Exploring the roles of RNAPII in genome organization

Dissertation

for the award of the degree

Doctor rerum naturalium (Dr. rer. nat)

of the Georg-August-Universität Göttingen

Within the Doctoral Program

International Max Planck Research School for Genome Science

of the Georg-August-University School of Science (GAUSS)

Submitted by

Shu Zhang

Goettingen, October 2022

Thesis Advisory Committee

Prof. Argyris Papantonis (Supervisor)

Department of Pathology, University Medical Center Göttingen

Prof. Tim Beißbarth

Department Medical Bioinformatics, University Medical Center Göttingen

Prof. Elizabeth Hessmann

Department of Gastroenterology, Gastrointestinal Oncology and Endocrinology, University Medical Center Göttingen

Members of the Examination Board

1st referee: **Prof. Argyris Papantonis (Supervisor)**

Institute of Pathology, University Medical Center Göttingen

2nd referee: **Prof. Tim Beißbarth**

Department Medical Bioinformatics, University Medical Center Göttingen

Further members of Examination Board

Dr. Marieke Oudelaar

Max Planck Institute for Biophysical Chemistry

Dr. Nico Posnien

Department of Developmental Biology, Johann-Friedrich-Blumenbach-Institute of Zoology and Anthropology, Georg-August-University Göttingen

Dr. Ufuk Günesdogan

Department of Developmental Biology, Göttingen Center for Molecular Biosciences

Date of oral examination: December 6th, 2022

Acknowledgements

In a flash, it is the end of my PhD study and I couldn't help but feel that the four years have passed so quickly. When the night I am about to finish the first draft of my thesis, I am listening to the rain outside and the four years start to come back to me. I had a good four years because I met nice people. Meeting nice people is a kind of luck, and a good luck makes a treasure of lifetime. Another what I have realized, the PhD study is an advanced course of learning how to learn, and how powerful scientific thinking and logic are. This journey has been a unique time in my life and I thank people I have met along the way, without your help and supports I wouldn't growth. Thank you all sincerely.

First of all, I would like to acknowledge my supervisor Prof. Argyris Papantonis for offering me the opportunity to join his team. Akis is a great mentor, and helped me a lot through my Ph.D. study. He not only supervised me how to do science, but also taught the wisdom of life. Thank you for your clear explanations of biology questions that I asked during the RNAPII projects. Thank you for your support of attention, sequencing, and rapid responses during the RNAPII projects. Thank you for solving problems of my Ph.D. study. The most important, you emphasize how importance of being happy in work and life. It reminds me of feeling my feelings, and to growth.

Fortunately, I have met the current and former members of the Papantonis lab, who are kind colleagues and good friends. I would like to specially thank Nadine, we have built a very productive group. Without your contributions, the projects can not be such smooth. Yajie, I am so happy to meet you and the inspiration your bravery has given me. Thanks Ting for helping me with my work and life. Spyridon, Natasa, Konstantinos, Tabea, Ioanna, Omar, Ilia, Adi, Judith, Mariano, Chris, and Juan, thank you for sharing the wonderful time together in the lab and out of the lab. Thanks for Sia supervised me how to do experiments, I am glad I know how to do Hi-C experiments with the kit. Of course, I would like to thank the rest

members of the lab, and the people who work in pathology have helped me. Thank you all sincerely.

I would like to thank my committee members, Prof. Tim Beißbarth and Prof. Elizabeth Hessmann. I am also grateful to Dr. Marieke Oudelaar, Dr. Nico Posnien, and Dr. Ufuk Günesdogan for agreeing to be on my examination committee.

I thank IMPRS-GS for creating such an interdisciplinary environment, holding wonderful retreat, and providing financial support. Also, I thank CSC provides me financial support.

I am very grateful for the friendships I meet in Göttingen. Xiaojuan, the times we stayed up night writing thesis together were very memorable. I am looking forward we kissing the Gänseliesel together. Mengyu, Le, and Xin, although you left Göttingen, it was a real pleasure time to be together with you. Of course, I would like to thank all my friends who have helped me in my life and spent wonderful time together. Thank you all sincerely.

Finally, words for my family. 我要感谢我的家人们。感谢我的妈妈爸爸，感谢你们的辛苦养育。感谢妈妈对我博士学习期间的支持和鼓励。我还要感谢所有关心爱护我的亲人们，是你们的关心让我面对困难时也能充满勇气。

Table of Contents

1. Summary

The role of RNAPII in 3D genome organization has been heavily debated over the last two decades. Here, *in situ* Hi-C and Micro-C data were used to conclusively address this debate. I discovered that: 1) using cell population *in situ* Hi-C data, longer loops with enhanced signal appear upon RNAPII depletion. Together with a rewired-loop pattern of gained loops in the absence of RNAPII, I can suggest that RNAPII may impede cohesin extrusion in human cells; 2) using *in situ* Hi-C data in G1-sorted cells, I observed loops with longer length, stronger signal, and stronger insulation, which are in line with the cell population *in situ* Hi-C data. Moreover, a diffuse signal pattern emerges upon RNAPII depletion in this data. Taken together, my data indicate RNAPII impeding cohesin extrusion in interphase. 3) using Micro-C data, transcription-level loops are captured, which show much weaker signal and insulation than CTCF loops. Moreover, on average, ~20% reduction in chromatin-bound CTCF and cohesin is seen in the absence of RNAPII. While this degree of loss does not affect loop detection in Micro-C maps, “transcription only” loops (especially enhancer-anchored loops) are severely weakened or lost in the absence of RNAPII. In addition, enhancer-bound cohesin is reduced by ~50%. Surprisingly, promoter-anchored loops are more stable than enhancer-anchored ones in the absence of RNAPII. Collectively, RNAPII is required for the formation of enhancer-anchored loops; and 4) using *in situ* Hi-C data from cells entering G1 after mitosis, ~20% reduction of chromatin-bound CTCF and cohesin is seen in the absence of RNAPII and loop formation is affected genome-wide during reentry into G1. Interestingly, as observed in Micro-C data, cohesin occupancy is reduced more at enhancers than promoters upon RNAPII degradation in this transition. Moreover, compartmentalization was also affected in the absence of RNAPII upon reentry G1. Thus, I can conclude that RNAPII is required to properly refold the genome during the mitosis to G1 transition.

2. Introduction

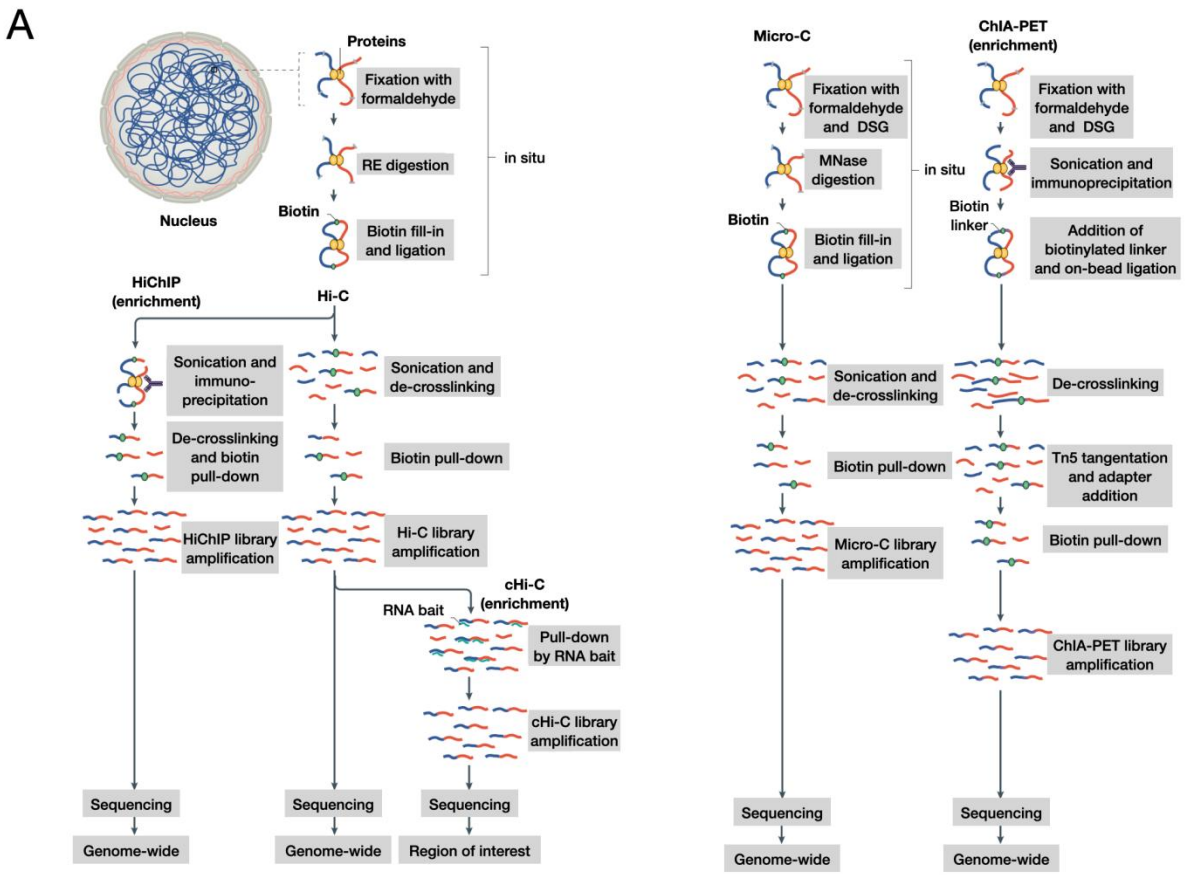
2.1 Layers and architectural factors of genome organization

Various concepts of nuclear architecture, such as lampbrush chromosomes and chromosome territories, were described much earlier than DNA was verified as the genetic material. Naturally, the nucleus, the most easily recognizable organelle inside a eukaryotic cell, and ideas of exploring its interior organization arose they were seen using microscopy. 100 years later, *in situ* hybridization (ISH), an imaging approach combining principles from biochemistry, successfully visualized "chromosome territories". According to these results, chromatin arrangement is non-random and each chromosome primarily occupies a distinct region of nuclear space during interphase. It was proposed that small chromosomes, gene-rich and active expressed regions prefer to locate at the nuclear interior, whereas large chromosomes, gene-poor and inactive regions prefer to locate at the nuclear periphery (Cremer and Cremer, 2010). An example of a chromosome territory linking structure to function is the Xa- and Xi-territory (Barr body). Using 3D FISH, it has been shown that the two are located in distinct parts in the nucleus (Teller et al., 2011; Giorgetti et al., 2016). Recently, RNA and DNA FISH of long and highly expressed genes also showed that transcribed genes are positioned towards the interior and inactive genes towards the nuclear periphery (Leidescher et al., 2022). However, two apparent limitations of the FISH technique are low-throughput and diffraction-limited resolution of light microscopy (the highest lateral resolution is ~250mn and the highest axial resolution is ~600nm with FISH; Jerkovic' and Cavalli, 2021). So, these limitations restrict high-throughput and fine-scale architecture be detected with FISH. Remarkable progress has been made recently with emerging of super-resolution microscopy, which drives studies of 3D genome organization to be more accurate (Liu et al., 2015; Sahl et al., 2017; Su et al., 2020; Takei et al., 2021). Such as single-molecule localization microscopy (SMLM) can reach a lateral resolution of 20

nm in biological samples (Jerkovic' and Cavalli, 2021).

Chromosome territories are also represented in cell-population or single-cell Hi-C maps (Lieberman-Aiden et al., 2009; Nagano et al., 2013), the whole-genome variant of the chromosome conformation capture (3C) technique (Dekker et al. 2002). Hi-C together with high-throughput DNA sequencing that can genome-wide capture proximal ligation. In C-based experiments, chromatin architecture is fixed by ~1% formaldehyde crosslinking, and then spatially adjacent fragments are ligated after being cut with a restriction enzyme. Contacts of ligated fragments are calculated, and it is assumed that space-closer fragments are ligated easier in 3D space. Hence, contact frequency matrices/heatmaps are built. Moreover, 3C is designed to capture interactions between two loci (Dekker et al. 2002), 4C (Circularized Chromosome Conformation Capture) is used to capture interactions of all loci with the anchor fragment (Zhao et al., 2006), 5C (Chromosome Conformation Capture Carbon Copy) is employed to capture interactions among the multiple anchor fragments (Dostie et al., 2006). In comparison, Hi-C is performed to capture interactions genome-wide (Lieberman-Aiden et al., 2009). Notably, the new method, Micro-C, which employs micrococcal nuclease (MNase) instead of restriction enzymes, which effectively captures transcription-level architecture (Hsieh et al., 2015; Hsieh et al., 2016; Hsieh et al., 2020; Krietenstein et al., 2020). Furthermore, protein-binding or region-specific enrichment methods (HiChIP, ChIA-PET, and capture-Hi-C), and ligation-free methods (SPRITE, ChIA-drop, and GAM) are designed for delivering aim-specific aspects (Fullwood et al., 2009; Hughes et al., 2014; Mumbach et al., 2016; Fang et al., 2016; Beagrie et al., 2017; Quinodoz et al., 2018; Zheng et al., 2019). C-based methods can be high-throughput, and then are complementary to microscopy techniques. The two kinds of techniques dominated studies of genome organization over the last two decades, and have offered tremendous opportunities and advancements in understanding genome organization. Parts of the sequencing-based methods can be found in **Figure 1**. In addition, computational analysis of

sequencing/microscopy data and mathematical modeling are also used broadly to interpret genome organization. Such as, genome-wide detection of loops, TADs, and compartment A/B are algorithm-needed, loop extrusion model was initially proposed theoretically based on molecular dynamics simulations (Fudenberg et al., 2016), and machine learning approaches can impute higher-resolution Hi-C maps from low-resolution to solve resolution problem (Zhang et al., 2018).



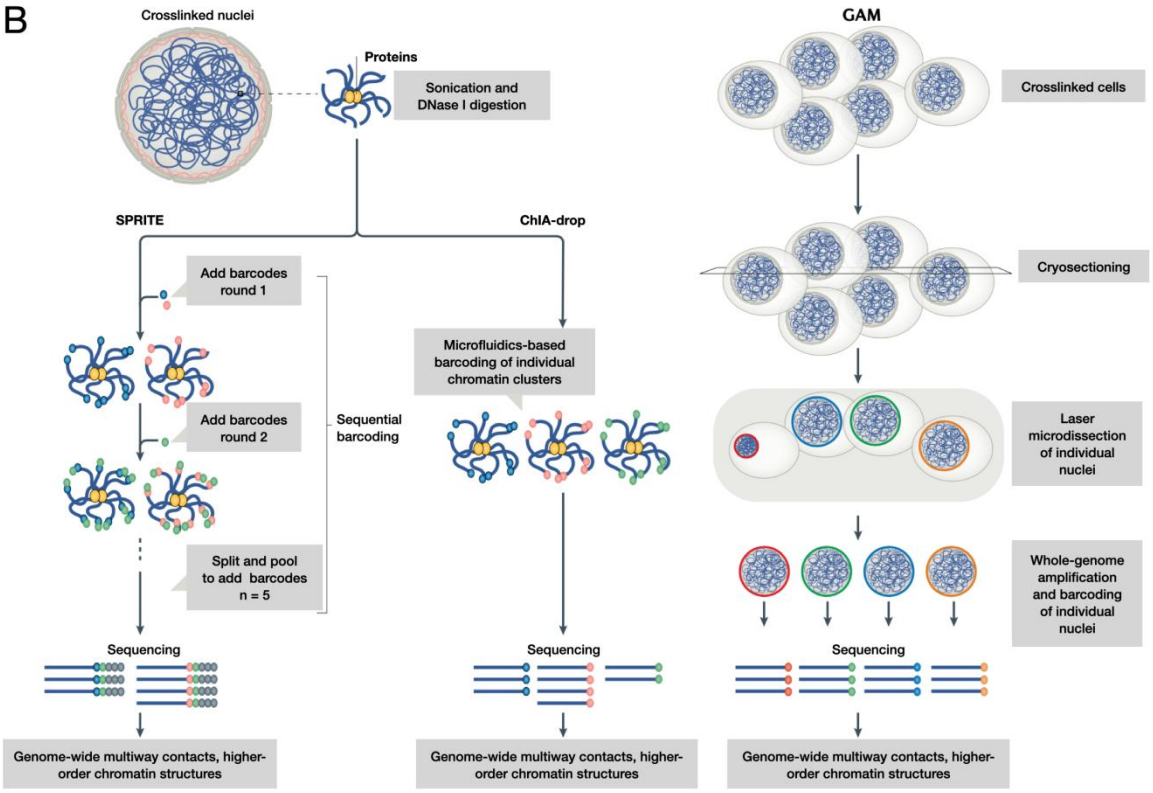


Figure 1. Sequencing-based methods for interrogation of 3D genome organization. (A) The plot shows examples of 3C-based methods of investigating genome organization. (B) The plot shows examples of ligation-free methods of investigating genome organization. Reprinted from (Jerkovic´ and Cavalli, 2021).

It is a grand challenge to fold ~2 m linear DNA into a nucleus with a diameter of 10μm. Furthermore, the organization of chromatin is principled and functional in space and time to fulfill periodic transcription, replication, repair, and mitosis programs correctly during cell cycles. Moreover, gene regulation need to be carried out properly during development, differentiation, and aging processes. Otherwise, improper genome organization and gene regulation can lead to dysfunction, disease, and cancer (Rada-Iglesias et al., 2018). The most commonly known DNA compaction event is the formation of nucleosomes that every ~146bp DNA wound around the histone octamer (Lawrence et al., 2016). Over the last two decades, the genome-wide hierarchical layers of genome organization were uncovered and defined with gradually higher resolutions of Hi-C/Micro-C contact maps due to the increasingly cheaper cost of sequencing (**Figure 2A**; **2B**). Substantial progress has been

made to explore genome architectural factors and has facilitated a better understanding of genome organization (**Figure 2C**).

Loops. Initially, the "looping" model was assumed to interpret enhancer and promoter interactions (E-P interactions) that a locus control region (LCR) is brought close to the *β-globin* locus, and unrelated regions are looped out. Then, the proximity of the E-P interaction was validated using 3C (Tolhuis et al., 2002). Furthermore, genome-wide detection of loops was done in 2014 using Hi-C sequencing data, and globally architectural loops (E-P loops are discussed in the next chapter) were mainly detected (Rao et al., 2014). The loop structure features as an apparent "dot" in a Hi-C map with ~5 kbp resolution (**Figure 2B**). Moreover, the loop structure highlights that two loci interact significantly more than the local background. Based on the feature, loop calling software is designed to detect the significantly enriched contacts from local background or uses imaging processed technology to blur background noise away (Durand et al., 2016; Roayaei-Ardakany et al., 2020; Rowley et al., 2020).

Loop formation during interphase was proposed to be associated with the loop-extrusion model. The model proposed that the extrusion factor can be loaded on chromatin and progressively reel its flanking regions into a loop, so linearly distant fragments can be close in 3D space. Loop-extrusion model was first pointed out to describe chromosome loop formation during mitosis through condensin (one member of the SMC family) and later was suggested to interpret interphase genome organization through cohesin (another member of the SMC family). Cohesin degradation papers (Rao et al., 2017; Wutz et al., 2017; Flyamer et al., 2017; Rhodes et al., 2020) showed that cohesin depletion led to almost all cohesin-CTCF loops and TADs loss, which indicated the role of cohesin as extrusion factor. Furthermore, similar observations were found in several parallel studies, such as knock out (KO) of cohesin loader NIPBL (Schwarzer et al., 2017) and MAU2 (Haarhuis et al., 2017), RNAi of cohesin acetyltransferase ESCO1 (Wutz et al., 2020), and ATP inhibition during recovery of

cohesin after its acute degradation (Vian et al., 2018). Recently, cohesin-mediated loop extrusion was observed directly through single-molecule imaging technique (Kim et al., 2019).

Moreover, compartmentalization was strengthened upon cohesin depletion from chromatin and weakened upon depletion of cohesin unloader WAPL, PDS5A, and PDS5B (Wutz et al., 2017). The observations indicate that compartmentalization and loop extrusion are two independent mechanisms of genome organization. It was reported that cohesin disrupted interactions between super-enhancers (SEs) in cancer cell lines (Rao et al., 2017) and polycomb-mediated interactions in mESCs (Rhodes et al., 2020). In addition, ATP inhibition experiments indicate that ATP is required for loop extrusion (Vian et al., 2018). Due to losing cohesin did not alter whole-genome level of gene expression (differential expression genes were fewer than 100 before and after degrading cohesin), which led to its role in gene expression regulation under discussion (Rao et al., 2017). However, recent studies showed contrary conclusions that hormone inducible transcription factor GR mediated gene expression was significantly downregulated upon cohesin degradation (Rinaldi et al., 2022). Another example is the same paper that TNF-α induced NF-κB mediated gene expression was significantly downregulated upon cohesin degradation (Rinaldi et al., 2022). These activity-induced conditions indicate that cohesin locally participates in E-P interaction, not globally.

From Hi-C analysis (Rao et al., 2014), CTCF is the most enriched boundary element to cohesin at loop anchors and TADs boundaries. Interestingly, CTCF at loop anchors are typically present in convergent orientations. It was reported that the N-terminal of CTCF binds on its 3' DNA binding site, which means CTCF face loops and TADs inwardly. Another study also reported that N- not C-terminal of CTCF interact with cohesin subunit SA2–SCC1 (Li et al., 2019), which is compatible with the loop extrusion model that CTCF blocks cohesin extrusion. Genome-wide acute degradation of CTCF led to the loss of loops, but only minor

gene expressions were affected (Nora et al., 2017). Unlike cohesin, losing CTCF did not increase compartmentalization, so the compartmentalization mechanism was thought to be CTCF-independent. In addition, it was proposed that CTCF not only insulates cohesin but also stabilizes it on chromatin with ESCO1. As well as, CTCF and ESCO1 inhibit WAPL unloads cohesin from chromatin to form longer loops (Wutz et al., 2020).

TADs. On a resolution of tens to hundreds of kilobases, a Hi-C map is partitioned into continuous triangles (**Figure 2B**). The triangle was defined as a "topologically associating domain" (TAD; Dixon et al., 2012; Nora et al., 2012), which has significantly enriched contacts within one TAD and a clear boundary between two adjacent TADs. CTCF was decisively related to TAD boundary due to its high occupancy (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014). Moreover, genome-wide acute degradation of CTCF reduced ~80% insulation of TADs boundaries (Nora et al., 2017). Loop extrusion was also indicated that underlies TAD formation. However, unlike loops that highlight interactions from two loci, TADs highlight that homogenous interactions could be detected within a region. TADs were suggested that emerged from multiple loops dynamic forming, and they were a cell population average phenomenon from polymer simulation (Fudenberg et al., 2016). Moreover, from single-cell Hi-C, TADs and TAD borders were thought to be highly dynamic structures and do not form fixed entities (Nagano et al., 2013). Furthermore, super-resolution microscopy observed TAD-like structures in single cells (Bintu et al., 2018). Interestingly, TAD-like structures exist even under cohesin depletion conditions. Whereas the position of the boundaries lies at CTCF is randomized. In addition, TADs were reported as functional units that constrain the proper enhancer and promoter within one TAD and insulate incorrect enhancers. Disrupting the boundaries of TADs can lead to gene misexpression (Lupiáñez et al., 2015; Rada-Iglesias et al., 2018).

Compartment A/B. Compartment A/B is another aspect of genome organization (Lieberman-Aiden et al., 2009). Genome is divided into A or B compartments, which exhibit a

checkerboard pattern in a Hi-C map with 1Mbp resolution, and they highly correlate with euchromatin and heterochromatin, respectively. Furthermore, discontinuous sub-regions of compartment A or B tend to homotypic gather together, and the mechanism of compartment A/B formation is still poorly known. Moreover, the driving force of compartmentalization should support *trans* interaction, which may be epigenetics condensates, owing to it being highly associated with features of euchromatin and heterochromatin. Hi-C analysis clusters compartment A/B into eight smaller subcompartments (A1.1, A1.2, A2.1, A2.2, B1.1, B1.2, B2.1, and B2.2) according to histone modifications, gene density, GC content, and algorithm-guided clusters of interaction (Liu et al., 2021). Briefly, subcompartments A1.1, A1.2, and A2.1 are associated with active transcription. Moreover, repressive chromatin mark H3K27me3 was enriched in the subcompartments B.1.1 and B.1.2. In contrast, the heterochromatin mark H3K9me3 was related to being in B.2.1 and B.2.2. In addition, polycomb-mediated space proximity was found both in *cis* and *trans* (Rada-Iglesias et al., 2018), for example, interactions between deactivated genes were observed upon heat shock in *Drosophila* (Li et al., 2015). Constitutive heterochromatin is typically located at the nuclear periphery and attached to the nuclear membrane to form lamina-associated domains (LAD). Heterochromatin protein 1 (HP1) proteins were suggested to regulate constitutive heterochromatin compaction and gene silencing through forming phase-separated droplets (Larson et al., 2017).

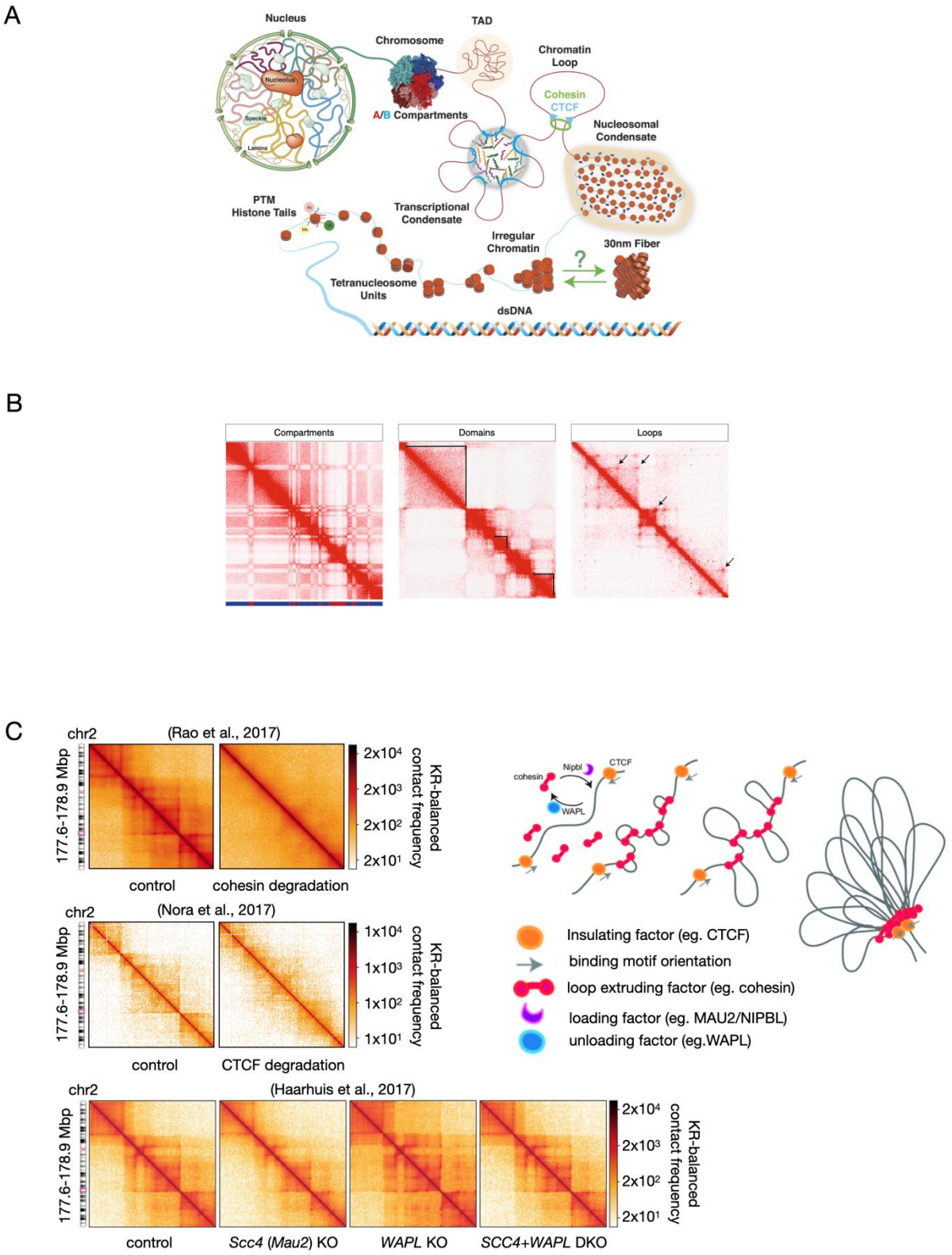


Figure 2. Different layers of genome organization and mechanisms of architectural factors. (A) The plot shows different layers of genome organization. (B) The plot shows different layers of genome

organization on Hi-C contact maps. (C) The plot shows effects of degradation architectural factors. Reprinted from (Lin et al., 2021; Zhang et al., 2021; Zhang et al., 2021)

2.2 How transcription shapes genome organization?

Study of transcription regulation has been one of the most intensive subject areas in biology research over the past decades. Transcription initiation lies recognition of promoter, which is *cis*-regulatory DNAs that is bound by *trans*-factors and generally defined as +-50-2000bp region around TSS in genomic data analysis. Using Cryo-EM, molecular structures of different RNA polymerase and their pre-initiation complex (PIC) have been elucidated unprecedentedly in human and yeast (Cramer, 2019). Moreover, another *cis*-regulatory DNAs, enhancer, does not decide where transcription happens but can positively regulate gene expression by interacting with its targeted promoter. Although promoters and enhancers were tacitly assumed to be distinct, their shared properties and functions have been considered due to the recent development of high-throughput DNA sequencing (Andersson and Sandelin, 2020). On the one hand, studies showed that enhancers contain promoter activity due to transcribing a few amounts of bidirectional enhancer RNAs (eRNA). As well as, ChIP-seq data of compositions of promoter PIC also exist at enhancer regions (De Santa et al., 2010; Kim et al., 2010). On the other hand, promoters also show enhancer activity that can increase gene expression level by forming promoter and promoter (P-P) interactions in both human and mouse, *in vivo* and *in vitro* (Diao et al., 2017; Dao et al., 2017; Engreitz et al., 2016; Mifsud et al., 2015). From mouse massively parallel reporter assays (MPRAs), it was reported that the regulatory DNAs intrinsically contain both enhancer and promoter activities, only the extent of transcription initiation and positive regulation are distinct (Nguyen et al., 2016). Interestingly, the mouse MPRAs found that promoters usually contain both strong initiation transcription and additive effects, whereas most enhancers contain limited promoter activity (Nguyen et al., 2016). Interestingly, many promoters contain CpG islands while only small part of enhancers contain it. Activating enhancers starts from

pioneer transcription factors bind on “closed” chromatin then recruit chromatin remodeling complex to open the chromatin (Ray-Jones and Spivakov, 2021). Subsequently recruiting sequence-specific TFs and coactivators such as P300-CBP, BRDs, and Mediator-complex to facilitate PIC assembling at enhancers. It has been 40 years since enhancer was discovered, and enhancers were thought to be independent of direction and distance of genes (Moreau et al., 1981; Banerji et al., 1981). However, recently, a paper proposed that enhancer action depends on genomic distance (Zuin et al., 2022). The functions of eRNAs are as yet unclear but were suggested that maintaining enhancer chromatin accessibility and interacting with factors (eg, BRD4; Ray-Jones and Spivakov, 2021). Jointly, more studies of enhancers and molecular determinants of regulatory potential are urgently needed.

Enhancers can be far from promoters on linear DNA sequence but be close with promoters in 3D space through favorable genome folding events, such as ZRS (enhancer of *Shh* that is within the intron of the *Lmbr1* gene) locates approximately 1 Mbp distant from *Shh* (Lettice et al., 2003). This characteristic brings a new perspective into transcription regulation, how long-distance genome organization influences gene expression? Although a rough model pointed out that RNAPII, TFs, mediators, and cohesin bind on enhancers and bring them to promoters, many detailed mechanisms are missing. Furthermore, recent observations offered contradictory questions to this challenging topic, such as whole-genome degradation cohesin mildly affected gene expression (Rao et al., 2017) and E-P interactions are still formed on Hi-C map upon mediator degradation. Understanding how E-P interactions are formed and regulate gene expression can drive tremendous progress in biology from theory to application. Here are some examples of urgently-answer questions, 1) What are the architectural molecular bases for E-P interactions? 2) Why enhancer can positively regulate gene expression? What are enhancers' functions to promoters during pre-initiation, initiation, and reinitiation? 3) Since enhancers contain limited promoter activity but promoters can contain relatively high enhancer activity (Nguyen et al., 2016), how do

enhancers and promoters interplay with architectural factors, respectively? Current opinions suggest several non-mutual-exclusive models that own their characteristics and properties.

Homotypic-interaction model. The first model indicates that E-P interactions are formed by physically interacting between enhancer and promoter through homotypic attraction factors, such as transcription factors SP1 and YY1 (Su et al., 1991; Weintraub et al., 2017). This model is the easiest and highlights that self-associated factors bring direct interactions. As well as self-interaction coactivators (e.g., LDB1) that bind on TFs can form this kind of loop. In myogenic progenitor cells, transcription factor PAX3 utilizes LDB1 to form E-P loops in many loci (Magli et al., 2019). Interestingly, coactivator LDB1, not its binding transcription factor GATA1, is required to form E-P loops at the *HBB* location in erythroid cells (Deng et al., 2012).

Mediator-complex model. The second model highlights how Mediator-complex regulates transcriptions and facilitates E-P interactions. In human, the total size of PIC is 4MDa, of which 1.4MDa is Mediator-complex which consists of 26 subunits. The structure of Mediator-complex is a quintessential example of intrinsically disordered regions (IDRs) and conformational allostery. Mediator-complex consists of three parts that “head” and “middle” can interact with RNAPII within PIC, whereas “tail” is more dynamic and interacts with TFs. The Mediator-complex is considered as a “signal integrator” due to binding diversely functional TFs then bridging them close to PIC, and shifting conformations facilitate and stabilize integrating processes. Mediator-complex was proposed that together with TFIID position PIC on promoter DNA, position RNAPII within PIC, position TFIIH within PIC, position XPB’s ATP-dependent translocase for promoter opening, and position CDK7 of TFIIF within PIC which can phosphorylate RNAPII (Richter et al., 2022). CDK-Mediator regulates transcription elongation through phosphorylating elongation factors using CDK8 or CKD19 in human. Not only Mediator-complex conformation dynamically shift, but also PIC adjusts conformation dynamically during these processes. TF activates transcription by

building favorable conditions, which recruits chromatin modifying factors such as CBP-p300 that can acetylate flanking histones to be H3K27ac, chromatin remodeling complexes such as SWI/SNF that can render chromatin accessibility, as well as recruits Mediator-complex. On the one hand, sequence-specific binding TFs recruit Mediator-complex and decide their locations on genome (for example, enhancers) due to subunits of Mediator-complex lack any discernible DNA-binding domains, and there is no specific-DNA binding motif of it. Moreover, it was reported that TFs also facilitate functional processes of Mediator-complex and stabilize Mediator-RNAPII interaction within PIC. On the other hand, Mediator-complex brings TFs to PIC and stabilizes this process. In addition, it was reported that CDK-Mediator can positively or negatively regulate TFs activities (Richter et al., 2022).

Interestingly and importantly, it was reported that Mediator-complex can compete TFs binding with DNA. Furthermore, unbound TFs not only compete with chromatin-bound ones but also compete with Mediator-bound ones. Due to binding competitions of high local concentration unbound TFs to DNA and Mediator-complex, Mediator-TF interactions may facilitate rebinding of TFs on chromatin, increase their residence time on DNA, and maintain enhancer occupancy. In addition, it was suggested that Mediator-complex bridges E-P together and positively regulates gene expression with a “transcription bursting” perspective. Transcription bursting requires transcription re-initiation after pioneer-initiation, which requires new RNAPII, PIC, Mediator-complex, and TFs to bind together on promoters again. Mediator-complex mediates E-P interactions may allow new-round assembly and recruitment through delivering Mediator-complex from enhancer to promoter, therefore, facilitates transcription re-initiation and increases gene expression.

Cohesin model. The third model highlights how cohesin participants in E-P interaction. Although cohesin is the master regulator in higher order genome organization, the role of cohesin plays in E-P interaction was under debate due to its whole-genome acute degradation did not alter global gene expression level (Rao et al., 2017). However, many

studies proposed that cohesin co-localized with inducible transcription factors at enhancers in tissue-specific samples, such as hormone inducible transcription factors glucocorticoid receptor (GR) and estrogen receptor (ER) (Schmidt et al., 2010; Rinaldi et al., 2022; D'Ippolito et al., 2018), microbial signals induced primary macrophages (Cuartero et al., 2018), calcium influx induced activated neutrophils (Zhu et al., 2021), as well as cohesin was reported that co-localized with mediator at enhancers in mESC and yeast (Kagey et al., 2010). Significant down-regulation genes were found under inducible-TFs mediate expression conditions upon cohesin depletion (Rinaldi et al., 2022). Interestingly, specific-TFs are more sensitively dropped at enhancers than promoters upon cohesin depletion (Zhu et al., 2021). Therefore, cohesin probably mediates E-P interaction follow cell/tissue-specific way and is required to proper expresion of inducible genes, and especially important at enhancers due to interact with TFs and mediator-complex.

According to the way cohesin mediates E-P tight connection with mediator, TFs, and enhancers, here, I describe the third model by adding cohesin to the second model. NIPBL was reported that directly interact with mediator-complex (Kagey et al., 2010), stabilizes TFs, therefore cohesin forms “stripes” from enhancer anchors to promoters’ (Vian et al., 2018).

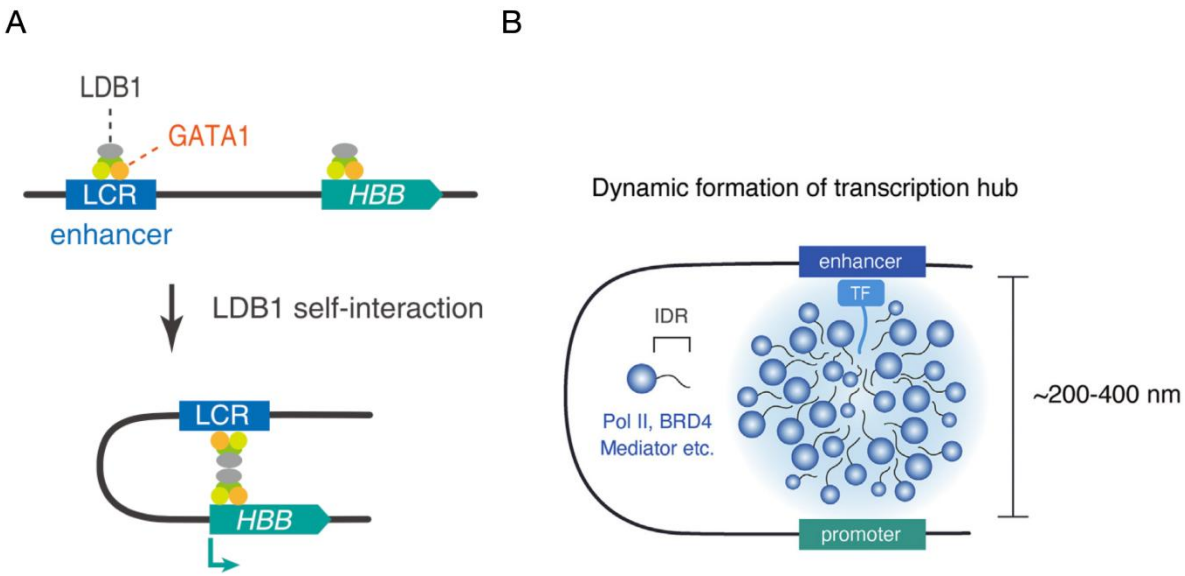
Condensates model. The above three models are discussing about architectural molecular mechanisms of E-P interactions, however, the forth model is discussing about organization of the many biomolecules that are involved in E-P interactions, such as molecular condensates/hubs/clusters formation that can facilitate interactions and reactions among biomolecules (Banani et al., 2017). Many condensates are formed through liquid–liquid phase separation (LLPS) that is driven by multivalent and cooperative interactions of intrinsically disordered low-complexity domains of biomolecules. The examples of natural condensates are, nuclear bodies, P-bodies, stress granules, and the nucleolus, which include hundreds to thousands proteins. It was reported that RNAPII transcription take places within transcription “factories” with fixed cells (Papantonis and Cook. 2013), and

RNAPII clusters together in live human cells (Cisse et al., 2013). Thus, understanding organization of proteins that are involved in RNAPII transcription is another angle to explore mechanisms E-P interactions. Such as transient or stable formation, compositions, and sizes (~100nm-400nm) are all features of condensates that can be regulated upon functions and affect functions. Furthermore, using imaging technique, substantial RNAPII transcription-associated condensates/hubs/clusters were observed (Cramer, 2019). It was reported that many *trans*-activation domains of TFs are low-complexity sequences that exhibit an intrinsically disordered conformation, and form transient TF-TF clusters that can be disrupted by hexanediols (Chong et al., 2018). Moreover, TFs can form phase separation that recruit the intrinsically disordered CTD of RNAPII (Kieffer-Kwon et al., 2013). In addition, TFs form phase separated condensates with Mediator-complex, for example, OCT4 and MED1 co-occupancy SEs of the key pluripotency genes *Esrrb*, *Nanog*, *Trim28*, and *Mir290* in ESCs (Boija et al., 2018). Mixing of 7 different TFs can incorporate into MED1 phase-separated droplets *in vitro* (Boija et al., 2018).

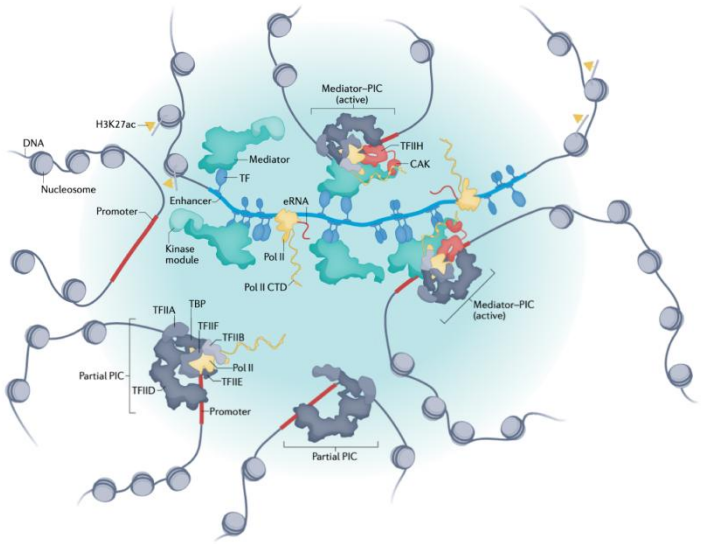
Moreover, BRD4 and MED1 were reported that form condensates at sites of SE-driven transcription (Sabari et al., 2018). Droplets formed by MED1-IDR are able to concentrate transcriptional machinery in a transcriptionally competent nuclear extract (Sabari et al., 2018). In addition, It was reported that CTD alone can undergo LLPS *in vitro* due to its intrinsically disordered in solution consists with low complexity of CTD sequences (Boehning et al., 2018). Interestingly, it was reported that CTD can undergo phase separation via at least two different mechanisms. On the one hand, interactions of unphosphorylated CTD-CTD forming hubs that may be recruited by coactivators that bind to enhancers (Gerber et al., 1995; Scafe et al., 1990; Kieffer-Kwon et al., 2013). On the other hand, phosphorylation was reported to disrupt CTD-CTD hub (Boehning et al., 2018), while $CTD^{phosphorylated}$ joins into transcriptional elongation factor, cyclin T1 that is positively charged and containing IDR, mediated condensates (Lu et al., 2018).

P-E interactions are reported not only *cis* but also *trans.* Forces driving of condensates support both *cis* and *trans*, which complement the cohesin mediated E-P interactions, and offer more mechanisms to explain E-P interactions. Many of E-P interaction loci are seen that separated by several hundreds of nm during transcriptional activation (Heist et al., 2019; Chen et al., 2018; Benabdallah et al., 2019; Alexander et al., 2019), which are easier explain by condensates model than directly interact models.

Gene expression regulation is an essential step to cell identity. It was reported that enhancers/SEs regulate gene expression are cell-type specific. The above, I described four models of enhancer-promoter interactions, which are not mutual exclusive. Indeed, cell-type specific genes need to be dynamically regulated, the molecular mechanisms could be individual and are more complex than the models. Thus, exploring architectural molecular basis of E-P interactions, and investigating transcription associated genome folding principles are urgent needed.



C



D

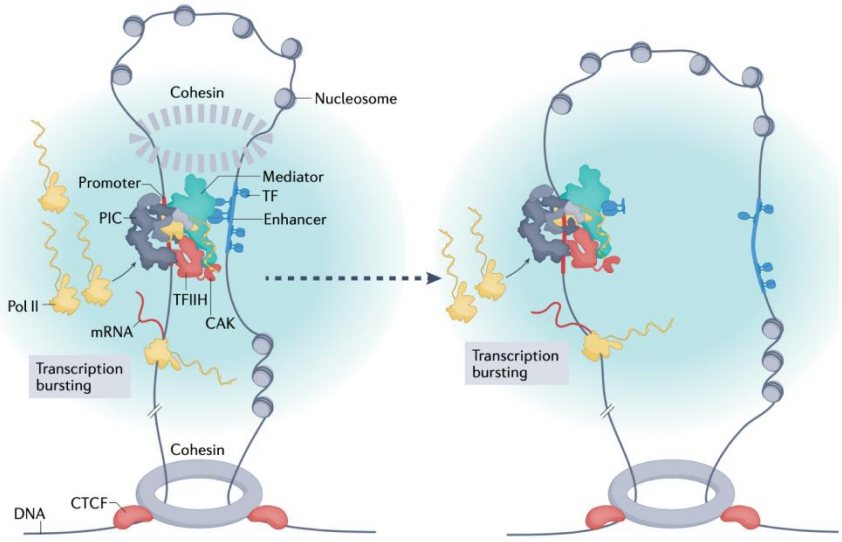


Figure 3. Molecular models to E-P interactions. (A) The plot shows homotypic interactions model of E-P interactions. (B) The plot shows condensates model of E-P interactions. (C) The plot shows Mediator-complex model of E-P interactions. (D) The plot shows cohesin model of E-P interactions. Reprinted from (Richter et al., 2022; Kota Hamamoto and Fukaya, 2022).

2.3 The aims of my study

As mentioned the above, sequencing-based proximity capturing and microscopy-based visualization techniques have renewed our understanding of the mechanisms and the functions of the three-dimensional genome organization in the last two decades. For example, CTCF-cohesin loops are thought to be a prominent feature of the loop layer of genome organization. The functions of related factors in loop extrusion model have been indicated, such as cohesin, CTCF, NIPBL, MAU2, WAPL, and ESCO1. However, how other chromatin-bound molecules interplay with the loop extrusion model on the busy chromatin are poorly understood. Moreover, RNAPII is not only chromatin-bound but also chromatin-translocating, which brings interesting questions that how mechanisms solve the collisions between RNAPII and cohesin. Moreover, RNAPII is an essential member in transcription, which brings more important questions how RNAPII/transcription shapes genome organization.

On the one hand, some recent examples support that transcription contributes to TAD and compartmentalization formation. In mouse neural progenitors (NPCs) and human GM12878 cells, inactive X chromosome lacks active/inactive compartments and topologically associating domains, except regions of genes that escape from X-chromosome inactivation (XCI; Darrow et al., 2016; Giorgetti et al., 2016). It was reported that TAD-like structures and compartments were shown around the loci of escape genes. Moreover, it was reported that TADs emerges during *Drosophila* embryogenesis, and establishment of the TAD boundaries are associated with active RNAPII loci (Hug et al., 2017). Furthermore, RNA was suggested that is involved in nuclear architectural organization, such as nuclear bodies (*Xist* lncRNA mediates Barr Body and *Neat1* lncRNA mediates paraspeckles; Rinn et al., 2014) and nucleolus (rRNAs are involved). Moreover, non-coding RNA was implicated that is involved in loop and compartmentalization formation (Isoda et al., 2019). For example, non-coding RNA ThymoD re-localizes the enhancer of *Bcl11* from lamina to the nuclear interior and

facilitates *Bcl11* E-P loop formation during T cells to DN2a differentiation (Isoda et al., 2019). Moreover, RNA and its RNP have been reported that are involved in phase separation (Lin et al., 2015). In addition, RNAPII was reported that impedes condensin extrusion in bacteria and affects cohesin positioning (Busslinger et al., 2017; Heinz et al., 2018; Brandão et al., 2019)

On the other hand, some reports conclude that transcription play a minor role in genome organization, which bring the topic under debate. Hi-C contact frequency maps showed that RNase treatment did not affect TAD boundaries, which indicates RNAs are not involved in TADs formation globally (Barutcu et al., 2019). However, compartmentalization is affected profoundly upon RNase treatment before crosslinking of cells (Barutcu et al., 2019). Furthermore, Hi-C experiments have been done upon transcription inhibition, and then only minor effects were seen on the insulation of TAD boundary (Barutcu et al., 2019; Hug et al., 2017). Moreover, depletion of RNAPII in mESCs showed minor effects in global genome organization using Hi-C data (Jiang et al., 2019). In addition, CTCF and cohesin depletion affected slightly on gene expression (Rao et al., 2017; Nore et al., 2017). However, transcription inhibition and RNAPII depletion are two different aspects due to inhibition of transcription has minor effects on RNAPII occupancy on chromatin.

Here, I used a modified human cancer cell line DLD-1-RPB1-mAID which can be reversible and rapid degradation of RNAPII. RNAPII is an robustness protein in transcription related processed, not only promoters but also enhancers are occupied by it. By comparing contact frequencies before and after degrading it to explore the role of RNAPII in genome organization. RNAPII is assumed that mainly play functions during G1-phase, here I use G1 Hi-C data to explore RNAPII functions in the G1 phase. As well, I use Micro-C data to study transcription-level structures. Interestingly, Reentry G1 cell cycle phase offers an opportunity to investigate genome re-organization (Zhang et al., 2019). Notably, it was reported that many CTCF-cohesin independent E-P interactions are emerged earlier than

structural loops. Hence, it is interesting to investigate the roles of RNAPII in genome re-establishment during mitosis to G1 transition.

3. Material and Methods

3.1 Generating Hi-C contact matrices

Hi-C contact matrices were produced using runHiC (0.8.4-r1) pipeline (http://xiaotaowang.github.io/HiC_pipeline/quickstart.html). runHiC is an easy-to-use pipeline and supports parallel computing. Matrices of replicates and replicates-merged one were generated separately through runHiC. Briefly, paired reads were separately mapped to human reference genome hg38 using BWA (>=0.7.17), and then ligation events were detected using "pairtools parse" (0.3.0) with core parameters: *--min-mapq 1 --walks-policy all --max-molecule-size 2000 --max-inter-align-gap 20*. Furthermore, unique alignment reads were kept, and then ligation events were verified by the cutting sites of restriction enzymes that were used in the experiments. Typical restriction enzymes (eg, HindIII) and Arima cocktail enzymes are both acceptable in runHiC. PCR duplicates were filtered. Finally, ICE-balanced *.cool* files were generated automatically, and KR-balanced *.hic* files were generated manually from *.pairs.gz* files. *.cool* files were visualized via HiGlass, and instructions for using HiGlass are at (https://github.com/higlass/higlass-docker). *.hic* files were visualized via JuiceBox, and the website version of it is at (https://aidenlab.org/juicebox/). Hi-C maps are shown from HiGlass (**Figure 11**; **Figure 17B**) or JuiceBox. Usually the 250kbp resolution for showing compartment A/B, the 10kbp resolution for showing loops. The "conda" is highly recommended to install runHiC.

The core and optimized code that handles this task:
Contents of Job.sh:

```
#!/bin/bash
# Slurm Parameters
#SBATCH -p fat,medium
#SBATCH --job-name=runHiC
#SBATCH --output=runHiC.%A_%a.out
#SBATCH --error=runHiC.%A_%a.err
```

#SBATCH -c 20
#SBATCH --mem=256GB
#SBATCH --time=48:00:00
runHiC pileup -m datasets.tsv -p <path of folder> -g hg38 -C hg38.chrom.sizes -f <name of folder> -F FASTQ -A bwa-mem -i hg38.fa -t 20 -O BAM --include-readid --include-sam --memory 200G --logFile <runHiC-mapping.log> --tmpdir <tmp_folder>
The format of datasets.tsv:
<reads name> <sample name> R1 Arima
<reads name> <sample> R2 Arima

3.2 Generating Micro-C contact matrices

Micro-C contact matrices were produced using commands of Dovetail Genomics (https://micro-c.readthedocs.io/en/latest/fastq_to_bam.html) and (https://micro-c.readthedocs.io/en/latest/contact_map.html). Unlike the runHiC pipeline that is hard to modify, the Dovetail Genomics offers naked commands that are easy to edit. The Micro-C experiments utilize micrococcal nuclease (MNase) instead of restriction enzymes, which is around five times more cutting frequency than 6bp-cutter restriction enzymes. Interestingly, ligation events of Micro-C were detected with slightly different parameters: *--min-mapq 40 --walks-policy 5unique --max-inter-align-gap 30*. However, I tested that multiple ligation events are few, so the parameter --walks-policy would not cause problems. The rules using “pairtools” to detect ligation events can be found at (https://pairtools.readthedocs.io/en/latest/parsing.html#overview). Micro-C maps are shown via HiGlass (**Figure 24A**; **Figure 31**; **Figure 37**) or JuiceBox.

The core and optimized code that handles this task:
Contents of Job.sh:
#!/bin/bash
 # Slurm Parameters
#SBATCH -p fat,fat+,medium
#SBATCH --job-name=MicroC
#SBATCH --output=MicroC.%A_%a.out
#SBATCH --error=MicroC.%A_%a.err

#SBATCH -c 20
#SBATCH --mem=256GB
#SBATCH --time=48:00:00
bwa mem -5SP -T0 -t <cores> <ref.fa> <MicroC.R1.fastq.gz> <MicroC.R2.fastq.gz> | pairtools parse --min-mapq 40 --walks-policy 5unique --max-inter-align-gap 30 --nproc-in <cores> --nproc-out <cores> --chroms-path <ref.genome> | pairtools sort --tmpdir=<full_path/to/tmpdir> --nproc <cores> | pairtools dedup --nproc-in <cores> --nproc-out <cores> --mark-dups --output-stats <stats.txt> | pairtools split --nproc-in <cores> --nproc-out <cores> --output-pairs <mapped.pairs> --output-sam - | samtools view -bS -@<cores> | samtools sort -@<cores> -o <mapped.PT.bam>; samtools index <mapped.PT.bam>
Generating matrices:
bgzip <mapped.pairs>
pairix <mapped.pairs.gz>
cooler cload pairix -p 16 hg38.genome:1000 mapped.pairs.gz matrix_1kb.cool
cooler zoomify --balance -p 16 -r 1000N matrix_1kb.cool

Notably, 1) -r 1000N is required to generate matrices with resolutions of 1000,5000,10000,... instead of the default resolutions 1000,2000,4000,8000,...; 2) The merged *.cool* file of replicates needs to be done manually with core and optimized parameters: *cooler merge <merged.cool> <rep1.cool> <rep2.cool>.* Importantly, the same <hg38.chrom.sizes> file need to be used during the whole processing of Micro-C data; 3) The merged *.hic* file of replicates needs to be done manually with core and optimized parameters: *pairtools merge -o <merge.pairs> --memory 200G --compress-program gzip --nproc 20 <rep1.pairs.gz> <rep2.pairs.gz>; pairtools sort -o <merge.pairs.sort> --memory 100G --nproc 15 <merge.pairs>; java -Xmx120000m -Djava.awt.headless=true -jar juicer_tools_1.22.01.jar pre --threads 16 <merge.pairs.sort> <merge.pairs.sort.hic> <hg38.chrom.sizes>.*

3.3 Interaction decay and interchromosome interaction plots

The script to plot intra-chromosomal average P(s) curves (**Figure 12A**; **Figure 18**; **Figure 24B**; **Figure 28B**; **Figure 41B**) that contact probability as a function of genomic distance, and to plot heatmaps of inter-chromosome interactions (**Figure 27B**; **Figure 41A**) are at

(https://cooltools.readthedocs.io/en/latest/notebooks/contacts_vs_distance.html). Here is an example to better understand decay plots (**Figure 4**). For different phases of a cell cycle, mitosis has a dramatically distinct pattern of contact frequencies compared to the other phases. Together with the morphology of mitotic chromosomes, it is easy to picture that the highly condensed polymers lose genome architecture of interphase and are compacted to be shorter. Therefore, the mitotic decay plot is a good example to interpret the decay plots. Moreover, the mitotic decay plot is a good reference to exhibit the reached level of the most significant changes of genome organization.

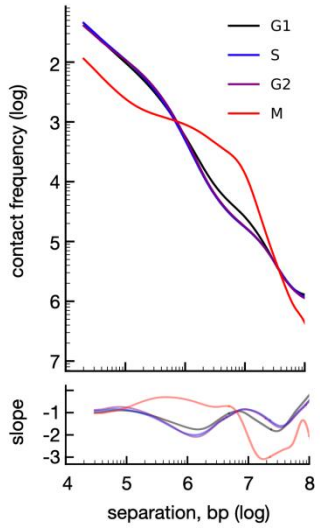


Figure 4. Interaction decay plots of different phases of a cell cycle. The plot shows interaction decay plots of G1-, G2-, S-, and M-phase. Data from (Wutz et al., 2017).

The core and optimized code that handles this task (only the optimized part is shown):

```
...
ctrl_cvd_smooth_agg = cooltools.expected_cis(
    clr=clr, view_df=None, smooth=True, aggregate_smoothed=True, nproc=20)
aux_cvd_smooth_agg = cooltools.expected_cis(
    clr=aux, view_df=None, smooth=True, aggregate_smoothed=True, nproc=20)
...
ax = axs[0]
ax.loglog(ctrl_cvd_merged['s_bp'], ctrl_cvd_merged['balanced.avg.smoothed.agg'],'-',
markersize=10,   color="black")
ax.set(ylabel='IC contact frequency', xlim=(1e4,1e8))
```

```
ax.set_aspect(1.0)
ax.grid(False)
ax.loglog(aux_cvd_merged['s_bp'], aux_cvd_merged['balanced.avg.smoothed.agg'],
      '-', markersize=10, color="red")
ax.legend(['sample1', 'sample2'],
     loc='upper right', frameon=False, labelspacing=1, fontsize=20)
ax.tick_params(direction="in", width=2.5, labelsize=16, length=12, which="major")
ax.tick_params(direction="in", width=1, labelsize=6, length=6, which="minor")
ax.spines["bottom"].set_linewidth(2.5)
ax.spines["left"].set_linewidth(2.5)
ax.spines["right"].set_visible(False)
ax.spines["top"].set_visible(False)
ax = axs[1]
ax.semilogx(ctrl_cvd_merged['s_bp'], ctrl_der, alpha=0.5, color="black")
ax.set(xlabel='separation, bp', ylim=(-1.7,0), ylabel='slope')
ax.tick_params(direction="in", width=2.5, labelsize=16, length=12, which="major")
ax.tick_params(direction="in", width=1, labelsize=6,length=6, which="minor")
ax.spines["bottom"].set_linewidth(2.5)
ax.spines["left"].set_linewidth(2.5)
ax.spines["right"].set_visible(False)
ax.spines["top"].set_visible(False)
ax.grid(False)
ax.semilogx(aux_cvd_merged['s_bp'], aux_der, alpha=0.5, color="red")
plt.savefig("out.pdf", dpi='figure', format=None, metadata=None)
...
```

For heatmaps of inter-chromosomal interactions, the optimized parameters are vmin and vmax.

3.4 Analysis of eigenvector

The eigenvector is the first principal component of Pearson's Hi-C matrix (Lieberman-Aiden et al., 2009). Eigenvector analysis was performed using (https://github.com/aidenlab/juicer/wiki/Eigenvector) with main parameters: *java -jar juicer_tools_1.22.01.jar eigenvector -p KR <input.hic> <chrom> BP 250000 <output>*. The output is a set of numbers and either >0 or <0. Signals of H3K27ac, H3K4me3, gene density, or GC content can be used to decide which one is active compartment (A), and then the

other one is compartment B. I used gene density with 250kbp resolution to decide it in my analysis.

Genomic subcompartment A/B analysis was done via CALDER(V0.1) with 50kbp resolution. The pipeline can be found at (https://github.com/CSOgroup/CALDER). The genome was clustered into A (A1.1, A1.2, A2.1, and A2.2) and B (B1.1, B1.2, B2.1, and B2.2). IGV can visualize regions of subcompartments.

The core and optimized code that handles this task (an example is showed): *CALDER_main(paste0("<KR_balanced_OB>"), chr="X", bin_size=50000, out_dir='<outname>', sub_domains=TRUE, save_intermediate_data=TRUE).*

3.5 Compartmentalization segregation strength and saddle plots

Compartmentalization segregation strength is designed to show segregation between compartment A and B (Zhang et al., 2019). Therefore, the top 20% subregions of compartment A and B are extracted out according to their EV1 value. Then the interactions among them are gotten from “dump” with main parameters: *java -jar juicer_tools_1.22.01.jar dump oe KR <input.hic> <chrom> <chrom> BP 250000 <output>*. The formula of A/B compartmentalization segregation strength of each chromosome is: (median(20% strongest AA)+median(20% strongest BB))/(median(20% strongest AB)+median(20% strongest BA)). The compartmentalization segregation strength of the entire genome is the average of all chromosomes’ (**Figure 12B**; **Figure 19**; **Figure 28A**; **Figure 41C**).

The main idea of displaying saddle plots is to aggregate interaction signals of B-B and A-A, respectively (Zhang et al., 2019). Therefore, reordering of genomic coordinates is necessary. First, all EV1 values of compartment B need to be adjusted to be negative, and all EV1 values of compartment A need to be positive. Second, I reordered genomic regions within each chromosome according to their EV1 value from small to big. Third, I built new contact matrices with reordered coordinates of each chromosome, and observed/expected values

from “dump” with 250kbp resolution were used. Last, each chromosome was aggregated into 50 sections to eliminate various chromosome sizes, and then the entire chromosomes matrix is merged from 50 sections of each chromosome. Therefore, all interactions of B-B are aggregated at left-upper, all interactions of A-A are enriched at right-lower (**Figure 12B**; **Figure 19; Figure 28A**; **Figure 41C**).

3.6 Loop calling of Hi-C matrices

A loop represents two loci are proximity in 3D space, and also means many cells from a cell population contain this proximal structure in a Hi-C map. Hence, loops represent significant-more contact frequencies than random interactions in Hi-C maps. Thus, there are two main ideas to call loops. One is calling significantly enriched contacts after normalizing by local expected value, like HiCCUPs (Hi-C Computational Unbiased Peak Search; https://github.com/aidenlab/juicer/wiki/HiCCUPS). The other one is using image analysis, like SIP (https://github.com/PouletAxel/SIP) and mustache (https://github.com/ay-lab/mustache). My Hi-C data is working better with image analysis method due to HiCCUPs requiring higher resolutions.

I used SIP (V1) to call loops from Hi-C matrices and the core and optimized parameters are: *-g 2.0 -mat 2000 -d 3 -res 10000 -t 2800 -min 2.0 -max 2.0 -sat 0.01 -nbZero 4 -factor 1 -fdr 0.01 -del true -cpu 1 -isAccurate false -isDroso false*.

3.7 Loop calling of Micro-C matrices

Micro-C loops were called with a multi-tool and multi-resolution approach (HiCCUPs (V1.22.01), cooltools (0.4.1), and mustache (V1.0.1); with 5- and 10-kbp resolution) as previously described (Hsieh et al., 2020; Krietenstein et al., 2020). Then, loop lists derived from each tool were merged using pgltools (2.2.0; Greenwald et al., 2017) as follows: dots from both 10- and 5-kbp resolution are retained if >10 read counts support them and kept at the higher resolution.

Core and optimized parameters for Micro-C loop calling are: *java -jar juicer_tools_1.22.01.jar hiccups -m 1024 -r 5000,10000 -k KR <input.hic> <output>; mustache -f <input.mcool> -o <output> -r <resolutions> -st <0.88> -pt <0.01> -p <10>; cooltools call-dots -p <10> --max-loci-separation <10000000> --fdr <0.01> --dots-clustering-radius <29000> -o <output> <input.mcool::/resolutions/resolutions>.*

3.8 Differential loop intensity analysis

The main idea to show the loop intensity of all loops in a Hi-C map is to show local background and centered loop together in a square of sub-matrices of 11x11 bins (**Figure 5**). Moreover, aggregating sub-matrices, and then averaging values of all squares (**Figure 5**). coolpup.py is the script to do it with core and optimized parameters: *coolpup.py --outname <outname> <input.mcool::/resolutions/10000> <loop.bedpe>*. The core and optimized parameters for plotting: *plotpup.py --n_cols <number of columns> --row_names <names> --col_names <names> -o <out.png>*. In addition, the size of the loop square can be adjusted as required via --pad, color can be adjusted via --cmap, and bins are involved in representing enrichment can be adjusted via --enrichment.

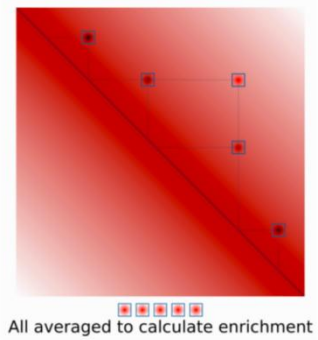


Figure 5. Schematic of loop intensity plots. The plot shows that the idea to display the loop intensity of all loops in a Hi-C map. Reprinted from (https://coolpuppy.readthedocs.io/en/latest/walkthrough.html).

I used two methods to search differential loops between loop lists. The first is HiCCUPSDiff (https://github.com/aidenlab/juicer/wiki/HiCCUPSDiff) which examines the loops detected in each *.hic* file on the other. However, it did not work efficiently on my data.

The second way is to classify shared or specific loops according to their coordinates. However, solely using coordinates is not enough. I also generated loop intensity plots at the same time to ensure shared and specific features between samples. By showing loop intensity using coordinates of sample A on contact matrix of sample B, and then the result shows the degree of same or different between A and B (**Figure13A**; **Figure 16B**; **Figure 20B**; **Figure 21A**; **Figure 22B**; **Figure 25B**; **Figure 26**; **Figure 29B**; **Figure 32**; **Figure 38A**; **Figure 42B and 42E**). Shared loops are presented in both, while sample B does not exhibit sample-A-specific loop signals. Related commands are shown. Notably, for deciding shared loops, exactly same coordinates are ineffective in my data, so I allow 20kbp or 30kbp extension of the coordinates.

Core and optimized parameters are: *pgltools intersect -a <control.bedpe> -b <degron.bedpe> -wa -d 30000 > <shared_inC>; pgltools intersect -a <degron.bedpe> -b <control.bedpe> -v -d 30000 > <special_inA>. An example of flexibility use pgltools to deal with coordinates of loops among three samples: pgltools intersect -a <control.loops> -b <auxin.loops> -d 30000 -v | pgltools intersect -stdInA -b <tri.loops> -d 30000 -v > <control.specific.loops>.*

The core code that shows Venn plots of shared and specific loops between two samples (**Figure 20A**; **Figure 25A**; **Figure 29A**; **Figure 42A**):

```
pdf( “venn.pdf", width = 5, height = 5 )
draw.pairwise.venn(<number1>, <number2>, <number3>, category = c("<sample1>", "<sample2>"), lty = rep("blank", 2), fill = c("light blue", "pink"), alpha = rep(0.5, 2), cat.pos = c(0,  0), cat.dist = rep(0.025, 2))
dev.off()
```

Loop length plots were generated for comparing loop length among samples (**Figure 15C**; **Figure 20D**; **Figure 22A**; **Figure 38C**; **Figure 42D**). Loop length extraction using command: *awk '{print $5-$2,sample_name}' OFS="\t" <loop.bedpe> > <loop.bedpe.length>*.

The core and optimized code that generates loop length plots:

```
library(ggplot2)
library(ggsignif)
file = read.table("length.txt", head = FALSE)
```

```
pplot = ggplot(file, aes(x = V2, y = log2(V1)))
pplot = pplot + geom_boxplot()
pplot = pplot + geom_signif(comparisons = list(c("name1", "name2")),map_signif_level = TRUE, tip_length = 0, y_position = 27)
pplot = pplot + theme_classic()
pplot = pplot + xlab("samples")
pplot = pplot + ylab("log2(loops-length)")
ggsave("legnth.pdf",device = "pdf")
```

3.9 Analysis of loop anchor insulation

Figure 6 is the idea to show insulation intensity genome-wide. I made insulation plots to compare insulation intensity of loop anchors among samples, to visualize insulation intensity of specific anchors, and to check stripe pattern of loop extrusion (**Figure 13B**; **Figure 20C**; **Figure 38B**; **Figure 39B and 39C**; **Figure 42C**). Notably, length of anchors need to be kept the same to generate a beautiful plot. Therefore, I cut all anchors and only took outer 5000bp to be as the input. I used coolpup.py to generate insulation values with core and optimized parameters: --local --pad 500 <::/resolutions/5000>, and used plotpup.py to generate plots. Notably, color patterns, gist_ncar and nipy_spectral can show minor changes of insulation.

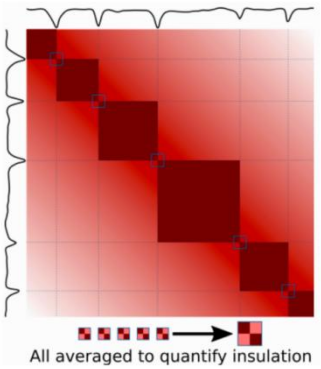


Figure 6. Schematic of loop anchor insulation plots. The plot shows that the idea to display the insulation of all loop anchors in a Hi-C. map. Reprinted from (https://coolpuppy.readthedocs.io/en/latest/walkthrough.html).

3.10 Analysis of interactions among peaks

Figure 7 is the idea to show interactions among ChIP peaks in 3D space. Core and optimized commands to do this task are (**Figure 36C and 36D**; **Figure 46**), *coolpup.py <::/resolutions/1000> <peak.bed> --outname <outfile> --pad 50 --maxdist <2000000> --mindist <20000>; plotpup.py <infile> --n_cols <number of columns> --col_names <ctrl,auxin> --row_names <ctrl,auxin> -o <out.png>.*

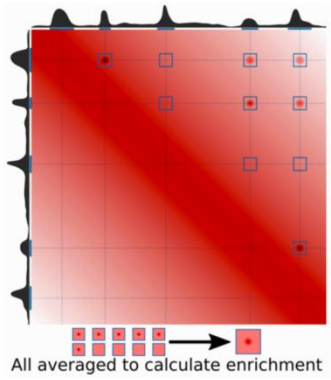


Figure 7. Schematic of plots of peak interaction intensity in 3D plots. The plot shows that the idea to display the interactions of ChIP-peaks in 3D space. Reprinted from (https://coolpuppy.readthedocs.io/en/latest/walkthrough.html).

3.11 Cleavage under targets and tagmentation (CUT&Tag) data analysis

CUT&Tag reads (**Table 1**) were processed according to the standard CUT&Tag pipeline (https://yezhengstat.github.io/CUTTag_tutorial/). Briefly, paired-end reads were aligned to human (hg38) reference genome using Bowtie2 (v2.3.4.1). Next, PCR duplications were removed using picard (v2.20.2), and then reads that mapping quality fewer than 30 were discarded using samtools (v1.11). Last, file format conversion from *.bam* to *.bedgragh*, and then peak calling was done with SEACR (v1.3). Adapters trimming step was skipped when reads length shorter than 40bp, which usually do not contain adapter sequences. The spike-in calibration step was also skipped when fewer than 50 reads mapped on E.coli

reference genome (ASM584v2), or there were no comparable spike-in mapping rates among samples. In addition, IgG control for thresholding was given up during peak calling when it was missed on auxin-treated samples.

Moreover, I summarized some tips for CUT&Tag data processing. It needs to be paid attention when duplication rate of material fragments is higher than 80%. The mitochondrial DNA contamination may cause this problem. Moreover, high-quality mapping and unique paired-end reads should be more than 1 million after all filtering and merging steps for a reliable peak calling result. Furthermore, almost no reads mapped on E.coli reference genome is a common phenomenon in my data, suggesting protein-A-Tn5 is highly effective utilized. Notably, IgG control normalization is a sensitive step, and skipping it is better than misusing it. I optimized peak calling results using a multi-FDR-tryout method (FDR < 0.01 to < 0.1) due to parameters of peak calling from standard pipeline did not meet my requirements. Acceptable FDRs could be vary between different datasets, but were always kept same for control and auxin-treated samples.

Thus, for CTCF, FDR <0.1 was selected after generating peak-signal visualization plots and peak-count plots with the FDR from 0.01 to 1.00 (**Figure 8**). Furthermore, for additional stringency, I only considered a CTCF-bound if it encompassed a canonical CTCF motif. In addition, all final kept CTCF peaks exhibit a similar nucleosome distribution pattern after MNase cutting without relating more or fewer reads counts within them, which also suggests optimized FDR <0.1 for this data is acceptable.

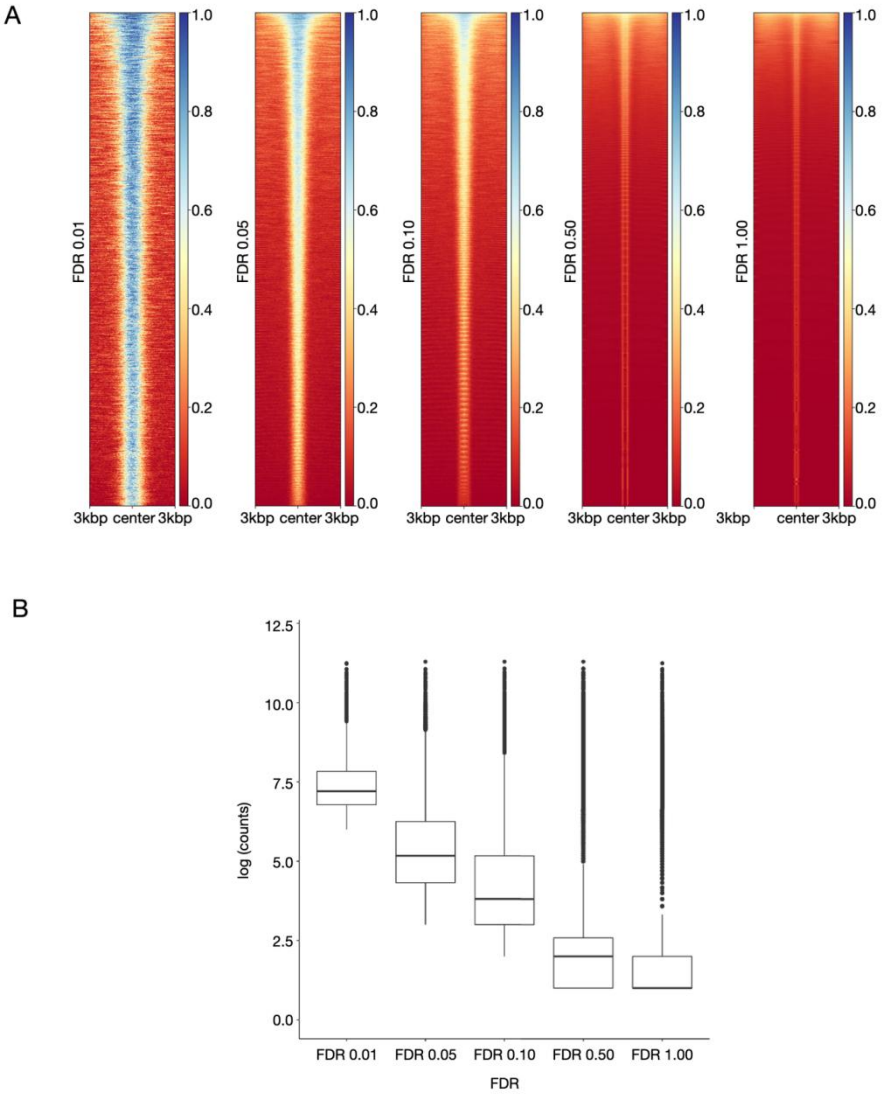


Figure 8. Analysis of CTCF CUT&Tag data. (A) The plot shows CTCF enrichment of the multi-FDR-tryout method. (B) The plot shows read counts of different FDR called unique CTCF peaks.

Related commands to generate above plots are : *bamCoverage -b <.bam> -o <.bw> --normalizeUsing <CPM>; computeMatrix reference-point -S <.bw> -R <.bed> --skipZeros --missingDataAsZero -o <.mat.gz> -p <20> -a <3000> -b <3000> --referencePoint ; plotHeatmap -m <.mat.gz> -o <.mat.gz.png> --sortUsing <sum> --startLabel <"Peak Start"> --endLabel <"Peak End"> --xAxisLabel <""> --yAxisLabel <""> --regionsLabel <CTCF> --samplesLabel <CTCF_ctrl>; bedtools getfasta -fi <hg38.fa> -bed <.bed> -fo <CTCF.peaks.bed.fa>; fimo --motif <CTCF_HUMAN.H11MO.0.A> --motif -o <CTCF_fimo_out> <HOCOMOCOv11_full_HUMAN_mono_meme_format.meme> <CTCF.peaks.bed.fa>; bedtools multicov -bams <.bam> -bed <.bed> | grep -v "chrM" | awk '{print $1,$2,$3,$11,$14}' OFS="\t" > <peaks.counts>*

Due to many CTCF and cohesin are co-occupancy, and CTCF peaks are credible with motif filtering. Thus, I mostly used coordinates of CTCF instead of SMC1A for loop anchor annotation. The problem of SMC1A peak calling is that peak numbers are not at the same level between control and auxin-treated samples even with FDR <0.01. For SMC1A, an FDR <0.01 was selected when did CTCF-independent analysis.

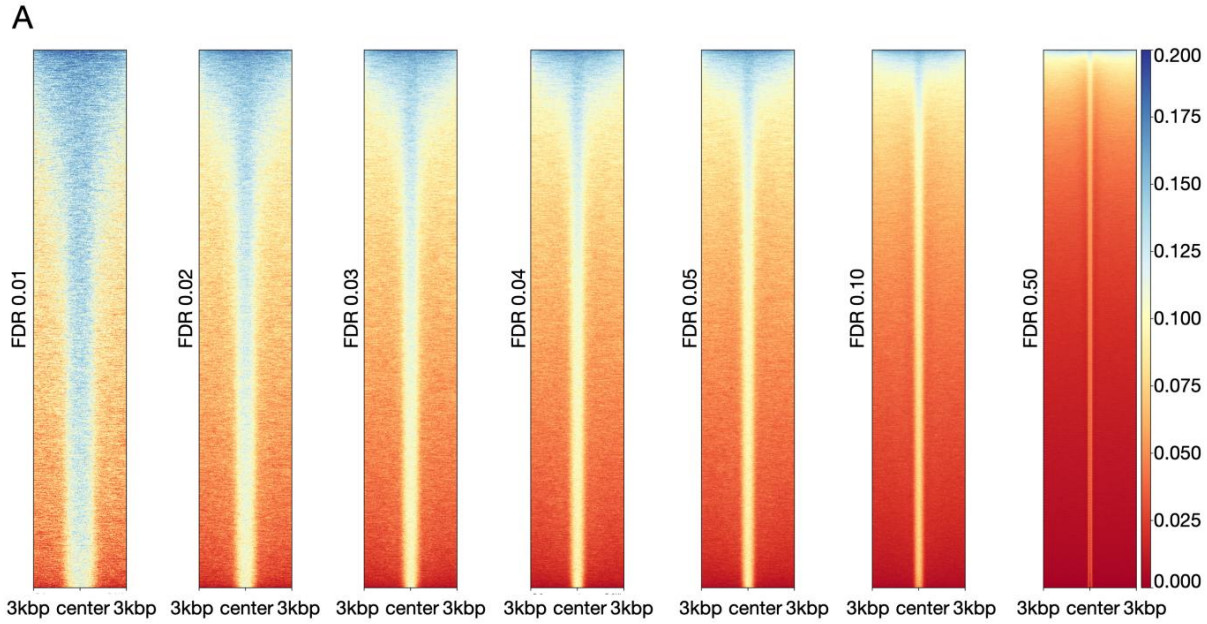
Similarly, for H3K27ac, FDR <0.025 was selected. H3K27ac peaks are considered as enhancer and promoter markers, which highly correlated with RNAPII occupancy. ChIPseeker (https://github.com/YuLab-SMU/ChIPseeker) program was used to annotate promoters and enhancers, and then TSS +- 2000bp is thought as promoter regions, the others are thought as enhancers. Notably, factory-RNA expression data was also used to define promoter regions, which +-2000bp of all expression genes' TSS are promoter regions (expression value >10). Importantly, isoforms are used instead of the whole genes when consider about coordinates of TSS.

The core code to annotate promoter and enhancers using H3K27ac peaks:

```
library(ChIPseeker)
library(TxDb.Hsapiens.UCSC.hg38.knownGene)
library(clusterProfiler)
library(annotate)
library(dplyr)
txdb <- TxDb.Hsapiens.UCSC.hg38.knownGene
samplefiles <- list.files("<path of folder>", pattern= "<name>", full.names=T)
samplefiles <- as.list(samplefiles)
names(samplefiles) <- c("H3K27ac")
peakAnnoList <- lapply(samplefiles, annotatePeak, TxDb=txdb, tssRegion=c(-1000, 1000), verbose=FALSE)
plotAnnoPie(peakAnnoList[["H3K27ac"]])
H3K27ac_annot <- as.data.frame(peakAnnoList[["H3K27ac"]]@anno)
write.table(H3K27ac_annot, "<outname>", append=FALSE, sep="\t", dec=".", row.names= FALSE, col.names=TRUE)
dev.off()
```

For H3K27me3, FDR < 0.01 was selected. Importantly, H3K27me3 is different as H3K27ac

that is an active histone mark. Moreover, Tn5 prefers to bind at active regions. Therefore, I checked whether this influences signals of H3K27me3. The results show there is no big problem like that. But still, with increasing of FDR, we see noise of open regions is gaining (**Figure 9**). In addition, material duplication rate of H3K27me3 usually relative lower than H3K27ac, CTCF, and SMC1A, but still higher than 20%. Together with suggestions from standard pipeline, I think keep duplication for H3K27me3 during analysis should not be a problem if it can optimize results. Standard pipeline contains <10% duplication and suggests keep them.



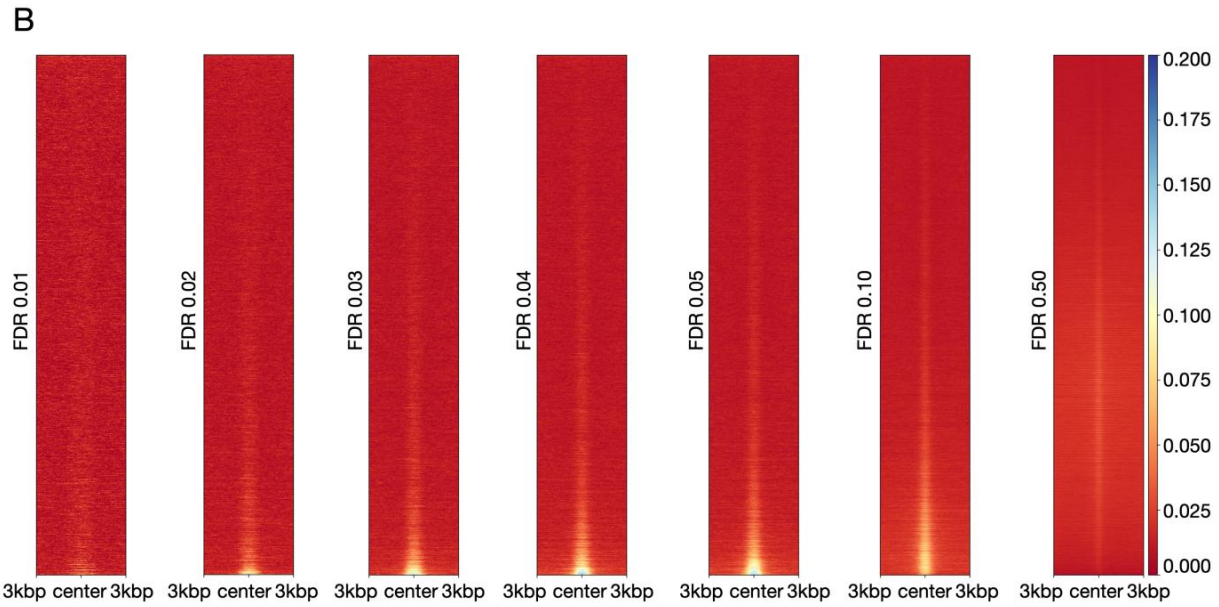


Figure 9. Analysis of H3K27me3 CUT&Tag data. (A) The plot shows H3K27me3 enrichment of multi-FDR-tryout method. (B) The plot shows H3K27ac enrichment of multi-FDR-tryout method using same coordinates in panel (A).

The script to handle differential peaks analysis (https://yezhengstat.github.io/CUTTag_tutorial/#VIII_Differential_analysis).

The core code is:
selectR = which(rowSums(countMat) > 5) ## remove low count genes
dataS = countMat[selectR,]
condition = factor(rep(histL, each = length(repL)))
dds = DESeqDataSetFromMatrix(countData = dataS, colData = DataFrame(condition), design = ~ condition)
DDS = DESeq(dds)
normDDS = counts(DDS, normalized = TRUE) ## normalization with respect to the sequencing depth
colnames(normDDS) = paste0(colnames(normDDS), "_norm")
res = results(DDS, independentFiltering = FALSE, altHypothesis = "greaterAbs")
countMatDiff = cbind(dataS, normDDS, res)

3.12 The annotation of loop anchors

Coordinates of peaks of CTCF, promoters, enhancers, and H3K27me3 were used to

annotate loop anchors using the command “pgltools intersect1D” (**Figure 14B**; **Figure 15A**; **Figure 17A**; **Figure 21B**; **Figure 23**; **Figure 30**; **Figure 36A and 36B**; **Figure 39A**; **Figure 43**). Examples of flexibility use pgltools to deal with the tasks are:

```
pgltools intersect1D -a <control.bedpe> -b <CTCF_ctrl.bed> -wa -allA | awk '$7=="A"' | awk '{print $1,$2,$3,$4,$5,$6}' OFS="\t" | sort -u | pgltools sort > tmp.left
pgltools intersect1D -a <control.bedpe> -b <CTCF_ctrl.bed> -wa -allA | awk '$7=="B"' | awk '{print $1,$2,$3,$4,$5,$6}' OFS="\t" | sort -u | pgltools sort > tmp.right
pgltools intersect -a tmp.left -b tmp.right -wa | pgltools sort | awk '{print $1,$2,$3,$4,$5,$6,"T","T"}' OFS="\t" > tmp_TT
pgltools intersect -a tmp.left -b tmp.right -v | pgltools sort | awk '{print $1,$2,$3,$4,$5,$6,"T","F"}' OFS="\t" > tmp_TF
pgltools intersect -a tmp.right -b tmp.left -v | pgltools sort | awk '{print $1,$2,$3,$4,$5,$6,"F","T"}' OFS="\t" > tmp_FT
cat tmp_TF tmp_TT tmp_FT | pgltools sort > tmp_CTCF
pgltools intersect -a <control.bedpe> -b tmp_CTCF -v | pgltools sort | awk '{print $1,$2,$3,$4,$5,$6,"F","F"}' OFS="\t" > tmp_FF
cat tmp_TT tmp_TF tmp_FT tmp_FF | pgltools sort > control.bedpe.CTCF
cut -f 7,8 control.bedpe.CTCF | sort | uniq -c
```

Notably, when deal with promoters, enhancers, and H3K27me3 peaks into loop anchors, parameter -d 10000 needs to be considered to facilitate the annotation to be more effective.

3.13 Cell culture of the DLD1-mAID-RPB1 cell line

Gene knockout, RNAi knockdown, and treatment with chemical inhibitors are typical methods to study functions of genes or proteins of interest. However, these methods contain apparent drawbacks, they are insufficient for studying various proteins. For example, knocking out essential genes can lead to death of cells, RNAi knowdown can induce target-off issues, and inhibitor treatment can cause unclear secondary effects. Here, I used a rapid and reversible protein degradation method, the AID (auxin-induced degradation) system, which was originally derived from plants and then widely utilized in non-plant eukaryotic cell lines. The AID domain can be poly-ubiquitylation by SCF-OsTIR1 E3 ubiquitin

ligase upon auxin existing, and then to acutely degrade by the proteasome. The colorectal cancer cell line DLD1-RPB1-mAID is a gift from the Maeshima laboratory (NIG, Japan), in which the mAID domain was inserted at the N-terminal of RPB1 (the biggest subunit of RNAPII) and Tet-OsTIR1 as well was inserted in the genome. In this cell line, after 24 hours of doxycycline-induced TIR1 expression, RPB1 can be acutely degraded upon adding auxin. Moreover, the recovery of RPB1 can be reached through washing off auxin and adding auxinole (an auxin antagonist in the mAID system; Nagashima et al., 2019).

DLD1-RPB1-mAID cells were cultured in RPMI 1640 medium (R8758, Sigma) supplemented with 10% fetal bovine serum (FBS) under 5% CO2. Degradation of RPB1 was initiated via treatment with doxycycline for 24 hours to induce OsTIR1 expression before adding 500 uM indole-3-acetic acid solution ("auxin," Sigma-Aldrich) for at least treatment for 2 hours. Details of treatment under different conditions are show in **Table 2**.

3.14 Hi-C experiments

Hi-C experiments (2h control, 2h+auxin, G1 control r1, 14h+auxin, and 14h+aux/+tript) were performed by myself using Arima Hi-C kits.

Firstly, cells without medium were crosslinked using 1% formaldehyde for 10 minutes under room temperature with shaking. Then cells were quenched on ices with 125mM Glycine for 3 minute after removing the formaldehyde. Furthermore, cells were harvested and centrifuged with 500 x G for 5 minutes. Moreover, cells were washed with cold PBS for twice and centrifuged then stored at -80°C.

The step of estimating input amount was performed according to the protocol of Arima Hi-C kits. Importantly, do not use more than 1 million cells when performing this step. Excessive cells can cause problem with the concentration of the output. Inaccurate estimates of DNA concentrations at this step can cause Hi-C experiments to use too many or too few cells,

which can affect the overall experiments. For my Hi-C experiments, using 1-2 million DLD1-mAID-RPB1 cells is enough to generate good Hi-C sequencing data. Using ~1.5 million DLD1-mAID-RPB1 cells get ~3μg DNA, and ~3μg DNA is recommended from the protocol of Arima Hi-C kits for each reaction.

The Hi-C experiments were performed according to the protocol of Arima Hi-C kits. There is no problem using the protocol with DLD1-mAID-RPB1 cells. However, some steps are adjustable to get better outputs with hard-to-lysis cells. The lysis of the cell is the essential step in Hi-C experiments. Insufficent lysis of cell membrane and nuclear membrane can cause insufficent cutting, thus affects the expensive experiments without any way to rescue. In the protocol of Arima Hi-C kits, the lysis step can extend to half hour. Moreover, treatment with conditioning solution can be extended to 15 minutes. Moreover, the step of enzyme cutting with 37°C can be extended to 60 minutes (step 5 in the protocol). Moreover, the ligation step can be extended to 25 minutes (step 9 in the protocol). Importantly, mixing well the QC beads by resuspending when doing QC1 quality control. All my Hi-C experiments passed QC1. In addition, enough sonication to break fragments to be ~200-600bp is also important. Eight-cycles sonication for DLD1-mAID-RPB1 cells are enough to break fragments to be small, but for some cell lines are not enough. Estimating the sizes of fragments is highly recommended before sequencing. The general statistics of the sequencing data are shown in **Table 3**.

Micro-C experiments were performed via the Dovetail Genomics, and general statistics of the sequencing data are shown in **Table 4**.

3.15 Downloading public datasets

Public Hi-C datasets were downloaded from GEO to facilitate analysis in the thesis. Details see the following:

(Rao et al., 2017).

https://ftp.ncbi.nlm.nih.gov/geo/series/GSE104nnn/GSE104333/suppl/GSE104333_Rao-2017-untreated_combined_30.hic
https://ftp.ncbi.nlm.nih.gov/geo/series/GSE104nnn/GSE104333/suppl/GSE104333_Rao-2017-treated_6hr_combined_30.hic

(Wutz et al., 2020).

https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM4106nnn/GSM4106789/suppl/GSM4106789_HeLa_G1_HiRes_control.hic
https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM4106nnn/GSM4106794/suppl/GSM4106794_HeLa_G1_CTCF_2reps.hic

(Wutz et al., 2017).

https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM2747nnn/GSM2747738/suppl/GSM2747738_HeLa_control_2reps.hic
https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM2747nnn/GSM2747743/suppl/GSM2747743_HeLa_S_2reps.hic
https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM2747nnn/GSM2747744/suppl/GSM2747744_HeLa_G2_2reps.hic
https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM2747nnn/GSM2747742/suppl/GSM2747742_HeLa_ProMeta_2reps.hic

(Nora et al., 2017).

https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM2644nnn/GSM2644945/suppl/GSM2644945_Untreated-R1.20000.cool.gz
https://ftp.ncbi.nlm.nih.gov/geo/samples/GSM2644nnn/GSM2644947/suppl/GSM2644947_Auxin2days-R1.20000.cool.gz

(Haarhuis et al., 2017).

https://ftp.ncbi.nlm.nih.gov/geo/series/GSE95nnn/GSE95014/suppl/GSE95014_Hap1.validPairs.txt.gz
https://ftp.ncbi.nlm.nih.gov/geo/series/GSE95nnn/GSE95014/suppl/GSE95014_WaplKO_1.14.validPairs.txt.gz
https://ftp.ncbi.nlm.nih.gov/geo/series/GSE95nnn/GSE95014/suppl/GSE95014_WaplKO_3.3.validPairs.txt.gz
https://ftp.ncbi.nlm.nih.gov/geo/series/GSE95nnn/GSE95014/suppl/GSE95014_SCC4KO.validPairs.txt.gz
https://ftp.ncbi.nlm.nih.gov/geo/series/GSE95nnn/GSE95014/suppl/GSE95014_DKO_3.3.validPairs.txt.gz

3.16 Tables

Table 1. General statistics of CUT&Tag sequencing data used in the thesis.

sample names :	total reads	reads after filtering	peak number (FDR 0.01)
CTCF +auxin rep1	3,672,670	1,211,308	5,782
CTCF +auxin rep2	4,586,487	1,690,828	8,037
H3K27ac +auxin rep1	13,260,333	7,154,977	18,858
H3K27me3 +auxin rep1	14,883,314	9,493,760	20,666
H3K27me3 +auxin rep2	13,505,102	7,099,102	10,333
SMC1A +auxin rep1	4,351,892	2,032,559	11,331
SMC1A +auxin rep2	2,873,213	1,524,101	8,833
CTCF ctrl rep1	5,241,343	1,871,395	6,827
CTCF ctrl rep2	5,932,344	1,133,832	4,863
H3K27ac ctrl rep1	13,883,545	6,387,579	16,749
H3K27me3 ctrl rep1	16,652,434	9,084,083	17,951
H3K27me3 ctrl rep2	15,370,147	7,720,848	10,563
SMC1A ctrl rep1	4,495,569	1,998,722	8,292
SMC1A ctrl rep2	3,415,345	1,470,669	6,948

Table 2. Information of Hi-C and Micro-C samples in the thesis.

Hi-C dataset:	+auxin treatment	+tript treatment	cell cycle	C-based
2h control	N/A	N/A	N/A	in situ Hi-C
2h +auxin	2 hours	N/A	N/A	in situ Hi-C
G1 control, r1	N/A	N/A	G1 sorted	in situ Hi-C
G1 control, r2	N/A	N/A	reentry G1	in situ Hi-C
14h +aux/+tript	14 hours	2 hours	G1 sorted	in situ Hi-C
14h +auxin	14 hours	N/A	G1 sorted	in situ Hi-C
reentry +aux, r1	8 hours	N/A	reentry G1	in situ Hi-C
reentry +aux, r2	8 hours	N/A	reentry G1	in situ Hi-C
control Micro-C	N/A	N/A	N/A	Micro-C
+auxin Micro-C	14 hours	N/A	N/A	Micro-C

Table 3. General statistics of Hi-C sequencing data used in the thesis.

Hi-C dataset:	total reads	% aligned	Hi-C contacts	% inter-chromo	% intra-chromo	% short-range	% long-range
2h control	570,638,421	89.91	387,244,720	10.44	57.42	21.94	35.47
2h +auxin	521,206,158	91.46	361,994,319	10.84	58.62	22.43	36.17
G1 control, r1	512,551,503	89.70	339,693,791	16.49	49.78	16.75	33.03
G1 control, r2	617,136,578	89.46	400,421,522	12.40	52.48	18.77	33.70
14h +aux/+tript	461,569,471	90.67	313,489,486	16.87	51.04	19.88	28.67
14h +auxin	516,427,195	89.04	338,221,913	17.69	47.80	17.65	30.14
reentry +aux, r1	538,423,263	90.05	355,346,372	24.38	41.61	16.90	24.70
reentry +aux, r2	1,153,253,737	89.37	714,520,317	22.96	39.00	15.78	23.21

Table 4. General statistics of Micro-C sequencing data used in the thesis.

	control DLD1 Micro-C	%	+auxin DLD1 Micro-C	%
total read pairs	1,648,350,580	100.00	1,524,930,442	100.00
mapped read pairs	1,355,832,740	82.25	1,249,138,654	81.91
duplicate pairs	448,059,906	27.18	402,268,327	26.38
valid *trans* read pairs	103,889,160	11.44	102,355,674	12.09
valid *cis* read pairs	803,883,674	88.56	744,514,653	87.91
long-range contacts	559,791,780	61.67	507,103,457	59.88
long-/short-contact ratio	2.29x	N/A	2.14x	N/A

4. Results

4.1 New loops are formed upon RNAPII depletion

To explore the roles of RNAPII in genome organization and the relationship between genome folding and genomic function, I used a modified colorectal cancer cell line, DLD1-RPB1-mAID, which allows for inducible and reversible RNAPII degradation without compromising cell viability (Yesbolatova et al., 2019; Nagashima et al, 2019; Zhang et al., 2021). By capturing and comparing whole-genome, long-distance chromatin interactions in the presence or absence of RNAPII, I investigated the contribution of RNAPII to different layers of genome folding.

4.1.1 The rapid degradation of RPB1

To assess the efficiency of the degradation system, I first examined RNAPII residual levels under different depletion times by western blot. As shown in **Figure 10**, obvious degradation already occurs within 2 hours, but degradation rates become slower soon after. Phosphorylation of RNAPII is dynamically regulated during transcription. It has been reported that the C-terminal domain (CTD) of RNAPII contains repeated heptapeptides, Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, of which Ser5 is phosphorylated upon transcription initiation, and Ser2 phosphorylated during transcriptional elongation. Interestingly, RNAPII^{ser5} levels in **Figure 10** are less affected by degradation and recover first upon auxin washout, which suggests that chromatin-bound phospho-RNAPII is most essential to cells. Similar results were obtained using an anti-RPB1-CTD antibody recognizing two RNAPII forms, the lower one thought to be soluble RNAPII with unphosphorylated CTD and the upper one phosphorylated. Notably, RNAPII recovery is unexpectedly slow and does not reach wt levels within 24 hours (**Figure 10**). Unlike cohesin that can be almost fully restored within

hours (Rao et al., 2017), only <70% RNAPII^{ser5} comes back after 24 hours (**Figure 10**); this needs to be considered when designing recovery experiments.

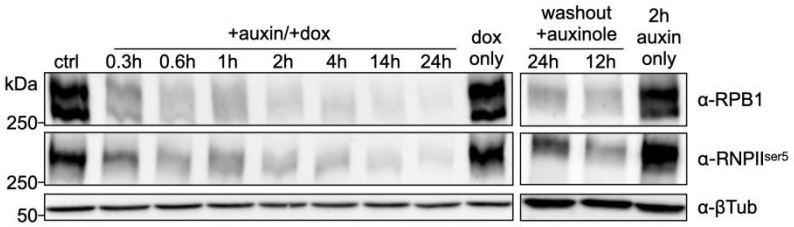


Figure 10. Western blots of DLD-1-RPB1-mAID degradation system. Western blots show degradation of total cell RPB1 or RNAPII^{ser5} on increasing treatment time of auxin, which is from 20 minutes to 24 hours (left), and recovery of RPB1 or RNAPII^{ser5} following auxin washout in the presence of auxinole (right); β-tubulin provides a loading control.

4.1.2 Overview of the genome organization in the absence of RNAPII

To further explore the aforementioned aims, I performed two Hi-C experiments using cells with RNAPII (control) and cells where RNAPII was degraded for 2 hours, whereby ~70% RNAPII was depleted (**Figure 10**). Moreover, Hi-C experiments, as described in the method part, can capture long-distance interactions genome-wide. Hi-C, together with high-throughput sequencing, is a prominent method to investigate the genome organization. After mapping the Hi-C sequencing data to reference genome, the Hi-C contact matrices or maps were generated. As a result, the Hi-C maps are clear and informational of the different layers of genome organization. For example, chromosome territories (**Figure 11A**), compartments (**Figure 11B**), TADs, stripes, and loops (**Figure 11C**).

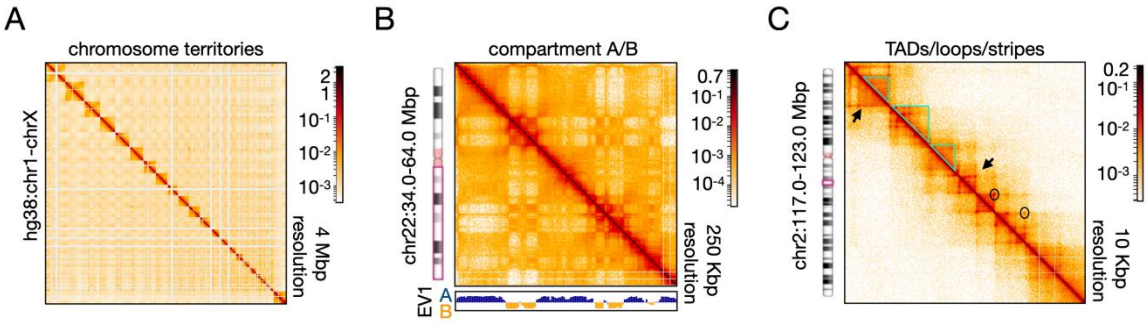


Figure 11. HiGlass visualizes the different layers of genome organization. (A) The plot shows chromosome territories that are presented as intrachromosomal interactions are much more than interchromosomal interactions. (B) The plot shows compartment A/B, in which blue-marked EV1 regions are active (compartment A) and yellow-marked EV1 regions are inactive (compartment B). Compartment A and B are homotypical clusters, which are presented as a checkboard pattern in a Hi-C map. (C) The plot shows examples of TADs (blue triangles), loops (black circles), and stripes (black arrows).

Furthermore, I generated interaction decay plots, in which contact frequency is plotted as a function of genomic distance, and their derivative slope plots, in which polymer features of the whole-genome organization are shown. In both, signals are not altered as a result of RNAPII degradation (**Figure 12A**). Moreover, A/B compartmentalization saddle plots, which show organization and segregation of active and inactive Mbp-size regions, are essentially the same between control and RNAPII degradation data (**Figure 12B**). Collectively, these results indicate that RNAPII degradation cell populations did not carry genome-wide alterations in 3D genome organization.

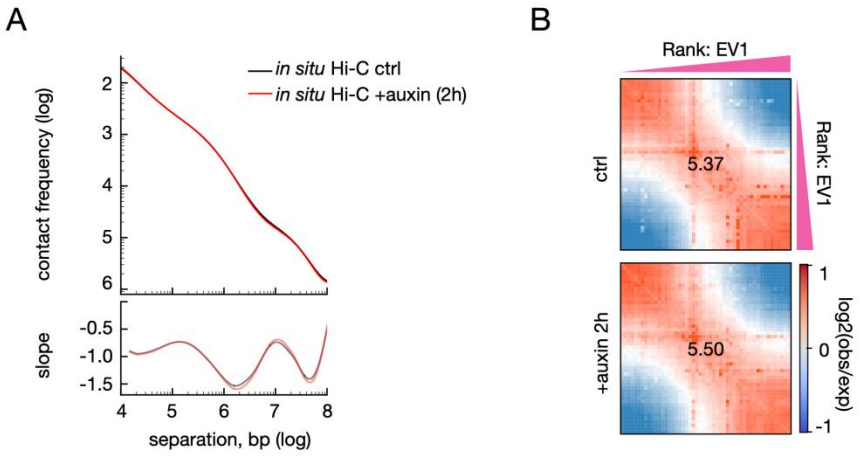


Figure 12. Interaction decay and saddle plots. (A) Interaction decay plots (upper) show Hi-C interaction frequency (log) as a function of genomic distance (log) of control (black) and RNAPII degradation (red) data. The derivative slope plots (lower) show polymer-like behaviors of genome folding. The plots were generated with 5kbp resolution. (B) A/B compartmentalization saddle plots show organization and segregation of active and inactive regions, in which signals at the left-upper represent B-B interactions and signals at the right-lower represent A-A interactions. The numbers in the center are compartment segregation strength. The plots were generated with 250kbp resolution.

4.1.3 Investigating formation of the new loops upon RNAPII depletion

4.1.3.1 Insulation of loop anchors do not alterations upon RNAPII depletion

Before data analysis, we assumed that enhancer and promoter interactions (E-P interactions) would be lost upon RNAPII degradation due to RNAPII occupying these regions. However, gained loops are more noticeable in the Hi-C maps than lost ones upon RNAPII degradation (**Figure 13A**). Interestingly, ~1000 loops are gained upon RNAPII degradation, which are shown in **Figure 13A** using loop signal plots. The method to generate loop signal plots was describe in the method part of the thesis. In brief, the loop signal plot presents the average loop intensity level of a sample. By comparing the plots in the presence and absence of RNAPII, the role of RNAPII underlying loop formation is estimated.

Furthermore, these gained loops are first assumed to be transcription-independent CTCF

loops (Rao et al., 2017; Nora et al., 2017). Moreover, the CTCF, a transcription factor, is also thought as insulator of the genome organization (Rao et al., 2014), which can insulate interactions between the adjacent chromatin regions. Thus, as the assumption described above, anchors of the gained loops should show specific and strong insulation. However, the insulation plots do not show differences between control and RNAPII degradation data at these loci (**Figure 13B**). The insulation plots, as described in the method part of the thesis, show the degree of segregation of the adjacent chromatin regions. Moreover, CTCF anchors were indicated that exhibit strong insulation capacity, whereas, transcription anchors and polycomb anchors were indicated that exhibit relative weaker insulation capacity than CTCF (Rao et al., 2014; Hsieh et al., 2020; Krietenstein et al., 2020; Rhoes et al., 2020). In **Figure 13B**, the degrees of insulation are exhibited as the blue/purple signals, which present fewer interactions between the adjacent chromatin regions.

In addition, although 3393 loops shared by control and RNAPII degradation show stronger loop signal in the latter, their insulation profiles remain almost the same (**Figure 13B**). The **Figure 13A** shows average loop intensity of whole genome, which is presented at the center bin in the square. Importantly, the bins around the center bin are also informational as they represent the stability of the loop. Moreover, ~1000 loops were weakened or lost upon RNAPII degradation. The insulation plots do not show differences between control and RNAPII degradation data at these loci as well (**Figure 13B**). Taken together, control and RANPII degradation *in situ* Hi-C data do not display changes at the level of loop anchor insulation.

In addition, the experimental biases of Hi-C, such as the efficiency of restriction enzyme cutting and proximity ligation, need to be considered when analyzing data. Thus, a shared loop by control and RNAPII depletion means that the distance between the coordinates is tolerated within three bins. The strategy is adopted throughout the thesis, unless otherwise stated.

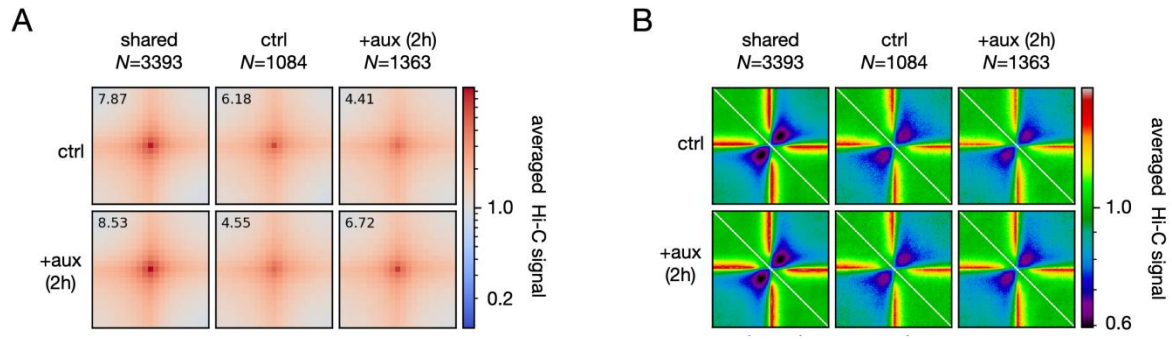


Figure 13. Loop signal and insulation plots. (A) The plots show averaged loop signal of the three loop categories, which 3393 (75%) loops are both exhibited in control and RANPII degradation samples, 1084 loops only present in the control sample, and 1363 loops only present in the RANPII degradation sample. Plots were generated with 10kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (B) As in panel A, but for averaged anchor insulation signal. Plots were generated with 5kbp resolution. The interactions among +-500kbp regions of loop anchors were selected to show on the squares.

4.1.3.2 The architectural CTCF loops dominate the loop layer

To further investigate the differences of loop formation in the presence or absence of RNAPII, I first did differential CTCF peaks analysis using CTCF CUT&Tag sequencing data. The result shows 67 up-regulated (1.6x increasing) and 157 down-regulated (64% decreasing) the peaks upon RNAPII degradation (**Figure 14A**). I used data of degrading for 14h because CTCF data of degrading for 2h is unavailable. The results suggest that RNAPII depletion does not significantly affect chromatin-bound CTCF (**Figure 14A**). Furthermore, I coordinated loop anchors with CTCF peaks to obtain CTCF loops. The results show that the CTCF occupancy at loop anchors are similar between control and RNAPII degradation samples (**Figure 14B)**. Briefly, in both control and RNAPII depletion data, ~86% loops are at least single CTCF anchored, and 4% more pair CTCF anchored loops upon RNAPII degradation **(Figure 14B)**. Together with **Figure 13,** the results indicate that the loop layer of genome organization is not significantly and genome-wide altered in the absence of RNAPII

using *in situ* Hi-C data with 5-/10-kbp resolution.

In addition, when dealing with CTCF data, the coordinates of the loops and the CTCF peaks need to fall in the same bin.

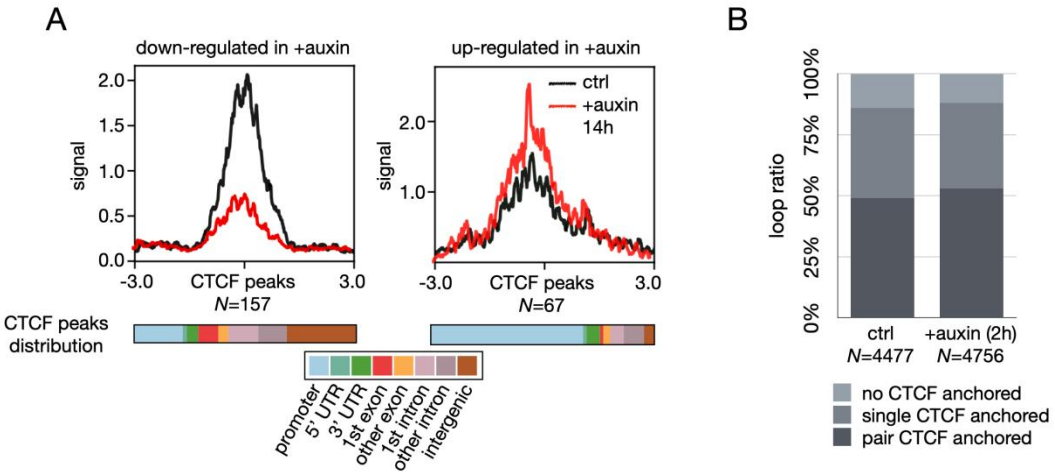


Figure 14. Analysis of CTCF CUT&Tag sequencing data. (A) The plots show averaged CTCF signal (+-3 Kbp) of down-regulated (left) peaks and up-regulated CTCF (right) peaks upon RANPII depleting for 14 h. And the distribution of differential CTCF peaks on the genome (lower). Data was normalized by the CPM method, and pvaule <0.01 was used in the analysis. (B) The plots show loop ratios of pair CTCF anchored, single CTCF anchored, and none CTCF anchored loops in control and RNAPII depletion data.

Furthermore, I examined whether the RNAPII depletion-specific loops (~1000) prefer CTCF at loop anchors than control (**Figure 15A**). Notably, the results show that shared loops by control and RNAPII degradation consist of more pair CTCF anchored loops (~3000) than sample-specific groups (54%>35%>43%; **Figure 15A**). Considering that the shared loops count for 75% of total loops, and chromatin-bound CTCF are not significantly affected upon RNAPII degradation. Collectively, these results indicate that transcription-independent, architectural CTCF loops dominate the loop layer of genome organization in the presence and absence of RNAPII. It is interesting to notice that two classes of loops in the loop layer of genome organization. One is transcription-independent, frame-like, architectural loops.

The other one is transcription-dependent, intra-domain/intra-frame, dynamically-regulated E-P loops (Hsieh et al., 2020; Krietenstein et al., 2020).

4.1.3.3 A pattern of rewired loops in the absence of RNAPII

Furthermore, the abundance of chromatin-bound CTCF, and minor alterations of it upon RNAPII degradation, prompted me that newly formed loops in the absence of RNAPII probably reuse anchors/CTCF rather than newly formed on chromatin. After analyzing, I found that 1) 48% (1249/2607) anchors of the depletion-specific loops are also loop anchors in the control; 2) 28% anchors (725/2607) of the depletion-specific loops are not loop anchors in the control but with CTCF binding (**Figure 15B**). Interestingly, only ~45% chromatin-bound CTCF are utilized as loop anchors in the control (18692/41226), which offers a flexibly rewiring possibility of loop formation. So, a pattern of rewired loops is involved in the loop formation in the absence of RNAPII.

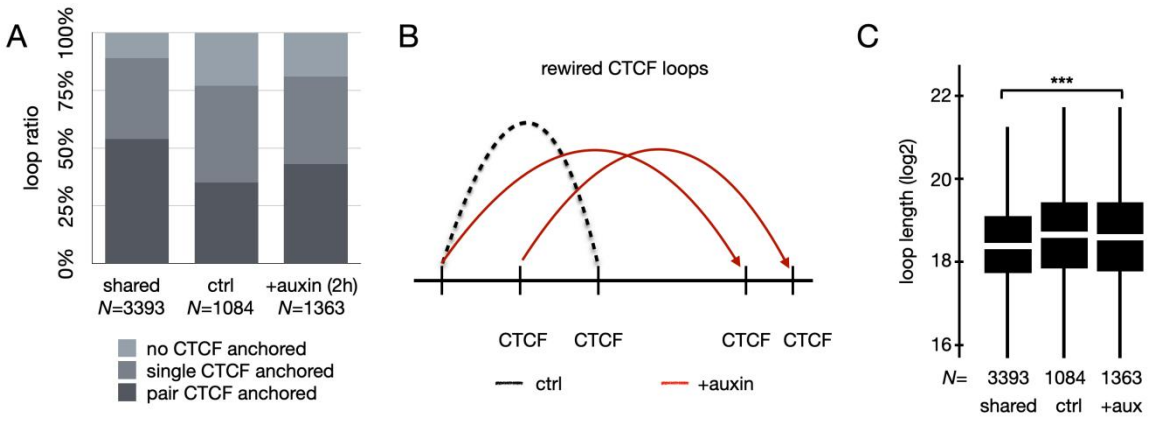


Figure 15. CTCF occupancy at anchors, the pattern of rewired loops, and loop length distributions. (A) The plots show loop ratios of pair CTCF anchored, single CTCF anchored, and none CTCF anchored loops in the groups of shared loop, control-specific loop, and depletion-specific loop. (B) The schematic shows the pattern of rewired loops upon RNAPII depletion, which anchors (CTCF or non-CTCF) of the depletion-specific loops are also anchors/CTCF in the control. (C) The loop length plots show loop length (log2) distributions in the panel (A).

Importantly, cohesin extrusion is important to rewire new loop formation upon RNAPII

depletion. However, how RNAPII affects cohesin is unclear. It was reported that RNAPII impedes and pushes condensin in bacteria (Brandão et al., 2019). It was reported that RNAPII affects cohesin positioning in eukaryotic cells (Busslinger et al., 2017; Heinz et al., 2018). It is unclear whether RNAPII impedes cohesin extrusion as well in my cell line. Thus, I assumed that the depletion-specific loops should be longer if RNAPII impedes cohesin in the DLD-1-RPB1-mAID cell line. Notably, the depletion-specific loops are longer than the shared loops, whose length medians are 400Kbp > 340Kbp (**Figure 15C**).

Taken together, the rewired anchors and the longer loops with enhanced signals upon RNAPII depletion suggest that RNAPII can impede cohesin extrusion in the DLD-1-RPB1-mAID cell line.

4.1.3.4 Investigating polycomb-mediated loops in the absence of RNAPII

In addition, it was reported that polycomb can mediate loop formation upon cohesin degradation (Rhoes et al., 2020). I further examined whether my observations of the gained loops upon RNAPII degradation were mediated via polycomb. To further estimate effects of polycomb on the depletion-specific loops, I first did differential H3K27me3 peaks analysis using H3K27me3 CUT&Tag sequencing data. The results show no significant differences between the control and the RNAPII depletion data with the parameter pvalue <0.01. I used data of RNAPII depleting for 14h because the data of depleting for 2h is unavailable. In addition, I annotated the distributions of H3K27me3 peaks on genome. Then the results show that only a small part (~13%) H3K27me3 locate at promoter regions (**Figure 16A**). Then, I coordinated anchors of the depletion-specific loops with the H3K27me3 peaks. I found 598 loops, which are at least single H3K27me3 anchored. However, “H3K27me3 only” loops without anchoring CTCF are only 108. The result suggests polycomb-mediated loops only take a small part (~8%) of the newly formed loops upon RNAPII depletion (**Figure 16B**).

In addition, when dealing with H3K27me3 data, the distances between the coordinates of

the loops and the H3K27me3 peaks are tolerated within two bins.

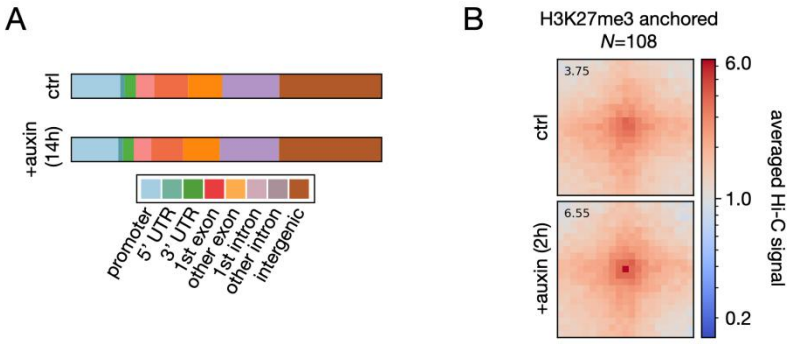


Figure 16. Analysis of H3K27me3 CUT&Tag sequencing data. (A) The plots show distribution of H3K27me3 peaks on the genome. (B) The plots show loop signal pattern of H3K27me3 mediating loops. The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of anchors were selected to show on the squares.

4.1.4 Investigating E-P interactions is ineffective using *in situ* Hi-C data

I further analyzed whether enhancers/promoters (E/P) are favored around anchors of control-specific loops. The ratios of E/P loops are pretty similar between the control and the RNAPII degradation data (~80%; **Figure 17A**). I further analyzed whether the anchors of depletion-specific loops around E/P is a result of CTCF enrich at E/P regions. Importantly, I found only 64 depletion-specific loops are “transcription only” without anchoring CTCF. Furthermore, I found that only 55 control-specific loops are “transcription only” without anchoring CTCF. Taking together that CTCF is a transcription factor, these results indicate that CTCF is intrinsically enriched at open-chromatin regions. Moreover, most of the “transcription only” loops are only algorithm detectable (**Figure 17B**), so exploring “transcription only” loops is impossible here. Thus, my *in situ* Hi-C data is not efficient enough in detecting transcription-level architecture.

In addition, when dealing with E/P data, the distances between the coordinates of the loops

and the E/P peaks are tolerated within two bins.

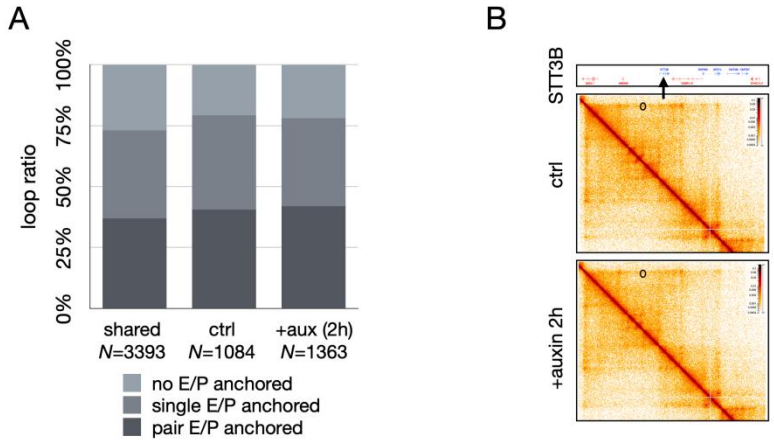


Figure 17. The E/P occupancy at loops anchors, and an example of "transcription only" loop. (A) The plots show loop ratios of pair E/P anchored, single E/P anchored, and none E/P anchored loops in groups of shared loop, control-specific loop, and depletion-specific loop. (B) The plots show a "transcription only" loop around *STT3B* location. The plots were generated with 5kbp resolution.

4.1.5 Short summary of 4.1

Taken together, the cell population in *situ* Hi-C data show only minor effects on the loop layer of genome organization upon RNAPII degrading for 2h. New loops are formed upon the depletion. These gained loops are longer and the loop signals are strong. Moreover, ~75% loops are architectural loops in DLD-1-mAID-RPB1 cell line, and they are dominated by CTCF and cohesin in the presence and absence of RNAPII. It is interesting that CTCF intrinsically prefers open-chromatin regions. Furthermore, RNAPII is suggested that impedes cohesin extrusion in the DLD1 cell line due to loops with longer size, strong signal, and the pattern of rewired anchors upon RNAPII depletion. Furthermore, polycomb-mediated loop formation upon the depletion may occur in my data, but they only take a minor part. However, no changes at the level of loop anchor insulation upon the depletion. Last, "transcription only" loops are too few to decipher features of E-P interactions. Also, results show that Hi-C data is not efficient enough in detecting transcription-level architecture.

4.2 RNAPII impedes cohesin extrusion in the G1 phase

To better investigate the roles of RNAPII in genome organization, I optimized the treatment time to 14 hours, whereby ~90% RNAPII was depleted (**Figure 10**). Moreover, only G1-sorted cells were collected to perform Hi-C experiments due to RNAPII primarily working in the G1 phase. In addition, one more Hi-C sample was made this time except control and auxin-treatment, which triptolide was added in the last two hours upon auxin, to strictly block functional RNAPII, and also to be a flexible replicate. Surprisingly, more apparent changes were recognized this time in many different layers of genome organization.

4.2.1 Overview of the genome organization in the absence of RNAPII

First, I generated interaction decay plots, in which contact frequency is plotted as a function of genomic distance, and their derivative slope plots, in which polymer features of the whole-genome organization are shown. In both, signals are slightly altered as a result of ~90% RNAPII depletion (**Figure 18**). However, the decay plot is sensitive enough with even minor changes to show distinctions of samples. The features of decay plots has been shown in many independent studies (**Figure 18**). For example, the cohesin degradation Hi-C shows lost interactions within ~2Mbp while gained longer-distance interactions due to enhancing compartmentalization (Rao et al., 2017). Moreover, CTCF degradation and WAPL degradation show gained interactions from ~500Kbp to ~10Mbp due to keeping cohesin extruding on chromatin (Haarhuis et al., 2017; Wutz et al., 2020). As shown from the RNAPII depletion data, decay plots show lost interactions within ~1Mbp, which is considered as losing transcription-mediated interactions (**Figure 18**). Whereas, longer-range interactions are gained upon the depletion, which may be because of enhancing compartmentalization of inactive regions.

In addition, the information of distances longer than 10Mbp in the interaction decay plots

were ignored due to only being suitable for longer chromosomes.

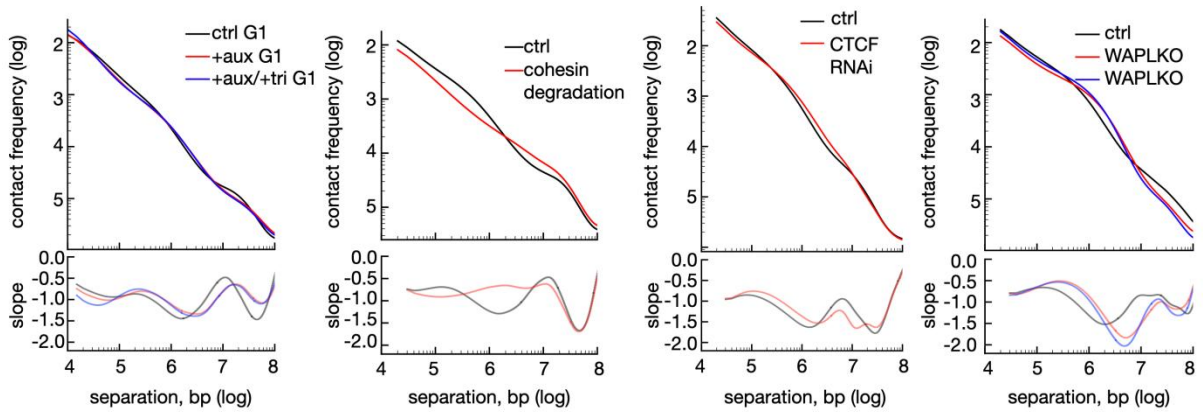


Figure 18. Interaction decay plots in independent studies. The interaction decay plots show Hi-C interaction frequency (log) as a function of genomic distance (log), and their derivative slope plots of decay plots (lower) show polymer-like behaviors of genome folding. From left to right are decay plots of my data (with 5kbp resolution), cohesin degradation data (Rao et al., 2017; with 10kbp resolution), CTCF RNAi data (Wutz et al., 2020; with 10kbp resolution), and WAPL knockout data (Haarhuis et al., 2017; with 10kbp resolution).

Second, A/B compartmentalization saddle plots, which show organization and segregation of active and inactive Mbp-size regions, are similar between control and the RNAPII depletion data (**Figure 19**). Interestingly, slightly decreased A-A interactions and minor enhanced B-B interactions in the depletion data are shown (**Figure 19**). Furthermore, increased H3K27me3 signals and decreased H3K27ac signals were observed with immunofluorescence experiments upon the depletion (Zhang et al., 2021). Also, reduced chromatin accessibility was shown upon the depletion (measured by ATAC-seq data; Zhang et al., 2021). Thus, independent experiments are in line with each other. The mechanism of compartmentalization formation is still poorly known, but forces of condensates are considered (Rada-Iglesias et al., 2018). It is interesting to recognize that active environment is weakened upon losing RNAPII, so interactions among compartment A are reduced in the G1 phase. While inactive environment is enhanced upon losing RNAPII, so interactions

among compartment B are increased in the G1 phase.

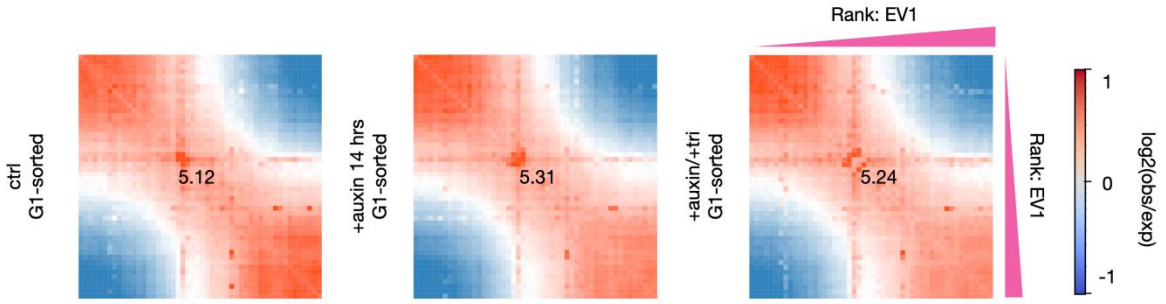


Figure 19. A/B compartmentalization saddle plots of G1-sorted samples. The A/B compartmentalization saddle plots show organization and segregation of active and inactive regions, in which signals at the left-upper represent B-B interactions and signals at the right-lower represent A-A interactions. The numbers in the center are compartment segregation strength. The plots were generated with 250kbp resolution.

4.2.2 Investigating the loop layer upon RNAPII depletion

4.2.2.1 Gained loop signal and anchor insulation upon RNAPII depletion

Third, although comparable loop numbers were detected in the two RNAPII depletion samples, loops in G1 control are fewer (**Figure 20A**). Moreover, all the three G1-sorted samples contain significantly fewer loops than the “2h” samples. This observation suggests cell population data contains S/G2-specific and more stable loops than G1-sorted data.

Furthermore, I generated loop signal plots, loop anchor insulation plots, and loop length distribution plots (**Figure 20**). Notably, loop signal and loop anchor insulation are stronger upon RNAPII degradation (**Figure 20B** and **20C**). Moreover, ~799 gained loops upon the depletion are significantly longer than the shared loops, whose medians are 410kbp > 350kbp; **Figure 20D**). Interestingly, for G1-sorted samples, the increased loop signal and loop anchor insulation upon the depletion are significantly more obvious than the “2h” data.

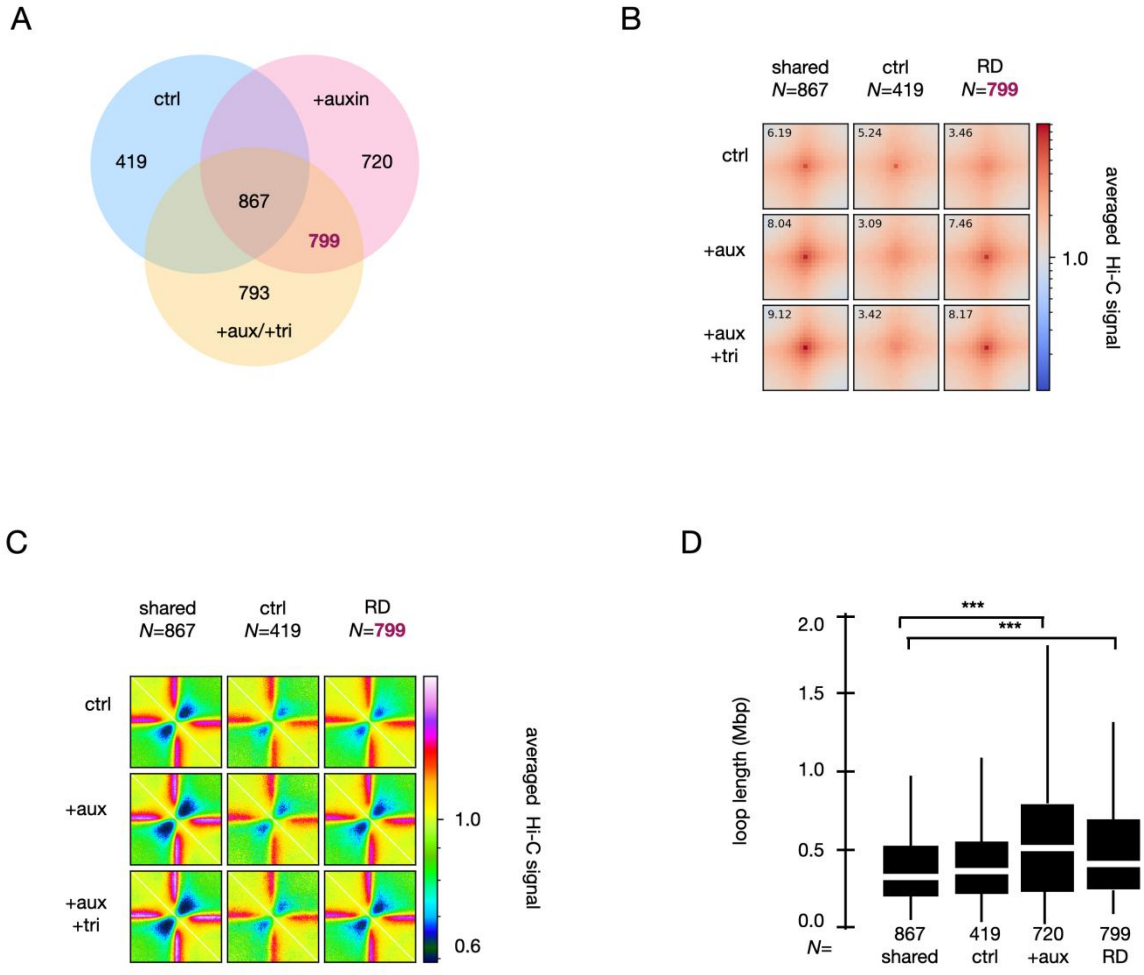


Figure 20. Loop numbers, loop signal, loop anchor insulation, and loop length distribution of G1-sorted samples. (A) The plots show loops called from G1-sorted samples, and they were classified as shared or specific loops. (B) The plots show averaged loop signal of the three loop categories, of which 867 loops present both in G1 control and RNAPII depletion samples, 419 loops only present in control-G1 sample, and 799 loops only present in the depletion samples. The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (C) As in panel (B), but for averaged loop anchor insulation signal. The plot was generated with 5kbp resolution. The interactions among +-500kbp regions of loop anchors were selected to show on the squares. (D) As in panel (B), but plots show loop length distribution.

4.2.2.2 Diffusing pattern emerges upon RNAII depletion during the G1 phase

Importantly, adjusting of the above mentioned strategy of coordinating between loops

shows interesting profiles. When I selected loop coordinates of the G1 control as a reference, and then aggregating the Hi-C signal of the depletion samples at these coordinates. The diffusing Hi-C signal emerges, which is not exhibited with the “2h” data (**Figure 21A**). Importantly, the diffusing pattern of loops means that highly-dynamic extrusion occurs upon RNAPII depletion. This result indicates that G1-sorted samples clearly show fine-scale dynamic extrusion process upon RNAPII depletion.

Furthermore, I checked CTCF occupancy at anchors of 867 shared loops that defined by the three bins tolerance strategy. Interestingly, 10% and 15% more loops are found that reaching CTCF upon RNAPII degradation, which is also not exhibited with the “2h” data (**Figure 21B**).

Taken together, these results indicate that during G1 phase, RNAPII impedes cohesin extrusion in the DLD1 cell line, and a dynamic-pattern of cohesin extrusion to reach CTCF is showed upon the depletion of RNAPII.

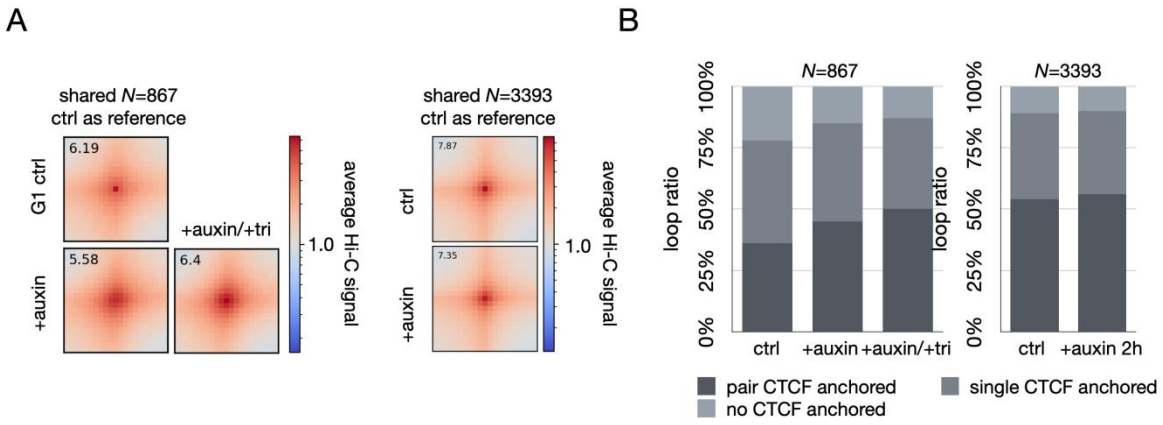


Figure 21. Diffusing pattern and CTCF occupancy at loop anchors. (A) The plots show averaged loop signal of 867 shared loops with coordinates of control-G1 as a reference, and the diffusing pattern is shown upon RNAPII degradation (left), but this is not exhibited in the cell population data (right). The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (B) The plots show loop ratios of pair CTCF anchored, single CTCF anchored, and none

CTCF anchored loops in the group of shared loop (left). Similar to the left, the “2h” cell population data is showing at the right.

Moreover, unlike the “2h” data, in which the depletion-specific loops exhibit a high anchor-rewired rate. The depletion-specific loops of G1 phase exhibit a relatively lower anchor-rewired rate, which is 25%. Collectively, the transcription-active G1 phase is thought to be the reason of lower anchor-rewired rate in the absence of RNAPII.

4.2.2.3 Investigating polycomb-mediated loops upon RNAPII depletion

In addition, auxin-specific (one of the two RNAPII depletion samples) loops exhibit a even longer loop length pattern than auxin/triptolide-specific loops, whose medians are 530 Kbp > 420 Kbp (**Figure 22A**). This result suggests that unclear mechanisms of loop formation are involved distinctly in the two RNAPII depletion samples. In addition, 163 and 140 “H3K27me only” loops without anchoring CTCF were found in auxin and auxin/triptolide data, respectively, which are considered as polycomb-mediated loops (**Figure 22B**). Interestingly, the loop length of these H3K27me loops are super long, whose medians are ~770Kbp and ~620Kbp, respectively (**Figure 22A**). In line with the “2h” data, polycomb-mediated loop formation may occur upon RNAPII depletion, but take a small part as well in the G1 phase.

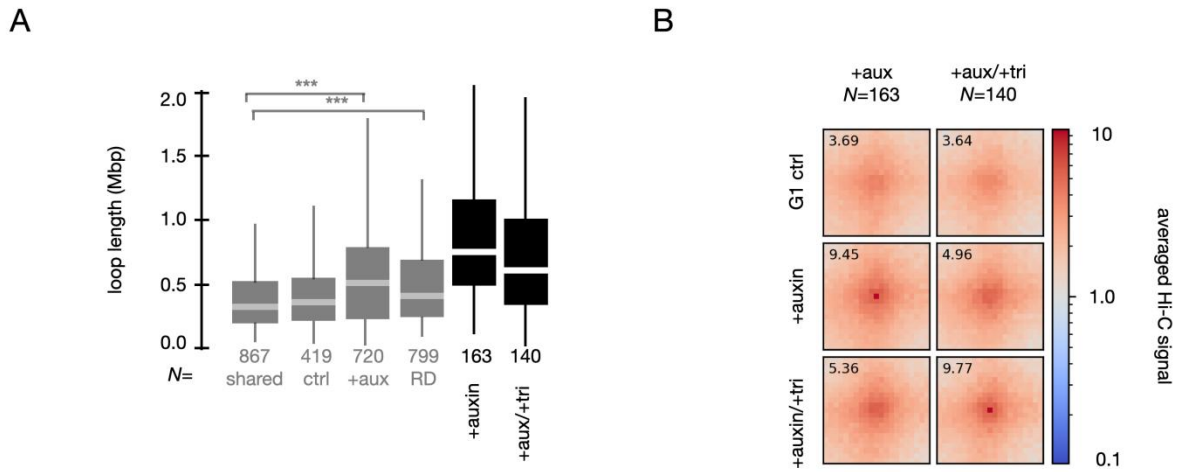


Figure 22. Polycomb-mediated loops in G1-sorted samples. (A) The plots show distributions of loop length of H3K27m3-anchored loops. (B) The plots show loop signal pattern of H3K27me3-mediates loops. The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares.

4.2.3 Investigating E-P interactions is ineffective using G1-sorted Hi-C data

In addition, anchors of the control-specific loops contain fewer CTCF caught my attention, which is a positive showing to find "transcription only" loops (**Figure 23**). Though only 48 "transcription only" loops are found in control-specific loop list, the ratio 11.5% (48/418) has improved compared with the "2h" data, which is 5%. While the "transcription only" loops without anchoring CTCF are too few to be explored. The loop signal is only algorithm detectable that is as similar as the "2h" data. So, study of E-P interactions is still impossible here.

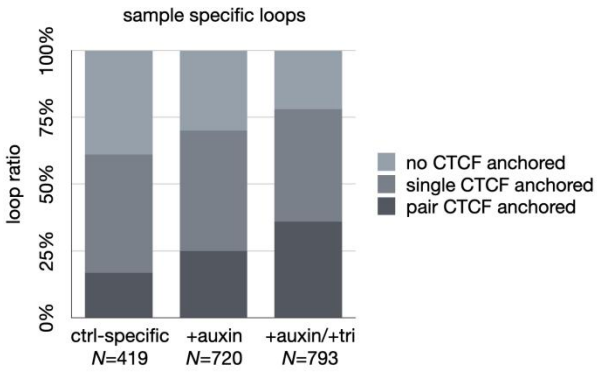


Figure 23. CTCF occupancy at loop anchors. (A) The plots show loop ratios of pair CTCF anchored, single CTCF anchored, and none CTCF anchored loops in gropus of control-G1-specific loop, auxin-specific loop, and auxin/triptolide-specific loop.

4.2.4 Short summary of 4.2

Collectively, distinctions of genome organization in the presence and absence of RNAPII are enhanced in G1 phase with sufficient degradation using *in situ* Hi-C data. Interaction decay

plots and saddle plots mildly show some lost/weaken transcription-mediated interactions and some gained interactions among inactive regions upon RNAPII depletion. However, in the loop layer, interesting and apparent profiles are shown. The longer loops with stronger loop and insulation signal, G1-sepcific diffusing pattern of loop signal, and increased CTCF occupancy at loop anchors upon RNAPII degradation, all indicate that RNPAII impedes cohesin extrusion in the G1 phase. Unfortunately, “transcription only” loops are still too few to be investigated.

4.3 Transcription-level genome organization in the DLD1 cell line

Previously, RNAPII impedes cohesin extrusion in DLD1 cell line was indicated using *in situ* Hi-C data. However, “transcription only” loops were few with only algorithm-detection signals, which is impossible to genome-wide explore the role of RNAPII in E-P interactions. While, a new member of the 3C family, Micro-C, came out recently, which is reported that capturing transcription-level interactions of genome organization (Hsieh et al., 2015; Hsieh et al., 2016; Hsieh et al., 2020; Krietenstein et al., 2020). Thus, I got G1-sorted Micro-C sequencing data from cells with RNAPII (control) and cells where RNAPII was degraded for 14 hours, whereby ~90% RNAPII was depleted (**Figure 10**).

4.3.1 Comparing between *in situ* Hi-C and Micro-C data

To further assess performances of Micro-C data, I first compared it with merged-G1 *in situ* Hi-C data. Surprisingly, fine intra-domain architectures were observed (**Figure 24A**), which did not capture by *in situ* Hi-C with a comparable sequencing depth (**Table 3**; **Table 4**). These observations indicate the high efficiency of Micro-C in studying fine-scale genome organization, which are most considered as transcription-mediated interactions (**Figure 24B**).

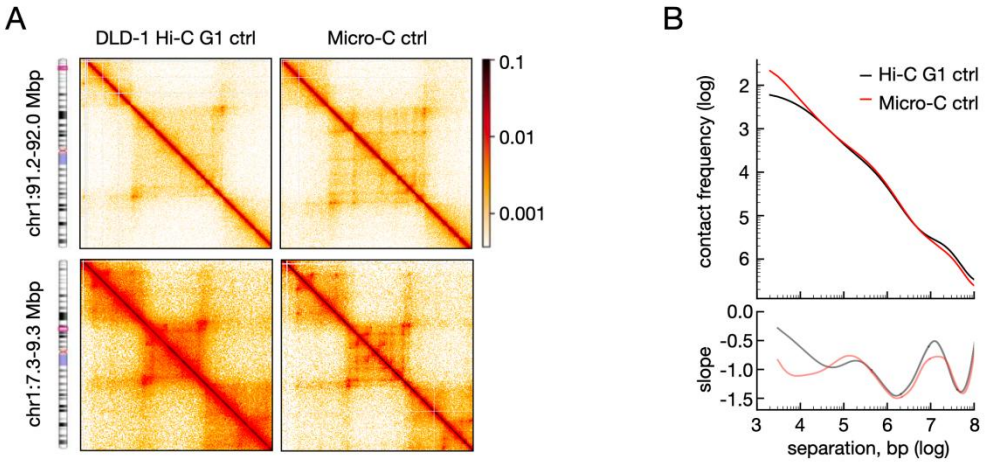


Figure 24. Comparisons between Micro-C and in situ Hi-C data. (A) The plots show that HiGlass visualized Micro-C-captured transcription-level architectures of genome organization around *CDC7* (upper) and *PARK7* (lower) regions. The upper region was visualized with 2kbp resolution, and the lower region was visualized with 5kbp resolution. (B) The interaction decay plots (upper) show Hi-C interaction frequency (log) as a function of genomic distance (log) in Hi-C (black) and Micro-C (red) data, and derivative slope plots of decay plots (lower) show polymer-like behaviors of genome folding. The plots were generated with 1kbp resolution.

Moreover, >80% (8212) Hi-C loops can be reproduced in Micro-C data, and >20,000 loops are "Micro-C only" (**Figure 25A**). Interestingly, the ~8,200 shared loops are more prominent than "Micro-C only" loops in Micro-C data, whose intensity values are 9.6 > 5.04 (**Figure 25B**). Notably, the ~8,200 more prominent loops are called from G1 phase. It was reported that loops (probably architectural loops) stably exist through G1 to G2 phase. However, RNAPII probably affects these loops in different cell cycle phases unequally.

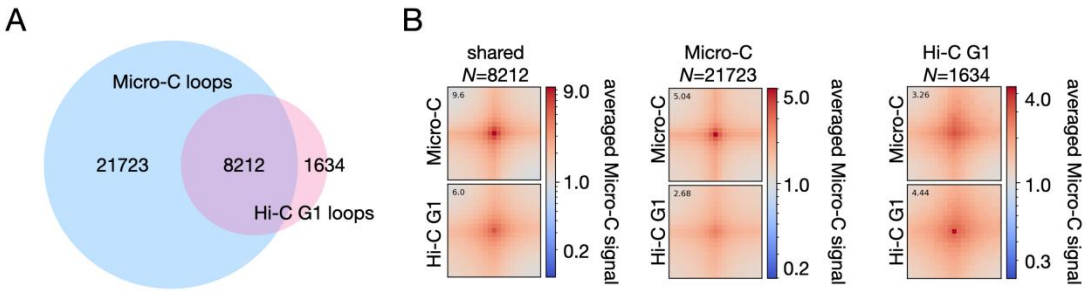


Figure 25. Comparisons between Micro-C and in situ Hi-C data. (A) The Venn plot shows numbers of loops that exist both in Hi-C and Micro-C data (8212), loops exist only in Micro-C (21723), and loops only exist in Hi-C (1634). (B) The plots show averaged loop signal in panel (A). The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares.

To further estimate the effects of cell cycle phases on loop exhibition, I first compared Hi-C data between the cell population and the G1 phase in the presence of RNAPII. Although the cell population Hi-C has twice fewer sequencing reads than the G1 Hi-C (**Table 3**), it shows stronger loop intensity than G1 data at the ~8200 loop loci (6.42 > 6.0; **Figure 26A**). This observation indicates that the ~8200 loops (probably most are architectural loops) are

enhanced in S/G2 phase. This observation may be as a result of reduced activity of transcription in S/G2 phase. Interestingly, the result is in line with enhanced loop signal upon RNAPII degradation.

Furthermore, I compared performances of the ~8200 loops between G1 Hi-C and Micro-C data in the absence of RNAPII. As shown in **Figure 26B**, a diffusing pattern emerges when using G1 Hi-C data of RNAPII degradation. Whereas, no diffusing pattern emerges when using G1-sorted Micro-C or cell population Hi-C data of RNAPII degradation (**Figure 26C**; **26D**). In addition, loop signal of G1-sorted Micro-C data is stronger than G1-sorted Hi-C data, which indicates Micro-C data has a higher effectiveness of cutting and ligating of the proximity (**Figure 26C**; **26D**).

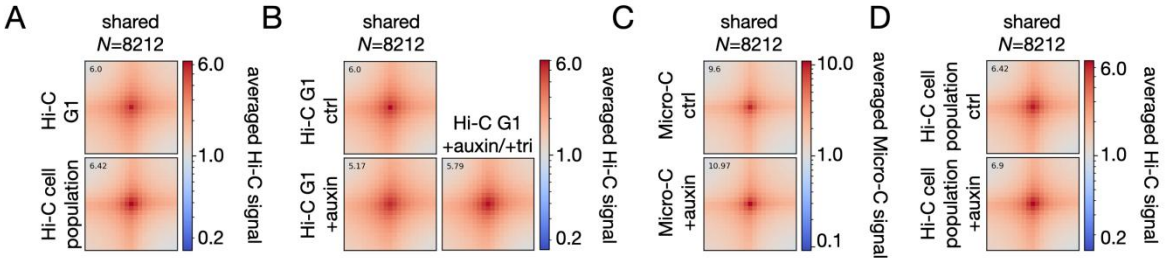


Figure 26. Effects of cell heterogeneity on Hi-C and Micro-C data. (A) The plots show loop signal patterns in G1 and cell population Hi-C data in the presence of RNAPII. The plots were generated with 10kbp resolution. The enrichment was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (B) The plots show loop signal patterns of G1 phase in Hi-C data in the presence and absence of RNAPII. The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bins. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (C) The plots show loop signal patterns of Micro-C data in the presence and absence of RNAPII. The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (D) The plots show loop signal patterns of cell population Hi-C data in the presence and absence of RNAPII. The plots were generated with 10kbp resolution. The enrichment value was calculated from the center bins. The interactions among +-100kbp regions of loop anchors were selected to show on the squares.

In addition, I found Micro-C data has a higher anchor usage rate than any previous Hi-C, which one anchor generates 0.696 loops (29935 loops with 42989 unique anchors; **Figure 27A**). This observation is in line with Micro-C contains more fine intra-domain loops. Moreover, interactions among inter-chromosomes are reduced in Micro-C data (**Figure 27B**). This is a result of abundant materials from MNase cutting decreasing background noise.

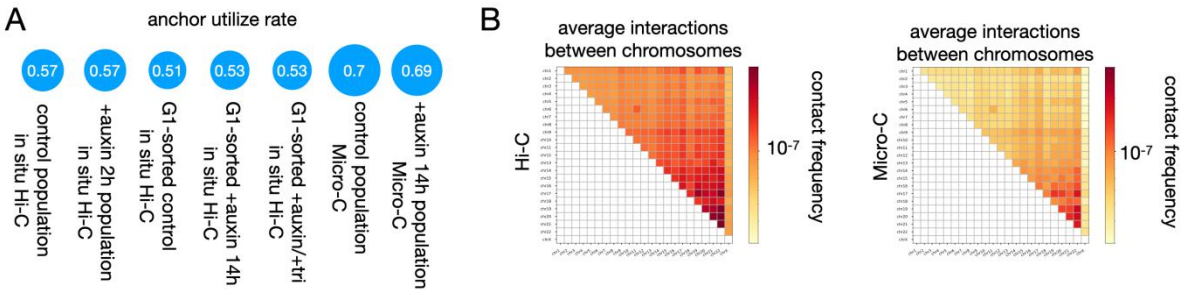


Figure 27. Anchor usage rates and interchromosomal interactions. (A) The plot shows anchor usage rates, which (loop number) / (unique loop anchor number) of *in situ* Hi-C and Micro-C data. (B) The plots show interactions between chromosomes.

4.3.2 Overview of the genome organization upon RNAPII depletion

Furthermore, A/B compartmentalization saddle plots, which show organization and segregation of active and inactive Mbp-size regions, which are a bit expansion from A to B upon RNAPII depletion (**Figure 28A**). Moreover, I generated interaction decay plots, in which contact frequency is plotted as a function of genomic distance, and their derivative slope plots, in which polymer features of the whole-genome organization are shown. In both, signals are not altered as a result of RNAPII degradation (**Figure 28B**).

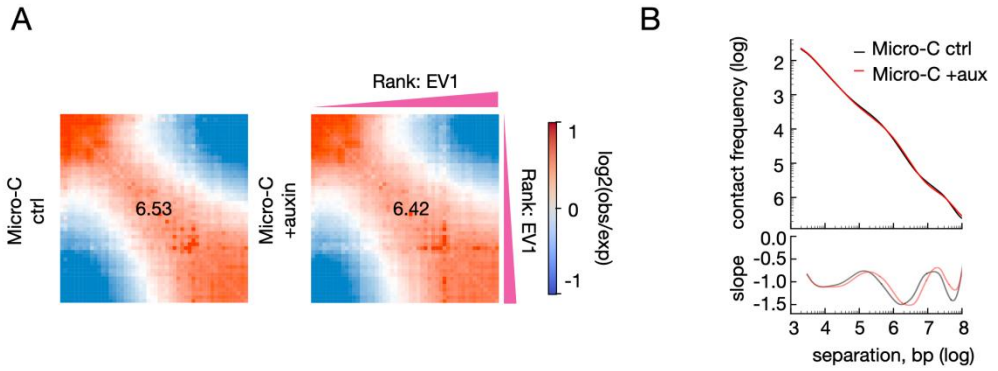


Figure 28. Saddle and interaction decay plots. (A) The A/B compartmentalization saddle plots show organization and segregation of active and inactive regions, which signals at the left-upper represent B-B interactions and signals at the right-lower represent A-A interactions. The numbers in the center are compartment segregation strength. The plots were generated with 250kbp resolution. (B) The interaction decay plots (upper) show Micro-C interaction frequency (log) as a function of genomic distance (log) in control (black) and RNAPII depletion (red) data, and the derivative slope plots of decay plots (lower) show polymer-like behaviors of genome folding. The plots generated with 1kbp resolution.

To further investigate how RNAPII affects loop formation using Micro-C data, I first classified loops between control and the depletion into the same-coordinate, the fine-shifting, and the sample-specific loops (**Figure 29A**; **Figure 29B**). Importantly, adjusting of the above mentioned strategy of coordinating between loops shows interesting profiles. The strategy is adjusted when analyzing Micro-C data. The same-coordinate loops between control and RNAPII depletion, and the nearby loops (fine-shifting) within three bins are extracted separately. Interestingly, the diffusing pattern emerges when using loops that are fine-shifting upon RNAPII depletion (**Figure 29B**). While, enhanced pattern emerges upon the depletion when using loops that are same coordinates (**Figure 29B**). These results indicate RNAPII impedes cohesin extrusion, and the architectural loops are enhanced upon RNAPII depletion. These results are in line with previous findings.

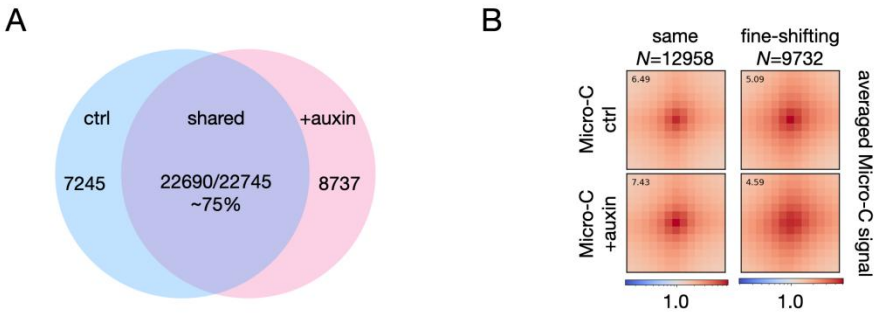


Figure 29. Numbers of loop, and loop signal plots. (A) The Venn plot shows numbers of loops that exist both in control and RNAPII depletion (count for ~75%, which 22690 in control and 22745 in the depletion), loops exist only in control (7245), and loops only exist in the depletion (8737). (B) The plots show loop signal patterns of 12958 same-coordinate loops and the 9732 fine-shifting loops of control and RNAPII depletion data. The plots were generated with 5kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-50kbp regions of loop anchors were selected to show on the squares.

Furthermore, I checked CTCF occupancy at the anchors of the same-coordinate and the fine-shifting loops, respectively. Interestingly, many fine-shifting loops reach CTCF upon RNAPII depletion (**Figure 30A**). Moreover, I found that fine-scale shifting loops consist of many transcription-anchored loops (>25%; **Figure 30B**). These results indicate that the same-coordinate loops between control and the depletion are architectural loops, while the fine-shifting loops are transcription-mediated loops.

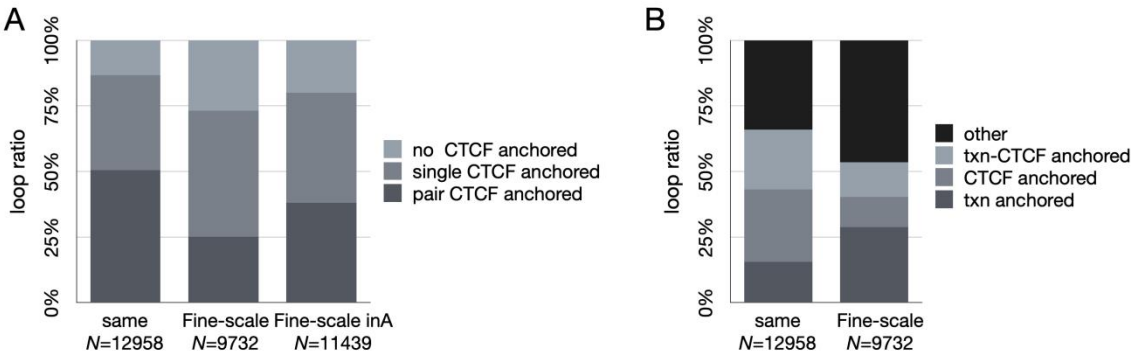


Figure 30. CTCF and E/P occupancy at loop anchors. (A) The plots show loop ratios of pair CTCF anchored, single CTCF anchored, and none CTCF anchored loops of the same-coordinate loops and the fine-shifting loops in control and RNAPII depletion, respectively. (B) The plots show loop ratios of

transcription anchored, CTCF anchored, and transcription-CTCF anchored loops of the same-coordinate loops and the fine-shifting loops. The other option (black marked) includes single transcription/CTCF anchored loops, and no transcription/CTCF anchored loops.

However, by checking Micro-C maps manually, I found lots of missing stripes upon RNAPII depletion, although loops are still there (**Figure 31**). Important, this phenomenon is not individual but relative common, which is in line with another study (Rhodes et al., 2020). I called stripes from Micro-C data, but unfortunately, these stripes are algorithm undetectable due to super high false positive rate. As shown in **Figure 31**, stripes, H3K27ac, and SMC1A are lost upon RNAPII depletion.

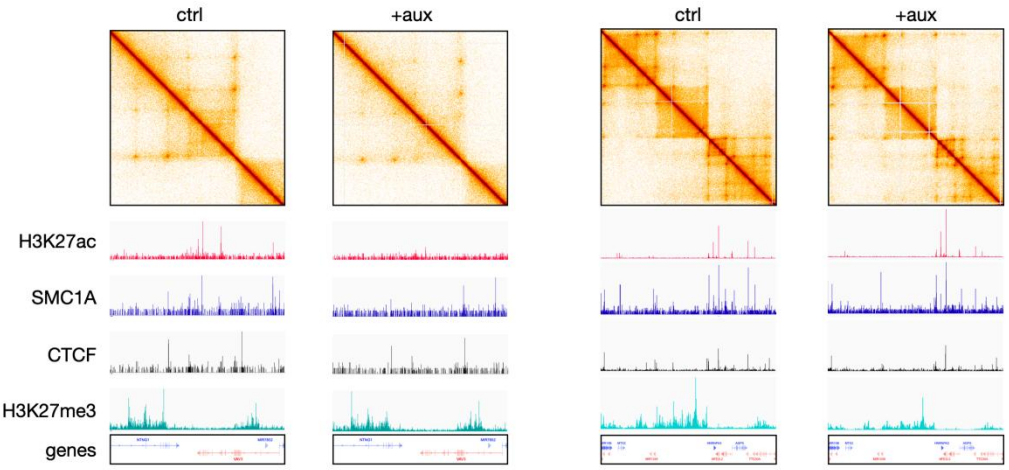


Figure 31. Missing stripes upon RNAPII depletion in Micro-C data. The pots show losing stripes in RNAPII depletion data around gene *VAV3, MTX2,* and *AGPS.*

4.3.3 Investigating E/P-anchored loops using Micro-C data

To further estimate performances of transcription-anchored loops in the absence of RNAPII, I extracted three kinds of loops from total loops of control, “transcription only” loops that paired E/P anchored without CTCF, “CTCF only” loops that paired CTCF anchored without E/P, and “transcription-CTCF” loops that paired E/P and CTCF anchored. Notably, “transcription only” loops show decreased Hi-C signal in average upon RNAPII depletion (4.64 < 5.26), whereas CTCF-related loops show increased Hi-C signal in average (7.52 >

7.35; 10.03 > 8.81; **Figure 32A**). Furthermore, in **Figure 32A**, “transcription only” loops show much weaker signals than CTCF-related loops (5.26 < 7.35 < 8.81). These two observations indicate that CTCF-related loops need to be filtered out when exploring roles of RNAPII in transcription-mediated loop formation. Furthermore, I extracted control-specific loops from the three loop categories, and generated loop signal plots to check performances of control-specific loops upon the depletion (**Figure 32B**). The results show CTCF-related control-specific loops are also lost or weakened upon RNAPII degradation in average. This indicates that not all CTCF-related loops are enhanced upon the depletion. ~18% (929/5195) “transcription-CTCF” and ~13% (239/1827) “CTCF only” loops are weakened or lost upon losing RNAPII. In addition, ~41% (929/2256) “transcription only” loops are significant weakened or lost upon RNAPII depletion. These observations indicate that RNAPII plays another role in genome organization besides antagonizing architectural loops.

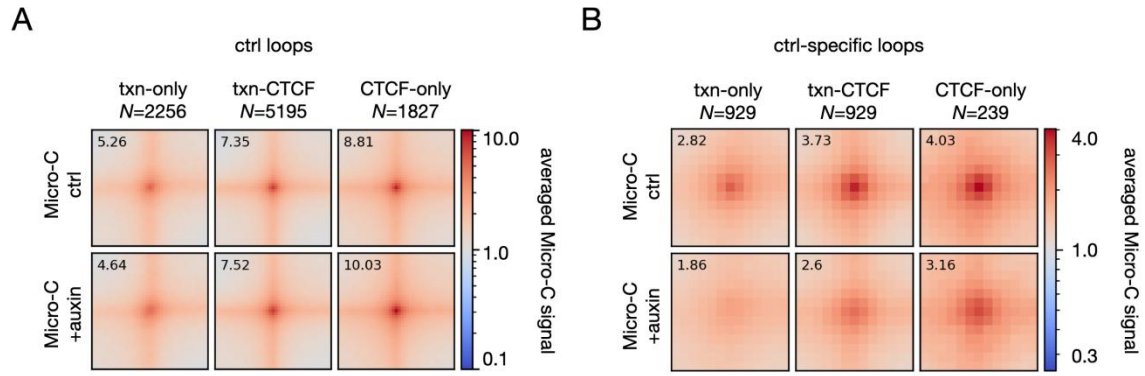


Figure 32. Plots of transcription-anchored and CTCF-related loops. (A) The plots show averaged loop signal of “transcription only” loops, “transcription and CTCF” loops, and “CTCF only” loops. The plots were generated with 5kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (B) The plots show control-specific signals of “transcription only” loops, “transcription and CTCF” loops, and “CTCF only” loops. The plots were generated with 5kbp resolution. The enrichment value was calculated from the center bin. The interactions among +-50kbp regions of loop anchors were selected to show on the squares.

To further explore performances of cohesin to loop formation in the presence and absence

of RNAPII, I first did differential SMC1A peaks analysis using SMC1A CUT&Tag sequencing data. SMC1A is a subunit of cohesin. The results show 1600 SMC1A peaks are down-regulated (~75% decreased) and 101 SMC1A peaks are up-regulated (~1.5X increased) upon RNAPII degradation (**Figure 33A**). Interestingly, most of the down-regulated SMC1A peaks locate at potential enhancer regions (**Figure 33A**). This observation suggests cohesin reduction at enhancers may be more severe. However, though the signals of shared loops are enhanced upon the depletion in the Micro-C maps, the signals of CTCF and SMC1A at the loop anchors are ~20% reduced in average (**Figure 33B**). This observation indicates that RNAPII depletion slightly, but genome-wide affects CTCF and cohesin on chromatin. However, the ~20% reduction of CTCF and cohesin does not affect exhibition of loops in the Micro-C maps.

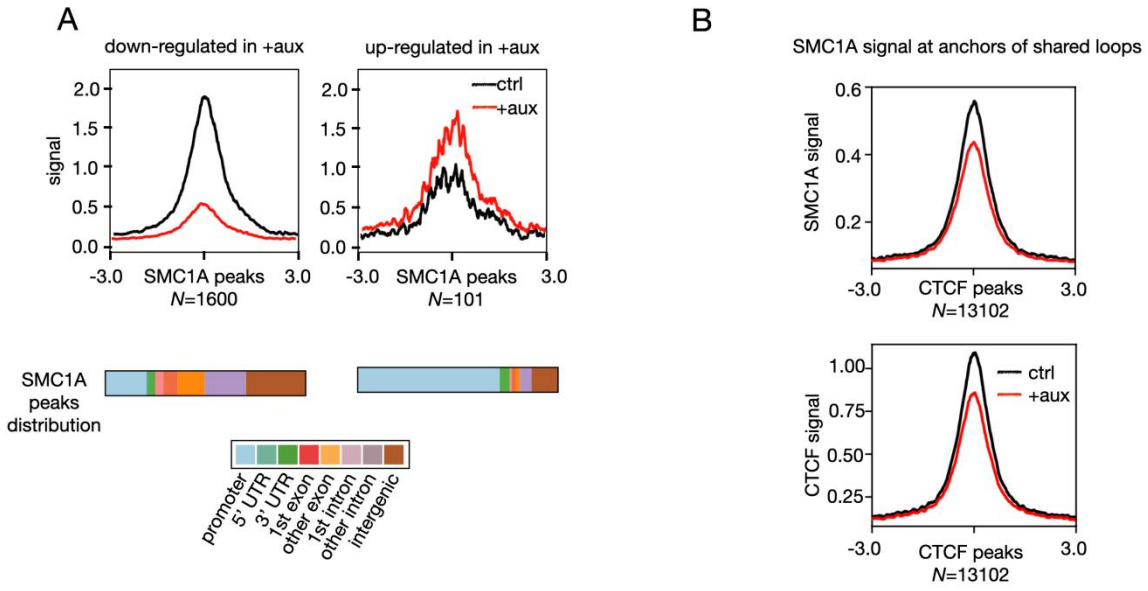


Figure 33. Differential cohesin peaks analysis. (A) The plots show averaged SMC1A signals of down-regulated (left) peaks and up-regulated (right) peaks upon RNAPII depletion, and the distribution of differential SMC1A peaks on the genome. Data was normalized by CPM method, and pvaule < 0.01 was used in the analysis. (B) The plots show averaged CTCF and SMC1A signals (+-3 Kbp) of anchors of shared loops.

To further explore functions of cohesin to E-P interactions, I checked the SMC1A signal at

anchors of the three loop categories, which are “transcription only,” “transcription-CTCF,” and “CTCF only”, respectively. Cohesin signals are decreased in all three categories (**Figure 34**), which is in line with **Figure 33B.** However, the signal is significantly decreased at anchors of “transcription only” loops (**Figure 34**). Together with the loop signal plots of the three categories (**Figure 32A**), the results indicate that ~20%-25% cohesin and CTCF loss do not affect loop exhibition in the Micro-C maps. While ~50% losing of cohesin does affect loop exhibition in Micro-C maps.

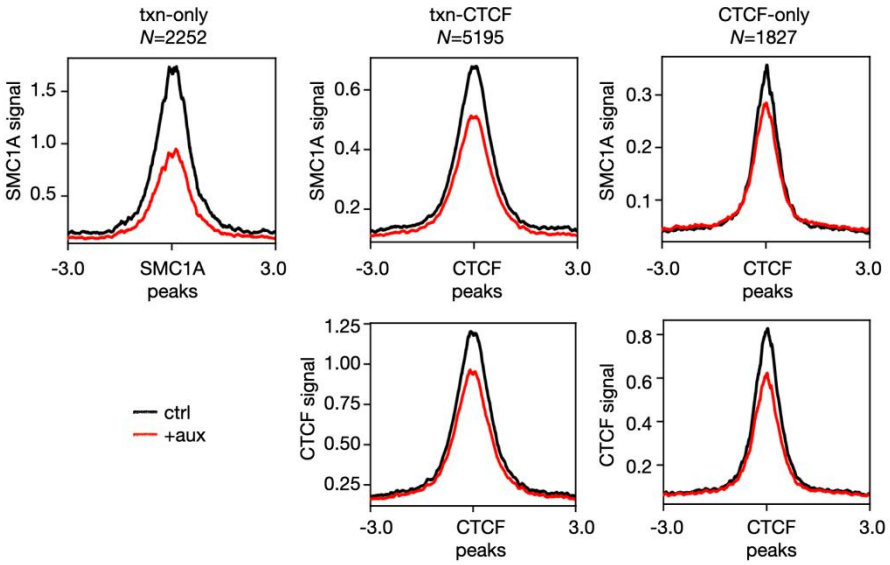


Figure 34. Decreased cohesin and CTCF signals at loop anchors. The plots show decreased cohesin and CTCF signals at anchors of “transcription only”, “CTCF only”, and “transcription and CTCF” loops upon RNAPII depletion.

Furthermore, I separately plotted CTCF and SMC1A signals around promoters and enhancers due to the significantly decreased SMC1A signals locating at potential enhancers (**Figure 33A**). To my surprise, CTCF and cohesin signals are reduced more at enhancers than promoters (**Figure 35**). Especially the cohesin signals at enhancers are ~50% down-regulated upon RNAPII degradation. Another surprising observation is H3K27ac signals show a similar reduction pattern as SMC1A that decreased more at enhancers than promoters (**Figure 35**). Thus, all observations indicate that enhancer regions are affected

more severe upon RNAPII depletion.

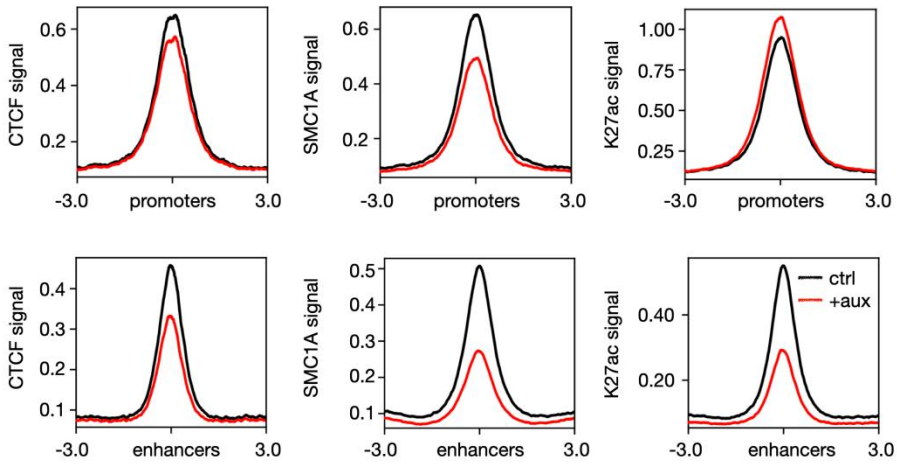


Figure 35. Decreased CTCF, SMC1A, and H3K27ac at promoters and enhancers. The plots show decreased signals of CTCF, SMC1A, and H3K27ac at promoters and enhancers upon RNAPII degradation, respectively.

In addition, I wonder why “transcription only” and enhancer-anchored loops are more sensitive upon RNAPII depletion. It may be RNAPII directly involved in E-P interactions, thus the depletion of it causes weakened or lost loops. However, how RNAPII contributes to E-P interactions is still unclear. Moreover, only ~40% “transcription only” loops are affected significantly in the absence of RNAPII. In addition, how RNAPII works with cohesin and CTCF locally for forming E-P loops is also unclear.

To answer these questions, I compared features between shared and control-specific loops of “transcription only.” I first checked promoter and enhancer occupancy on the two loop groups (**Figure 36A**). The results show control-specific loops consist of more enhancer-anchored loops (64% > 51% > 45%). I separately extracted the same-coordinates and the fine-shifting loops of the “transcription only” due to the different features between the two, as described the above (**Figure 29B**; **Figure 30**). Furthermore, I checked CTCF occupancy at the nearby-one bin of enhancer anchors. I did this to examine whether nearby-one bin CTCF affects stabilization of enhancer-anchored loops upon RNAPII

depletion. Notably, the results show the lost/weakened enhancer anchors upon RNAPII degradation apparently lack nearby-one bin CTCF (80% > 54% > 55%; **Figure 36B**). These observations suggest that RNAPII could act as boundaries to enhancer-anchored loops.

However, the RNAPII does not work as same as CTCF that directly interacts with cohesin. How RNAPII works as boundaries is poorly understood, here is just an assumption. RNAPII can impede cohesin extrusion but not absolutely block cohesin. Moreover, NIPBL and Mediator-complex were reported that directly interact with cohesin at enhancers (Kagey et al., 2010). It could be RNAPII interacts with NIPBL and Mediator-complex, thus indirectly affects cohesin upon its depletion. In addition, transcription-mediated condensates are formed at enhancers, in which RNAPII and Mediator-complex are involved (Kieffer-Kwon et al., 2013). Thus, it is possible that RNAPII impede cohesin as well at enhancers through condensates or physical binding on chromatin.

However, how CTCF at nearby-one bin of enhancer anchors stabilize “transcription-only” loops in the absence of RNAPII? The all “transcription-only” loops are weakened in average upon RNAPII depletion, though >80% the loops are non-significantly weakened (**Figure 32**). This observations indicate CTCF, which is at nearby-one bin of “transcription-only” loops, does not work as same as it is at architectural loops. Moreover, interactions among enhancers are shown in **Figure 36D**. Interestingly, signals of enhancer-anchored “transcription only” loops are almost lost upon RNAPII depletion, and signals of enhancer-anchored “transcription-CTCF” loops are weaken. It is still poorly understood how CTCF work locally for enhancer-anchored loops. Thus, RNAPII is required to form enhancer-anchored loop, not CTCF. However, why promoter-anchored loops are more stable than enhancer-anchored loops (**Figure 36C**)? This is a good question waiting to be answered. It may the mechanisms of promoter-anchored loops feature more similar to architectural loops, which are enhanced upon RNAPII depletion.

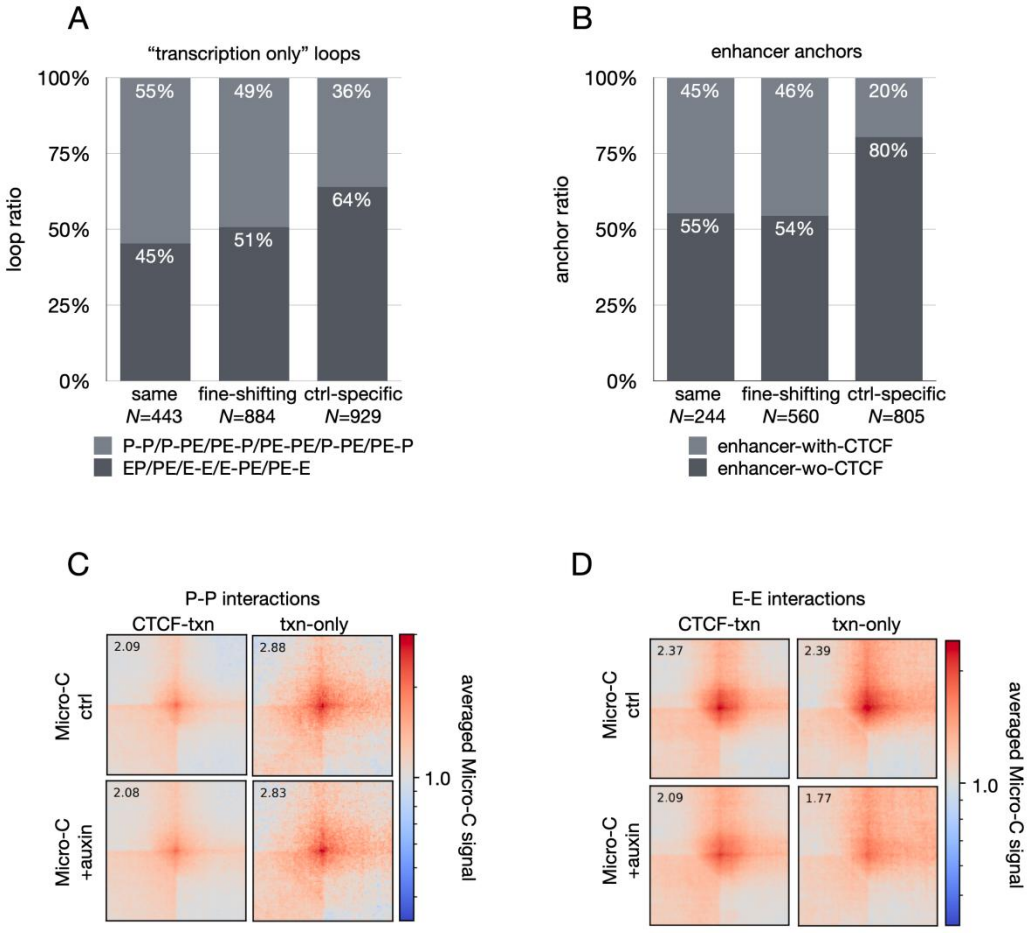


Figure 36. Analysis of promoter-anchored and enhancer-anchored loops. (A) The plots show loop ratios of P-P anchored and E-E/E-P anchored loops of "transcription only" loops. (B) The plots show loop ratios of CTCF occupancy at nearby-one bin of enhancer-anchored "transcription only" loops. (C) The plots show interactions of promoter-promoter in "transcription only" and "transcription-CTCF" loops. (D) Similar in panel (C), but showing interactions of enhancer-enhancer.

4.3.4 Investigating new loops upon RNAPII depletion using Micro-C data

4.3.4.1 Investigating polycomb-medaited loops in the absence of RNAPII

Moreover, polycomb-mediated loops are observed (**Figure 37)**. It was reported that cohesin can disrupt polycomb-mediated loops, while cohesin degradation leads to reformation of these loops (Rhodes et al., 2020). My observation is in line with the report. Polycomb-mediated loops are reformed upon RNAPII depletion which leads to cohesin

reduction from chromatin.

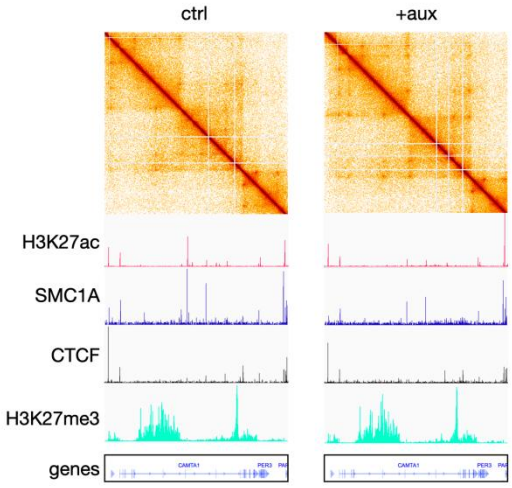


Figure 37. Polycomb-mediated loop formation upon RNAPII depletion. Plots show polycomb mediated loop formation upon RNAPII depletion.

4.3.4.2 Investigating new loops in the absence of RNAPII

Furthermore, I investigated the depletion-specific loops to check whether anything new with Micro-C data. I generated loop signal plots (**Figure 38**). The control-specific loops are unique but relatively weaker (4.19 < 7.71), and the depletion-specific loops are pretty strong but weaker than the shared loops (6.31 < 9.55). Then, stronger insulation than control at the depletion-gained loop loci are shown in **Figure 38B**. Moreover, the depletion-specific loops are significantly longer than the other loop categories (**Figure 38C**), whose medians (same-coordinates: 260kbp; fine-shifting: 245kbp; control-specific: 195kbp; depletion-specific: 460kbp). Collectively, these observations are in line with previous results using *in situ* Hi-C data.

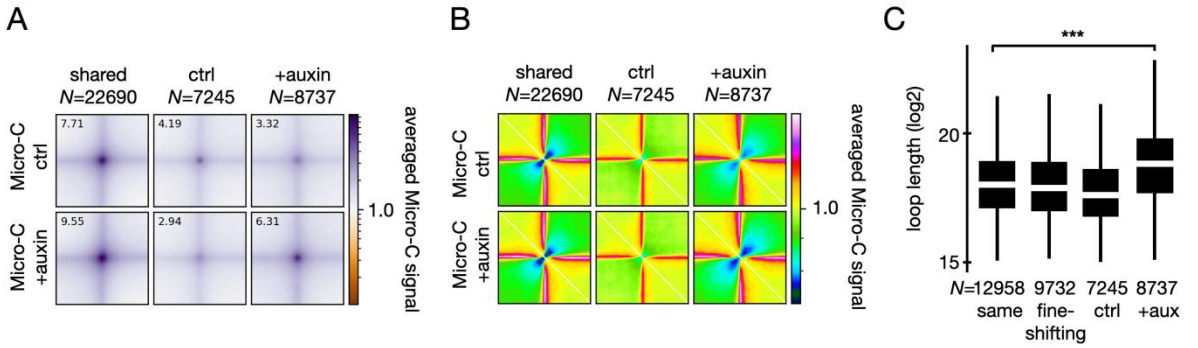


Figure 38. New loops upon RNAPII depletion. (A) The plots show averaged loop signals of the three loop categories, of which shared loops are present both in control and RNAPII depletion samples, loops only present in control, and loops only present in RNAPII depletion. The plots were generated with 5kbp resolution. The enrichment was calculated from the center bin. The interactions among +-100kbp regions of loops anchors were selected to show on the squares. (B) As in panel A, but for averaged anchor insulation signals. The plot was generated with 5kbp resolution. The interactions among +-500kbp regions of loops anchors were selected to show on the squares. (C) Loop length plots show loop length (log2) distributions.

In addition, the anchor-rewired rate of the depletion-specific loops reaches 57%. Moreover, results show that the depletion-specific loops contain obviously more CTCF at anchors than control-specific loops, especially pair anchored CTCF ratio is 14% higher (**Figure 39A**). Next, I extracted rewired and new-formed (non-rewired) anchors from depletion-specific loops, and then generated insulation plots of them (**Figure 39B**; **Figure 39C**). To my surprise, rewired anchors show much stronger insulation than new-formed anchors (**Figure 39B**). Indeed, new-formed anchors seem to contain the bundle pattern within 700kbp, which is considered as polycomb-mediated loops (**Figure 39C**). I soon checked H3K27me3 occupancy at the anchor loci, but results show that rewired and new-formed anchors of depletion-specific loops contain similar occupancy rates, which is ~25%. As well as, enhancer and promoter occupancy rates are similar, which rewired anchors are 76% and newly-formed anchors are 69%. However, CTCF occupancy at the rewired anchors is significantly higher than at the new-formed (~70% > ~36%), at least it is in line with the stronger insulation profile (**Figure 39C**). Last, 837 newly formed, H3K37me3-anchored loops without CTCF were found using Micro-C data in the absence of RNAPII.

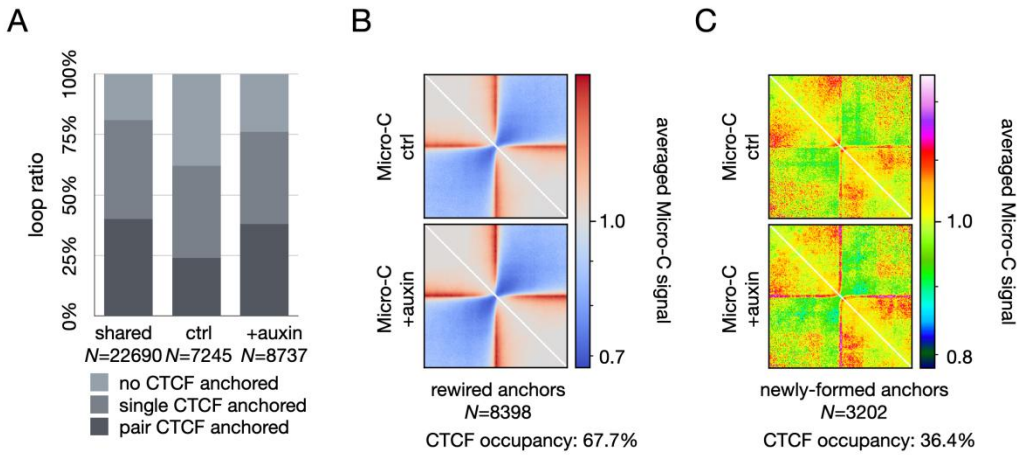


Figure 39. Analysis of anchors of RNAPII depletion specific loops. (A) The plots show loop ratios of pair CTCF anchored, single CTCF anchored, and none CTCF anchored loops of the three loop lists. (B) The plot shows insulation of rewired anchors of the depletion-specific loops. The plots were generated with 5kbp resolution. The interactions among +-500kbp regions of loops anchors were selected to show on the squares. (C) The plots show insulation of newly-formed anchors of the depletion-specific loops. The plots were generated with 5kbp resolution. The interactions among +-500kbp regions of loops anchors were selected to show on the squares.

In addition, I analyzed where rewired and new-formed anchors locate in sub-compartments in the presence and absence of RNAPII. Interestingly, the result show that some anchors of the depletion-specific loops were re-clustered to A2.2 and B2.1 upon RNAPII depletion (**Figure 40**), which suggests some loop-domain regions are clustering to facultative chromatin upon RNAPII depletion.

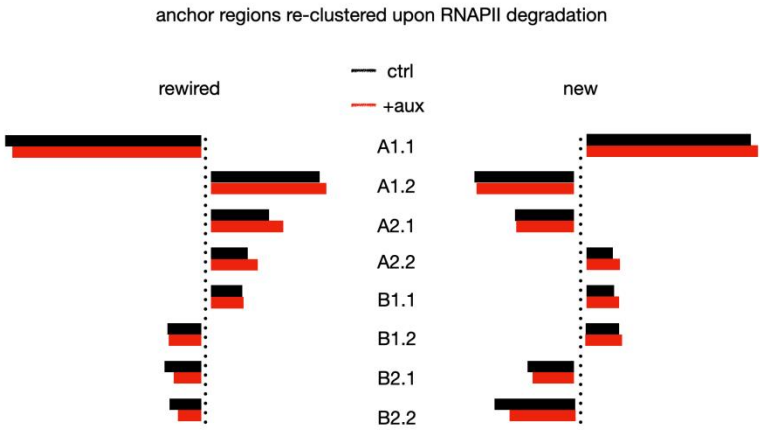


Figure 40. Re-clustering of the depletion-specific loop upon RNAPII depletion. The plots show anchors of depletion-specific loops are re-clustered to different sub-compartments upon RNAPII degradation.

4.3.5 Short summary for 4.3

The understanding of the loop layers of genome organization is substantially updated using Micro-C data. Micro-C data and *in situ* Hi-C data indicate the roles of RNAPII on genome organization. First, RNAPII impedes cohesin extrusion. Moreover, cell population Hi-C has stronger loop signals than G1-sorted Hi-C with twice fewer sequencing reads due to S/G2 loops may be transcription-independent. This observation is in line with loops are enhanced upon RNAPII depletion. Notably, Micro-C can capture transcription-level architecture, which *in situ* Hi-C can not. The other role of RNAPII in genome organization was indicated, which RNAPII is required in E-P loop formation. Moreover, RNAPII depletion affects cohesin and CTCF on chromatin with an average degree of ~20%-25%, and this degree of reduction does not affect the loop exhibition in the Micro-C maps. While, when the degree of losing cohesin increase to ~50%, obviously decreased loop signals are observed. In addition, cohesin on chromatin is affected more severely at enhancer than promoter in the absence of RNAPII.

4.4 Roles of RNAPII in genome reorganization during mitosis to G1 transition

It was reported that many E-P interactions were formed before CTCF-cohesin loops during mitosis to G1 transition (Zhang et al., 2019). Thus, what are roles of RNAPII to genome reorganization during mitosis to G1 transition? To further explore this question, I analyzed *in situ* Hi-C data from synchronized reentry G1 cells in the presence and absence of RNAPII.

4.4.1 Overview of genome reorganization in the absence of RNAPII

Notably, inter-chromosomal interactions are increased upon RNAPII depletion during mitosis to G1 transition (**Figure 41A**). This observation suggests that chromosome territories were affected in the absence of RNAPII during mitosis to G1 transition. Importantly, intra-chromosomal interactions are decreased, which is a significant alteration upon RNAPII depletion (**Figure 41B**). Notably, A/B compartmentalization is weakened in the absence of RNAPII (**Figure 41C**). As shown in **Figure 41C**, organization and segregation of compartment A (euchromatin) and B (heterochromatin) are both affected upon RNAPII depletion, which indicates RNAPII is essential in compartmentalization during mitosis to G1 transition. Interestingly, these observations are much more apparent than any G1 or cell population data, which indicates RNAPII is very important and required in the genome reorganization of mitosis to G1 transition.

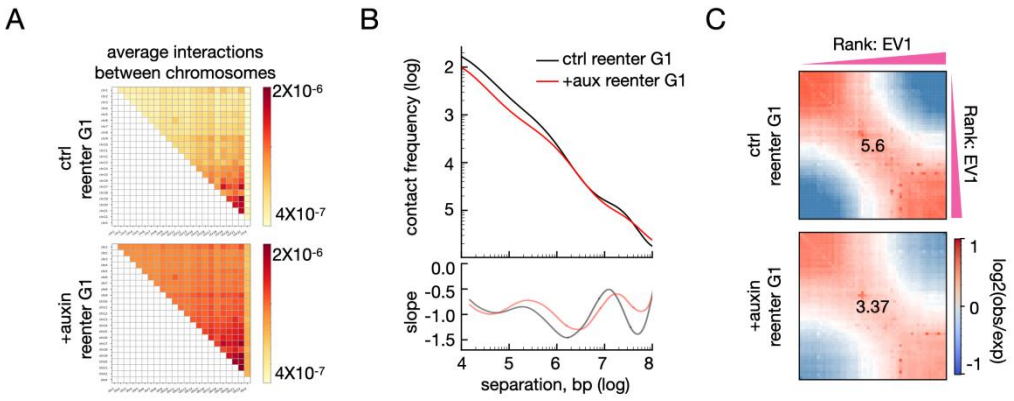


Figure 41. Apparent distinctions between control and RNAPII degradation data of genome reorganization. (A) The plots show average interactions between chromosomes. (B) Interaction decay plots (upper) show Hi-C interaction frequency (log) as a function of genomic distance (log) in control-reenter-G1 (black) and RNAPII depletion reentry G1 (red). The slope plots of decay plots (lower) show polymer-like behaviors of genome folding. The plots were generated with 5kbp resolution. (C) A/B compartmentalization saddle plots show organization and segregation of active and inactive regions, in which signals at the left-upper represent B-B interactions and signals at the right-lower represent A-A interactions. The numbers in the center are compartment segregation strength. The plots were generated with 250kbp resolution.

4.4.2 Effects of loop formation in the absence of RNAPII

Furthermore, I investigated how RNAPII affects loop formation during genome reorganization. Interestingly, 2443 shared loops are weakened in the absence of RNAPII (5.81 < 6.43; **Figure 42B**). Moreover, insulation strength of loop anchors are reduced upon RNAPII depletion (**Figure 42C**). While the depletion-specific loops are still longer than shared loops, whose medians are 610kbp > 420kbp (**Figure 42D**). In addition, I checked the diffusing pattern that is special to G1 data, and the results show signal is much weaker in the absence of RNAPII (**Figure 42E**). All observations suggest loop formation was affected upon RNAPII degradation during mitosis to G1 transition. Interestingly, unlike loop formation in the interphase, RNAPII especially affects loop formation in the reentry G1 phase.

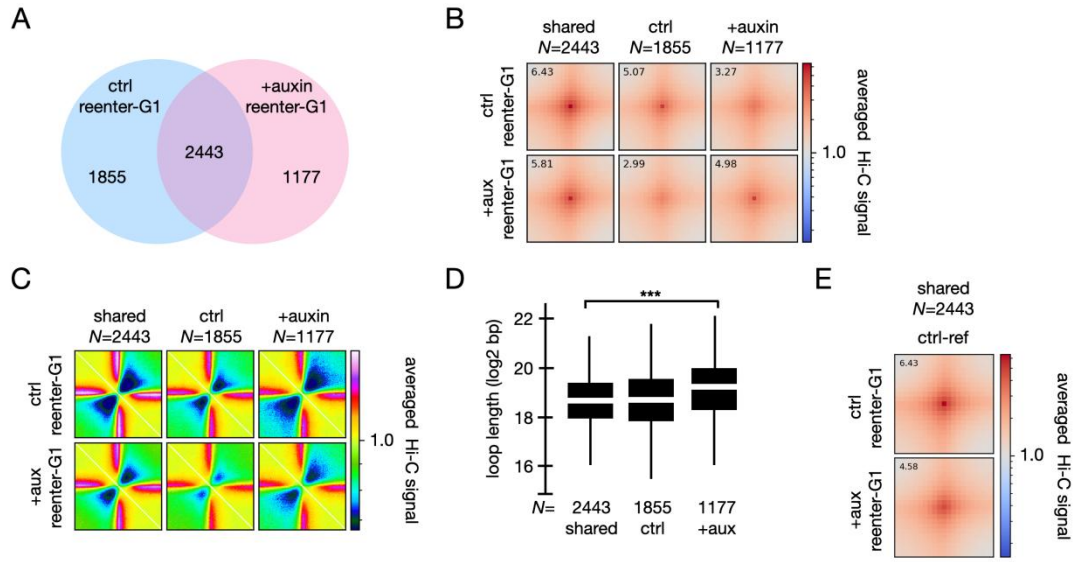


Figure 42. Numbers, signals, insulation, length, and diffusing pattern of reentry G1 loops. (A) Venn plot shows numbers of loops that exist both in control and the depletion (2443), loops exist only in control (1855), and loops only exist in the depletion (1177). (B) The plots show loop signals of panel (A). The plots were generated with 10kbp resolution. The enrichment was calculated from the center bin. The interactions among +-100kbp regions of loop anchors were selected to show on the squares. (C) The plots show insulation of anchors of panel (A). Plots were generated with 5kbp resolution. The interactions among +-500kbp regions of loop anchors were selected to show on the squares. (D) The plots show loop length distribution of panel (A). (E) The plots show the diffusing pattern in reentry G1 data.

To further estimate how RNAPII affects loop formation during mitosis to G1 transition, I first checked CTCF (**Figure 43A**), promoter (**Figure 43B**), and enhancer (**Figure 43C**) occupancy at loop anchors of different loop categories. Unlike previous findings (**Figure 21B**; **Figure 30**), CTCF occupancy at loop anchors does not increase upon RNAPII depletion. This observation indicates the prominent role of RNAPII is not impeding cohesin extrusion during mitosis to G1 transition. RNAPII affects loop formation with unclear mechanisms during mitosis to G1 transition.

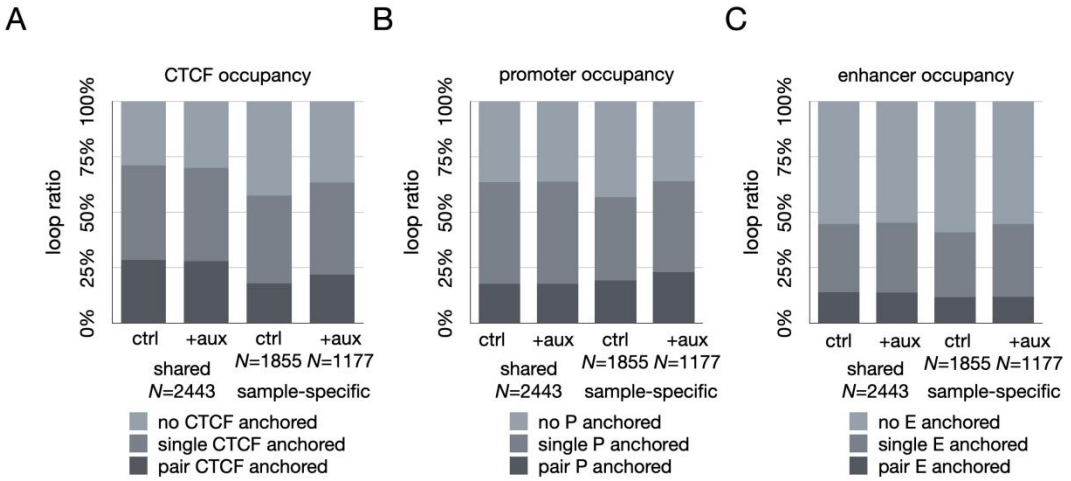


Figure 43. CTCF, promoter, and enhancer occupancy at loop anchors. (A) The plots show loop ratios of pair CTCF anchored, single CTCF anchored, and none CTCF anchored loops. (B) The plots show loop ratios of pair promoter anchored, single promoter anchored, and none promoter anchored loops. (C) The plots show loop ratios of pair enhancer anchored, single enhancer anchored, and none enhancer anchored loops.

Furthermore, I checked CTCF signals at loop anchors (**Figure 44A**), and SMC1 signals at the coordinates of CTCF peaks (**Figure 44B**). The results show that ~20% CTCF and SMC1A are reduced at loop anchors upon RNAPII depletion during reentry G1. While, this time 20% losing of them affects loop exhibition in Hi-C maps, the signals are decreased. This indicates RNAPII is required to loop formation during mitosis to G1 transition.

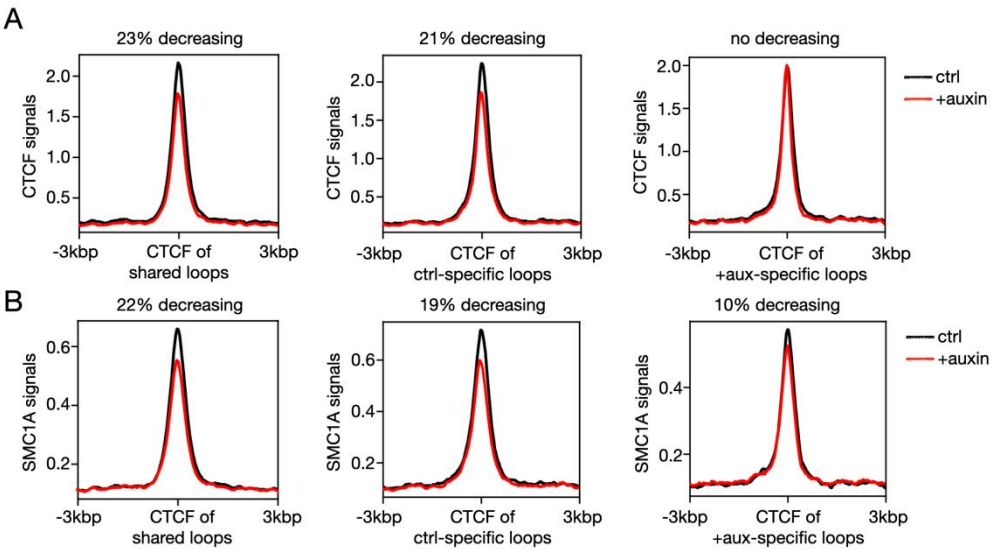


Figure 44. CTCF and SMC1A signals at loop anchors. (A) The plots show averaged CTCF signals of anchors of shared loops, control-specific loops, and RD-specific loops. (B) The plots show averaged SMC1A signals of anchors of shared loops, control-specific loops, and the depletion-specific loops.

Moreover, I checked whether reentry G1 data also exhibits the patterns of Micro-C data, which SMC1A more decreased at enhancers than promoters. To my surprise, the results show ~20% CTCF, SMC1A, and TBP decreasing at promoter regions (**Figure 45A**), while ~50% signals are decreased at enhancer regions (**Figure 45B**).

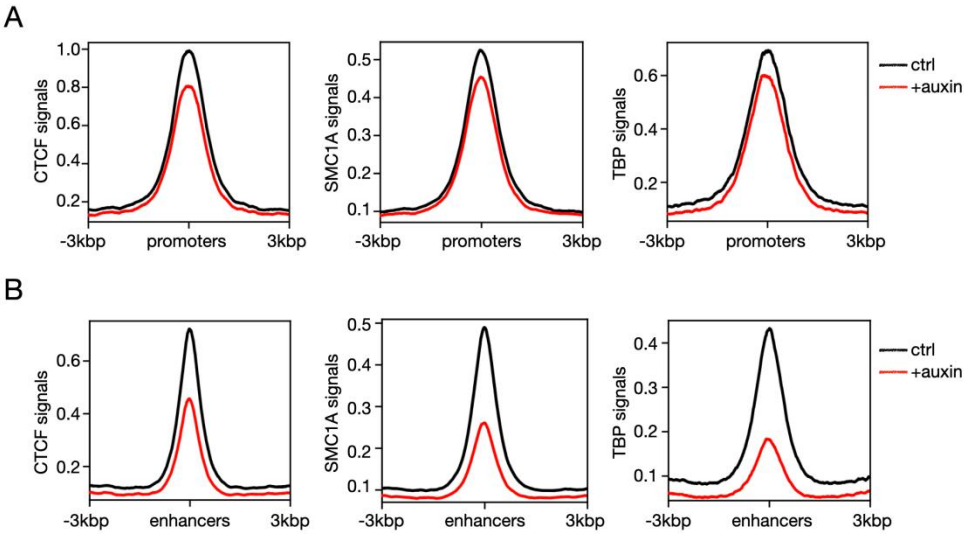


Figure 45. CTCF, SMC1A, and TBP signals at promoters and enhancers. (A) The plots show averaged CTCF, SMC1A, and TBP signals at promoters. (B) The plots show averaged CTCF, SMC1A, and TBP signals at enhancers.

Furthermore, I checked interactions among promoters and enhancers using Hi-C data. Importantly, unlike using Micro-C data, which interactions among promoters were not affected upon RNAPII depletion, interactions among promoters were mildly affected during mitosis to G1 transition (**Figure 46A**). Notably, interactions among enhancers are decreased upon RNAPII depletion (**Figure 46B**), which is in line with previous results.

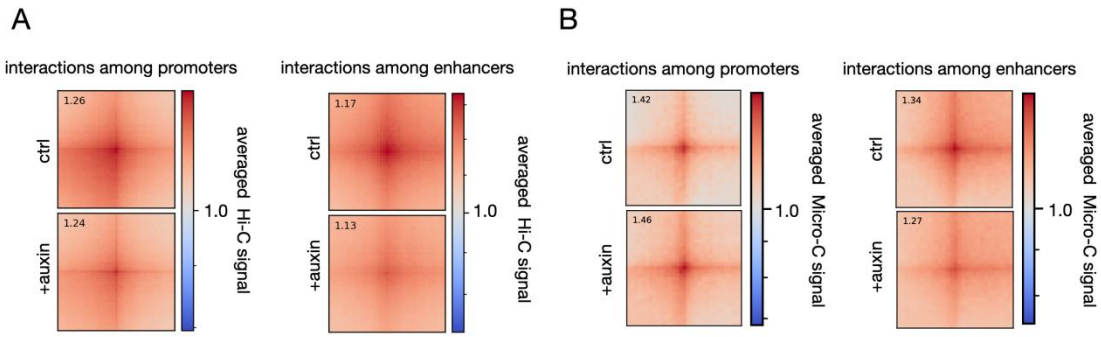


Figure 46. Interactions among promoters and enhancers. (A) Plots show interactions among promoters, and among enhancers before and after RD with reenter-G1 Hi-C data. Plots were generated with 5kbp resolution, enrichments were calculated from the center bins, and interactions among +-100kbp regions of loops were selected to show on the squares. (B) Plots show interactions among promoters, and among enhancers before and after RD with Micro-C data. Plots were generated with 1kbp resolution, enrichments were calculated from the center bins, and interactions among +-15kbp regions of loops were selected to show on the squares.

4.4.3 Short summary of 4.4

Notably, the different layers of genome organization exhibit apparent alterations upon RNAPII depletion during mitosis to G1 transition. 1) The increased inter-chromosomal interactions; 2) The decreased intra-chromosomal interactions, which is in line with the lost or weakened loops; 3) The reduction of organization and segregation of compartmentalization. Interestingly, loop formation is affected severely in the absence of RNAPII during reentry G1 process. However, these observations are not found in the interphase data. The results are interesting, it suggests E-P interactions may play a role in architectural loop formation during mitosis to G1 transition. However, how RNAPII affects cohesin is still unclear. Moreover, CTCF and SMC1A signals are reduced more at enhancers than promoters. This observation is in line with the previous findings. Thus, RNAPII is required to enhancer-anchored loops during mitosis to G1 transition, and loop formation is effected severely in the absence of RNAPII during reentry G1 process.

5. Discussion

5.1 Comparing our data with RNAPII degradation effects in mESCs

In my data, I found both lost and gained loops upon RNAPII depletion. However, compared to a previous study, in which RNAPII degradation was performed in mESCs, the conclusions were not identical (Jiang et al, 2020; Zhang et al., 2021).

5.5.1 Similarities between the two independent studies

On the one hand, although no gained loops and insulation were shown upon RNAPII depletion by Jiang et al., many results in the two independent studies were similar. For example, neither cell viability nor cell cycle were affected upon rapid depletion of RNAPII in both studies. Interestingly, no dramatic alteration of global mature mRNAs level in mESCs could be detected following RNAPII depletion. Similarly, only ~1400 down-regulated genes in our factory-RNA-seq data were detected upon RNAPII degradation. These could be linked to biological processes of chromatin assembly, DNA packaging, gene silencing, heterochromatin organization, and gene expression regulation.

Moreover, obviously decreased chromatin accessibility (measured via ATAC-seq) and cohesin signals at promoters and enhancers were observed in both asynchronous mESCs and DLD-1 cultures upon sufficient (>90%) depletion of RNAPII. Furthermore, although Jiang et al. concluded that RNAPII play a modest role in mediating transcription-related interactions, their data did show significantly decreased contacts at RNAPII-occupied loci, which are similar to my results.

5.5.2 Differences between the two independent studies

On the other hand, diverging observations between mESC and DLD-1 cells in the absence

of RNAPII appear interesting. For example, my data shows considerable reduction of cohesin at enhancers compared to promoters in the absence of RNAPII, which is not the case in mESCs. Notably, my cell model is a human cancer cell line, whereas mESCs feature a faster cell cycle and particular gene expression patterns. Especially, enhancer-related interactions can vary between cell types (Rao et al., 2017; Rhodes et al., 2020). SE-mediated loops were enhanced upon cohesin degradation in HCT-116 cells, whereas polycomb-mediated loops emerged instead of SE-mediated loops upon cohesin degradation in mESCs. Interestingly, SE-mediated interactions were mildly enhanced upon RNAPII degradation in mESCs potentially as a result of losing cohesin.

It is important to note that RNAPII-related loops are dynamic and cell-type specific. But why did Jiang et al., not show stronger loop signals in the absence of RNAPII indicative of RNAPII impeding cohesin extrusion? Furthermore, how RNAPII impedes cohesin extrusion remains unclear. It is probable that it represents a physical obstacle, although cohesin can disrupt compartmentalization (Rao et al., 2017). Moreover, the transcriptional pattern of the mESCs differs from human cancer cells. For example, mESCs require a large number of repressive markers (eg, H3K27me3) to maintain stem cell identity and cohesin can disrupt interactions formed by polycomb (Rhodes et al., 2020). Thus, cell-type specific attributes may the reason for some of the different observations between DLD1 and mESCs upon RNAPII depletion.

5.5.3 Micro-C is more efficient at capturing transcription-level architecture

We now know that Micro-C is more efficient at capturing transcription-level architecture than Hi-C. Many independent studies have shown that using MNase instead of restriction enzymes in C-based experiments produce more pronounced transcription-associated 3D interactions (Krietenstein et al., 2020; Aljahani et al., 2022; Dequeker et al., 2022; Preprint: Hsieh et al., 2021; Preprint: Barshad et al., 2022; Preprint: Ramasamy et al., 2022). Our findings from DLD-1 that RNAPII impedes cohesin extrusion and mediates E-P interaction,

are probably to be applicable to other cell lines as well. Thus, our studies provide potentially generalisable principles of 3D genome organization. As a result, the contribution of RNAPs in 3D genome folding, reported to be at best modest in previous work (Jiang et al, 2020), can now be put in a different light. I expect more and more attention to now focus on such transcription-based interactions and their functions.

5.2 How to interpret gained loop signal upon RNAPII degradation

5.2.1 RNAPII impedes cohesin extrusion during interphase

How RNAPII impedes cohesin extrusion is unclear, probably through physical obstacles. Moreover, RNAPII was reported that pushes and impedes condensin translocation in bacteria (Brandão et al., 2019). Also, RNAPII was reported that affect cohesin positioning in eukaryotic cell lines (Busslinger et al., 2017; Heinz et al., 2018).

Notably, transcription inhibition studies did not show enhanced architectural loop intensity (Krietenstein et al., 2020; Barutcu et al., 2019). This indicates chromatin-bound RNAPII is essential to impede cohesin extrusion. Although transcription inhibitor triptolide was reported to remove RNAPII from chromatin, my western blots show the effect of triptolide is minor, and much weaker than auxin induced depletion in my experimental system (**Figure 47**).

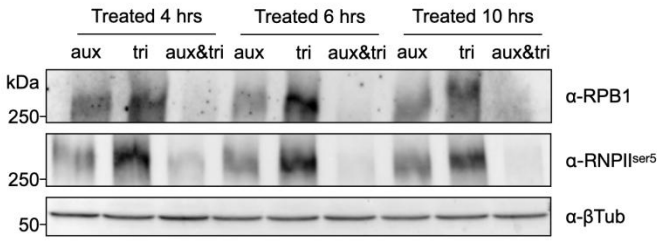


Figure 47. Western blots of DLD-1-RPB1-mAID degradation system. Western blots show degradation of total cell RPB1 or RNAPII^{ser5} on increasing treatment time of auxin only, triptlode only, and auxin plus trip; β-tubulin provides a loading control.

Recently, MCM-complex was reported that can impede cohesin extrusion. Furthermore, Hi-C and Micro-C data of inhibiting MCM-complex on chromatin showed the similar results as my data in the absence of RNAPII (Dequeker et al., 2022). These results indicate that chromatin-bound macromolecules can impede cohesin extrusion through physical obstacles. Furthermore, it was reported that MCM contains YDF motif that is functionally to

interact with SA2-SCC1 subunits of cohesin. Thus, MCM block cohesin through directly interact with it. Importantly, the N-terminal of CTCF also contains YDF motif (Li et al., 2020). It was reported that cohesin can pass through molecules larger than its diameter (Preprint: Pradhan et al., 2021). Thus, how RNAPII impedes cohesin is still a question needed to be answer due to no evidence showing RNAPII directly interact with cohesin.

My results indicate that RNAPII impedes cohesin extrusion because following results have been observed using Micro-C and Hi-C data. 1) loop signals are strengthened in average upon RNAPII degradation; 2) loops are extruded to be longer in the absence of RNAPII; 3) insulation strength of loop anchor is enhanced in average upon RNAPII degradation, and many anchors are rewired to CTCF in the absence of RNAPII; 4) diffusing pattern of loop signals emerges in the absence of RNAPII. These results indicate cohesin extrusion is easier to reach CTCF in the absence of RNAPII.

In addition, cell population Hi-C, which contains twice fewer reads than G1 Hi-C, shows slightly stronger loop signals than G1 Hi-C in the presence of RNAPII. This observation also indicates that architectural loops are enhanced upon reducing transcription activity in S/G2 phase, which is in line with my findings.

5.2.2 Gained loop signals with losing architectural factors

However, in average ~20% CTCF and cohesin are lost from chromatin in the absence of RNAPII, how to interpret lost architectural factors with increased loop intensity? Indeed, loops from Hi-C/Micro-C data are averaged signals over an entire cell population. A more red dot on Hi-C/Micro-C contact frequency map does not mean more binding of CTCF/cohesin at the location in one cell. While it means many cells contain the same interaction (of course, it should be a reason for it) within the cell population. Moreover, fixed number of cells contain fixed number of materials, once fragments are ligated they are not available for other ligation any more. The loops that can be detected on the Hi-C maps are

supported by a big proportion of cells. However, transcription process is dynamic, many transcription-mediated loops can not be detected in Hi-C/Micro-C maps due to being supported by a small proportion of cells. Thus, increased CTCF loop numbers and signals in Hi-C/Micro-C maps indicate the loop layer of the genome organization converges to unity (losing dynamic transcription activity) within a cell population. Interestingly, ~75% loops of DLD1 are transcription-independent architectural loops. These loops dominate the loop layer of genome organization, and RNAPII-dependent loops are almost intra-domain interactions. Importantly, the hierarchical organization of the loop layer of genome folding is interesting, which outer loops are transcription-independent and inner loops are transcription-dependent.

It is still unclear how RNAPII affects CTCF and cohesin on chromatin. There is no observation of RNAPII directly interact or recruit CTCF/cohesin on chromatin until now. Moreover, CTCF could be affected via chromatin accessibility in the absence of RNAPII. However, CTCF was reported that exhibited a dose-dependent insulation level, and TADs insulation exhibition on Hi-C map was slightly affects upon < 85% CTCF losing (Nora et al., 2017). Thus, ~20% CTCF losing in my data may not cause exhibition problem. In addition, it is unclear whether cohesin is also dose-dependent.

In addition, cohesin and CTCF are reduced more at intra-domain loops than at architectural loops. And many intra-domain loops are weakened or lost in the absence of RNAPII.

Altogether, for architectural loops, gained contacts upon losing active transcription are dominate the exhibition of loops in Hi-C/Micro-C maps. However, for transcription-dependent intra-domain loops, cohesin and CTCF are severely affected in the absence of RNAPII, as well as RNAPII-dependent loop formation was affected.

5.3 How RNAPII affect E-P interactions?

The results of Hi-C and Micro-C data have shown that RNAPII plays a role as physical barrier that impedes cohesin extrusion genome-wide. However, the loss or weakening of transcription-anchored loops, many stripes, and even CTCF-related loops upon RNAPII depletion indicate that RNAPII has another role in genome organization besides antagonizing architectural loops.

5.3.1 Role of cohesin in E-P interactions

Furthermore, cohesin was reduced twice more at anchors of “transcription only” loops than the average upon the depletion in my data. Although cohesin is the globally architectural extrusion factor of genome organization, the role of it in building local promoter-enhancer interactions is unclear. Since Rao et al., 2017 showed that only minor gene expressions were altered upon cohesin degradation, the role of cohesin in E-P interactions has been under debate. However, cohesin was reported that directly interact with NIPBL and mediator at enhancers (Kagey et al., 2010), which suggests its role in mediating E-P interactions. Recently, cohesin was reported that is necessary to form E-P interactions of some activity-induced genes, and its depletion down regulate gene expression (D'Ippolito et al., 2018; Cuartero et al., 2018; Zhu et al., 2021; Rinaldi et al., 2022).

Conversely, HindIII-based promoter Capture Hi-C data in HeLa cells suggests that E-P/P-P interactions are largely unaffected upon cohesin depletion (Thiecke et al., 2020). However, a study of cohesin degradation in mESCs using Tiled-MCC data shows that depletion of cohesin reduced E-P interactions, as well as down regulate gene expression (Aljahani et al., 2022). However, comparing with deletion the enhancer regions of the genes, the degree of down-regulated expressions were much less upon cohesin depletion.

Thus, these results suggest enhancer-promoter interactions were partly affected upon

cohesin degradation. Also, cohesin participates E-P interaction gene specific. Therefore, unlike architectural loops require cohesin, E-P interactions probably require more complex mechanisms.

5.3.2 How cohesin is affected locally at E/P in the absence of RNAPII

Importantly, in my data, averaged loop signal plots show clear reduction of signals with all “transcription only” loops in the absence of RNAPII. So, RNAPII is essential in forming E-P interactions. Moreover, it is common to find the missing stripes but not the final loops upon the depletion when manually check in Hi-C maps. This observation arises a question how does the depletion affect chromatin-bound cohesin? And this question is directly related to the role of RNAPII in forming E-P interactions. Interestingly, cohesin reduction is uneven to architectural loops and E-P loops in the absence of RNAPII. It was reported that cohesin-SA1 contributes to the architectural loops, and cohesin-SA2 contributes to tissue-specific transcription (Kojic et al., 2018). This report is in line with my results. Thus, it could be the cohesin-SA2, not the cohesin-SA1, is involved in formation of “transcription only” loops in DLD1 cells.

Furthermore, RNAPII degradation could affect cohesin-SA2 in many aspects, such as, cohesin loading, cohesin positioning, and cohesin stabilization. There is no reports about RNAPII directly interact with cohesin-SA2 until now. While cohesin was reported that directly interact with mediator-complex and NIPBL at enhancers (Kagey et al., 2010). Recently, a preprint about depleting MED14 in HCT-116 cell line also showed reductions of cohesin at enhancers but not CTCF (Preprint: Ramasamy et al., 2022). It also showed ~34% reduction of interaction strength between promoters and mediator-occupied super-enhancers upon MED14 depletion. In addition, NIPBL was reported that interact with sequencing-specific TFs (Rinaldi et al., 2022). Together with the four models that were described in the introduction part, it could be mediator-TF mediate the loading of cohesin-SA2 on chromatin.

Moreover, cohesin-SA2 was reported that contributes to specific E-P interactions in a CTCF-independent way (Kojic et al., 2018). Thus, it could be RNAPII also act as physical barrier to anchor cohesin-SA2 at enhancers. Moreover, RNAPII could facilitate mediator-TF stability at enhancers. Interestingly, all these assumptions are in line with the results that lost or weakened anchors of “transcription only” loops upon RNAPII degradation are most CTCF-independent enhancer anchors (**Figure 27A**; **27B**).

5.3.3 How condensates affect E-P interactions

In addition, disrupting of transcription-related condensates can reduce E-P interactions (Preprint: Barshad et al., 2022), which arises another question that how RNAPII-related condensates affect E-P interactions. Condensates can stabilize molecules, thus it could be condensates may facilitate RNAPII stabilize mediator-TF at enhancers. RNAPII can be involved in various condensates, such as unphosphorylated related condensates at enhancers, transcription initiation related condensate, and transcription elongation related condensates (Boehning et al., 2018; Cramer et al., 2019). Interestingly, disrupting of CTD-CTD condensates only affected enhancer activity not Sp1 regulated promoter activity (Gerber et al., 1995). Moreover, super-enhancers were affected more profound than typical enhancers upon disrupting of BRD4 and MED1 condensates (Sabari et al., 2018). Thus, factors (eg, RNAPII and MED1) could exhibit various functions at promoters, enhancers, and super-enhancers through various condensates. In addition, RNAPII does not solely contribute to E-P interaction, and the mechanisms of E-P are complex and involve cooperation among factors. Therefore, it could be RNAPII play different roles at promoters and enhancers in forming E-P interactions. For example, my results that P-P interactions among “transcription only” loops are affected less upon RNAPII depletion, while E-E interactions among the loops are affected profound (**Figure 27C**; **27D**). Interestingly, to these E-E interactions, the reduction of the interactions is CTCF-independent upon RNAPII depletion (**Figure 27C**; **27D**).

5.3.4 Why P-P interactions are affected less in the absence of RNAPII

There are many aspects to answer the question why P-P interactions are less affected in the absence of RNAPII while E-E interactions are not. The question is related to another question, are there different mechanisms of loop formation at enhancers and promoters? Enhancers and promoters are different *cis*-regulatory elements. Differences are shown between the two, not only at loop formation aspect but also at other RNAPII related aspects. For example, eRNAs are minor amount when comparing with mRNA, and they both transcribe from RNAPII. Furthermore, talking back to loop formation. Cohesin variants that cohesin-SA1 and cohesin-SA2 show different preferences. Moreover, CTCF is abundant at open-chromatin regions. While cohesin-SA2 was reported as CTCF-independent. Also, CTCF-independent reduction of interactions are observed in my data upon RNAPII depletion. In addition, CTD of RNAPII participates various condensates at promoters and enhancers. RNAPII is recruited by sequencing-specific TFs at enhancers. All of the observations indicate, enhancers can be dynamically and flexibly regulated at the loop formation aspect. This is in line with current opinion that mis-regulated enhancers induced cancers. P-P interactions are less affected because the mechanism of it did not affect profoundly upon RNAPII depletion. In addition, architectural loops are also less affected in the absence of RNAPII. It could be the mechanism of P-P interactions is more similar to architectural loops. In addition, the number of P-P loops without anchoring CTCF is ~1000, which is a small proportion. Moreover, H3K27ac signals at promoters are also less affected in the absence of RNAPII, while H3K27ac at enhancers are affected profoundly. Altogether, RNAPII involved loop formation mechanisms may be different for promoters and enhancers.

6. References

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., & Weiner, O. D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. eLife, 8, e41769.

Aljahani, A., Hua, P., Karpinska, M. A., Quililan, K., Davies, J., & Oudelaar, A. M. (2022). Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. Nature communications, 13(1), 2139.

Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. Nature reviews. Genetics, 21(2), 71–87.

Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nature reviews. Molecular cell biology, 18(5), 285–298.

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell, 27(2 Pt 1), 299–308.

Barshad, James J. Lewis1, Alexandra G. Chivu1, Abderhman Abuhashem2,3,4, Nils Krietenstein5, Edward J. Rice1, Oliver J. Rando6, Anna-Katerina Hadjantonakis2,4 and Charles G. Danko1,7,*RNA polymerase II and PARP1 shape enhancer-promoter contacts

Barutcu, A. R., Blencowe, B. J., & Rinn, J. L. (2019). Differential contribution of steady-state RNA and active transcription in chromatin organization. EMBO reports, 20(10), e48068.

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C., Chotalia, M., Xie, S. Q., Barbieri, M., de Santiago, I., Lavitas, L. M., Branco, M. R., Fraser, J., Dostie, J., Game, L., Dillon, N., Edwards, P. A., Nicodemi, M., & Pombo, A. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature, 543(7646), 519–524.

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., Grimes, G. R., Therizols, P., & Bickmore, W. A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. Molecular cell, 76(3), 473–484.e7.

Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N., & Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science (New York, N.Y.), 362(6413), eaau1783.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A. S., Yu, T., Marie-Nelly, H., McSwiggen, D. T., Kokic, G., Dailey, G. M., Cramer, P., Darzacq, X., & Zweckstetter, M. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nature structural & molecular biology, 25(9), 833–840.

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., Li, C. H., Shrinivas, K., Manteiga, J. C., Hannett, N. M., Abraham, B. J., Afeyan, L. K., Guo, Y. E., Rimel, J. K., Fant, C. B., Schuijers, J., Lee, T. I., Taatjes, D. J., & Young, R. A. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell, 175(7), 1842–1855.e16.

Brandão, H. B., Paul, P., van den Berg, A. A., Rudner, D. Z., Wang, X., & Mirny, L. A. (2019). RNA polymerases as moving barriers to condensin loop extrusion. Proceedings of the National Academy of Sciences of the United States of America, 116(41), 20489–20499.

Busslinger, G. A., Stocsits, R. R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., & Peters, J. M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature, 544(7651), 503–507.

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., & Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. Nature genetics, 50(9), 1296–1303.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., & Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science (New York, N.Y.), 361(6400), eaar2555.

Cisse, I. I., Izeddin, I., Causse, S. Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., & Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. Science (New York, N.Y.), 341(6146), 664–667.

Cramer P. (2019). Organization and regulation of gene transcription. Nature, 573(7772), 45–54.

Cremer, T., & Cremer, M. (2010). Chromosome territories. Cold Spring Harbor perspectives in biology, 2(3), a003889.

Cuartero, S., Weiss, F. D., Dharmalingam, G., Guo, Y., Ing-Simmons, E., Masella, S., Robles-Rebollo, I., Xiao, X., Wang, Y. F., Barozzi, I., Djeghloul, D., Amano, M. T., Niskanen, H., Petretto, E., Dowell, R. D., Tachibana, K., Kaikkonen, M. U., Nasmyth, K. A., Lenhard, B., Natoli, G., … Merkenschlager, M. (2018). Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. Nature immunology, 19(9), 932–941.

Dao, L., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., Alomairi, J., Martin, D., Torres, M., Fernandez, N., Soler, E., van Helden, J., Puthier, D., & Spicuglia, S. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. Nature genetics, 49(7), 1073–1081.

Darrow, E. M., Huntley, M. H., Dudchenko, O., Stamenova, E. K., Durand, N. C., Sun, Z., Huang, S. C., Sanborn, A. L., Machol, I., Shamim, M., Seberg, A. P., Lander, E. S., Chadwick,

B. P., & Aiden, E. L. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proceedings of the National Academy of Sciences of the United States of America, 113(31), E4504–E4512.

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. Science (New York, N.Y.), 295(5558), 1306–1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., & Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell, 149(6), 1233–1244.

Dequeker, B., Scherr, M. J., Brandão, H. B., Gassler, J., Powell, S., Gaspar, I., Flyamer, I. M., Lalic, A., Tang, W., Stocsits, R., Davidson, I. F., Peters, J. M., Duderstadt, K. E., Mirny, L. A., & Tachibana, K. (2022). MCM complexes are barriers that restrict cohesin-mediated loop extrusion. Nature, 606(7912), 197–203.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C. L., & Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS biology, 8(5), e1000384.

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K. C., Huang, H., Liu, T., Marina, R. J., Jung, I., Shen, Y., Guan, K. L., & Ren, B. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nature methods, 14(6), 629–635.

D'Ippolito, A. M., McDowell, I. C., Barrera, A., Hong, L. K., Leichter, S. M., Bartelt, L. C., Vockley, C. M., Majoros, W. H., Safi, A., Song, L., Gersbach, C. A., Crawford, G. E., & Reddy, T. E. (2018). Pre-established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids. Cell systems, 7(2), 146–160.e7.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B.

(2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380.

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D., & Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome research, 16(10), 1299–1309.

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell systems, 3(1), 95–98.

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., & Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature, 539(7629), 452–455.

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., & Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. Cell research, 26(12), 1345–1348.

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., Razin, S. V., Mirny, L. A., & Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature, 544(7648), 110–114.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell reports, 15(9), 2038–2049.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., Chew, E. G., Huang, P. Y., Welboren, W. J., Han, Y., Ooi, H. S., Ariyaratne, P. N., Vega, V. B., Luo, Y., Tan, P. Y., Choy, P. Y., … Ruan, Y. (2009). An

oestrogen-receptor-alpha-bound human chromatin interactome. Nature, 462(7269), 58–64.

Gerber, H. P., Hagmann, M., Seipel, K., Georgiev, O., West, M. A., Litingtung, Y., Schaffner, W., & Corden, J. L. (1995). RNA polymerase II C-terminal domain required for enhancer-driven transcription. Nature, 374(6523), 660–662.

Giorgetti, L., Lajoie, B. R., Carter, A. C., Attia, M., Zhan, Y., Xu, J., Chen, C. J., Kaplan, N., Chang, H. Y., Heard, E., & Dekker, J. (2016). Structural organization of the inactive X chromosome in the mouse. Nature, 535(7613), 575–579.

Greenwald, W. W., Li, H., Smith, E. N., Benaglio, P., Nariai, N., & Frazer, K. A. (2017). Pgltools: a genomic arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. BMC bioinformatics, 18(1), 207.

Haarhuis, J., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., Krijger, P., Teunissen, H., Medema, R. H., van Steensel, B., Brummelkamp, T. R., de Wit, E., & Rowland, B. D. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell, 169(4), 693–707.e14.

Hamamoto, K., & Fukaya, T. (2022). Molecular architecture of enhancer-promoter interaction. Current opinion in cell biology, 74, 62–70.

Heinz, S., Texari, L., Hayes, M., Urbanowski, M., Chang, M. W., Givarkes, N., Rialdi, A., White, K. M., Albrecht, R. A., Pache, L., Marazzi, I., García-Sastre, A., Shaw, M. L., & Benner, C. (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell, 174(6), 1522–1536.e22.

Heist, T., Fukaya, T., & Levine, M. (2019). Large distances separate coregulated genes in living Drosophila embryos. Proceedings of the National Academy of Sciences of the United States of America, 116(30), 15062–15067.

Hsieh, T. H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., & Rando, O. J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell, 162(1), 108–119.

Hsieh, T. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R., & Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Molecular cell, 78(3), 539–553.e8.

Hsieh, T. S., Fudenberg, G., Goloborodko, A., & Rando, O. J. (2016). Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. Nature methods, 13(12), 1009–1011.

Hug, C. B., Grimaldi, A. G., Kruse, K., & Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell, 169(2), 216–228.e19.

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., & Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nature genetics, 46(2), 205–212.

Isoda, T., Morio, T., & Takagi, M. (2019). Noncoding RNA transcription at enhancers and genome folding in cancer. Cancer science, 110(8), 2328–2336.

Jerkovic, I., & Cavalli, G. (2021). Understanding 3D genome organization by multidisciplinary methods. Nature reviews. Molecular cell biology, 22(8), 511–528.

Jiang, Y., Huang, J., Lun, K., Li, B., Zheng, H., Li, Y., Zhou, R., Duan, W., Wang, C., Feng, Y., Yao, H., Li, C., & Ji, X. (2020). Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III. Genome biology, 21(1), 158.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L.,

Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., & Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature, 467(7314), 430–435.

Kieffer-Kwon, K. R., Tang, Z., Mathe, E., Qian, J., Sung, M. H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., Vian, L., Nelson, S., Zare, H., Hakim, O., Reyon, D., Yamane, A., Nakahashi, H., Kovalchuk, A. L., Zou, J., Joung, J. K., … Casellas, R. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell, 155(7), 1507–1520.

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., & Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature, 465(7295), 182–187.

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J., & Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. Science (New York, N.Y.), 366(6471), 1345–1349.

Kojic, A., Cuadrado, A., De Koninck, M., Giménez-Llorente, D., Rodríguez-Corsino, M., Gómez-López, G., Le Dily, F., Marti-Renom, M. A., & Losada, A. (2018). Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. Nature structural & molecular biology, 25(6), 496–504.

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T. S., Parsi, K. M., Yang, L., Maehr, R., Mirny, L. A., Dekker, J., & Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. Molecular cell, 78(3), 554–565.e7.

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., & Narlikar, G. J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature, 547(7662), 236–240.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet. 32, 42–56.

Leidescher, S., Ribisel, J., Ullrich, S., Feodorova, Y., Hildebrand, E., Galitsyna, A., Bultmann, S., Link, S., Thanisch, K., Mulholland, C., Dekker, J., Leonhardt, H., Mirny, L., & Solovei, I. (2022). Spatial organization of transcribed eukaryotic genes. Nature cell biology, 24(3), 327–339.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human molecular genetics, 12(14), 1725–1735.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, N.Y.), 326(5950), 289–293.

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C. T., Cubeñas-Potts, C., Hu, M., Lei, E. P., Bosco, G., Qin, Z. S., & Corces, V. G. (2015). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. Molecular cell, 58(2), 216–231.

Lin, X., Qi, Y., Latham, A. P., & Zhang, B. (2021). Multiscale modeling of genome organization with maximum entropy optimization. The Journal of chemical physics, 155(1), 010901.

Lin, Y., Protter, D. S., Rosen, M. K., & Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Molecular cell, 60(2), 208–219.

Liu, Y., Nanni, L., Sungalee, S., Zufferey, M., Tavernari, D., Mina, M., Ceri, S., Oricchio, E., & Ciriello, G. (2021). Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. Nature communications, 12(1), 2439.

Liu, Z., Lavis, L. D., & Betzig, E. (2015). Imaging live-cell dynamics and structure at the single-molecule level. Molecular cell, 58(4), 644–659.

Li, Y., Haarhuis, J., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M. S., Willems, L., Teunissen, H., Muir, K. W., de Wit, E., Rowland, B. D., & Panne, D. (2020). The structural basis for cohesin-CTCF-anchored loops. Nature, 578(7795), 472–476.

Lu, H., Yu, D., Hansen, A. S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., & Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature, 558(7709), 318–323.

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell, 161(5), 1012–1025.

Magli, A., Baik, J., Pota, P., Cordero, C. O., Kwak, I. Y., Garry, D. J., Love, P. E., Dynlacht, B. D., & Perlingeiro, R. (2019). Pax3 cooperates with Ldb1 to direct local chromosome architecture during myogenic lineage specification. Nature communications, 10(1), 2316.

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., LeProust, E., Follows, G. A., Fraser, P., Luscombe, N. M., & Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nature

genetics, 47(6), 598–606.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., & Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic acids research, 9(22), 6047–6068.

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., & Chang, H. Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nature methods, 13(11), 919–922.

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., & Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature, 502(7469), 59–64.

Nagashima, R., Hibino, K., Ashwin, S. S., Babokhov, M., Fujishiro, S., Imai, R., Nozaki, T., Tamura, S., Tani, T., Kimura, H., Shribak, M., Kanemaki, M. T., Sasai, M., & Maeshima, K. (2019). Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. The Journal of cell biology, 218(5), 1511–1530.

Nguyen, T. A., Jones, R. D., Snavely, A. R., Pfenning, A. R., Kirchner, R., Hemberg, M., & Gray, J. M. (2016). High-throughput functional comparison of promoter and enhancer activities. Genome research, 26(8), 1023–1033.

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., & Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell, 169(5), 930–944.e22.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., & Heard,

E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature, 485(7398), 381–385.

Papantonis, A., & Cook, P. R. (2013). Transcription factories: genome organization and gene regulation. Chemical reviews, 113(11), 8683–8705.

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M. M., Shishkin, A. A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., & Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell, 174(3), 744–757.e24.

Rada-Iglesias, A., Grosveld, F. G., & Papantonis, A. (2018). Forces driving the three-dimensional folding of eukaryotic genomes. Molecular systems biology, 14(6), e8214.

Ramasamy S, Aljahani A, Karpinska MA, Cao TBN, Cruz JN, Oudelaar MA. The Mediator complex regulates enhancer-promoter interactions. bioRxiv doi:10.1101/2022.06.15.496245 (2022).

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell 171, 305-320.e24.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680.

Ray-Jones, H., & Spivakov, M. (2021). Transcriptional enhancers and their communication with gene promoters. Cellular and molecular life sciences : CMLS, 78(19-20), 6453–6485.

Rhodes, J., Feldmann, A., Hernández-Rodríguez, B., Díaz, N., Brown, J. M., Fursova, N. A.,

Blackledge, N. P., Prathapan, P., Dobrinic, P., Huseyin, M. K., Szczurek, A., Kruse, K., Nasmyth, K. A., Buckle, V. J., Vaquerizas, J. M., & Klose, R. J. (2020). Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. Cell reports, 30(3), 820–835.e10.

Richter, W. F., Nayak, S., Iwasa, J., & Taatjes, D. J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. Nature reviews. Molecular cell biology, 1–18. Advance online publication.

Rinaldi, L., Fettweis, G., Kim, S., Garcia, D. A., Fujiwara, S., Johnson, T. A., Tettey, T. T., Ozbun, L., Pegoraro, G., Puglia, M., Blagoev, B., Upadhyaya, A., Stavreva, D. A., & Hager, G. L. (2022). The glucocorticoid receptor associates with the cohesin loader NIPBL to promote long-range gene regulation. Science advances, 8(13), eabj8360.

Rinn J. L. (2014). lncRNAs: linking RNA to chromatin. Cold Spring Harbor perspectives in biology, 6(8), a018614.

Roayaei Ardakany, A., Gezer, H. T., Lonardi, S., & Ay, F. (2020). Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. Genome biology, 21(1), 256.

Rowley, M. J., Poulet, A., Nichols, M. H., Bixler, B. J., Sanborn, A. L., Brouhard, E. A., Hermetz, K., Linsenbaum, H., Csankovszki, G., Lieberman Aiden, E., & Corces, V. G. (2020). Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. elegans to mammals. Genome research, 30(3), 447–458.

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., Roeder, R. G., … Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control.

Science (New York, N.Y.), 361(6400), eaar3958.

Sahl, S. J., Hell, S. W., & Jakobs, S. (2017). Fluorescence nanoscopy in cell biology. Nature reviews. Molecular cell biology, 18(11), 685–701.

Scafe, C., Chao, D., Lopes, J., Hirsch, J. P., Henry, S., & Young, R. A. (1990). RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. Nature, 347(6292), 491–494.

Schmidt, D., Schwalie, P. C., Ross-Innes, C. S., Hurtado, A., Brown, G. D., Carroll, J. S., Flicek, P., & Odom, D. T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. Genome research, 20(5), 578–588.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., Haering, C. H., Mirny, L., & Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature, 551(7678), 51–56.

Su, J. H., Zheng, P., Kinrot, S. S., Bintu, B., & Zhuang, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. Cell, 182(6), 1641–1659.e26.

Su, W., Jackson, S., Tjian, R., & Echols, H. (1991). DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes & development, 5(5), 820–826.

Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C. L., Guttman, M., Yuan, G. C., & Cai, L. (2021). Integrated spatial genomics reveals global architecture of single nuclei. Nature, 590(7845), 344–350.

Teller, K., Illner, D., Thamm, S., Casas-Delucchi, C. S., Versteeg, R., Indemans, M., Cremer, T., & Cremer, M. (2011). A top-down analysis of Xa- and Xi-territories reveals differences of higher order structure at ≥ 20 Mb genomic length scales. Nucleus (Austin, Tex.), 2(5),

465–477.

Thiecke, M. J., Wutz, G., Muhar, M., Tang, W., Bevan, S., Malysheva, V., Stocsits, R., Neumann, T., Zuber, J., Fraser, P., Schoenfelder, S., Peters, J. M., & Spivakov, M. (2020). Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. Cell reports, 32(3), 107929.

T.-H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, X. Darzacq, R. Tjian, Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1. bioRxiv (2021), p. 2021.07.14.452365.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell, 10(6), 1453–1465.

Vian, L., Pękowska, A., Rao, S., Kieffer-Kwon, K. R., Jung, S., Baranello, L., Huang, S. C., El Khattabi, L., Dose, M., Pruett, N., Sanborn, A. L., Canela, A., Maman, Y., Oksanen, A., Resch, W., Li, X., Lee, B., Kovalchuk, A. L., Tang, Z., Nelson, S., … Casellas, R. (2018). The Energetics and Physiological Impact of Cohesin Extrusion. Cell, 173(5), 1165–1178.e20.

Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S., & Young, R. A. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell, 171(7), 1573–1588.e28.

Wutz, G., Ladurner, R., St Hilaire, B. G., Stocsits, R. R., Nagasaka, K., Pignard, B., Sanborn, A., Tang, W., Várnai, C., Ivanov, M. P., Schoenfelder, S., van der Lelij, P., Huang, X., Dürnberger, G., Roitinger, E., Mechtler, K., Davidson, I. F., Fraser, P., Lieberman-Aiden, E., & Peters, J. M. (2020). ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. eLife, 9, e52091.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. 36, 3573–3599.

Yesbolatova, A., Natsume, T., Hayashi, K. I., & Kanemaki, M. T. (2019). Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. Methods (San Diego, Calif.), 164-165, 73–80.

Zhang, D., Lam, J., & Blobel, G. A. (2021). Engineering three-dimensional genome folding. Nature genetics, 53(5), 602–611.

Zhang, H., Emerson, D. J., Gilgenast, T. G., Titus, K. R., Lan, Y., Huang, P., Zhang, D., Wang, H., Keller, C. A., Giardine, B., Hardison, R. C., Phillips-Cremins, J. E., & Blobel, G. A. (2019). Chromatin structure dynamics during the mitosis-to-G1 phase transition. Nature, 576(7785), 158–162.

Zhang, S., Übelmesser, N., Josipovic, N., Forte, G., Slotman, J. A., Chiang, M., Gothe, H. J., Gusmao, E. G., Becker, C., Altmüller, J., Houtsmuller, A. B., Roukos, V., Wendt, K. S., Marenduzzo, D., & Papantonis, A. (2021). RNA polymerase II is required for spatial chromatin reorganization following exit from mitosis. Science advances, 7(43), eabg8205.

Zhang, Y., An, L., Xu, J., Zhang, B., Zheng, W. J., Hu, M., Tang, J., & Yue, F. (2018). Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. Nature communications, 9(1), 750.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., & Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nature genetics, 38(11),

1341–1347.

Zheng, M., Tian, S. Z., Capurso, D., Kim, M., Maurya, R., Lee, B., Piecuch, E., Gong, L., Zhu, J. J., Li, Z., Wong, C. H., Ngan, C. Y., Wang, P., Ruan, X., Wei, C. L., & Ruan, Y. (2019). Multiplex chromatin interactions with single-molecule precision. Nature, 566(7745), 558–562.

Zhu, Y., Denholtz, M., Lu, H., & Murre, C. (2021). Calcium signaling instructs NIPBL recruitment at active enhancers and promoters via distinct mechanisms to reconstruct genome compartmentalization. Genes & development, 35(1-2), 65–81.

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., & Giorgetti, L. (2022). Nonlinear control of transcription through enhancer-promoter interactions. Nature, 604(7906), 571–577.