 hours after auxin treatment – with the aim of examining direct consequences, rather than indirect effects due to changes in cell state.

To look for signs of ectopic activation, we examined the 14,853 genes that were not expressed (RPKM<0.5) in untreated cells. Of these genes, 1% (216) were activated after treatment ($p<0.05$, >30% change in RPKM, RPKM>0.5 in treated cells). Thus, while cohesin plays a role in preventing ectopic activation, most genes remain inactivated even in the absence of cohesin and loop domains.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We next looked for changes in the 12,222 genes that were expressed (RPKM>0.5) in untreated cells (Fig. 5B). Here again, most genes (87%, 10,615) exhibited similar levels of transcription after cohesin degradation (RPKM changed by less than 30%). Strong effects were infrequent: 64 genes (0.5%) showed a 2-fold change, and 2 genes showed a 5-fold change (Fig. 7B). While the quantitative impact may seem modest, such changes can have important biological impacts (Flavahan et al., 2016).

Of genes that exhibited a strong change in transcription, more were downregulated than upregulated (61% vs. 39%)—suggesting that cohesin-associated loops may both facilitate activation of promoters by their distal enhancer elements and block activation by inappropriate enhancers, with the former being somewhat more common.

We wondered how cohesin facilitated these promoter-enhancer contacts. We noticed that many of the genes that were downregulated (by >1.75-fold) were located within 500kb of superenhancers (23 of 49, 4.8-fold enrichment, Fig. 7C,D; S7A–B). Of these genes, 29% were located with 500kb of one of the top 100 superenhancers (8.5-fold enrichment). Strikingly, these superenhancers were often located at the anchors of the cohesin-independent links seen in treated cells (8 of 19, a 13.7-fold enrichment).

The above results are interesting from the standpoint of transcriptional regulation. In the absence of cohesin, superenhancers associated with the downregulated genes exhibit a strong tendency to form links with one another. By contrast, in the presence of cohesin, the majority of these superenhancers were located in the interior of cohesin-associated loop (in 13 of 19 cases) and the long-distance cohesin-independent links were much weaker.

DISCUSSION

Here, we explore the 4D Nucleome of a human colon cancer cell line during cohesin loss and recovery, achieving Hi-C map resolutions of 5kb with a time resolution of 20 minutes. We find that cohesin is required for the establishment and maintenance of loop domains. After cohesin loss, we also find that: (i) histone marks are unchanged; (ii) compartment structure is strengthened in the absence of cohesin, as loop domains spanning multiple compartment intervals lead to mixing among loci in different compartments; (iii) only a small subset of genes exhibit large changes in transcription level. As auxin is withdrawn, cohesin levels recover, and we are able to measure the rate of formation for nearly every loop domain, genome-wide. Loop domains that recover quickly are much more likely to span superenhancers and binding sites of the cohesin loading factor NIPBL. Finally, we identify a class of cohesin-independent loops, links and hubs connecting superenhancer loci on many chromosomes.

Comparison with other studies

It is important to set our results in the context of other studies related to loop formation. While there is clear agreement that deletion of individual CTCF binding sites can result in a loss of cohesin binding and can abolish loops and contact domains (Guo et al., 2015; Sanborn et al., 2015; de Wit et al., 2015), there have been conflicting reports about the effects of depleting cohesin or CTCF.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Early Hi-C studies of cohesin and CTCF depletion, using both gene knockouts and proteolytic cleavage, reported that contact domains remained (Seitan et al., 2013; et al., 2013; Zuin et al., 2014). The discrepancy may be due to (i) the fact that low resolution Hi-C analysis cannot not distinguish between loop domains, which are sensitive to cohesin depletion, and compartment domains, which are not; and (ii) the possibility that the cohesin depletion was incomplete (Fig. S2C). More recently, CTCF depletion followed by higher-resolution Hi-C (Nora et al., 2017) revealed the disappearance of a subset of contact domains. Our data for cohesin depletion is consistent with is study.

Two recent studies have also sought to examine the effects of depletion of *NIPBL*, which encodes a cohesin loader protein. They report opposite conclusions. The (Schwarzer et al., 2016) preprint reported, on the basis of new experiments, that the genome-wide aggregate signal from loop domains is absent following *NIPBL* deletion. By contrast, a recent publication reported the continued presence of individual loop domains after the near-complete depletion of *NIPBL*, although these loop domains were abnormally small (Haarhuis et al., 2017). In any case, it is unclear what effect NIBPL depletion would be expected to have on loop domain formation because, although *NIPBL* facilitates cohesin loading, it may not be essential for cohesin loading. Indeed, cohesin loading independent of NIPBL has been observed *in vitro,* albeit at low efficiency (Davidson et al., 2016; Stigler et al., 2016).

Finally, an exciting recent experiment demonstrated that deletion of WAPL, a cohesin antagonist that removes cohesin from chromatin, results in the formation of thousands of new loops and loop domains, which are larger than those found when WAPL is intact (Haarhuis et al., 2017). Our results are consistent with these findings.

Two mechanisms that guide genome folding

Our results highlight two distinct mechanisms that guide genome folding.

The first is the cohesin-dependent formation of loop domains. The data presented above are consistent with several models of this process. We (Sanborn et al., 2015) and others (Alipour and Marko, 2012; Fudenberg et al., 2016; Nasmyth, 2001) have hypothesized that the underlying physical process is the formation of loops by extrusion. In this model, loop domains form when a hypothetical cohesin-based extrusion complex (“Xcom”), which comprises two physically tethered subunits, binds chromatin at a particular location; subsequently, the subunits slide in opposite directions until they arrive at a bound CTCF protein. Thus, the disappearance of cohesin can eliminate all loop domains without influencing CTCF binding. Other models include the initial formation of loops via 3D diffusion of anchor loci, followed by cohesin-mediated stabilization. See (Sanborn et al., 2015) for a fuller discussion.

The second mechanism is the cohesin-independent compartmentalization of chromatin intervals with similar histone marks (Lieberman-Aiden et al., 2009; Rao et al., 2014). This observation is also compatible with several models: histone modifications might drive the formation of compartments (i.e., “phase separation” (Hnisz et al., 2017; Jost et al., 2014; Di

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pierro et al., 2016)); compartmentalization might lead to histone remodeling; or both processes might be caused by a third mechanism.

The data presented above – specifically, the fact that cohesin loss does not affect histone modifications, but does cause long-range contact patterns to better match patterns of histone marks – is more consistent with histone patterns governing genome compartmentalization, rather than the reverse. This phase separation process could involve histone-tail interactions, or the binding of reader proteins that target modified histones to specific locations in the nucleus (Wijchers et al., 2016). Interestingly, two recent studies have provided evidence that the protein HP1a, which binds the heterochromatic H3K9me mark, forms liquid droplets *in vivo* via phase separation, such that H3K9me heterochromatin is contained within the droplets (Larson et al., 2017; Strom et al., 2017). Of course, it remains possible that independent mechanisms may shape both histone mark patterns and genome compartmentalization.

The speed of loop extrusion

We show that loop domains disappear shortly after auxin-induced cohesin loss, and reappear shortly after auxin withdrawal, implying that (i) they require cohesin both for formation and maintenance; and (ii) they do not represent stable states of chromosome condensation.

In the loop extrusion model, the two physically tethered subunits of the Xcom bind chromatin at a single site, and then slide in opposite directions along chromatin. Measurements of loop re-formation enable estimates of extrusion speed. For instance, in figure 2E, we show a ~900kb loop that is restored within 40 minutes of auxin withdrawal. Thus, this loop is extruded at an average rate of no less than 375 bp/s, with each Xcom subunit sliding at no less than 188 bp/s. Our estimates are lower bounds, as they ignore the time needed for auxin to disappear and for cohesin to re-form and be loaded on chromatin. (Given the correlation between loop reformation speed and Nipbl, loading time may be a significant consideration.)

These estimates are similar to ones obtained studying the SMC complex in *B. subtilis* [500–1000 bp/s] (Wang et al., 2017).

The rate estimates bear on the protein motors involved when Xcom subunits slide. For instance, it is possible that cohesin itself serves as a motor during the extrusion process. However, single molecule studies of cohesin translocation *in vitro*, have yielded an estimated sliding rate of only 1–2bp per second on chromatin (Davidson et al., 2016; Stigler et al., 2016). These could indicate that cohesin alone – under the conditions probed in those experiments – is unlikely to be the principal source of translocase activity. Similarly, RNA polymerase II can push cohesin along DNA (Davidson et al., 2016) but elongation rate estimates (9–90 bp/s; (Jonkers and Lis, 2015)) are slower than what we observe. This suggests that other translocases may be involved in loop extrusion.

Two classes of loops

We observe a population of loops that are frequently anchored at superenhancers and do not depend on cohesin.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

One explanation is that these loops, too, form by extrusion, but using alternative protein complexes, such as condensin, instead of cohesin. However, the superenhancer anchors also form links with one another when they lie on different chromosomes, whereas loop extrusion cannot form interchromosomal links. Our data is therefore less consistent with a model where where superenhancer loops form by extrusion, and more consistent with the presence of an alternative mechanism, perhaps based on some form of facilitated diffusion or phase separation (Sanborn et al., 2015).

In particular, superenhancer links may represent compartmental co-segregation of small, H3K27-acetylated intervals (Fig. 5G), which accounts for why these links can join loci on different chromosomes, why they are weaker in the presence of cohesin, and why their anchors form large cliques and highe-rorder hubs.

Notably, loops and links between superenhancers increase in strength rapidly following cohesin loss, reaching a plateau within hours. This implies that compartmentalization is capable of inducing intrachromosomal loops and interchromosomal links at rapid rates, comparable to those of loop-domain formation. (Our findings may be related to those of other studies, which have noted enhanced interactions between higher-order intrachromosomal interactions between domains containing superenhancers (Beagrie et al., 2017).)

The interplay between loop extrusion and compartmentalization

Using our high-resolution contact maps, we are also able to examine the ways in which loop domain formation and compartmentalization interact. It is commonly thought that compartment intervals are typically megabases in length, and are subdivided into smaller domains in a hierarchical fashion (Dixon et al., 2012; Nora et al., 2012). Here, we demonstrate that compartment intervals can be as short as tens of kilobases, and can overlap loop domains in complex ways. For instance, we observe numerous examples of loop domains spanning multiple compartment intervals.

In such cases, we find that loop extrusion, by facilitating contacts between all loci in the loop domain, can enhance the contact frequency of loci that would ordinarily lie in different subcompartments. Thus, the long-range contact pattern seen for each locus is a mixture of the pattern that would ordinarily be seen for loci in the corresponding subcompartment, and the pattern seen for other loci in the loop. This mixing disappears upon cohesin depletion (Fig. 7E). Similarly, deletion of WAPL appears to increase the processivity of the Xcom, and thus increases the size of loops (Haarhuis et al., 2017). Consistent with our observations, these larger loops are associated with extensive mixing, which obscures long-range compartment patterns.

The interplay between cohesin and gene regulation

Many studies have proposed that cohesin facilitates interactions between enhancers and promoters, thereby upregulating the transcription of many genes (Kagey et al., 2010; Merkenschlager and Nora, 2016). Moreover, studies have also suggested that loop domains formed between CTCF and cohesin binding sites create insulated regulatory neighborhoods —partially protecting genes with a loop domain from the influence of enhancers outside the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

domains (Flavahan et al., 2016; Lupiáñez et al., 2015). Our study, combining rapid depletion of cohesin and measurement of nascent transcription using PRO-Seq, allows us to more clearly dissect the direct effects of cohesin loss on transcription.

We find that a very small set of genes, often lying near superenhancers, becomes strongly downregulated after cohesin loss. However, most genes are not strongly affected. This suggests that cohesin-dependent loop domains themselves play at most a modest role in facilitating or disrupting interactions between promoters and enhancers. Of course, we cannot dismiss modest effects on overall level of transcription as unimportant. For example, modest increases in the expression of receptor tyrosine kinase genes can have meaningful effects on cell proliferation (Flavahan et al., 2016).

Nevertheless, it is particularly interesting to compare the above findings with earlier studies, using similar methods, that showed a strong correlation between the presence of a loop domain and many-fold increases in the expression of genes at the loop domain anchor (Kagey et al., 2010; Rao et al., 2014). Taken in isolation, these earlier results are consistent with a model where the formation of loop domains routinely causes many-fold changes in gene expression. By contrast, the data presented here are more consistent with a model where the formation of loop domains influences gene transcription, but rarely causes many-fold changes.

One possibility is that both processes might be independent consequences of upstream regulatory events. For example, changes in accessibility of a gene promoter might facilitate both transcription factor binding, activating the gene, and cohesin arrest, activating the loop domain. Alternatively, large increases in gene expression may alter the accessibility of CTCF motifs at the promoter and cause loop domain formation.

Our study suggests a model where cohesin-associated looping, by increasing the frequency of contact between loci within loop domains and by disturbing patterns of compartmentalization, facilitates mixing between elements (such as genes and superenhancers) that would otherwise be segregated. Thus, compartmentalization and extrusion – through independent and complementary mechanisms – interact to shape transcription.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

All requests for information, reagents and resources should be directed to the Lead Contact, Erez Lieberman Aiden (erez@erez.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

HCT-116 cells—We obtained HCT-116-CMV-OsTir1 and HCT-116-RAD21-mAID-mClover cells (HCT-116 RAD21-mAC) from (Natsume et al., 2016). The cells were cultured in McCoy's 5A medium supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100ug/ml streptomycin at 37C with 5% CO2. Degradation of the AID-tagged RAD21 was induced by the addition of 500uM indole-3-acetic acid (IAA; Sigma

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Aldrich). For our standard *in situ* Hi-C, ChIP-Seq, and PRO-Seq experiments on untreated cells and cells treated for 6 hours, medium was aspirated at t=0, and either replaced with fresh medium (untreated) or medium containing 500uM IAA. The cells were then washed, trypsinized and processed for downstream experiments at t=6hrs.

For our time course experiments, cells were treated with 500uM IAA and crosslinked with 1% formaldehyde directly in wells of a 6-well plate at various time points after treatment (20, 40, 60, 240, 360 minutes). For the auxin withdrawal experiments, after 6 hours treatment with 500um IAA, the cells were trypsinized, washed twice in fresh media and replated in 6-well plates in fresh media. They were then crosslinked with 1% formaldehyde directly in the 6-well plates at various time points after auxin withdrawal (20, 40, 60, 180, 360, 1080, 1440 minutes).

In order to ensure that our results were not due to the HCT-116 cells stalling in mitosis, we also repeated our Hi-C experiments after cell synchronization and arrest of the cells at the G1/S boundary. First, we added 2mM thymidine to arrest the cells in S-phase and incubated the cells for 12 hours. We then trypsinized and spun down the cells and replated in fresh media, allowing the cells to grow for 12 hours to exit from S-phase. We then added mimosine to a final concentration of 400uM and incubated the cells for 12 hours to arrest the cells at the G1/S boundary. We then replaced media with either complete media + 500uM IAA + 400uM mimosine (treated cells) or complete media + 400uM mimosine (untreated cells) and incubated the cells for 6 hours before processing for downstream experiments.

METHOD DETAILS

Microscopy—Live HCT116 RAD21-mAC cells in growth medium without phenol red were added to a chambered coverglass (Lab-Tek #155409) 24 hours prior to imaging and incubated at 37°C, 5%CO2, allowing them to attach to the coverglass. One hour before imaging, the growth medium was replaced with 2 μg/ml of Hoechst 33342 in phosphate-buffered saline (PBS) to visualize nuclei. Time-lapse widefield fluorescence microscopy was performed on a DeltaVision OMX microscope (GE Healthcare) equipped with a 37°C incubation chamber, using a 60× oil immersion objective. Cells were treated with 500μM of IAA immediately before imaging. Images were collected every 10 minutes from 0 to 60 minutes following treatment (DAPI filter at 5%T, 100ms for Hoechst; FITC filter at 100%T, 100ms for mClover), and deconvolved using the built-in SoftWoRx software. The Hoechst images were adjusted in Photoshop by increasing brightness by 92 and contrast by 92 in legacy mode (applied equally to the entire image for all timepoints). The mClover images were adjusted in Photoshop by increasing brightness by 164 and contrast by 123 in legacy mode (applied equally to the entire image for all timepoints). The images were merged in Photoshop using the “screen” function.

In situ Hi-C—We generated 69 *in situ* Hi-C libraries using the MboI restriction enzyme following the protocol described in(Rao et al., 2014) without modifications. In brief, the *in situ* Hi-C protocol involves crosslinking cells with formaldehyde, permeabilizing them with nuclei intact, digesting DNA with a suitable 4-cutter restriction enzyme, filling the 5’-overhangs while incorporating a biotinylated nucleotide, ligating the resulting blunt-end

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fragments, shearing the DNA, capturing the biotinylated ligation junctions with streptavidin beads, and analyzing the resulting fragments with paired-end sequencing.

We generated 7 libraries each for our main maps (untreated HCT-116 RAD21-mAC cells and HCT-116 RAD21-mAC cells treated for 6 hours with IAA) comprised of two sets of biological replicates each (three and four technical replicate libraries per biological replicate). In addition, we generated four technical replicate libraries each for untreated and treated HCT-116 RAD21-mAC cells after cell synchronization and arrest. We also generated an additional 5 libraries from untreated HCT-116 RAD21-mAC cells and 4 libraries from HCT-116 RAD21-mAC cells treated for 6 hours with IAA (two additional biological replicates each) that were not included in our main maps, but were used for analysis of higher-order contacts (see below). We generated four libraries (two biological replicates) for each of our loop-resolution maps along the auxin withdrawal time course (20, 40, 60 and 180 mins after auxin withdrawal). Finally, we generated two technical replicate libraries per time point of our auxin treatment and withdrawal time course. Similar results were obtained with Hi-C libraries from synchronized and arrested cells (Data. S1, I), so for all analyses presented in the main text and figures of the manuscript (other than the time course analyses), we utilized our high resolution maps from the unsynchronized cells. Further details about the Hi-C libraries and details about which experiments were used in which figures are provided in Table S1.

ChIP-Seq—ChIP-Seq for H3K27Ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3, H3K9me3, H4K16Ac, H4K20me3, H3K79me2, and H2.AZ was performed using a native ChIP-Seq protocol. Chromatin from untreated HCT-116 RAD21-mAC cells or cells treated for 6 hours with 500uM IAA was digested with Mnase (Sigma) in digestion buffer (50 mM Tris-HCl, pH7.6, 1 mM $CaCl_2$, 0.2% Triton X-100, butyrate 5 mM) for 5’ at 37°C and dialyzed against RIPA buffer for 2hrs at 4°C. Five microgram of respective antibody was incubated with 40 μl of Dynabeads Protein A (or G) for 40 min at room temperature. Antibody-bound beads were added to 500 μl of sonicated or Mnase-digested chromatin, incubated at 4°C overnight, and washed twice with RIPA buffer, twice with RIPA buffer containing 0.3M NaCl, twice with LiCl buffer (0.25 M LiCl, 0.5% Igepal-630, 0.5% sodium deoxycholate), once with TE (pH 8.0) plus 0.2% Triton X-100, and once with TE (pH 8.0). ChIP DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. Libraries were prepped for Illumina sequencing and 50bp single-end reads were sequenced on a HiSeq2000 or 2500 (Illumina). We also performed ChIP-Seq for RAD21 and CTCF following the same protocol as above, except that cells were fixed with 1% formaldehyde for 10 minutes at 37°C and fixation was quenched by the addition of glycine to a final concentration of 125mM for 10 minutes. The fixed cells were sonicated using a Branson sonifier at amplitude 35%, 12 cycles of 20 seconds of sonication and 30 seconds of pause. Antibody was added to the sonicated chromatin as above and the samples were further processed as above.

We also performed ChIP-Seq for SMC1 and an additional replicate for CTCF following the protocol outlined by the ENCODE consortium (2012).

All ChIP-Seq experiments were processed in parallel with whole cell extract input controls.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

PRO-Seq—To measure changes in transcription resulting from cohesin loss, we performed precision run-on sequencing (PRO-Seq) (Jonkers and Lis, 2015), a variant of global run-on sequencing (GRO-Seq), using a single biotinylated nucleotide (biotin-11-CTP) as previously described (Engreitz et al., 2016). We made one modifications to the protocol: at the end of each biotin enrichment, we eluted biotinylated RNAs from the streptavidin-coated magnetic beads by heating beads in 25 μl of 20 mM Tris-HCl pH 7.5, 10 mM EDTA, 2% N-lauroylsarcosine at 95°C for 5 minutes, followed by a magnetic-bead nucleic acid purification with 20 μl of MyONE SILANE beads. During the nuclei preparation step, we processed pairs of RAD21-mAC cells with and without auxin treatment in parallel. In addition, we performed PRO-Seq on HCT-116 CMV-OsTIR1 cells, the parental cell line of RAD21-mAC containing the *OsTIR1* gene integrated at the AAVS1 locus and no mAID tags integrated on any protein. By performing PRO-Seq on CMV-OsTIR1 cells with and without auxin treatment, we could control for transcriptional effects of the auxin treatment itself on HCT-116 cells, as well as any consequences of tagging the RAD21 protein.

QUANTIFICATION AND STATISTICAL ANALYSIS

Hi-C Data Processing—All Hi-C libraries were sequenced either on an Illumina NextSeq500 (either 80 or 85bp paired-end reads) or a HiSeqX (150bp paired-end reads). All resulting data was processed using Juicer (Durand et al., 2016b; Rao et al., 2014). The data was aligned against the hg19 reference genome. All contact matrices used for further analysis were KR-normalized with Juicer. Loops were annotated in both untreated and treated maps using HiCCUPS (Durand et al., 2016b; Rao et al., 2014).

Loops were called at 5kb, 10kb, and 25kb resolutions and merged as described in (Rao et al., 2014). Default parameters as described in (Durand et al., 2016b; Rao et al., 2014) were used with the exception that an additional enrichment filter was added. We noted that due to karyotypic abnormalities in the HCT-116 cell line, many rearrangements were annotated in both the untreated and treated maps. Since rearrangements appear as very intense pixels off-diagonal, we removed any peak calls that displayed an observed/expected enrichment of >4.5. Empirically, this max threshold removed peak annotations due to rearrangements; notably, nearly the same number of annotated peaks were removed from the untreated and the treated annotations, 277 and 269 respectively. In the end, we annotated 3,170 loops in our untreated maps and 81 loops in our treated maps.

Domains were annotated in both untreated and treated maps using Arrowhead (Durand et al., 2016b; Rao et al., 2014). Domains were called at 5kb and 10kb resolutions using default parameters and merged (retaining the 5kb domain annotation for any pair of domains annotated in both the 5kb and 10kb annotations). We annotated 9,845 domains in our untreated maps and 2,090 domains in our treated maps.

ChIP-Seq Data Processing—All ChIP-Seq data was aligned to hg19 with BWA (Li and Durbin, 2010) (Li and Durbin, 2010) (Li and Durbin, 2010), deduplicated using PicardTools, and analyzed with MACS 2.0 (Liu, 2014). All data was normalized against the corresponding input control using the '-c' option of MACS 2.0. ChIP-Seq peaks were called using the 'callpeak' function of MACS 2.0 with default parameters. For H3K4me3 and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

H3K4me1, we additionally used the '–broad' option of MACS 2.0 and merged close by peaks to identify broad peaks. Finally, for H3K4me1 and H3K27Ac, we filtered peaks called by MACS by requiring an adjusted -log10 p-value of at least 7 and 5 respectively in order to ensure that we had a high quality peak annotation. Finally, in order to annotate broad-source domains, we utilized RSEG (Song and Smith, 2011) using the following flags: '-b 25000', '-i 20', '-mode 2'.

Signal tracks were calculated by using the 'bdgcmp' option of MACS 2.0 with the 'FE' (fold-enrichment) method. All data for downstream analysis was averaged and extracted from these tracks.

PRO-Seq Data Processing—For analysis of PRO-Seq data, we aligned 30-bp paired-end reads to the hg19 reference (bowtie2 v2.1.0, (Langmead and Salzberg, 2012)), removed duplicate reads (Picard http://picard.sourceforge.net), and discarded reads with MAPQ < 30. We counted reads overlapping RefSeq genes (collapsed by gene symbol to the longest isoform) — this quantification procedure includes signal both at the paused position (near the TSS) as well as in the gene body. We identified genes showing significant differences in transcription with DESeq2 (Love et al., 2014), excluding genes with zero coverage in all samples and calling significance at Benjamini-Hochberg corrected p-value < 0.05.

To determine whether there were global changes in the total amount of transcription (up or down) that would affect the normalization and analysis of these experiments, we included a spike-in control in three of the four PRO-Seq replicates for each of untreated and treated RAD21-mAC and CMV-OsTIR1 cells. Specifically, we added ~500,000 Drosophila S2 cells at the beginning of the protocol. Upon sequencing of these libraries, we counted the number of spike-in reads by aligning to the Drosophila genome (dmel3) with bowtie2 v2.1.0. We observed similar fractions of reads mapping to the Drosophila spike-in in the matched pairs of degron-expressing and control replicate experiments, indicating that there are not significant global changes in the total amount of transcription upon cohesin loss.

Random Shuffle Annotations—When performing quantitative analyses on our feature annotations, it was frequently desirable to have a "random control" for the feature annotation in question. We generated such annotations through a random permutation procedure. For one-dimensional features, such as peak loci, we randomly placed the one-dimensional features throughout the genome such that (1) the number of features on any one chromosome stayed the same; (2) the random features did not overlap any gaps in the assembly (i.e. centromeres, telomeres, etc.). Similarly, for two-dimensional features (domains, peaks), we randomly placed the two ends of the features across the genome such that (1) the size distribution of the two-dimensional features stayed the same; (2) the number of features on any one chromosome stayed the same; (3) the interval between the ends of the randomized two-dimensional features did not overlap any gaps in the assembly.

Analysis of CTCF and cohesin binding—In order to confirm that degradation of RAD21 resulted in abrogation of full cohesin complex binding to chromatin, we performed ChIP-Seq for RAD21 and SMC1 (see above for experimental details). We visually confirmed that cohesin binding was significantly diminished (see Fig. 1C). We also analyzed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the RAD21 and SMC1 signal at all ChIP-Seq peaks called in our data from untreated cells using MACS 2.0. The heatmaps clearly showed that cohesin binding was eliminated upon degradation of RAD21. We saw an average 81% reduction in binding strength of RAD21 (mean enrichment = 12.58 in untreated HCT-116 RAD21-mAC cells; mean enrichment = 2.39 in treated cells) and 75% reduction of SMC1 (mean enrichment = 16.23 in untreated HCT-116 RAD21-mAC cells; mean enrichment = 4.00 in treated cells). Moreover, 97% of peaks called in untreated cells (from a merged list of peaks binding both RAD21 and SMC1) were not called after auxin treatment (Fig. S1C). These results demonstrate that we were able to quickly abrogate cohesin binding to chromatin to near completion using our auxin-inducible degron system.

We also performed ChIP-Seq for CTCF to establish whether CTCF binding was dependent on cohesin binding or not. Visual inspection and analysis of signal at all peaks called in data from untreated cells using MACS 2.0 clearly demonstrated that CTCF remained bound after RAD21 degradation (Fig. 1C,D). The average enrichment at all CTCF binding sites called by ENCODE was 10.89 in untreated cells and 8.93 in treated cells (The difference in enrichment was likely due to differential quality of the immunoprecipitation rather than true biological differences; one of our replicate experiments showed an average increase of CTCF binding after auxin treatment and the other showed an average decrease.) This data supports a model in which cohesin is a member of an extrusion complex that slides across DNA, whereas CTCF finds it binding sites independently of the extrusion complex and acts as an oriented brake to halt extrusion complex translocation.

Analysis of histone modification patterns—We calculated average signal over 5kb bins across the entire genome and correlated signal between ChIP-Seq experiments before and after auxin treatment for each of CTCF, H3K27me3, H3K9me3, H3K4me1, H3K4me3 and H3K27Ac. All modifications showed high correlations before and after auxin treatment (Spearman's r = 0.80 [CTCF]; 0.95 [H3K27me3]; 0.95 [H3K9me3]; 0.94 [H3K4me1], 0.70 [H3K4me3]; 0.90 [H3K27Ac]; 0.96 [H2.AZ]; 0.94 [H3K36me3]; 0.96 [H3K79me2]; 0.89 [H4K20me3]; 0.95 [H4K16Ac]).

Additionally, we examined overlaps of our ChIP-Seq peak annotations before and after auxin treatment for H3K4me3, H3K4me1 and H3K27Ac. High overlap (90%, 85% and 81% respectively) was seen for all three marks, indicating that the positions of active promoters and enhancers are largely unchanged after cohesin loss (Fig. S1D).

Finally, we examined the positions and enrichments of broad source marks (H3K27me3, H3K9me3, H2AZ, H3K36me3, H4K17Ac, H3K79me2). We called broad domains in our data from untreated cells using RSEG (see above), and identified a high-quality list of boundaries by retaining only the broad domains that were >=100kb in size and restricting to boundaries of these domains where at least a 50% change in signal across the boundary was seen (when averaging over the 50kb upstream and downstream of the boundary). We identified 2907 such boundaries for H3K27me3, 2647 boundaries for H3K9me3, 6356 boundaries for H2.AZ, 4829 boundaries for H3K36me3, 2515 boundaries for H4K16Ac, and 3634 boundaries for H3K79me2. We then extracted the data +/− 200kb from the boundary in both untreated and auxin-treated cells for each mark. The heatmaps and average profiles

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

across the boundaries called in untreated cells indicated that the positions of the boundaries was unchanged after auxin treatment (i.e. there was no spreading of histone modifications) and the average enrichments of histone modifications over the broad enriched domains were unchanged (Fig. S1E).

Taken together, these results suggest that histone modification patterns are largely unaffected by loss of cohesin.

Evaluation of loops and loop domains—We used HiCCUPS to calculate local enrichments on treated maps for the 3,170 loops we annotated with HiCCUPS in untreated maps. No loop showed at least 1.3-fold enrichment over local backgrounds and <30% FDR q-value. This clearly demonstrates that the vast majority of looping is lost after cohesin degradation.

We identified loop domains as in (Rao et al., 2014), by searching for loop-domain pairs where the peak pixel was within the smaller of 50kb or 0.2 of the length of the domain at the corner of the domain. Using this procedure, we identified 2,140 loop domains in untreated cells and only 9 in treated cells. Of the 9, 8 were false positives due to rearrangements in HCT-116 cells and one was a false positive due to extensive compartmentalization that was mistakenly annotated as a loop by HiCCUPS. Notably, the high false discovery rate after auxin treatment is due to the very small number of true positives (in this case, no detectable true positives). The false discovery rates of HiCCUPS and Arrowhead before auxin treatment were comparable to the FDRs documented in (Rao et al., 2014). This clearly demonstrates that loop domains are lost after cohesin degradation.

We also assessed the loss of loop domains and loops via aggregate peak analysis (APA). We used default parameters at 10kb resolution, excluding loop domains and loops within 300kb of the diagonal to avoid distance decay effects and extracting a 200kb by 200kb submatrix around every loop domain or loop. In aggregate, the signal from loop domains and loops was clearly and completely lost after auxin treatment: the APA score (fold-enrichment of the peak pixel over the mean value of the 36 pixels in the 6×6 box in the lower left of the aggregate matrix) went from 2.10 to 0.78 for loop domains and 2.09 to 0.78 for all loops. (The APA scores <1 after treatment are expected since random pixels would show an APA score <1 because of the contact probability distance decay.) All visual signs of looping and domain formation were also lost in the aggregate matrices (Fig. S2C). In addition, we confirmed that the loss of loops/loop domains was not a result of any cell cycle effects; we visually observed loss of loop domains after auxin treatment in G1/S arrested cells (Data S1, I.H-L) and observed similar results to above via APA (Fig. S2C).

In order to assess the dynamics of loop and domain formation we used APA and aggregate domain analysis (ADA) to assess loop, loop domain, and domain strength across a time course of auxin treatment and withdrawal.

ADA uses the same principle of aggregating submatrices across a feature list, but instead of calculating a score representing the focal enrichment of a peak pixel against pixels to its lower left (the APA score), we calculate a score representing the enrichment of contacts just

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

inside the domain boundaries over the contacts just outside the boundary, i.e. a gradient across the boundary. More specifically, we compare the average contacts in the pixels [i+3, j-13:j-3], [i+4,j-12:j-2], [i+5,j-11:j-1], [i+1:i+11,j-5], [i+2:i+12,j-4], [i+3:i+13,j-3] (the inside domain pixels) to the pixels [i-5,j-21:j-11], [i-4,j-20:j-10], [i-3,j-19:j-9], [i+11:i+21,j +5], [i+10:i+20,j+4], [i+9:i+19,j+3] (the outside domain pixels) where (i,j) is the center of the aggregate matrix (i.e. the corner of the domains). Here, we extracted a 200kb by 200kb matrix at 5kb resolution around every domain corner.

For APA on the time course experiments, as with the APA on our deep maps, we used default parameters at 10kb resolution.

In Fig. 2B, we show the APA scores for all loop domains greater than 300kb in size. The APA scores demonstrate that after cohesin is degraded, loop domains are completely lost between 40–60 minutes after treatment. From our own imaging (Fig. S1A,B) and imaging performed in Natsume et al. (2016), we know that the half life of cohesin after auxin treatment is about 20 minutes. Thus, loop domains are lost within minutes of cohesin degradation, indicating that cohesin is required for active maintenance of loop domain structures, not just establishment. After withdrawal of auxin, loop domains form similarly quickly, with strong loop domain signal by 60 minutes after withdrawal. This time includes the time for cohesin levels to recover and thus likely represents a very conservative upper bound on the time required for loop domain formation. Additionally, it indicates that loop domain structures are dynamically maintained during interphase.

In Fig. S2B, we show similar results for all loops greater than 300kb in size called in untreated cells. Additionally, we show ADA scores for all contact domains larger than 300kb in size called in untreated cells. While the ADA scores for all domains show a sharp decline in the first 60 minutes after auxin treatment, they plateau above 1, indicating some residual domain signal from al domains. Since, as we noted in (Rao et al., 2014), not all domains are loop domains, this suggest residual retention of non-loop domains. As we noted in (Rao et al., 2014), non-loop domains are usually created by compartment intervals. The retention of compartment domains is discussed below.

Analysis of previous cohesin-depletion Hi-C data sets—Previous Hi-C studies after cohesin or CTCF depletion showed limited effects, with both contact domains and compartments present after depletion (Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014). However, in these studies, the authors performed low resolution Hi-C experiments, raising the possibility that either (i) the authors could not resolve the difference between loop domains (which disappear after cohesin loss) and compartment domains (which remain) due to resolution issues, or (ii) incomplete depletion of cohesin or CTCF led to modest phenotypes. The authors in all three studies acknowledge the possibility that the limited effects they see may have been due to incomplete depletions.

To test this hypothesis, we re-analyzed data from these three studies. We downloaded the raw fastqs for all Hi-C experiments performed in the studies and processed them with Juicer (in exactly the same way that we processed all the Hi-C data generated for this study. Although the experiments did not have sufficient resolution to visualize individual loops, we

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

looked for the statistical signal of loop enrichment in aggregate using APA (Durand et al., 2016b; Rao et al., 2014). For the mouse data sets generated in Seitan et al. and Sofueva et al., we used a loop list we had previously generated in CH12-LX mouse lymphoblast cells (Rao et al., 2014) with the added filter that we removed loops with >4.5 enrichment over local background in order to stay consistent with the methods used in this study (see above). For the human data sets generated in Zuin et al., we used the loop list of 3,170 loops in untreated HCT-116 RAD21-mAC cells described above. We observed positive APA scores (>1) and visible focal enrichment in all experiments generated in previous studies, before and after cohesin or CTCF depletion (Fig. S2C). By contrast, our maps after auxin treatment show complete loss of APA signal and no visible focal enrichment, even when APA is performed on low resolution data sets (Fig. 2B, S2C). In previous studies, the APA score was weaker after cohesin or CTCF depletion but still clearly visible and notably, positive APA signal was seen in every replicate experiment performed in previous studies. Taken together, this suggests that a major confound of previous studies was the incomplete depletion of cohesin or CTCF, and along with the low resolution of the Hi-C experiments, likely explains the limited effects seen.

Evaluation of the dynamics of loop domain formation—In order to better understand the process of loop domain reformation, we utilized loop-resolution Hi-C datasets generated after 6 hours of auxin treatment and 20, 40, 60 or 180 minutes of auxin withdrawal. We sequenced 951M, 952M, 909M, and 949M reads from each of the timepoints respectively. We then converted our list of loop domains identified in untreated RAD21-mAC cells (generated at 5kb, 10kb and 25kb resolutions) into a standard 25kb resolution, by identifying the 25kb pixel containing the peak pixel as well as the 8 25kb pixels around it, and assigning the new 25kb peak pixel as the one of nine containing the most contacts. (We analyzed the 9 25kb pixels rather than just the 25kb pixel containing the original peak pixel in order to account for small amounts of noise in the peak localization.)

Using this list of 1,988 loop domains at 25kb resolution (leaving out loop domains <100kb in size which are more difficult to analyze at 25kb resolution), we used HiCCUPS to identify the observed and local expected values for all peak pixels in our untreated, auxin treated, 20 min withdrawal, 40 min withdrawal, 60 min withdrawal, and 180 min withdrawal data sets. In order to compare between datasets of different sizes, we scaled the observed and expected values for each map down by a scalar factor equal to the (# of contacts in the map/# of contacts in the 20 min withdrawal map). (The 20 min withdrawal map was our lowest sequencing depth map.) For each loop domain, we constructed a recovery curve, setting the number of observed contacts in the untreated map as 1, the number of observed contacts in the auxin treated map as 0, and (observed – [6hr treat observed])/([no treat observed] – [6hr treat observed]) as the value at every other time point (Fig. 2C). (Similar results were observed if we used an observed/local expected metric, or a z-score metric [(observed-local expected)/sqrt(local expected)]; data not shown.)

To rank loop domains by speed of recovery, we utilized a metric where we calculated the difference between the loop domain's recovery at a particular time point and the recovery of the median loop domain at that time point. We calculated this value for all loop domains at the 40 minute and 60 minute time points and summed the two to get a recovery score for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

each loop domain. To identify features associated with variation in loop domain recovery time, we stratified loop domains by the above recovery score, and compared loop domains from different quantiles to a number of features (NIPBL binding, promoters, enhancers, histone modifications). For punctate features (such as NIPBL binding sites, promoters and enhancers), we calculated the density of peaks across each domain (i.e. peaks/Mb). For broad source features (such as H3K36me3 and H3K27me3), we calculated the average enrichment across each domain. For each quantile, we calculated the enrichment of a feature by comparing the average peak density or signal enrichment of the quantile to the average peak density or signal enrichment across all loop domains (Fig. 2D). Additionally, we included superenhancers (as annotated by Hnisz, et al Cell 2013) and strong NIPBL binding sites (the top 10% of binding sites called in a merged list of peaks called in both untreated and auxin treated data sets) in our feature list. We found that activating marks, and in particular superenhancers and strong NIPBL binding sites, were enriched in loop domains that recovered quickly (fast loop domains) vs loop domains that recovered slowly (slow loop domains).

We also stratified loop domains by their density of NIPBL ChIP-Seq peaks and H3K27Ac ChIP-Seq peaks. We found that the top 10% of loop domains by density of NIPBL peaks (>=24 peaks/Mb) showed 38% recovery by 40 minutes, while the bottom 10% of loop domains by density of NIPBL peaks (0 peaks/Mb) showed only 11% recovery by 40 minutes. Similarly, the top 10% of loop domains by density of H3K27Ac peaks (>=54.4 peaks/Mb) showed 38% recovery by 40 minutes, while the bottom 10% of loop domains by density of H3K27Ac (<=2.76 peaks/Mb) showed only 8% recovery by 40 minutes (Fig. S2D).

We also repeated the above analyses after restricting to loop domains that exhibited no NIPBL or H3K27Ac binding within 50kb of either loop anchor, in order to test whether features in the interior of a loop domain could affect the formation of loops. We observed similar enrichments of activating factors, and especially strong NIPBL sites and superenhancers at the loop domains that recovered more quickly vs those that recovered more slowly. The top 10% of loop domains by NIPBL density (>=12 peaks/Mb) showed 27% recovery by 40 minutes vs. 9% recovery by 40 minutes for the bottom 10% of loop domains (0 peaks/Mb). The top 10% of loop domains by H3K27Ac density (28 peaks/Mb) showed 29% recovery by 40 minutes vs. 7% recovery by 40 minutes for the bottom 10% of loop domains (0 peaks/Mb). The association between NIPBL and enhancers and faster loop domain recovery, even when restricting to domains that only exhibit these features in the interior of the domain far from the anchors, is highly suggestive of an extrusive process; the ability of a feature far from the anchors of a loop to modulate loop domain formation time is more consistent with an extrusion model rather than a 3D diffusion model. Of course, these features may be simply correlated with other features at the anchors, or may modify the flexibility of the chromatin fiber in a way that affects the rate of 3D diffusion between the anchors; further experiments are needed to elucidate the relationship between NIPBL, superenhancers and loop domain formation speed.

Evaluation of genome compartmentalization—The most common method used for classifying Hi-C patterns is the principal component (PC) approach, which we introduced in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Lieberman-Aiden et al., 2009). In this approach, each intrachromosomal contact matrix is converted to an observed/expected matrix, and the first principal component of this matrix is used to bifurcate the data into two clusters. We showed in (Rao et al., 2014) that this method does not capture compartment structure accurately at high resolutions; however it is useful for comparing gross compartmentalization patterns.

We first calculated the first three principal components of the 25-kb resolution observed/expected matrix for each chromosome (constructed using Juicer) using scikit learn's RandomizedPCA function. We chose the principal component most correlated with GC content and assigned sign such that the vector was positively correlated with the GC content vector. We then calculated the correlation of the eigenvector for each chromosome between untreated and treated maps. The mean correlation was 0.968.

To identify transitions in compartment state at higher resolution, we used a combination of techniques. First, we calculated an edge score using an algorithm similar to Canny edge detection. For every 25kb locus in the genome, we looked at the corresponding column of the 25kb log2(observed/expected) matrix. For every pixel (i,j) in column j, we calculated a gradient = [i,j:j+3] – [i,j-4,j-1]. We then searched for stretches of at least 7 pixels in the column with a gradient × such that abs(x) was greater than 0.5. We then extended the edges by including pixels adjacent to an edge that had a gradient of at least 0.3. Finally, we summed the number of pixels in a column belonging to an edge to calculate the edge score for a locus. We then called local peaks in this track which could correspond to compartment state transitions since a compartment state transition at locus I will create an edge between locus i-1 and i.

Since loci in the same compartment will exhibit the same rises and falls in contact probability as one slides along the genome, we reasoned that adjacent pixels should exhibit high correlations of the derivative of their contact patterns and low correlations could indicate a compartment state transition. As described in Section V.a.3 of Rao et al., 2014, this is akin to measures in finance that correlate returns of prices to identify similarities between stocks. To calculate this sliding derivative correlation score, we calculated the gradient in the log2(observed/expected) matrix over every boundary called in our edge score track. More specifically, for every locus i, and all boundaries j in boundary set J that were within 15Mb of i, we calculated the difference of mean([j:j+5,i]) and mean([j-6:j-1,i]). We then calculated the Spearman correlation coefficient of these two vectors (one vector for the gradients at all boundaries j in J for the pixels upstream of i, and one vector for the pixels downstream of i). We excluded the derivative signal at pixels not located at compartment state transitions as defined by the edge score to reduce noise, reasoning that pixels inside compartment intervals were unlikely to contribute meaningful rises/falls in contact probability. Similarly, we only included pixels within 15mb of i to reduce the noise arising from sparsity far off the diagonal.

We then identified compartment boundaries by calling local peaks in the edge score track and local valleys in the sliding derivative correlation score track and merging the two peak call lists.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We identified 4,325 boundaries in untreated cells for a median compartment size of 425 kb and 4,424 boundaries in treated cells for a median compartment size of 475 kb. These are very likely conservative upper bounds on the true median compartment size, since we utilized stringent peak calling and compartment structure can be difficult to detect in maps that are not extremely high resolution (Rao et al., 2014).

To assess the presence and strength of contact domains after auxin treatment, we used the Arrowhead algorithm (Durand et al., 2016b; Rao et al., 2014). The Arrowhead algorithm calculates a corner score for every pixel, where higher corner score values represent a higher likelihood that a pixel is at the corner of a domain (see Section IV.a.3 of the Extended Experimental Procedures of Rao et al., 2014). For our list of 9,845 contact domains identified by the Arrowhead algorithm in untreated RAD21-mAC cells, we compared the corner scores of the contact domains to the corner scores of random pixels with an identical chromosome and length distribution. The median corner score in untreated cells for all domains called in untreated cells was the 97th percentile of random corner scores. We then calculated the corner scores in treated cells for the list of contact domains annotated in untreated cells, as well as the corner scores for the random control. Here, the median corner score for annotated contact domains was only the 86th percentile of random corner scores. (Notably, the distributions of scores for random pixels did not change, see Fig. 3B.) This indicates that contact domains were significantly weakened after auxin treatment. However, there was still some residual signal.

Since we knew that loop domains were completely eliminated from our previous analyses and that compartment structure remained after treatment, we reasoned that the residual signal was arising from retained compartment domains (contact domains whose boundaries overlap compartment interval boundaries). To test this, we identified 974 contact domains whose boundaries overlapped a compartment interval boundary (within 25 kb), i.e. compartment domains. Additionally, we identified 410 contact domains whose boundaries were not within 100 kb of a compartment boundary even after using a relaxed threshold for identifying compartment boundaries, i.e. a high confidence set of non-compartment domains. We then analyzed the corner scores for each of these sets of domains separately in treated cells and found that while the median score for compartment domains was 89th percentile of the random corner scores, the median score for non-compartment domains was only 72nd percentile of the random corner scores. This indicates that the residual signal stems from retained compartment domains. Thus, while loop domains are completely eliminated, contact domain structure arising from genome compartmentalization is still present after auxin treatment, although the domains are weaker than those found in untreated cells.

It is commonly thought in the literature that contact domains and compartment intervals form a hierarchy, with compartment intervals often being subdivided into multiple contact domains, but each contact domain belonging to only one compartment interval. Having determined that loop domains and compartmentalization formed via independent mechanisms, we wondered whether loop domains and compartment intervals shared characteristic hierarchical relationships or whether they truly formed independently in the genome.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To assess whether compartment boundaries could be spanned by loop domains, we intersected our loop domain annotation and our compartment boundary annotation. Specifically, we identified compartment boundaries in our treated maps that were contained within a loop domain called in untreated cells and >100kb away from either loop anchor (obviously this excludes loop domains smaller than 200kb from the analysis). We identified 349 such boundaries. Visual examination also confirmed that these boundaries were true compartment state transitions lying inside loop domains (Fig. 3C,D). Note that this is a lower bound on the number of compartment boundaries spanned by loop domains, as we used stringent distances from loop anchors to reduce false positives and our compartment boundary annotation has false negatives as well. This demonstrates that there is no true hierarchy between compartmentalization and loop domain formation, contrary to what has been suggested in the literature. We wondered what happen to compartment strength at these boundaries when loop domains were eliminated. To analyze this, we calculated the average sliding derivative correlation score (see above) for the 1Mb intervals centered on the 349 compartment boundaries contained within loop domains before and after auxin treatment.

We observed that the boundaries contained within loop domains showed a strong increased in compartment strength (larger dip in the sliding correlation score) after the elimination of loop domains: 0.10 decrease in the sliding correlation score in untreated cells vs. 0.31 in treated cells. In contrast, when we identified 389 compartment boundaries in treated cells that were positioned at loop domain anchors annotated in untreated cells (within 25kb), we found that there a much more modest increase in compartment strength after treatment: 0.35 decrease in the sliding correlation score in untreated cells vs. 0.53 in treated cells (Fig. 3E). This indicates that cohesin facilitates mixing of distinct compartment states and causes decreases in compartmentalization unless it is halted at the compartment boundary.

The results were similar when we examined compartment boundaries inside all loops: we identified 593 compartment boundaries in treated cells that were spanned by loops and at least 100kb away from either loop anchor, and we identified 503 compartment boundaries in treated cells that were positioned at loop anchors. We saw an 0.11 decrease in the sliding correlation score in untreated cells vs. 0.37 decrease in treated cells for compartment boundaries spanned by loops, and an 0.38 decrease in the sliding correlation score in untreated cells versus an 0.54 decrease in treated cells for compartment boundaries at loop anchors (Fig S3D).

To assess whether the changes in compartmentalization seen after treatment corresponded to epigenetic activity, we performed a similar analysis except instead of calling compartment boundaries, we identified transitions in broad histone modification state for H3K27Ac and H3K27me3. Since histone modifications have been shown to very closely correlate with compartmentalization (Lieberman-Aiden et al., 2009; Rao et al., 2014; Sexton et al., 2012), we reasoned that changes in histone modification within loop domains and loops should show greater changes in compartmentalization to better match the histone modification pattern compared to changes in histone modification status at loop anchors. We identified changes in H3K27Ac status by creating a 25kb binary track that was either 0 if the enrichment was less than 0.35 or 1 if the enrichment was greater than 0.35. We then calculated the absolute value of a smoothed gradient (using the kernel [1 1 1 -1 -1 -1]) and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

called local peaks to identify changes in histone modification status. We identified 264 H3K27Ac transitions spanned by loop domains (same definition as above) and 307 H3K27Ac transitions positioned at loop domain anchors. The H3K27Ac signal in the 1Mb intervals around these transitions did not change after auxin treatment (Fig. 3F). However, while there was very little change in the compartmentalization strength at transitions at loop domain boundaries (0.41 dip in sliding correlation in untreated vs. 0.49 in treated), there was a dramatic increase in compartmentalization strength at transitions spanned by loop domains (0.02 dip in sliding correlation in untreated vs. 0.19 in treated). This indicates that removal of loop domains by cohesin loss leads to genome compartmentalization that more closely matches histone modification patterns.

Similar results were seen for H3K27Ac transitions spanned by all loops: we identified 426 H3K27Ac transitions in untreated cells that were spanned by loops and at least 100kb away from either loop anchor, and we identified 381 H3K27Ac transitions in untreated cells that were positioned at loop anchors. The H3K27Ac signal in the 1Mb intervals around these transitions did not change after auxin treatment (Fig. 3F). We saw an 0.41 decrease in the sliding correlation score in untreated cells vs. 0.50 decrease in treated cells for H3K27Ac transitions spanned by loops, and an 0.10 decrease in the sliding correlation score in untreated cells versus an 0.26 decrease in treated cells for H3K27Ac transitions at loop anchors (Fig S3B).

We also performed this analysis for H3K27me3. We calculated the gradient at every 25kb locus i the genome by taking the absolute value of the difference between the summed log2 fold-enrichment for pixels i-8 to i-1 and the summed log2 fold-enrichment for pixels i+1 to i +8. We called local peaks on this gradient track to identify loci where the broad H3K27me3 modification status changed. We identified 209 H3K27me3 transitions spanned by loop domains (same definition as above) and 384 H3K27me3 transitions positioned at loop domain anchors. The H3K27me3 signal in the 1Mb intervals around these transitions did not change after auxin treatment (Fig. S3A). However, while there was very little change in the compartmentalization strength at transitions at loop domain boundaries (0.29 dip in sliding correlation in untreated vs. 0.33 in treated), there was a stronger increase in compartmentalization strength at transitions spanned by loop domains (0.01 increase in sliding correlation in untreated vs. 0.03 dip in treated).

Similar results were seen for H3K27me3 transitions spanned by all loops: we identified 391 H3K27me3 transitions in untreated cells that were spanned by loops and at least 100kb away from either loop anchor, and we identified 469 H3K27me3 transitions in untreated cells that were positioned at loop anchors. The H3K27me3 signal in the 1Mb intervals around these transitions did not change after auxin treatment (Fig. S3C). We saw an 0.27 decrease in the sliding correlation score in untreated cells vs. 0.31 decrease in treated cells for H3K27me3 transitions spanned by loops, and an 0.03 decrease in the sliding correlation score in untreated cells versus an 0.12 decrease in treated cells for H3K27me3 transitions at loop anchors (Fig S3C).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Taken together, these results suggest that cohesin facilitates mixing of chromatin with different histone modification states and loss of cohesin leads to better correspondence of genome compartmentalization with histone modification patterns and gene activity.

Annotation and analysis of cohesin-independent links—We first annotated loops in our maps for auxin-treated RAD21-mAC cells using default HiCCUPS parameters for 5,10, and 25kb resolutions (Durand et al., 2016b; Rao et al., 2014) with the additional requirement that the peak pixel show less than 4.5-fold enrichment over local expecteds (in order to remove as many false positives as possible due to rearrangements and assembly issues, see above). Using this procedure, we annotated 81 loops in treated RAD21-mAC cells. When we visually examined these loops, we found that 66 were false positives, with 55 of the false positives due to assembly issues, issues with repetitive elements or structural rearrangements. The false discovery rate for HiCCUPS is much higher in treated cells because the number of true positives is dramatically lower. As mentioned above, the false discovery rate in untreated cells was comparable to the rates described previously in (Rao et al., 2014); in fact, as one might expect false positives to arise from artifacts in the data that are independent of cohesin-mediated looping, the reduction by nearly 98% of numbers of loops called by HiCCUPS after auxin treatment is a powerful proof of its accuracy. When we examined the 15 true positive loops annotated by HiCCUPS, we found that they had a dramatically different distance distribution than cohesin-associated loops: where the median size of a cohesin-associated loop was 275kb, the median size of these 15 loops was 1.75Mb. We also noticed that the anchors involved in these 15 loops were often forming long-range loops at distances of tens of megabases and hundreds of megabases. We reasoned that HiCCUPS using default parameters for loop detection was missing many of these extremely long-range loops because of the extra stringency of the HiCCUPS lambda chunking procedure for multiple hypothesis testing for pixels with low counts (i.e. pixels far off the diagonal). To call more of these long-range loops, we decided to modify the HiCCUPS parameters similar to make the parameters more similar to those used to identify the extremely long-range “superloops” on the inactive X chromosome (Rao et al., 2014; Darrow et al., 2016).

We decided to annotate loops in auxin-treated RAD21-mAC cells with the parameters used in to annotate superloops on the inactive X chromosome (which also form between loci tens to hundreds of megabases apart). More specifically, we annotated loops by running HiCCUPS at 50 and 100kb resolutions with the following parameters: p = 2,1; w = 4,2; fdr = 10%, 10%. We additionally filtered loops that were within 5 Mb of the diagonal, had less than a 2-fold observed/expected for any of the local expected, and had fewer than 3 pixels clustered into the peak pixels (see section VI.a.5 of Rao et al., 2014). This annotation yielded 88 loops. After visual examination, we found that 46 of these loops corresponded to true positives while the other 42 were false positives (22 were due to issues with repetitive regions and 15 were due to other forms of structure in the contact map, for instance interactions between broad compartment intervals). Combining these 46 loops with the 15 loops annotated with high resolution HiCCUPS, we obtained a final curated list of 61 intrachromosomal cohesin-independent loops.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We first identified the loop anchors contributing to the cohesin-independent loops. We merged all adjacent loci involved in one of the 61 loops annotated above. We then expanded all loop anchor loci to be 100kb in size, yielding a list of 64 loop anchor loci.

To assess the presence and orientation of CTCF at loop anchor loci for both cohesin-associated and cohesin-independent loop anchors, we followed the procedure exactly from section VI.e.7 of (Rao et al., 2014). In order to use comparable loop anchor sizes, we collapsed each 100kb cohesin-independent loop anchor to the 15kb interval in the center of the 100kb interval. We found that while 90% of cohesin-associated loop anchors were associated with CTCF binding, only 20% of cohesin-independent loop anchors were associated with CTCF binding. More over, while 95% of unique CTCF motifs in cohesin-associated loop anchors pointed towards the interior of the loop (consistent with the convergent rule), the unique CTCF motifs in cohesin-independent loops did not exhibit any such bias (56% pointing towards the interior of the loop) (Fig. 4C). This strongly suggests that cohesin-independent loops form via a mechanism other than extrusion.

To analyze enrichment of proteins bound at cohesin-independent loop anchors, we reproduced the analysis from section VI.e.7 of (Rao et al., 2014), using the 100kb loop anchors and comparing to the average of 100 randomly shuffled loop anchor lists (see the section on Random Shuffle controls above). We downloaded peak calls for 36 DNA-binding proteins or histone modifications in HCT-116 cells from ENCODE (ENCODE Consortium, 2012). We also utilized an annotation of stitched and ranked (by H3K27Ac enrichment) superenhancers and enhancers from (Hnisz et al., 2013). For each of the proteins or histone modifications, we calculated the percentage of loop anchors that overlap the feature as well as the enrichment over the percentage of random anchors overlapping the feature. We found that strong H3K27Ac sites and superenhancers (especially the strongest 100 superenhancers) were very strongly enriched at cohesin-independent loop anchors (Fig. 4D). We also wondered whether broad H3K4me3 peaks were enriched at cohesin-independent loop anchors. In order to assess this, we first created an annotation of broad H3K4me3 peaks. In brief, we called peaks using MACS 2.0 with the '— broad' option enabled, and then merged peaks that were within 5kb of each other. We then retained peaks from this merged list that were >10kb long to yield a final list of 549 broad H3K4me3 peaks. We analyzed enrichment of broad H3K4me3 peaks at cohesin-independent loop anchors as we did for all other protein peak calls (see above). Notably, we observed that 36% (23/64) of cohesin-independent loop anchors overlap a broad H3K4me3 peak, a 21-fold enrichment over random chance. We also performed the analyses listed above on automated lists of cohesin-independent loops without any manual curation. We found that the results showing a lack of CTCF binding at cohesin-independent loop anchors and a lack of CTCF orientation preference were similar (Fig. S4A).

We also found that superenhancers were strongly enriched at loop anchors generated from the 88 loop list automatedly called with low resolution HiCCUPS; the top 100 superenhancers were 47-fold enriched (present at 30/115 loop anchors). See figure S4B. This indicates that our results were not biased by our use of a manually curated loop list.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We noticed that our 64 cohesin-independent loop anchors determined from our 61 loop intrachromosomal list often formed focal interchromosomal links between pairs of loop anchors and that there were large cliques of interactions between anchors (Fig. 4A,E, 5A). This is in stark contrast to cohesin-associated loop anchors, which show no such enrichment for extremely long Intrachromosomal interactions or interchromosomal interactions, either when examined individually or in aggregate via APA (Fig S2A). This strongly suggests that cohesin-independent loops and links form via a mechanism other than extrusion, since extrusion cannot occur on two topologically distinct molecules..

To annotate these interchromosomal links between pairs of cohesin-independent loop anchors, we used HiCCUPS to calculate local enrichments at 100kb resolution for all possible interchromosomal pairs of cohesin-independent loop anchors. We then identified enriched focal interchromosomal interactions by filtering for links that were enriched at least 5.5-fold over local background (empirically chosen to ensure a <10% false discovery rate). Using this procedure, we identified 203 interchromosomal cohesin-independent links. This likely underestimates the true number of interchromosomal cohesin-independent links, as evidenced by Fig. 5A.

We analyzed the change in strength of cohesin-independent links after auxin treatment by using APA at 100kb resolution. APA analysis clearly demonstrated that while cohesin-independent links (both intra and interchromosomal) were weakly present before auxin treatment, they were ~2-fold strengthened after auxin-treatment (Fig S4C,D). This result was robust to using either our manually curated lists (of 61 intra and 203 inter chromosomal links) or automatedly generated lists (all intrachromosomal pairs of the 47 superenhancers overlapping anchors in the 88-loop automated list from above and all interchromosomal pairs of the 47 superenhancers overlapping anchors in the 88-loop automated list from above) (Fig S4C,D).

We also analyzed induction of cohesin-independent links across an auxin treatment and withdrawal time course. This analysis was performed as above with the cohesin-associated loops and loop domains, but at 100kb resolution instead of 10kb resolution and for both our 61 intrachromosomal links and our 203 interchromosomal links. The opposite pattern of cohesin-associate loop formation was seen; APA scores for cohesin-independent links rapidly increased upon auxin treatment and rapidly dropped upon auxin withdrawal (Fig. 4E). Similar results were seen upon performing the time course APA at 100kb resolution using all intrachromosomal pairs of the 47 superenhancer overlapping anchors in the 88-loop automated list from above (Fig. S4E).

Analysis of higher order contacts—In order to assess whether the large cliques between superenhancers that we observed after cohesin loss corresponded to higher order hubs, we utilized higher order contacts present at low frequencies in *in situ* Hi-C data. The Juicer pipeline (Durand, et al, 2016b) separately outputs abnormal chimeric read pairs (i.e. read pairs that map to more than 2 loci). We combined all of our data from untreated RAD21-mAC cells and all our data from treated RAD21-mAC cells (unsynchronized and synchronized/arrested), parsed the chimeric abnormal reads and deduped them (using the same deduping procedure used in Juicer, only applied to three, four or five positions, rather

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

than just two), retaining unique reads that mapped to three or more positions with MAPQ>=10. With this procedure, we obtained 32M triples, 18M quadruples, and 75K quintuples in our untreated data set, and 25M triples, 14M quadruples and 55K quintuples in our treated data set.

Given the sparsity of our higher-order data set, we were unable to call individual hubs. However, we reasoned that just as APA enabled us to examine the aggregate enrichment of loops in low-resolution pairwise Hi-C data sets, we could perform 3D-APA to examine the aggregate enrichment of trio hubs in a low-resolution triple tensor. In order to perform 3D-APA, we identified 131 intrachromosomal trios of cohesin-independent loop anchors, where each pair of loop anchors in the trio were at least 10Mb apart (we merged loop anchors that were within 1 Mb of each other, in order to avoid double counting in the 3D-APA aggregate tensor). We then extracted a 3.9Mb × 3.9Mb × 3.9Mb sub-tensor at 300kb resolution, centered on the 300kb × 300kb × 300kb voxel containing each trio, and summed these cubes to get an aggregate 3D-APA sub-tensor. When summing the 131 cubes, the cubes were always oriented so that the upstream locus was on the z-axis, the middle locus was on the x-axis and the downstream locus was on the y-axis.

While we did not observed many quintuples, we did observe high rates of quadruples, so we devised several strategies to project quadruples onto triple space in order to use them along with our triples in searching for hubs. A naive strategy would be to extract all 4 choose 3 triples from each quadruple and count each one separately. However, using this method does not allow us to utilize Poisson statistics to calculate expected models, as the four triples extracted from a given quadruple are not independent. The most stringent way to handle this issue is by randomly discarding one of the four loci. However, this tends to underutilize the information contained in quadruples. For instance, if one has a quadruple ABCD, where the triple ABC falls within my 3D-APA sub-tensor but D is outside of all of the sub-tensors, using the random projection method, there is a 75% chance of completely discarding the ABCD quadruple, despite the fact that it contains some information relevant to the hypothesis being tested in 3D-APA. In order to maximize utilization of the information contained quadruples but simultaneously ensure that Poisson statistics were still applicable, we developed a “Poisson-projection” method. Namely, for a quadruple ABCD, we would examine all four contained triples (ABC, ABD, ACD, BCD). If one and only one fell inside our 3D-APA sub-tensor (say ABC), then we would include that as a triple for further analysis. If more than one of the four fell within our 3D-APA sub-tensor, we would randomly choose one of the triples that fell inside our 3D-APA sub-tensor to include in the analysis. By randomly choosing among the contained triples that fell within the bounds of our 3D-APA sub-tensor, we can avoid double counting and maintain the independence of events necessary for Poisson statistics to apply.

Using our triple dataset as well as the Poisson projection of our quadruples, we identified 11 contacts that fell within the center voxel of the 3D-APA sub-tensor in our auxin-treated dataset, where as we saw none in our untreated dataset. No other voxel in the sub-tensor for either dataset (out of 4394 voxels) contained more than 5 reads (Fig. 5D,E). We also extracted the aggregate 3D-APA sub-tensors corresponding to shifting one or more the loci in each trio by 3.9Mb. No other voxel in the sub-tensor for either dataset (out of 118,638

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

voxels) contained more than 8 reads (Fig. 5F). In order to assess the statistical significance of seeing 11 contacts in the center voxel of our 3D-APA sub-tensor after auxin treatment, we also calculated a number of local expecteds (Darrow, et al 2016). The center voxel of our 3D-APA tensor after auxin treatment was strongly enriched relative to all expected models. We also tested statistical significance against a local expected model that accounts for 2D bias, that is, the fact that three loci that show pairwise enrichments in the 2D matrix (i.e. they form loops) will show enrichments in the 3D tensor that corresponds simply to the product of their 2D enrichments and not to any higher order simultaneity; the 11 contacts we observe in the center voxel after auxin treatment is still significantly enriched (Fig. S5B, bottom model).

Taken together, these results highlight that superenhancers at cohesin-independent loop anchors interact simultaneously in higher order hubs after cohesin loss (Fig. 5G).

Simulations of extrusion and compartmentalization—Simulations were run for 200,000 timesteps with only Lennard-Jones intermonomeric forces and then for 800,000 timesteps with 8 extrusion complexes. In the HOOMD-blue molecular dynamics package (Glaser et al., 2015) (Anderson et al., 2008; Glaser et al., 2015) (Anderson et al., 2008; Glaser et al., 2015), temperature is set to 2.0 and gamma (viscosity) is set to 0.02. Contact maps and globules are shown from the final frame of simulation. In simulations of the auxin-treated condition, the final 800,000 timesteps were simulated without extrusion. All other parameters are as described in (Sanborn et al., 2015).

CTCF and cohesin binding strengths were determined by integrating a Gaussian fit to ChIP-Seq data around every CTCF motif. Simulated extrusion binding strengths were determined by taking the geometric mean of the CTCF and cohesin binding strengths and renormalizing to a binding probability, as described in (Sanborn et al., 2015).

Each monomer was assigned to either an "A" or a "B" type. Lennard-Jones forces between different-type monomers was set to 98% the strength of LJ forces between same-type monomers. Because compartment transitions can only be defined in Hi-C maps at coarse resolutions (25kb and above), the compartment transition of each simulation replicate was varied randomly within 30kb (30 monomers) of defined transition points.

Compartment transitions were determined in one of two ways: (1) A/B compartment states were annotated by hand for the regions that were simulated using the treated Hi-C maps or (2) 9 histone modifications (H3K27me3, H3K9me3, H3K36me3, H2.AZ, H3K79me2, H4K17Ac, H3K4me1, H3K27Ac, H4K20me3) were clustered into 6 clusters using k-means clustering, the clusters were further collapsed into two clusters (A or B) based on whether each cluster had a positive enrichment for H3K36me3 (A) or not (B). This two cluster track was then used as input for simulation. For the k-means clustering, the histone modification data was first converted to a z-score value for each mark in order to account for differences in the dynamic range between marks. The latter input was used for the simulation shown in Fig. 6A,B, demonstrating our ability to recapitulate all the major features of Hi-C datasets (loops, domains, and compartments) using only ChIP-Seq data as input. Both hand

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

compartment annotation and automated compartment annotation simulations are shown in Data S1, VI for comparison.

Oligonucleosome Resolution Simulations of Chromatin Fibers—Segregated interactions between short chromatin intervals, such as the cohesin-independent loop anchors we observed, have not previously been reported, and require the strength of the interaction between similarly-decorated nucleosomes to overcome the stiffness of the local chromatin fiber.

To explore whether such aggregation or segregating mechanisms are physically feasible on the oligonucleosome level, we simulated a coarse-grained mesoscale chromatin fiber (Fig. S6A,B) of 100 nucleosomes (~20kb) with NRL = 200bp without linker histone, consisting of either wild-type nucleosomes, nucleosomes in which the histone tails are rigid, reflecting the known effect of acetylation, as in the H4K16Ac mark, or 4 intervals of equal length, with the intervals alternating between wild-type nucleosomes and nucleosomes in which the histone tails are rigid. (The biophysical consequences of the H4K16Ac mark have been studied in detail, and are well understood in terms of more rigid tails, which in turn inhibit tail/tail internucleosome interactions and thus disrupt crucial stabilization of condensed chromatin fiber). In brief, the mesoscale chromatin model (Bascom and Schlick, 2017) represents the DNA as coarse-grained beads (each ~9bp) using a worm-like chain model; the nucleosome core particle with the wrapped DNA but without the histone tails is treated as a rigid body with ~300 pseudocharges that mimick the electrostatic environment of the nucleosome; and the histone tails are coarse grained to beads of about 5 amino acids using united-atom polymer chain models to mimick atomistic behavior. (See full details in Bascom, Kim, and Schlick 2017, and Grigoryev et al, 2016). Monte Carlo sampling of 40 million or more steps are performed for three ensembles of 100-nucleosome systems: wildtype fiber, all-folded control, and alternating construct where the pattern of 25-wt, 25-folded tails was repeated twice. From ensembles of up to 35 trajectories, contact maps were produced, and resulting chromatin configurations analyzed.

As seen in the contact maps and corresponding images in Fig. S6E, demarcated zones emerge in the ensemble of the alternating constructs: the wt nucleosomes segregate from the other nucloeosmes as well tend to associate with one another, while the folded-tail nucleosomes segregate separately. The alternating construct tends to adopt an overall figure-8 shaped fold or hierarchical loop (Grigoryev et al., 2016). The contact maps of these alternating constructs reveal these checkerboard patterns in marked contrast to the control wt and all-folded tail systems (Fig. S6C–E). The most dense regions in the contact maps of the alternating constructs come from wt/wt local and nonlocal interactions. These are followed by the local folded/folded and wt/folded interactions.

That such a striking segregation effect arises from simple alternating fiber constructs suggests that short chromatin intervals separate in sequence can spontaneously separate in space in specific patterns due to charge effects that alter the intrinsic histone-tail flexibility of specfiic nucleosomes. Already, we have shown that such domain segregation naturally emerges from nucleosome-free or depleted regions (Bascom, Kim, and Schlick 2017), but here the nucleosome spacing is uniform. Such alterations in tail flexibility can be caused by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

chemical modifications of the histone tails, DNA, or linker histones, as well as by protein anchoring, which can restrict the range of interactions to specific domains. It is likely that protein binding could amplify intrinsic segregation as well as induce domain aggregation of the fiber. Results with only H4 tails folded (mimicking H4K16Ac) are very similar to the case of all-folded tails (data not shown). In vivo and in vitro experiments on this length scale are also needed to probe these effects further.

Assessment of changes in transcription after cohesin loss—To look for signs of ectopic activation, we examined the 14,853 genes that were not expressed (RPKM<0.5) in untreated cells. We identified 2,145 genes that were significantly (adjusted p<0.05) changed by DESeq2. Of these genes, 1% (216) were ectopically activated after treatment (p<0.05, >30% change in RPKM, RPKM>0.5 in treated cells). In addition, 7% of these genes (1063) exhibited “leaky” transcription in treated cells: a larger PRO-Seq signal (p<0.05, >1.3 fold change difference) that fell short of the threshold for an expressed gene (i.e., RPKM was still below 0.5). 1.4% of these genes were significantly downregulated (>1.3-fold change), but it is unclear what reductions in expression at such low levels of expression mean biologically.

We next looked for changes in the 12,222 genes that were expressed (RPKM>0.5) in untreated cells (Fig. 6B). We identified 4,196 genes that were significantly changed (adjusted p<0.05) changed by DESeq2. Here again, most genes (87%, 10,615) exhibited similar levels of transcription after cohesin degradation (RPKM changed by less than 30%). The remaining genes (13%, 1607) showed a larger transcriptional effect (p<0.05, >30% change in RPKM). Stronger effects were seen, but less frequently: 64 genes (0.5%) showed a 2-fold change, and 2 genes showed a 5-fold change (Fig. 7B).

We identified 49 genes that were 1.75-fold downregulated with p<0.05 after auxin treatment. We noticed that many of the genes that were downregulated (by >1.75-fold) were located within 500kb of superenhancers (23 of 49, 4.8-fold enrichment compared to randomly shuffling the positions of the TSS of the 49 genes across the genome, Fig. 7C,D). Of these genes, 29% (14 of 49) were located with 500kb of one of the top 100 superenhancers (8.5-fold enrichment compared to randomly shuffling the positions of the TSS of the 49 genes across the genome). The overall distribution of distance to the nearest superenhancer was shifted significantly closer compared to randomly selected genes (Fig. 7D). Strikingly, these superenhancers were often located at the anchors of the cohesin-independent links seen in treated cells (8 of 19, a 13.7-fold enrichment).

To rule out the possibility that changes in gene expression were due to the auxin hormone itself, we performed PRO-Seq on HCT-116-CMV-OsTIR1 cells (HCT-116 cells with OsTIR1 at the *AAVS1* locus but no mAID tag on any protein) before and after auxin treatment. Only 105 genes were detected as significantly different, and only 56 genes were detected as significantly different with at least a 1.3-fold change. This indicates that our results are not confounded by the auxin hormone itself.

To rule out the possibility that tagging RAD21 itself led to significant transcriptional consequences, we compared our auxin-treated PRO-Seq data to a control of untreated HCT-116-CMV-OsTIR1 cells. The following paragraphs are the analyses from above except

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with the numbers from the CMV-OsTIR1 control. Analogous plots to those shown in Fig. 7B and 7D for the CMV-OsTIR1 control are shown in Fig. S7C–D.

To look for signs of ectopic activation, we examined the 14,884 genes that were not expressed (RPKM<0.5) in untreated cells. We identified 2,284 genes that were significantly (adjusted p<0.05) changed by DESeq2. Of these genes, 1% (255) were ectopically activated after treatment (p<0.05, >30% change in RPKM, RPKM>0.5 in treated cells). In addition, 7% of these genes (1179) exhibited “leaky” transcription in treated cells: a larger PRO-Seq signal (p<0.05, >1.3 fold change difference) that fell short of the threshold for an expressed gene (i.e., RPKM was still below 0.5). 1.8% of these genes were strongly downregulated (>1.3-fold change), but it is unclear what reductions in expression at such low levels of expression mean biologically.

We next looked for changes in the 12,191 genes that were expressed (RPKM>0.5) in untreated cells (Fig. 5B). We identified 4,251 genes that were significantly changed (adjusted p<0.05) changed by DESeq2. Here again, most genes (85%, 10,330) exhibited similar levels of transcription after cohesin degradation (RPKM changed by less than 30%). The remaining genes (15%, 1861) showed a larger transcriptional effect (p<0.05, >30% change in RPKM). Stronger effects were seen, but less frequently: 86 genes (1%) showed a 2-fold change, and 3 genes showed a 5-fold change (Fig. S7C).

We identified 43 genes that were 2-fold downregulated with p<0.05 after auxin treatment. We noticed that many of the genes that were downregulated (by >2-fold) were located within 500kb of superenhancers (28 of 43). Of these genes, 49% (21 of 43) were located with 500kb of one of the top 100 superenhancers. The overall distribution of distance to the nearest superenhancer was shifted significantly closer compared to randomly selected genes (Fig. S7D).

Our previous analyses (Rao et al 2014) have suggested that a subset of cell-type specific loops is associated with very strong gene activation (>10-fold upregulation of gene expression in the cell type where the loop is present). We repeated the same analysis from Rao et al (2014) with our HCT-116 untreated map from this study and our GM12878 map from our previous study in order to identify cases where cell-type specific loops appeared in HCT-116 and genes were simultaneously upregulated to then assess the effects of loop loss. To our surprise, we only identified 68 cell type specific loops (compared to ~600 each per pair of cell types in Rao et al 2014). In order to increase statistical power, we identified 518 loops that had been called in one of our cell types in Rao et al 2014 (HMEC, IMR90, K562, HUVEC, HeLa, NHEK) but not in GM12878, or vice versa and had been associated with 10-fold upregulation of a gene whose promoter lay at the loop anchor. We sought to then identify cases where these loop-gene pairs were present in HCT-116 to then examine the results of loop loss. However, only 15 of the 518 loops were conserved in HCT-116. Of those 15 loops and the 12 genes associated with them, 9 were not expressed in HCT-116 (of the other three, two were downregulated, and one was upregulated). These data are consistent with a number of possibilities: First, it is possible that HCT-116 is somehow distinct from all of our previously examined cell types and exhibits different distal regulatory principles. Alternatively, it is possible that the tagging of cohesin and resulting

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

potential loss of stability results in preferential loss of cell type specific loops. However, even if this were the case, the fact that we do not see large-scale expression changes between our untagged HCT-116 cells (above) and the auxin-treated RAD21-mAC cells (only 3 genes with a greater than 5-fold change in expression), suggests that the association between loop appearance and strong (>10-fold) gene activation may not be such that loop formation causes gene activation. In fact, it may be the case that some other regulatory event catalyzes both loop formation and gene activation, or that gene activation itself enables loop formation.

DATA AND SOFTWARE AVAILABILITY

All datasets reported in this paper are available at the Gene Expression Omnibus (GEO), series accession number GSEXXXX.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a Paul and Daisy Soros Fellowship, a Fannie and John Hertz Foundation Fellowship and a Stanford Medical Scholars Fellowship to S.S.P.R., NIGMS award R01GM055164 to T.S., and an NIH New Innovator Award (1DP2OD008540-01), an NSF Physics Frontier Center Grant (PHY-1427654, Center for Theoretical Biological Physics), the NHGRI Center for Excellence for Genomic Sciences (HG006193), the Welch Foundation (Q-1866), an NVIDIA Research Center Award, an IBM University Challenge Award, a Google Research Award, a Cancer Prevention Research Institute of Texas Scholar Award (R1304), a McNair Medical Institute Scholar Award, an NIH 4D Nucleome Grant (U01HL130010), an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375), and the President's Early Career Award in Science and Engineering to E.L.A. B.E.B. holds equity in Fulcrum Therapeutics. We thank Masato Kanemaki for sharing the HCT-116 RAD21-mAID-mClover cell line; Roger Kornberg, Olga Dudchenko, Peter Geiduschek and Miriam Huntley for their thoughtful comments on the manuscript; Neva Durand for assistance with computation; and Sigrid Knemeyer, Nathaniel Musial, and Lynn Zhu for assistance with figures. The experiments in this study were informed by discussions with E. Nora regarding his experiments on domain structure after CTCF degradation, and we thank him and his collaborators. All contact maps reported here can be explored interactively via Juicebox at http://www.aidenlab.org/juicebox/.

References

Alipour E, Marko JF. Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. 2012; 40:11202–11212. [PubMed: 23074191]

Anderson JA, Lorenz CD, Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units. Journal of Computational Physics. 2008; 227:5342–5359.

Bascom G, Schlick T. Linking chromatin fibers to gene folding by hierarchical looping. Biophysical Journal. 2017; 112:434–445. [PubMed: 28153411]

Bascom GD, Kim T, Schlick T. Kilobase Pair Chromatin Fiber Contacts Promoted by Living-System-Like DNA Linker Length Distributions and Nucleosome Depletion. J Phys Chem B. 2017; 121:3882–3894. [PubMed: 28299939]

Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas L-MM, Branco MR, et al. Complex multi-enhancer contacts captured by genome architecture mapping. Nature. 2017; 543:519–524. [PubMed: 28273065]

Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang S-CC, Sanborn AL, Machol I, Shamim M, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. U.S.A. 2016; 113:E4504–12. [PubMed: 27432957]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Davidson IF, Goetz D, Zaczek MP, Molodtsov MI, Huis In 't Veld PJ, Weissmann F, Litos G, Cisneros DA, Ocampo-Hafalla M, Ladurner R, et al. Rapid movement and transcriptional re-localization of human cohesin on DNA. EMBO J. 2016; 35:2671–2685. [PubMed: 27799150]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. 2016a; 3:99–101. [PubMed: 27467250]

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 2016b; 3:95–98. [PubMed: 27467249]

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature. 2016; 539:452–455. [PubMed: 27783602]

Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2016; 529:110–114. [PubMed: 26700815]

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016; 15:2038–2049. [PubMed: 27210764]

Glaser J, Nguyen TD, Anderson JA, Lui P. Strong scaling of general-purpose molecular dynamics simulations on GPUs. Computer Physics. 2015

Grigoryev SA, Bascom G, Buckwalter JM, Schubert MB, Woodcock CL, Schlick T. Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. Proc. Natl. Acad. Sci. U.S.A. 2016; 113:1238–1243. [PubMed: 26787893]

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, 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. [PubMed: 26276636]

Haarhuis JHIH, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHLH, Teunissen H, Medema RHH, van Steensel B, et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell. 2017; 169:693–707. e14. [PubMed: 28475897]

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. [PubMed: 24119843]

Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. Cell. 2017; 169:13–23. [PubMed: 28340338]

Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 2015; 16:167–177. [PubMed: 25693130]

Jost D, Carrivain P, Cavalli G, Vaillant C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. Nucleic Acids Res. 2014; 42:9553–9561. [PubMed: 25092923]

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–435. [PubMed: 20720539]

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 2012; 9:357–359. [PubMed: 22388286]

Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature. 2017; 547:236–240. [PubMed: 28636604]

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. [PubMed: 20080505]

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. [PubMed: 19815776]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Liu T. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol. Biol. 2014; 1150:81–95. [PubMed: 24743991]

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. [PubMed: 25516281]

Lupiáñez DGG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015; 161:1012–1025. [PubMed: 25959774]

Merkenschlager M, Nora EPP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu Rev Genomics Hum Genet. 2016; 17:17–43. [PubMed: 27089971]

Nasmyth K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annu. Rev. Genet. 2001; 35:673–745. [PubMed: 11700297]

Natsume T, Kiyomitsu T, Saga Y, Kanemaki MT. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. Cell Rep. 2016; 15:210–218. [PubMed: 27052166]

Nora EPP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012; 485:381–385. [PubMed: 22495304]

Nora EPP, Goloborodko A, Valton A-LL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 2017; 169:930–944. e22. [PubMed: 28525758]

Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines PS, Narisu N, Black BL, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. U.S.A. 2013; 110:17921–17926. [PubMed: 24127591]

Di Pierro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JNN. Transferable model for chromosome architecture. Proc. Natl. Acad. Sci. U.S.A. 2016; 113:12168–12173. [PubMed: 27688758]

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159:1665–1680. [PubMed: 25497547]

Sanborn AL, Rao SS, Huang S-CC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U.S.A. 2015; 112:E6456–65. [PubMed: 26499245]

Schwarzer W, Abdennur N, Goloborodko A. Two independent modes of chromosome organization are revealed by cohesin removal. bioRxiv. 2016

Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, Lenhard B, Giorgetti L, Heard E, Fisher AG, et al. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. 2013; 23:2066–2077. [PubMed: 24002784]

Sofueva S, Yaffe E, Chan W-CC, Georgopoulou D, Vietri Rudan M, Mira-Bontenbal H, Pollard SM, Schroth GP, Tanay A, Hadjur S. Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J. 2013; 32:3119–3129. [PubMed: 24185899]

Song Q, Smith AD. Identifying dispersed epigenomic domains from ChIP-Seq data. Bioinformatics. 2011; 27:870–871. [PubMed: 21325299]

Splinter E, Heath H, Kooren J, Palstra R-J, Klous P, Grosveld F, Galjart N, de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes & Development. 2006; 20:2349–2354. [PubMed: 16951251]

Stigler J, Çamdere GÖ, Koshland DE, Greene EC. Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. Cell Rep. 2016; 15:988–998. [PubMed: 27117417]

Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. Nature. 2017; 547:241–245. [PubMed: 28636597]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Wang X, Brandão HB, Le T, Laub MT. Bacillus subtilis SMC complexes juxtapose chromosome arms as they travel from origin to terminus. Science. 2017

Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 2008; 451:796–801. [PubMed: 18235444]

Wijchers PJ, Krijger PH, Geeven G, Zhu Y, Denker A, Verstegen MJ, Valdes-Quezada C, Vermeulen C, Janssen M, Teunissen H, et al. Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. Mol. Cell. 2016; 61:461–473. [PubMed: 26833089]

De Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell. 2015; 60:676–684. [PubMed: 26527277]

Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPP, van de Werken HJ, Knoch TA, van IJcken WF, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. SciU.SA. 2014; 111:996–1001.

An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- We track the 4D Nucleome at 10kb/20min resolution during cohesin loss and recovery.
- After cohesin loss, loop domains disappear; genome compartments are strengthened.
- Superenhancers form loops, interchromosomal links, and higher-order hubs.
- During cohesin recovery, loop domains form in minutes, consistent with fast extrusion.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

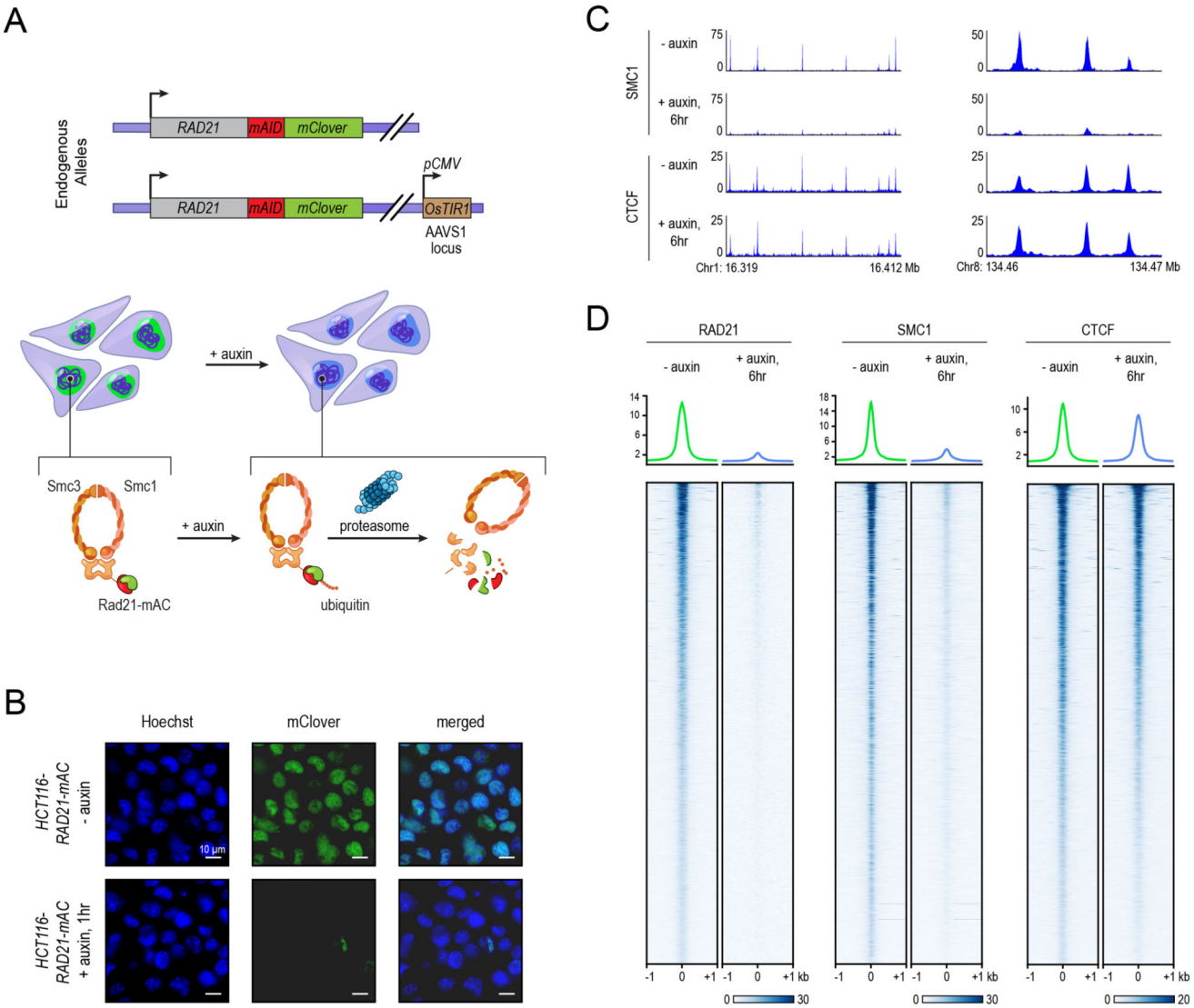


Figure 1. Tagging of endogenous *RAD21* with an auxin-inducible degron allows for rapid, near complete cohesin loss

(A) In HCT-116-RAD21-mAC cells, both *RAD21* alleles are tagged with auxin-inducible degrons and an mClover reporter, and the *OsTIR1* gene is integrated at the *AAVS1* locus. Auxin treatment leads to proteasomal degradation of RAD21.

(B) Live cell imaging after Hoechst 33342 staining to label nuclei. Nuclear mClover fluorescence corresponding to tagged RAD21 was lost after 1 hour of auxin treatment. (See Fig. S1.)

(C) SMC1 and CTCF ChIP-Seq signal with and without auxin treatment.

(D) RAD21, SMC1 and CTCF ChIP-Seq signal (left, middle, right) across all peaks called for each of the proteins in untreated RAD21-mAC cells. (Top) Average enrichments for each protein. After RAD21 degradation, the cohesin complex no longer binds to chromatin. CTCF binding is unaffected.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

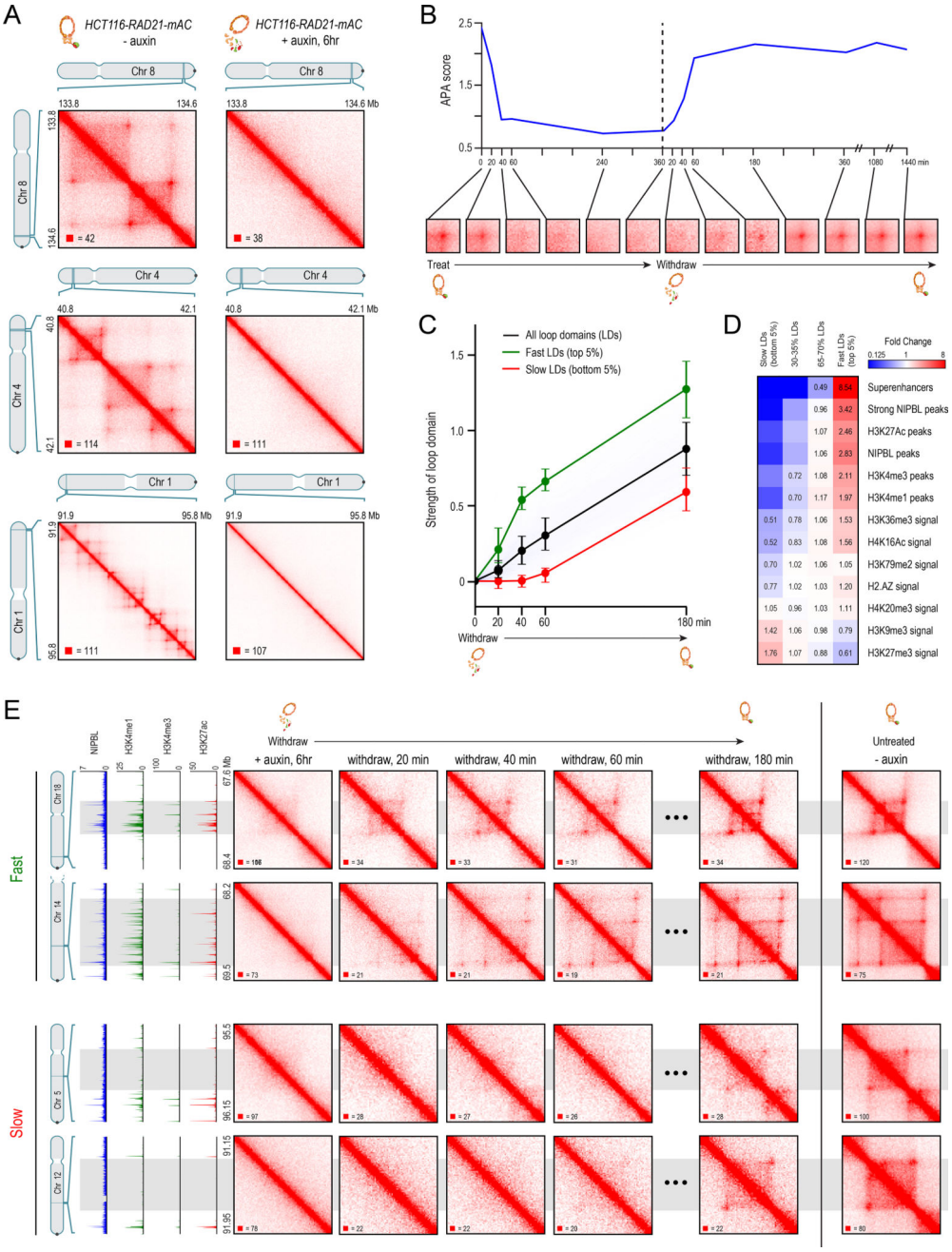


Figure 2. Cohesin degradation eliminates loop domains

(A) Contact matrices show that loop domains in untreated RAD21-mAC cells (top) disappear after auxin treatment (bottom). Three representative loci are shown (at 10kb resolution): chr8:133.8–134.6Mb (left), chr4:40.8–42.1Mb (middle) and chr1:91.9–95.8Mb (right).

(B) Aggregate peak analysis (APA) was used to measure the aggregate strength of the links associated with all loop domains in low-resolution Hi-C maps generated across a time course of auxin treatment and withdrawal. (Top) APA scores; values greater than 1 indicate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the presence of loops. (Bottom) APA plots; loop strength is indicated by the extent of focal enrichment at the center of the plot (See Fig S2B).

(C) Individual loop reformation curves for each of 1,988 loop domains (blue lines); the number of contacts in the untreated map corresponds to a value of 1, and the number of contacts in the auxin-treated map corresponds to 0. We highlight the media (black), the 5th percentile (red) and the 95th percentile (green) in terms of speed of recovery, see Methods. Error bars indicate 25th and 75th percentile within each subset.

(D) Enrichment of epigenetic features within a loop domain vs. speed of recovery. Enrichment is with respect to all intervals spanned by loop domains.

(E) Regions containing fast loop domains (1st row: chr18:67.6–68.4Mb; 2nd row: chr14:68.2–69.5Mb) and slow loop domains (3rd row: chr5:95.5–96.15Mb; 4th row: chr12:91.15–91.95Mb) are shown, along with ChIP-Seq tracks (from auxin-treated cells) for NIPBL, H3K4me1, H3K4me3, and H3K27Ac. For fast loop domains, reformation is apparent by 20–40 minutes after auxin withdrawal, whereas for slow loop domains, reformation is not seen until 3 hours after auxin withdrawal.

An interactive version of this figure is available at: http://bit.ly/2wl14TE

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

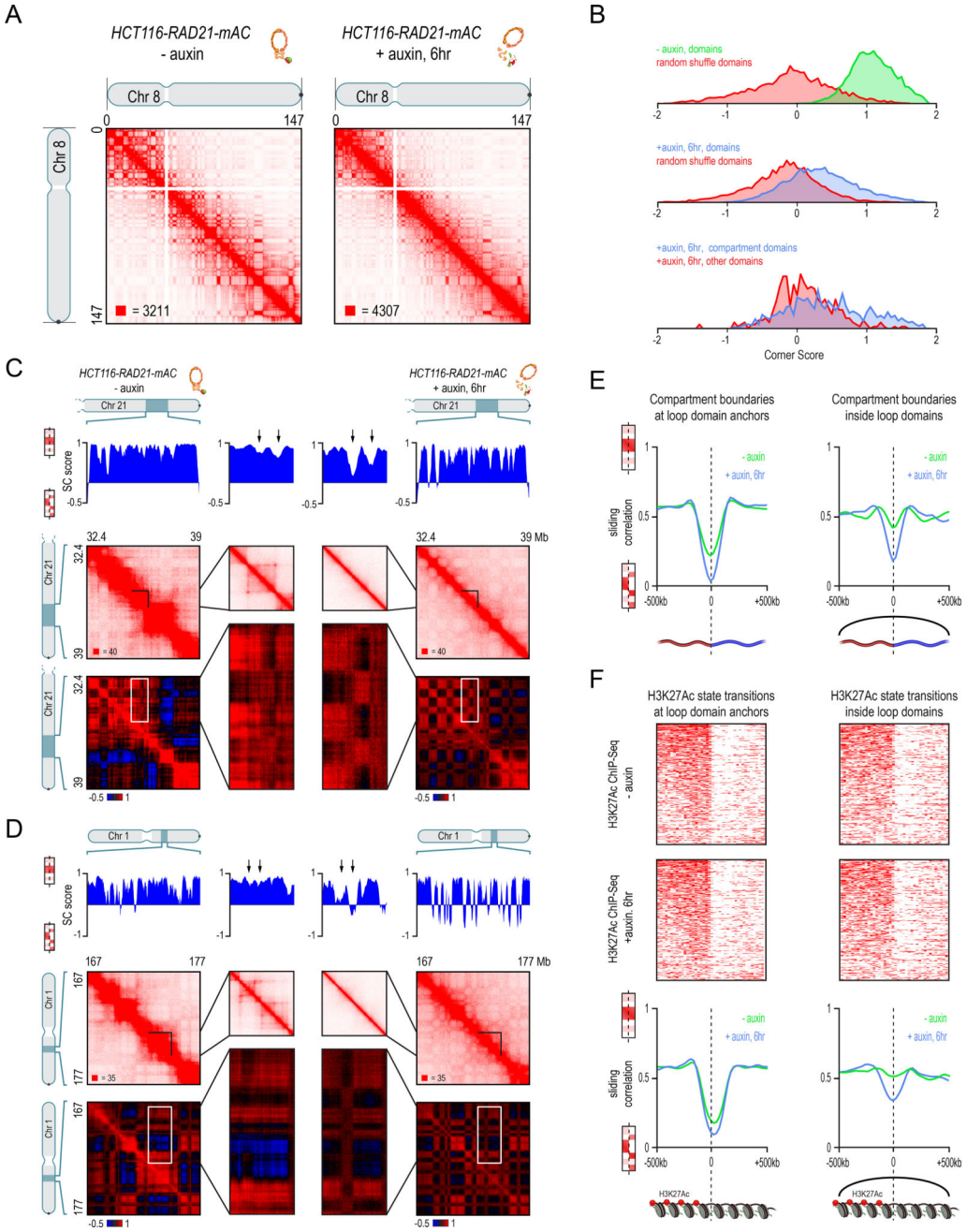


Figure 3. Genome compartmentalization is strengthened after cohesin degradation

(A) Contact matrices of chromosome 8 at 500kb resolution. The plaid pattern in the Hi-C map, indicating compartmentalization, is preserved after auxin treatment.

(B) Strength of contact domains called in untreated cells versus random intervals measured using the corner score (see Methods) in untreated (top) and treated cells (middle). Contact domain strength is reduced, but does not disappear. The remaining signal comes from compartment domains (bottom). The signal in treated maps from contact domains where both boundaries are contained completely inside a compartment interval ('other domains') is not enriched vs. random pixels.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(C,D) Examples (C: chr21:32.4–39Mb and D: chr1:167–177Mb) showing that the loss of cohesin-associated loops after auxin treatment results in increased fine-scale compartmentalization. Top: Sliding correlation scores; valleys imply strong differences in long-range contact pattern observed at a locus as compared to neighboring loci, indicating a change in compartment (see Methods). Middle: Observed contact matrices. Bottom: Pearson's correlation maps for the local region shown (see Methods). Deeper valleys in the sliding correlation score and increased plaid patterning in the observed and Pearson's correlation maps indicate stronger fine-scale compartment interactions after auxin treatment. Blowouts: loss of a loop domain results in strengthening of a compartment boundary spanned by the loop. Blown-out regions are indicated on zoomed out maps for both the observed (black upper triangle) and Pearson's correlation maps (white rectangle). Observed and Pearson's correlation maps are both shown at 25kb resolution for the zoomed out matrices and 10kb and 25kb resolution respectively for the blown-out matrices.

(E) Sliding correlation scores before and after auxin treatment for compartment boundaries which either coincide with loop domain anchors (left) or are located in the interior of a loop domain (right).

(F) Sliding correlation scores before and after auxin treatment for H3K27ac boundaries in untreated cells which either coincide with loop domain anchors (left) or are located in the interior of a loop domain (right). H3K27Ac modification patterns are unchanged after auxin treatment (top and middle).

Interactive figure: http://bit.ly/2vhBT7u

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

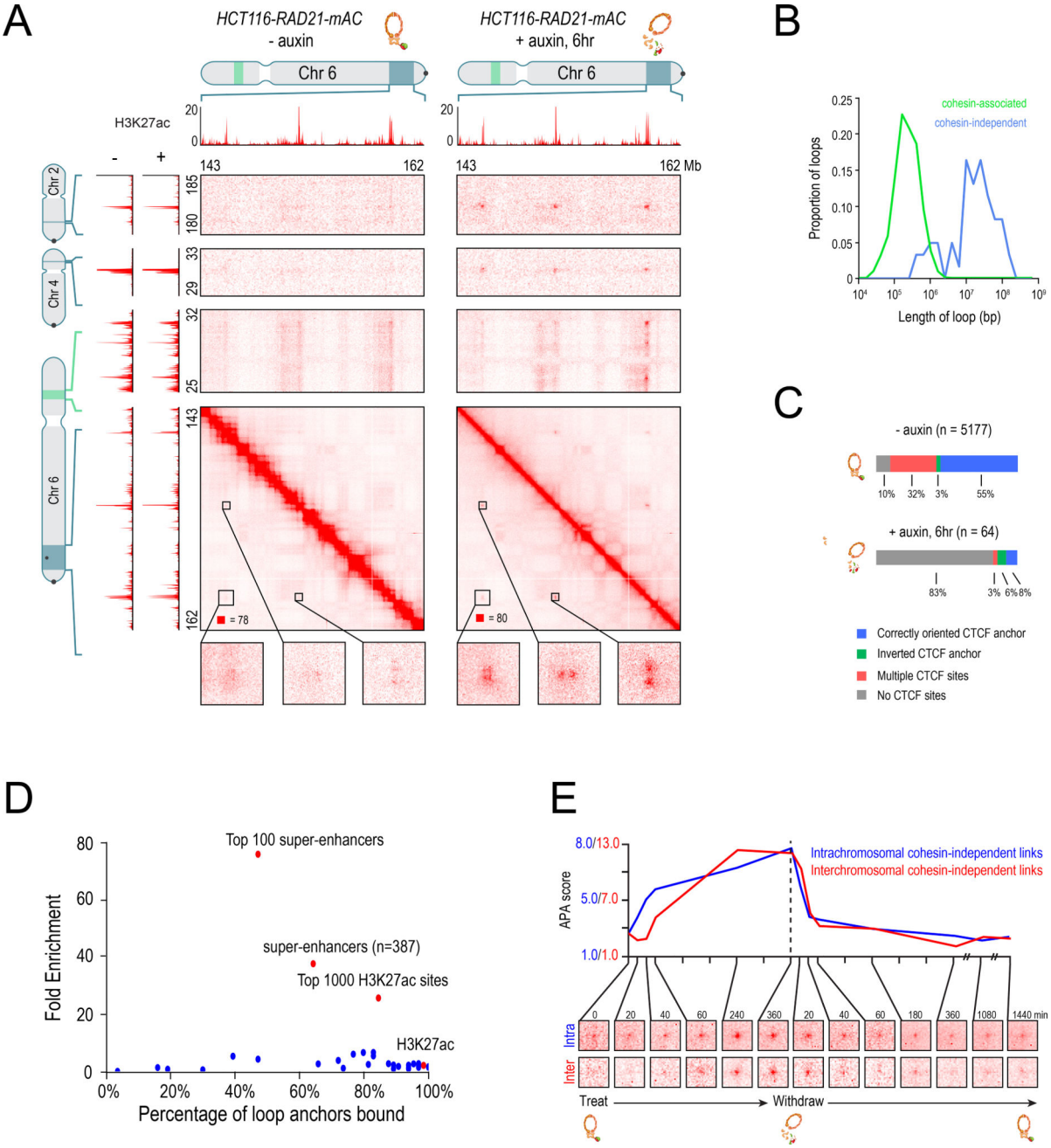


Figure 4. Cohesin loss causes superenhancers to co-localize, forming hundreds of links within and across chromosomes

(A) A network of intra- and interchromosomal cohesin-independent links between superenhancers on chr6, chr4, and chr2. H3K27 acetylation does not change with auxin treatment, but cohesin-independent links are significantly strengthened upon treatment. Intrachromosomal matrices are shown at 25kb (on-diagonal) and 50kb (off-diagonal) resolutions; interchromosomal matrices are shown at 100kb resolution. Maximum color intensities are 28 reads for the offdiagonal intrachromosomal matrices and 20 reads for the interchromosomal matrices.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(B) Length distribution of cohesin-associated loops (green) versus cohesin-independent loops (blue).

(C) CTCF binding patterns at cohesin-associated (top) versus cohesin-independent loop anchors (bottom).

(D) Percent of cohesin-independent loop anchors bound versus fold enrichment for 36 DNA-binding proteins and histone modifications.

(E) APA for intrachromosomal (blue) and interchromosomal (red) cohesin-independent links across a time course of auxin treatment and withdrawal (top: APA scores; bottom: APA plots).

Interactive figure: http://bit.ly/2vhEFts

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

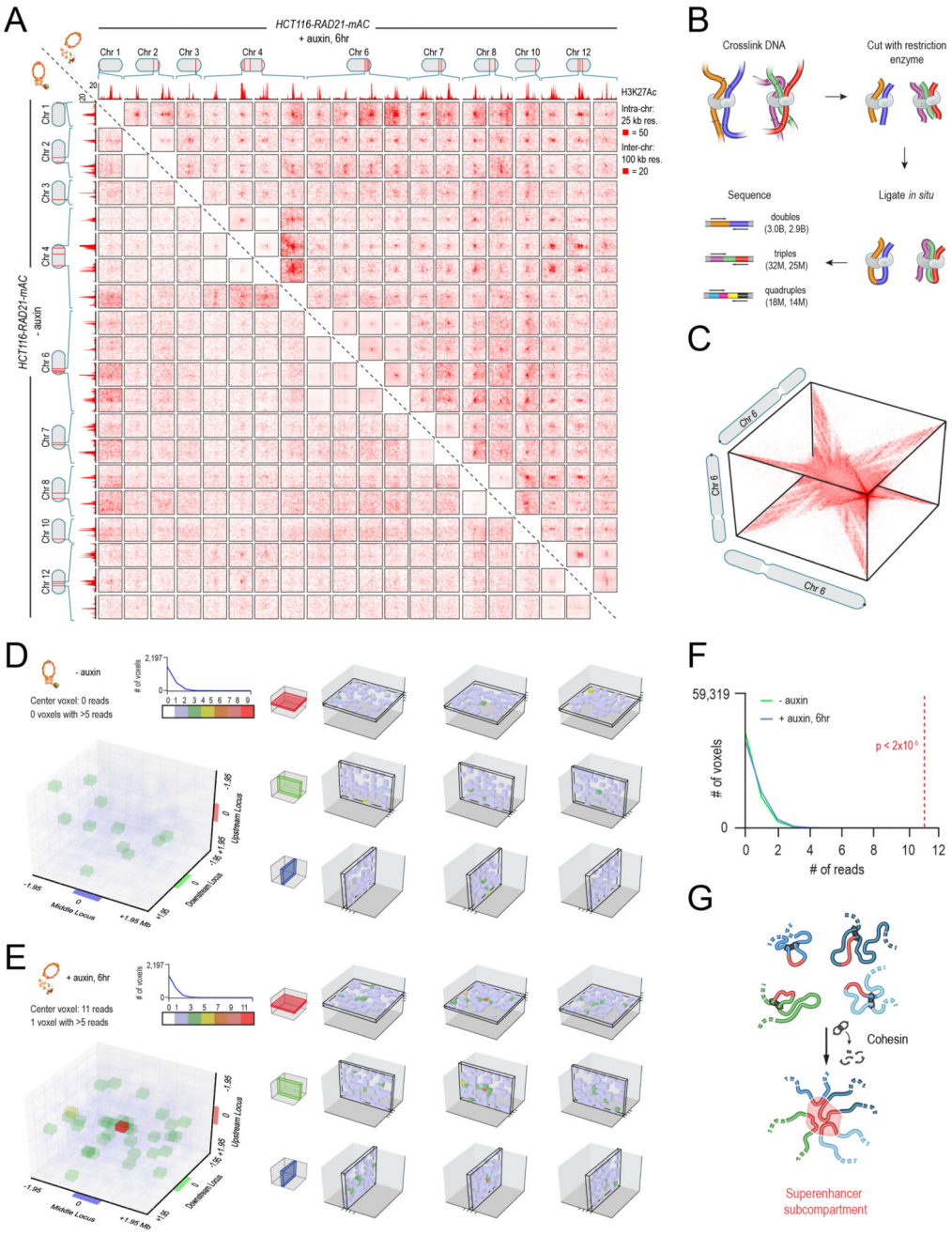


Figure 5. In the absence of cohesin, a clique spanning more than 20 superenhancers forms pairwise links and higher-order hubs

(A) The interactions between 20 cohesin-independent loop anchors spread across 9 chromosomes are shown before (lower triangle) and after (upper triangle) auxin treatment. Each matrix shows a 2 Mb by 2 Mb matrix centered on the respective anchors. Intrachromosomal interactions are shown at 25kb resolution; interchromosomal interactions are shown at 100kb resolution. The anchors are strongly enriched for H3K27 acetylation both before and after auxin treatment. (ChIP-Seq data is shown at 25kb resolution.) Cohesin loss causes the anchors to form a clique, with focal interactions seen between nearly all pairs of loop anchors, regardless of whether they lie on the same chromosome.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(B) In addition to pairwise contacts, *in situ* Hi-C generates concatemers spanning three or more fragments. There are millions of triples (chimeric reads which align to three loci) and quadruples (chimeric reads which align to four loci) in both our untreated and auxin-treated *in situ* Hi-C data sets for RAD21-mAC cells. The numbers in parentheses indicate the number of n-mer contacts observed in the untreated (left) and auxin-treated (right) data.

(C) 3D tensor showing collisions between three loci on chromosome 6 at 1Mb resolution (see Methods).

(D) (Left) 3D aggregate peak analysis (APA) using the untreated *in situ* Hi-C data for all 131 intrachromosomal trios of cohesin-independent loop anchors, chosen so that each anchor in a trio lies on the same chromosome as the other two anchors, but no two anchors in a trio lie within 10Mb of one another. To create a 3D APA cube, we excise a 3.9 × 3.9 × 3.9Mb subtensor centered on each trio, and superimpose the results. The cube is shown at 300kb resolution (i.e., each voxel corresponds to all collisions between three loci, each 300kb in length). The subtensors are oriented such that the locus closest to the p-terminus of a chromosome is always located on the z-axis, the one closest to the q-terminus is located on the y-axis, and the locus in between is located on the x-axis. The number of collisions in a voxel is indicated by its color; the histogram above the color scale shows the number of voxels of each color. No voxel contains more than 5 collisions, and the center voxel – reflecting all collisions between three cohesin-independent loop anchors – contains no collisions at all. (Right) Top Row: The central cross-section in z is shown, flanked by the two adjacent cross-sections. Middle Row: The central cross-section in y, flanked by the adjacent cross sections. Bottom Row: The central cross section in x, flanked by the adjacent cross sections. There is no enrichment at the center of the 3D APA cube.

(E) The preceding analysis is repeated using the auxin-treated data. Now, the center voxel contains 11 collisions, whereas no other voxel contains more than 5 collisions. These findings indicate that, in the absence of cohesin, cohesin-independent loop anchors tend to co-localize to form hubs containing three or more anchors.

(F) Histogram of number of voxels vs. number of collisions for the two 3D-APA cubes shown in 5D and 5E, as well as for 52 control 3D-APA cubes obtained by shifting one or more of the loci in each of the above trios by 3.9Mb. With the exception of the central voxel in the auxin-treated 3D APA cube, which contains 11 collisions, no voxel contains more than 8 collisions. This indicates that the observation of 11 collisions purely by chance is exceedingly unlikely.

(G) Under normal circumstances, loop extrusion facilitates short-range contacts between superenhancers and neighboring loci. Upon cohesin loss, superenhancers begin to co-localize, even when located on different chromosomes, and thereby form a subcompartment. Interactive figure: http://bit.ly/2x9penF

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

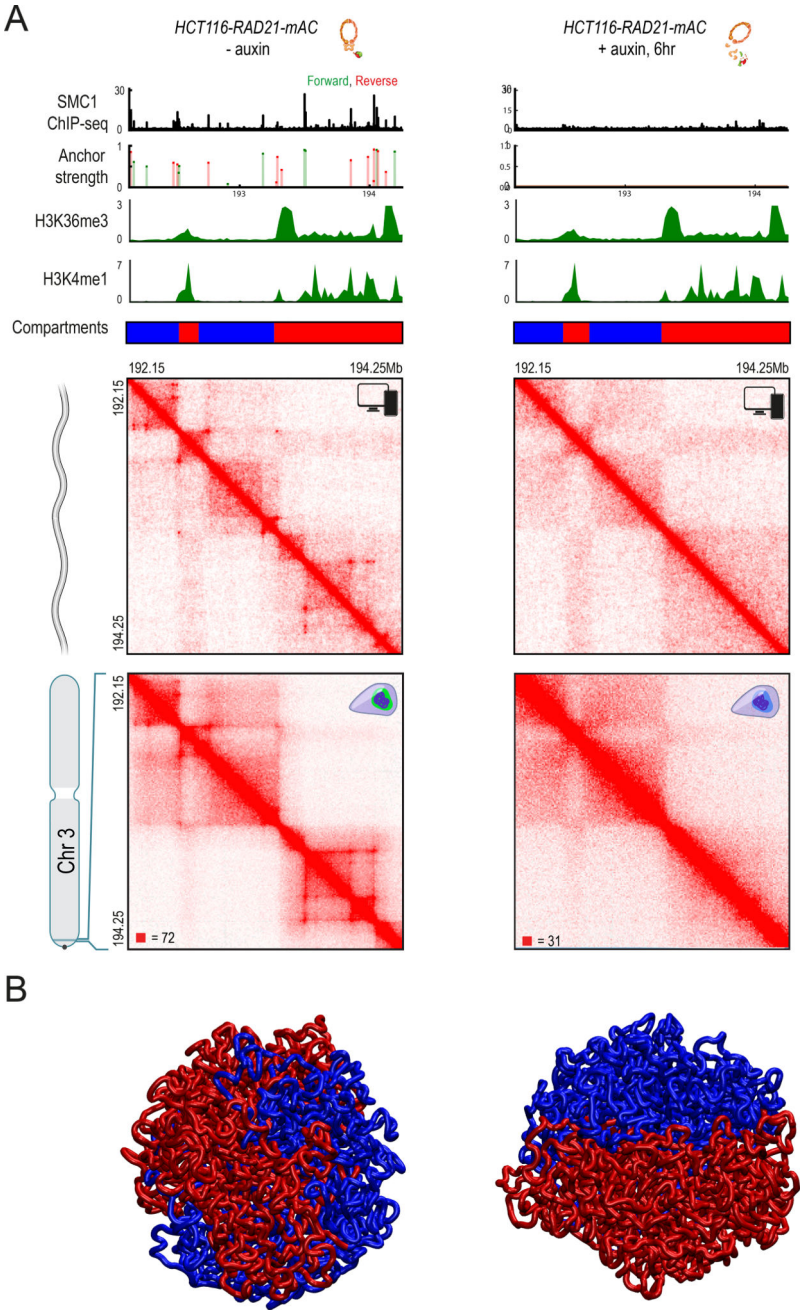


Figure 6. Molecular dynamics simulations combining extrusion and compartmentalization can recapitulate Hi-C experimental results

(A) We use loop extrusion and compartmentalization to simulate a 2.1 Mb region on chromosome 3 in RAD21-mAC cells before (left) and after (right) auxin treatment. CTCF and SMC1 ChIP-Seq signals are normalized and converted into binding probabilities for the simulated extrusion complex (first and second rows). Each peak is assigned a forward (green) or reverse (red) orientation based on the corresponding CTCF motif. ChIP-Seq data for 9 histone modifications were used to classify loci into two compartments (red and blue, fifth row). Histone modification data for H3K36me3 and H3K4me1 is shown, illustrating the correspondence between the classification tracks and the underlying ChIP-Seq signals (third

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and fourth rows). The simulations yield an ensemble of polymer configurations. We show contact maps from the simulated ensemble (top) and from the corresponding Hi-C experiments (bottom).

(B) Examples of globules from simulations of compartmentalization with extrusion (left) and without (right). The globule without extrusion shows stronger segregation of compartment types.

Interactive figure: http://bit.ly/2vsfSDC

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

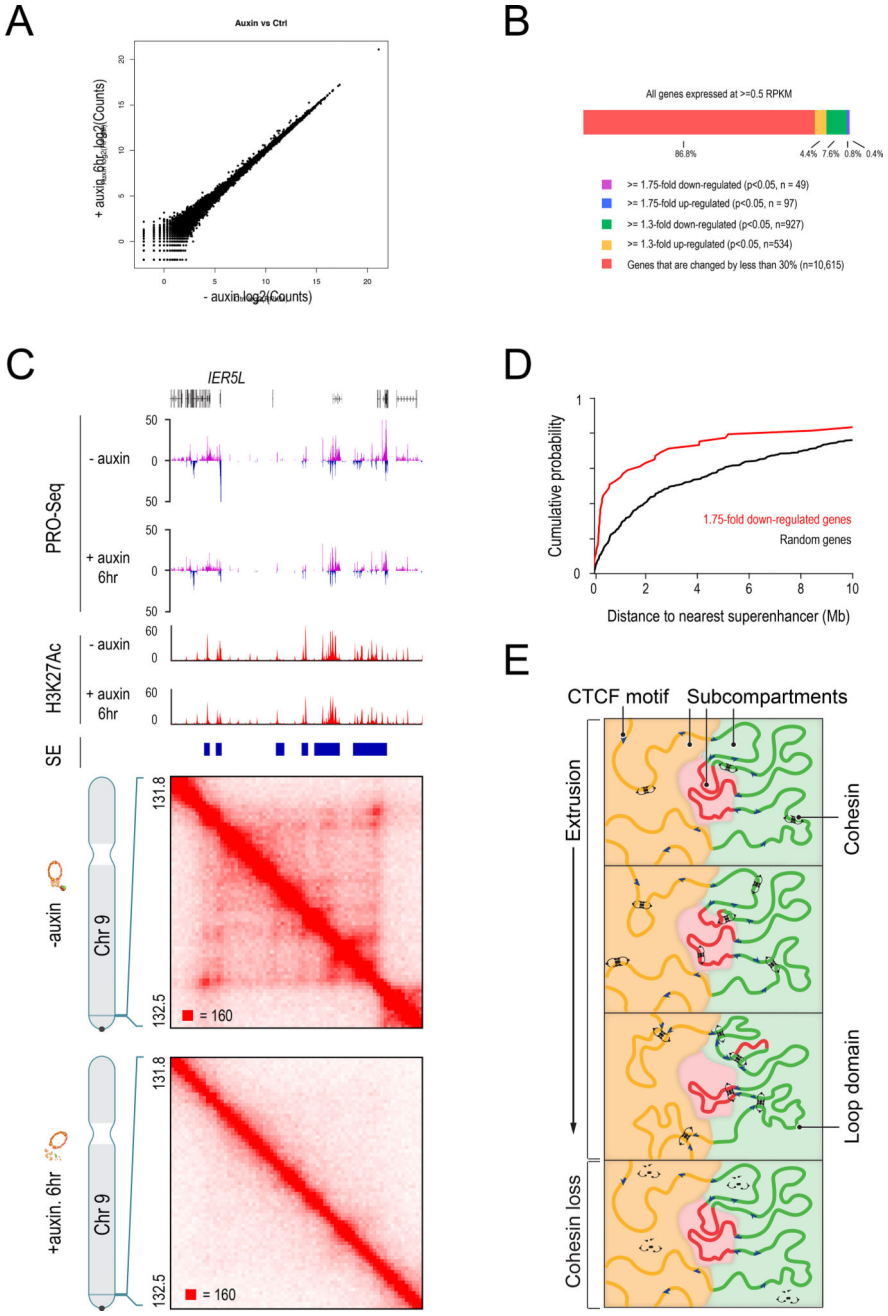


Figure 7. Cohesin degradation results in strong down-regulation of genes near superenhancers but does not result in widespread ectopic gene activation

(A) Scatter plot of gene-wide PRO-Seq counts in RAD21-mAC cells before (x-axis) and after (y-axis) treatment.

(B) Genes that are expressed in untreated cells rarely undergo substantial changes in expression level after cohesin loss.

(C) An example of a strongly down-regulated gene near a superenhancer. In untreated cells, a series of cohesin-associated loops form between the *IER5L* promoter and nearby superenhancers. Upon auxin treatment, these loops are lost and *IER5L* expression is 2.6-fold down-regulated.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(D) Cumulative probability distributions of distances to the nearest superenhancer for 1.75-fold down-regulated genes after auxin treatment (red) versus random genes (black).

(E) A model of how extrusion and compartmentalization combine to shape the spatial organization of the genome inside the nucleus. Intervals of chromatin with similar patterns of histone modification co-localize in nuclear subcompartments. Loop extrusion facilitates short-range contacts between nearby loci as the two subunits of the cohesin-based extrusion complex translocate in opposite directions on chromatin. The extrusion subunits halt at CTCF motifs facing inward, thus forming a loop domain between a pair of motifs in the convergent orientation. Loop domains represent dynamic structures that are maintained by cohesin; only a subset of them may be present at any given time. When the loop anchor motifs span multiple compartment intervals, the dynamics of loop extrusion interfere with compartmentalization by facilitating contacts between loci in different compartments. Loss of cohesin leads to the disappearance of loop domains and to a closer correspondence between genome compartmentalization patterns and histone modification patterns.

Interactive figure: http://bit.ly/2uiu514