8-2019

Dissecting The Roles Of Human Smc Complexes In Transcription Regulation And Chromatin Organization

Ruoyu Wang

DISSECTING THE ROLES OF HUMAN SMC COMPLEXES IN TRANSCRIPTION REGULATION AND CHROMATIN ORGANIZATION

by

Ruoyu Wang, B.S.

APPROVED:

_____
Wenbo Li, Ph.D.
Advisor

_____
Darren Boehning, Ph.D.

_____
Leng Han, Ph.D.

_____
Chunru Lin, Ph.D.

_____
Dung-Fang Lee, Ph.D.

APPROVED:

_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

DISSECTING THE ROLES OF HUMAN SMC COMPLEXES IN

TRANSCRIPTION REGULATION AND CHROMATIN ORGANIZATION

A

THESIS

Presented to the Faculty of

The University of Texas MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

by

Ruoyu Wang B.S.

Houston, Texas

August, 2019

Acknowledgement

Firs of all, I want to express my most sincere appreciation and gratitude to my advisor, Dr. Wenbo Li, for all his support and help during the last two years. Under Dr. Li's guidance, I learned the meaning and significance of doing biomedical research, and how to really enjoy the fun and rewarding of doing science. He taught me how to tackle these biomedical questions from different aspects, and he instructed me these research approaches in a hands-on and comprehensive way. He is the most enthusiastic and encyclopaedic person in science that I have ever met, his endless passion on scientific research has inspired me very much to pursue biomedical research as a life-long career.

Secondly, I want to thank all of the lab members in Dr. Li's lab. Li lab is a united of creative and passionate scientists. I want to thank Dr. Joanna Krakowiak for her immeasurable help in this thesis project. I want to deeply thank Drs. Feng Xiong, Joo-Hyung Lee, Xiaoyu Zhu, and Ms. Guojie Li for their contributions to my projects. Also, I want to thank Dr. Jiaofang Shao for all the suggestions and discussions in bioinformatics.

During the past two years, I have the fortune to be also advised by many great scientists. I want to express my high gratefulness again to all of my advisory committee members, Drs. Darren Boehning, Leng Han, Chunru Lin, and Dung-Fang Lee. At the

same time, they are also our great collaborators. Especially, I want to thank the lab members in Dr. Han's lab and Dr. Lee's lab, Drs. Zhang Zhao, Jian Tu, Zjiun Huo, and Mo Liu for all of their help and discussions during my research.

Science is about the spirit of sharing and teamwork. I want to thank Drs. Masato Kanemaki, and Toyokai Natsume (National Institute of Genetics, Japan) for developing the mAID system to open a window to assess the direct effects of protein after their acute depletion, and for generously sharing the RAD21-mAID cells with us. I also want to thank Dr. Masatoshi Takagi (RIKEN, Japan) for kindly providing us the NCAPH-mAID cells to study Condensin I complex.

Lastly, I want to thank my parents, my family members, and my friends for their love and support. I owe you all a debt of gratitude which I will never be able to pay back.

Abstract

Metazoans utilize a constellation of distal regulatory elements to control gene transcription, and therefore they have to forge highly complex chromatin loops to spatially bridge these regulatory elements and genes in the three-dimensional (3D) genome. However, the hierarchy of chromatin contacts and their underlying mechanisms are not well-understood. SMC complexes including Cohesin complex and Condensin complex has been widely proposed to organize 3D genome structure, and further regulate metazoans' gene transcription. Here, we aim to dissect the direct functions of SMC complexes (both Cohesin and Condensin) in transcriptional regulation and 3D genome organization, by utilizing an inducible protein degradation system. Nascent transcriptome analysis revealed that Cohesin acute depletion impacts the nascent transcription at promoter regions rather than at the entire gene bodies, indicating a potential role of Cohesin in RNA polymerase II pause-release. Combined analysis with MYC degradation nascent transcriptome showed Cohesin complex and MYC are co-regulating a large portion of MYC targeted genes. Moreover, our nascent transcriptome with different signals' stimulus demonstrated Cohesin complex is generally not required for TNF-alpha and heat-shock stimulated transcriptional events. To further understand the hierarchy of 3D genome, high-resolution H3K27ac HiChIP analysis after Cohesin depletion revealed that there is a large portion of regulatory chromatin loops persist after Cohesin depletion. This is a significant finding suggesting that Cohesin is not a universal organizer of short-range enhancer-gene loops. We further showed that Cohesin-antagonized extra-long-distance super-enhancer loops are mediated by BRD4 phase separation, validating an emerging hypothesis that LLPS (Liquid-liquid phase separation) can drive chromatin contacts. For Condensin complex, our nascent transcriptome data after Condensin I depletion indicated that Condensin I modulates TNF-alpha stimulated transcription, but not the basal transcription program. Also, Hi-C analysis after Condensin I depletion revealed that Condensin I has a role in counteracting

A/B compartmental interaction. Our work has provided a functional dissection of roles played by human SMC complexes in transcription regulation and 3D genome organization.

Table of Contents

List of Illustrations

Abbreviations:

UT : Untreated

IAA : Auxin

Hi-C: High through-put chromosome conformation capture

HiChIP: Protein-centric chromatin conformation

SE: Super-enhancers

H3K27ac: The acetylation modification on Histone 3 Lysine 27

H3K4me: The mono-methylation modification on Histone 3 Lysine 4

TNF-alpha: Tumor Necrosis Factor – alpha

HS: Heat-shock

PRO-Seq: Precision run-on sequencing

TT-Seq: Transient Transcriptome

Introduction

1.1 Metazoan transcription regulation

The central dogma of molecular biology depicts how genetic information is transferred and conveyed. As the first step of the central dogma, transcription processes play a critical role in determining the diversity and activities of different cell types during most biological processes (1). The regulation of transcription was first well-established in prokaryotes (2). Bacteria utilize a controlling system called “Operon Model” to regulate the gene transcription, in which operon is a proximal DNA regulatory element that could be occupied by specific protein factors to boost/inhibit gene’s transcription. The DNA binding affinities of these factors can be controlled by small molecules to achieve the purpose of operating gene expression. With such transcription regulatory machinery, bacteria are able to cope with different environment and stimulus (2). Even though bacteria is considered as a relatively simple organism, it is quite complex when it comes to the transcriptional regulation. Metazoan possesses versatile cell types to exert much various and complexed cellular functions compared with bacteria. Thus, a naïve regulation system like “Operon Model” couldn’t meet the extreme high demanding for complexity and robustness by metazoans. Indeed, metazoans have evolved a collection of premier mechanisms which we are just starting to understand them. In metazoan cells, the nuclear genetics materials-DNA are wrapped into a basic chromatin unit-nucleosome with an octamer protein called histone. Not like bacteria’s DNA which is directly accessible for RNA polymerase apparatus, nucleosome structures make the binding between metazoan DNA and transcription factors quite effortful. The metazoan cells have devised many tactful mechanisms to make the DNA

accessible, including histone modifying and chromatin remodeling machinery (3). Like bacteria, metazoan cells also utilize gene-proximal DNA elements called promoters to assemble the whole transcription apparatus and drive the transcription initiation of genes. Given the fact that these core promoters themselves are not sufficient to provide such regulatory complexity for metazoan and the metazoan's genome size is huge, it is not hard to reckon that there are other types of regulatory elements in metazoan's genomes, function in trans or in long-distance. The distal regulatory-element idea was first demonstrated in 1981. (4) showed that a virus SV40 DNA element named enhancer later could enhance the expression of the beta-globin gene even this element is 10kb away from the beta-globin gene. Since then, mounting evidence have been showed that enhancers are pervasive and distinctive mechanisms in controlling metazoan gene expression during development. The most famous case is the eve locus in Drosophila, in which enhancers are responsible for body segmentation (5). Moreover, enhancers can even drive multiple oncogenes' dysregulating expression during tumorigenesis, like MYC's overexpression in leukemia (6). With the recent advancement of high throughput sequencing technologies and deeper understandings of these enhancer elements, we are able to map putative enhance elements genome-widely and test these enhancer elements' potential function through functional genomics approaches. Based on ENCODE project results, there are at least 400,000 putative enhancer elements in our human genome. ChIP-Seq (Chromatin Immunoprecipitation with sequencing) experiments of histone modifications and transcription factors have revealed several distinctive features of enhancers. First, distal enhancer elements are enriched for specific transcriptionally-active histone modifications like

H3K27 acetylation and H3K4 mono-methylation, rather than other transcriptional-active histone modification enriched on promoters like H3K4 tri-methylation; Second, a variety of transcription factors/co-activators (Foxa1, ER-a) will also reside on these distal elements suggesting enhancers' role as delivering these co-factors to core promoters; Third, a good amount of distal enhancers are also occupied by active RNA polymerase apparatus, producing a large amount of RNA molecules which mostly don't have the protein-coding capacity. These RNA molecules produced from enhancers are called eRNAs (enhancer RNAs), taking up a large portion of nuclear non-coding transcriptome. Whether these eRNAs largely dictate enhancers' function on gene control still remains debatable, several recent evidences suggest eRNAs could facilitate the communication between enhancers and promoters (7, 8). Whereas, other reports showed that the transcription activities on enhancers per se enhance gene expression rather than enhancers' transcription products (9). Beyond these well-distributed distal enhancer elements along the genome, the archipelago of regulatory elements named super-enhancers or stretch enhancers have been unmasked recently, which they are widely thought to be crucial to control key genes involved in cell identity and diseases development (10). The clustering of multiple enhancer elements provides an ideal platform for assembling and concentrating transcription factors to largely boost their cognate genes' transcription. The underlying mechanism of super-enhancers' formation and organization still remain unclear. Nevertheless, distal enhancer element and their clustering form are distinctive and indispensable features of metazoan transcription regulation.

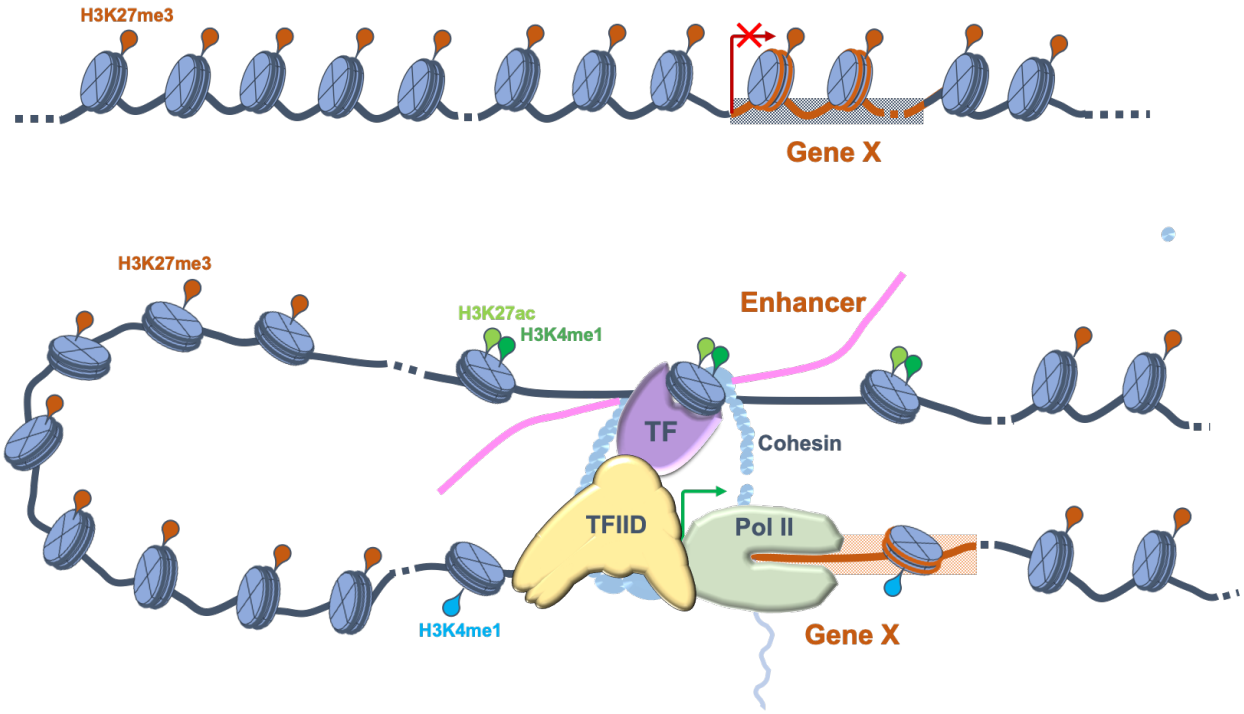

Figure 1. A Model of Metazoan gene transcription regulation

A Model of Metazoan gene transcription regulation, which includes key features of distal regulatory elements-enhancers: 1). Enhancers are marked by specific histone marks, like H3K27ac, H3K4me1; 2). Enhancers are bound by TFs (Transcription factors); 3) Enhancers are occupied by active transcription apparatus, thereby producing a large portion of non-coding RNAs called enhancer RNAs. * This figure was re-made based on Figure 4 in (11), with the permission from the journal via Copyright Clearance Center.

1.2 3D genome organization

According to the estimated numbers of enhancers and genes in our human genome, enhancers are ten times more pervasive than protein-coding genes(12), since one gene could utilize different enhancers to control their expression at different conditions and development stages. However, this raised up a fundamental, unsolved question. How these distal elements precisely find their cognate gene promoters which are sometimes even 100kb away, and convey their regulatory information to these targets at certain cell status or developmental stage? The work in the past fourty years have provided a significant amount of evidences that support a chromatin-looping model that distal enhancer elements need to have physical contacts with their cognate promoters through the proper folding and movement of chromatin fibers (13). Studies showed that engineering a forced enhancer-promoter contact at beta-globin locus only could result in gene activation (14), indicating enhancer-promoter looping is sufficient to induce gene activation. However, we still cannot understand the specificity of enhancer-promoters' communication till this decade. With the great advancement of chromosome conformation capture assay and high throughput sequencing, a genome-wide mapping of chromatin interactions is achievable in different cells and species (15–18). Since then, the study of transcription regulation has entered a new 3D genome era. These 3D chromatin structure studies have revealed that higher-order chromatin folding inside nuclear is organized into different hierarchical layers (Figure 2). The most unexpected finding from genome-wide chromatin interaction maps is that long chromatin fibers are spatially organized into ~1Mb size chromatin domains

named Topological Associating Domains (TADs) (18). The triangle shape structure on Hi-C heatmap (Hi-C: High throughput chromosome conformation capture) represent each topological chromatin domain. Enhancer-promoter communications are well-restricted inside each TADs, preventing outside genes from misregulation by intra-TAD enhancers. To understand the domain segmentation pattern on the genome and underlying mechanisms of TADs formation, integrative sequence analysis of domain boundary regions revealed that TAD boundaries are enriched in CTCF/Cohesin binding sites and highly-transcribed genes (18). Experimental evidences have showed that deleting TAD boundaries could result in alterations of nearby gene expression (19). The pathological consequences of breaking TAD boundaries include several congenital diseases (20) and cancer progression(21). Comparative analysis of TADs in different cell types and during different cellular processes revealed that TADs are generally cell-type invariant (17, 22), although there are recent evidences that suggest 10%-40% TADs have different boundary strength in different cell types(1). Further Hi-C analysis in different species showed that chromatin domains structure (e.g. TADs in mammals) exists in a wide range of organisms, from yeast to human(23). Thus, these comparative analyses unveiled that chromatin domain is an evolutionary conserved and fundamental structural unit of 3D genomes. Beyond this domain level, another unexpected layer of higher-order chromatin organization inferred from Hi-C maps is A/B compartment (Active/Inactive). The plaid-like pattern in large-scale Hi-C map (usually greater than 20Mb) represents the A/B compartment structure. The first component of Principal component analysis (PCA) of multi-dimensional Hi-C map is normally utilized to classify the genome into two compartments (15). Transcriptionally

active genes and chromatin regions (euchromatin) are enriched in A compartment, while transcriptionally inactive genes and chromatin regions (heterochromatin) are enriched in B compartment. A previous study showed that recruitment of certain transcription factors to a specific locus is able to induce A/B compartment transition, suggesting transcription factors are important for shaping compartment structures(24). Unlike TADs, A/B compartment are widely considered cell-type specific and dynamic during cell fate transitions (1). Combining with transcriptomic data, the transitions of compartment between cell types are shown to be correlated well with gene expression changes, suggesting chromatin organization's role in transcriptional regulation (15). While A/B compartment and TADs are relatively large-scale chromatin structures, a more fine-scale structure emerged on high-resolution in situ Hi-C maps is chromatin loop, which is usually represented by the focal point pattern in Hi-C data (16). Besides enhancer-promoter loops we discussed above, there is a great portion of chromatin loops considered as domains loops. Domains loops' anchors are mostly located in domain boundaries and highly occupied by CTCF/Cohesin. It is expected that these domain loops are generally cell-type invariant, since chromatin domains/TADs are usually not cell-type specific (16). In contrast, a large portion of enhancer-promoter loops is notably cell-type dynamic (25) (22). Although previous studies suggest that Cohesin, YY1, and other factors are mediating these enhancer-promoter loops, the underlying mechanisms of these cell-type specific loops are not well understood.

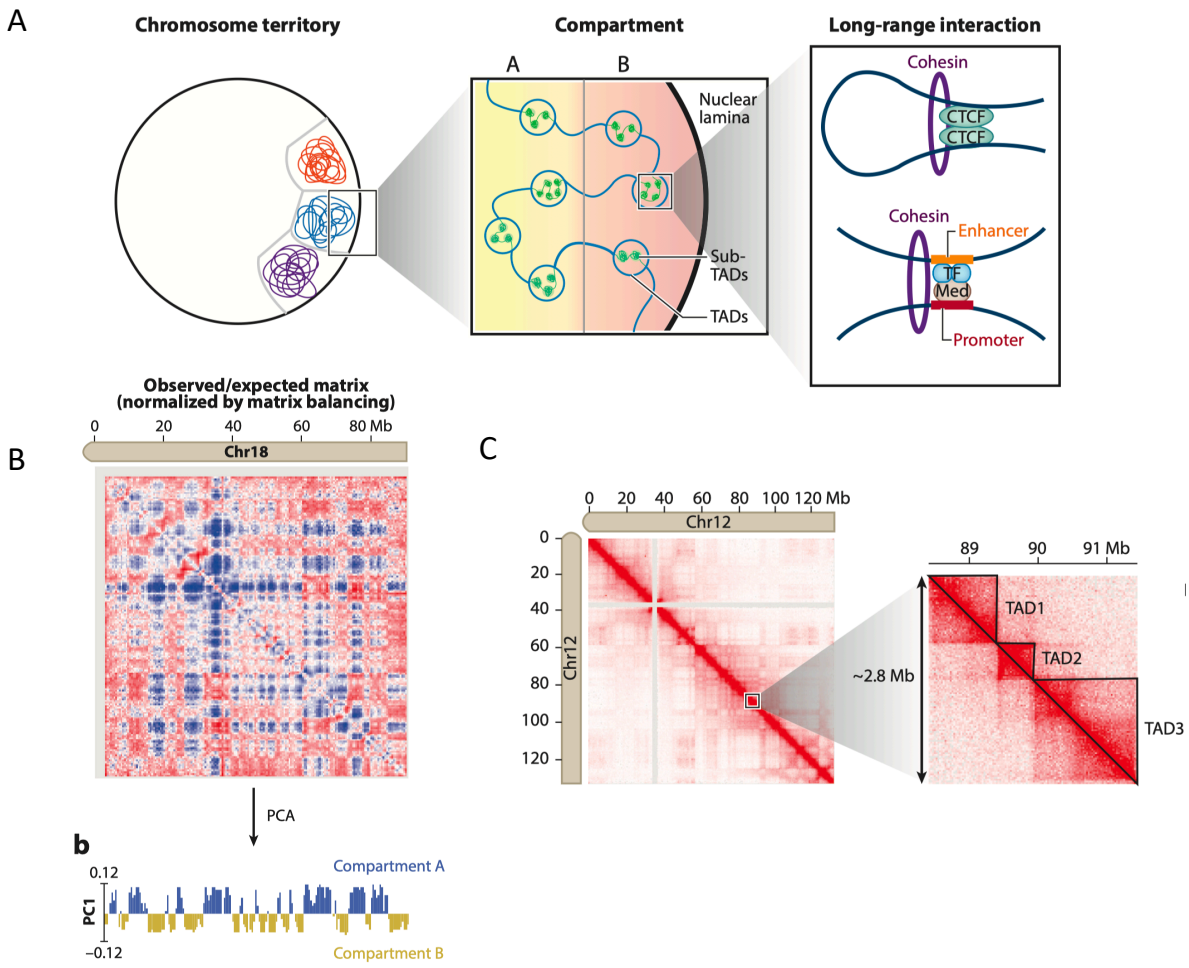

Figure 2. The introduction to different layers of 3D genome hierarchy

The introduction to different layers of 3D genome hierarchy: A) A schematic model of different models of hierarchical 3D genome organizations. The left panel depicts the largest structure – chromosome territories; the middle panel described the emerging concepts of 3D genome -A/B compartments and TADs; the right panel showed two different types of long-range chromatin loops- domain loops and enhancer promoter loops. * This figure is adapted modified from Figure 2 and Figure 3 in (26), used with permissions of the Annual Review Dev. Cel. Bio. I via Copyright Clearance Center.

1.3 SMC complexes: Cohesin and Condensin

The SMC (Structural Maintenance Complex) complexes are a family of complexes that are widely involved in sister chromatid cohesion, mitotic chromosome condensation, and several other DNA related nuclear events. And inside this family, there are three major highly-conserved SMC complexes – Cohesin, Condensins, and SMC5-6 complexes (27). The first SMC protein was discovered in E. coli almost thirty years ago, when the first SMC gene-mukB was showed to regulate the chromosome segregation (28). Later on, extensive work on different organisms have identified homologous SMC complexes in different species, which share a very conserved function in regulating chromosome segregation (27). Specifically, Cohesin complexes were demonstrated to regulate the chromosome segregation through tethering the sister chromatids together through G2 phase, namely Cohesion function (29). And this cohesion function is crucial for equally dividing the replicated genetic material into daughter cells. Condensin complexes have two major subtypes in mammals, Condensin I and Condensin II (Figure 3). Both of them are responsible for the rod-shape chromosome condensation during mitotic stage (30), but they also have distinctive roles in this short condensation processes. Electron microscopy studies showed that Both Cohesin and Condensin complexes have a striking structure feature, that they formed a ring-like structure (31). This unique ring-like structure provide a vivid explanation of Cohesin's role in holding two sister chromatids together during interphase. Also, the model that the ring-like structure of SMC complexes can hold two chromatin fragments is extended to transcriptional regulation. It is widely hypothesized that SMC

complexes, specifically Cohesin, plays a pivotal role in forming and strengthening chromatin loops between enhancers and promoters (19). With the emergence of Hi-C data, a model termed loop extrusion by SMC complexes has been widely proposed to explain how chromatin loops' Hi-C pattern is formed in a biophysics manner (32, 33). Experimentally, multiple independent studies have demonstrated complete depletion of Cohesin complex, or its associating proteins can diminish the domain (TAD) and domain loop pattern of Hi-C data, consolidating the long-standing hypothesis that SMC complexes are mediating the interphase chromatin looping (34). Compared to Cohesin complexes, Condensin's role in interphase genome organization has been less studied in a genome-wide manner. ChIP-Seq analysis of Condensin I and II showed that both Condensin complexes can be recruited to interphase chromatin regions, specifically enhancers (35, 36). Another study has revealed that Condensin binding is enriched at TAD boundaries, suggesting Condensin is also helping to organize the chromatin structure in TAD level (37). However, there is no direct evidence that suggests depleting Condensin has a dramatic effect on 3D genome organization in mammals.

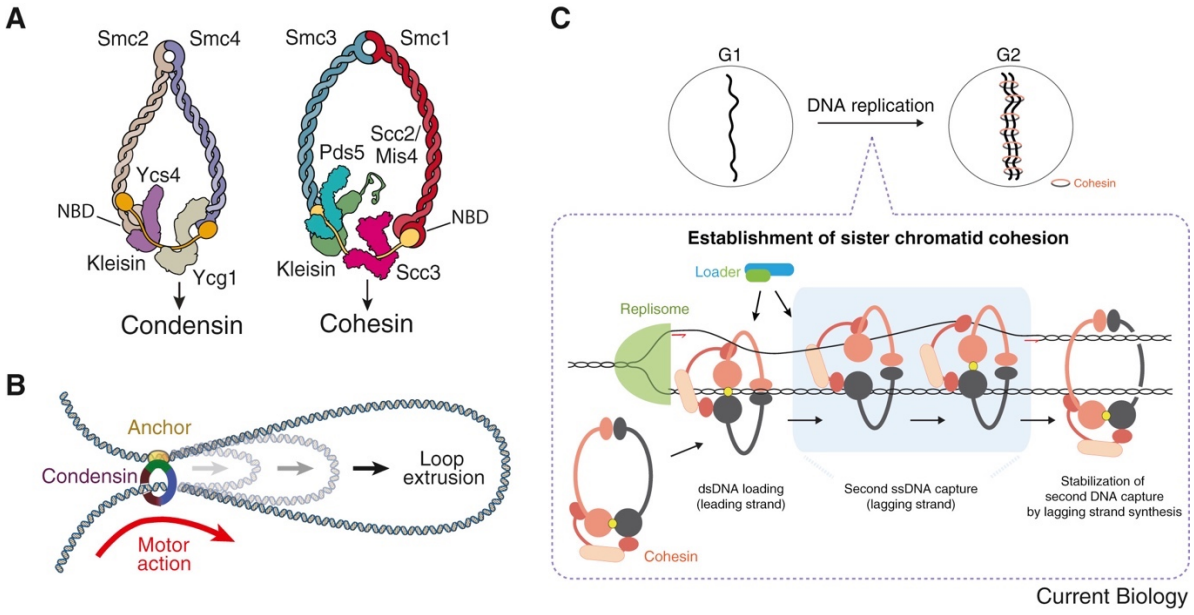

Figure 3, The Introduction to SMC complexes

The introduction to SMC complexes: A). There are two major types of SMC complexes well-characterized in mammals, Condensin (left) and Cohesin (right) complexes. Both of them have distinctive ring-like structures and multiple subunits. B). In this widely proposed loop extrusion model, Cohesin/Condensin complexes are extruding on the chromatin until stopped by certain anchors, thereby driving the formation of loops. C). A schematic diagram describing how G2 phase sister chromatid cohesion is established by Cohesin complexes. * This figure is adapted from (38). Copyright by Current Biology, used with permissions of Current Biology via Copyright Clearance Center.

2. Assessing the direct effect of acutely depleting Cohesin Complex on transcription

2.1 Introduction to nascent transcriptome

Conventional transcriptome analysis method like bulk cell total RNA-Seq measures the steady-state RNA abundance, which is overall correlated with genes' expression propensities. However, the steady-state RNA abundance of a particular gene is determined by multiple complexed processes, including transcription, splicing, nuclear export, post-transcriptional modifications, etc. Thus, the RNA abundances quantified from whole-cell transcriptome analysis could not precisely represent the true transcriptional events of genes, and whole-cell transcriptome is not able to use to study the dynamics of transcription in cells given certain stimulus. To study the transcriptional regulation, a method to direct measure the nascent transcriptome is necessitated. There is a growing body of efforts have been made in the transcription fields to develop methods for quantifying nascent transcriptome. In this work, we majorly discuss and utilized two methods here to study the SMC complex's role in transcription regulation. The first method discussed here is GRO-Seq/PRO-Seq (39). In GRO-Seq/PRO-Seq, a nuclear run-on was performed to let active RNA Pol II incorporate Br-UTP into the transcripts, and Br-UTP (5-bromo-2-deoxyuridine) labeled RNAs were hydrolyzed, and further pull-downed with anti-Br-UTP beads. By sequencing these RNAs, we will be able to get a genome-wide map of actively transcribing RNA polymerase. However, the GRO-Seq/PRO-Seq method is a semi *in vivo* method, which re-activate the RNA Pol II in extracted nuclei and may cause some artifacts or over-activation during this process. The second method is a newly developed method

called TT-Seq (Transient Transcriptome Sequencing) (40). In TT-Seq methods, we incorporated another labeled nucleotide named 4sU (4-thiouridine) into culturing cells and efficiently labelled newly transcribed RNAs with 4sU (Figure 3). These 4sU labelled RNAs were further converted into biotinylated RNAs and pull-downed by streptavidin beads. By sequencing these RNAs, the *in vivo* transcription rates in cells can be precisely measured genome-widely.

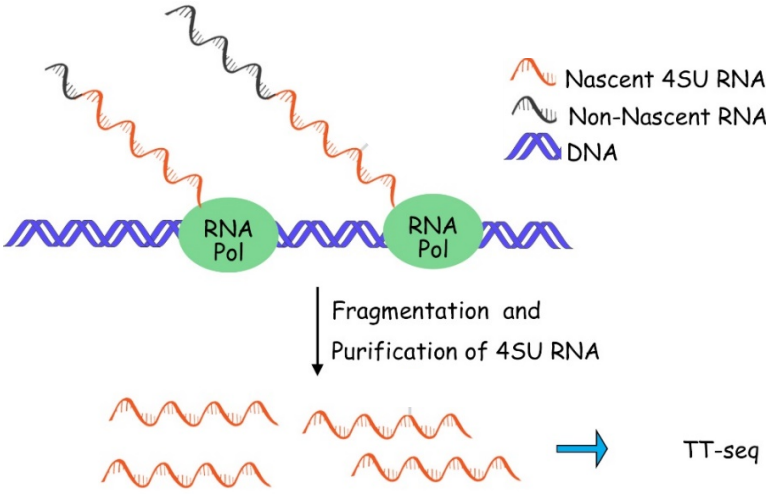

Figure 4, The schematic description of nascent transcriptome analysis

The schematic workflow of TT-Seq: 4sU were incorporated into culturing cells and labelled for 5 mins, and 4sU-labelled RNAs were further converted and fragmented.

2.2 Materials and Methods

Cell Culture and Treatment:

Rad21-mAID HCT-116 (#555) cells were obtained from Dr. Masato T. Kanemaki lab at NIG, Japan. And these cells were cultured in McCoy's 5A medium supplemented with 10% FBS in a 5 % CO2 humidified cell culture incubator. The auxin (IAA, Sigma) were treated at a final concentration of 500 uM for 6 hours. The TNF-alpha was given at a final concentration of 25 ng/ul of 1 hour. The heat-shock of cells were performed at 43 degree for 20 mins in a 5% CO2 cell culture incubator.

TT-Seq (Transient Transcriptome Sequencing)

4SU labeling and cell harvest. 4SU was diluted to pre-warmed fresh culture medium to 700uM. Then change the medium to 4SU-containing medium, put in incubator for 5mins. The medium was aspirated and TRIzol was directly added to the cell. Shake the dish to ensure all cells were covered by TRIzol. Cells were incubated at RT for at 5mins. Then cell lysate was transferred to a 15ml or 50ml tube. 1/5 volume chloroform was added to the tube, shaken vigorously, and incubated at RT for at 5mins. Samples were centrifuge at 12,000g for 15mins, with low deacceleration speed. Then we carefully transferred upper aqueous phase to a new tube and added 1/10 volume of 3M pH5.2 NaOAC to the new tube. An equal volume of isopropanol was further added and, mixed well. All samples were precipitated at -20 overnight. RNA Biotinylation: Centrifuge total RNA sample at 21,000g 4degree for 20mins. The supernatant was aspirated and the pellet were washed once with 1ml 75% ethanol (Centrifuge at

21,000g, 4 degree for 5mins.) Aspirate the ethanol and air dry the RNA pellet. RNA pellets were further dissolved in proper amount of DEPC water. The RNA concentrations were measured with Nanodrop, and further diluted with DEPC water and 10x Biotinylation buffer to make final concentration of RNA at 0.4ug/ul in 1x Biotinylation buffer. A 1/4 volume of dissolved MTSEA-XX was added to RNA solution. Rotate in dark for 2hrs. Add equal volume of Phenol-Chloroform to RNA solution, shake vigorously. Incubate at RT for 5 mins. Centrifuge at 12,000g for 10mins, with low de-acceleration speed. The upper aqueous phase was transferred to a new tube. A 1/10 volume of 3M ph5.2 NaOAc was added and mixed well. And a equal volume of isopropanol was further added, mixed and precipitated at -20 overnight. Biotin pulldown of nascent RNA: Centrifuge biotinylated RNA at 21,000g, 4 degree for 20mins. Aspirate the supernatant and wash the pellet once with 1ml 75% ethanol. Centrifuge at 21,000g, 4 degree for 5mins. Aspirate the supernatant and air dry the RNA pellet. RNA pellet was dissolved in proper amount DEPC water (200 ul for 10cm dish). Dilute RNA with DEPC water and 10x fragmentation buffer to make final concentration of RNA at 1.5ul/ul in 1x fragmentation buffer. Aliquot RNA to 1.6ml tube with 1ml in each tube. Incubate at 70c for 15mins. Invert tube several times every 2mins. Add 10x stopping buffer to each tube and put samples on ice. Samples were combined into one tube, and equal volume of 2x HSB buffer were added. Wash C1 beads with HSW buffer 3 times and resuspend in HSW buffer. Washed C1 beads were added to RNA samples, and rotated at RT for 30mins. Then beads were washed with three times HSW buffer (700 ul) at RT for 2 mins, two times TET buffer at RT for 2mins, one -ime TE buffer at 65 degree for 1.5 mins, and one-time TE buffer at RT for 2mins. After wash, 150ul

freshly prepared 5% b-ME were added to each tube, and rotate in dark for 15min. Put on magnet rack and transfer the supernatant to a new tube as elution1. Pre-warm 5% b-ME at 55c. Add 150ul b-ME to C1 beads, briefly vortex. Put on magnet rack and collect the second elution. Combine with elution1. 30ul 3M NaOAc and 2ul Glycoblue were added and mixed by vortex. Further add 300ul isopropanol, briefly vortex and precipitate at -20c overnight. The RNA pellets were dissolved in 10ul DEPC water and proceeded to NEB RNA-Seq library making.

TT-Seq/PRO-Seq data analysis:

After sequencing the library in illumina NextSeq 550, we de-multiplexed our sequencing samples through bcl2fastq. The sequencing reads of TT-Seq/PRO-Seq were mapped to reference genome hg19 by STAR 2.6.0. (41). The FPKM values of gene bodies were calculated by HOMER (42). The functional enrichment analysis of regulated genes was done by Metascape (43). Other analyses were performed by custom scripts.

2.3 Result and Discussion I: Cohesin's role in basal transcriptional regulation

2.3.1 TT-Seq (Transient Transcriptome) analysis after acute depleting Cohesin complex:

To fully understand Cohesin complex's role in human transcription regulation, we want to examine the direct effect on nascent transcriptome after acutely depleting Cohesin complex. Rather than conventional protein depletion method like siRNA/shRNA, we utilized a previous established auxin-inducible target protein degradation system to quickly degrade Cohesin key subunit – RAD21 completely in less than 1 hour (44). This protein degradation system can rule out other indirect or secondary effect that may be caused by long-term depletion of Cohesin or defected cell cycle. After treating auxin (IAA) to our RAD21-mAID cells, the mClover-tagged RAD21's level got completely depleted in 6 hours (Figure 5).

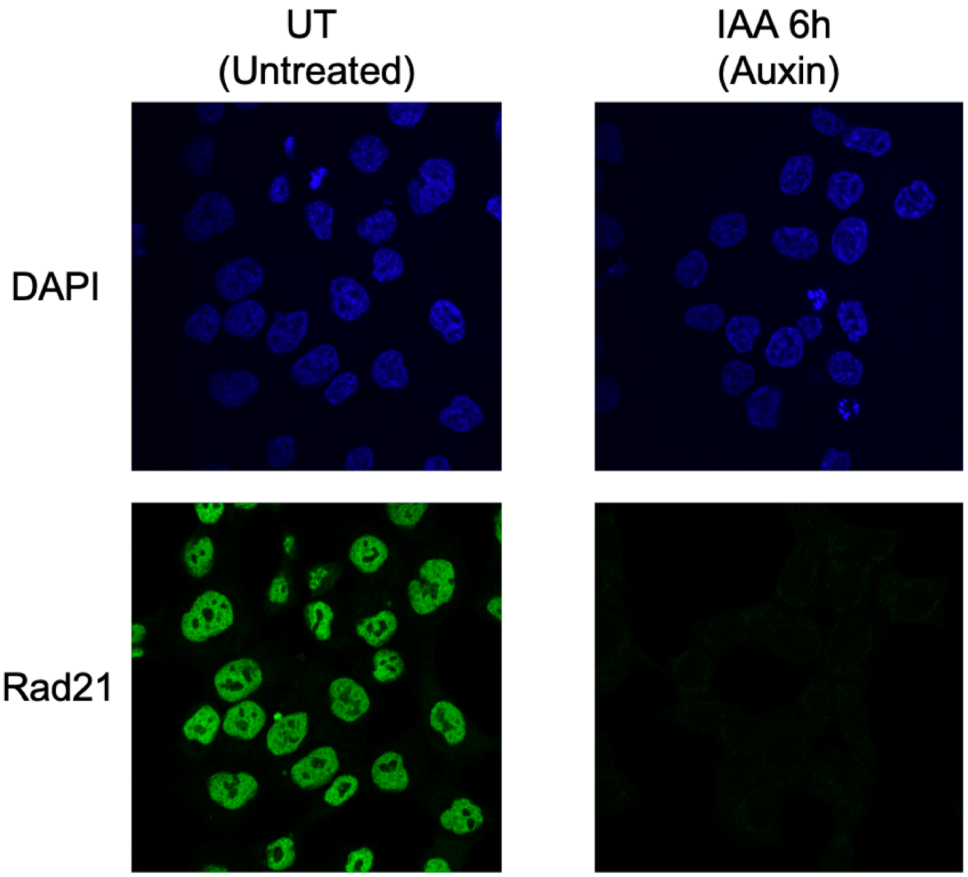

Figure 5, The depletion effect of Rad21 after treatment of auxin in 6 hours

The depletion effect of Rad21 after treatment of auxin in 6 hours: The blue signal here stands for DAPI (DNA), the green signal stands for mClover tagged RAD21-mAID.

After confirming the acute depletion of Rad21 in our systems, we sought to measure the direct transcriptional effect by performing TT-Seq (Transient Transcriptome Sequencing) after 6hr IAA treatment, and comparing its nascent transcriptome with untreated (UT) cells. After TT-Seq performed, we first quantify the nascent RNA abundances of all protein-coding genes by calculating the FPKM value along the gene body (from TSS to TES, include intron regions). Overall, we observed a nuanced effect on nascent transcriptome after acutely depleting Cohesin in a genome-wide manner, which is consistent with previous study's conclusion (Figure 6a). We also validated our TT-Seq results by performing PRO-Seq in UT and IAA 6h treated cells and observed a similar nuanced genome-wide effect on transcription. (Figure 6b)

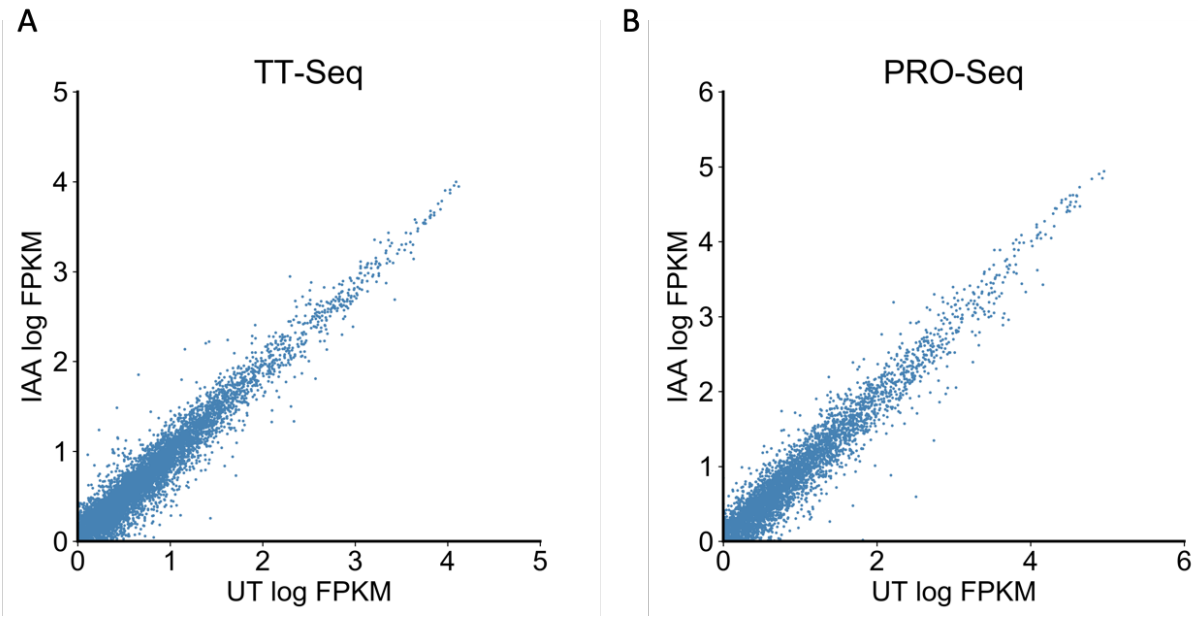

Figure 6, The genome-wide nuanced effect measured by both TT-Seq and PRO-Seq after acute depletion of Cohesin complex

The genome-wide nuanced effect measured by both TT-Seq and PRO-Seq after acute depletion of Cohesin complex: A). The scatter plot of each protein-coding gene's TT-Seq FPKM in UT and IAA; B). The scatter plot of each protein-coding gene's PRO-Seq FPKM in UT and IAA samples. Each blue dot stands for one protein-coding gene.

We further sought to check the specific up and down-regulated genes numbers after acute depletion of Cohesin. We first considered genes that have an average FPKM greater than 1 in two samples as expressed genes and got 6186 expressed genes in TT-Seq data and 5990 expressed genes in PRO-Seq data. Here, we set a rather loose cut-off that is fold-change of FPKM greater than 1.3, considering that the whole gene body regions we are quantifying are extremely long. In total, we identified 250 genes that are up-regulated and 522 genes that are down-regulated by Cohesin depletion, in terms of nascent RNA level measure by TT-Seq data. For PRO-Seq, we generally identified more changed genes (Figure 7). By visualizing the TT-Seq and PRO-Seq data on the genome browser, we can validate these genes are indeed up-regulated or down-regulated (Figure 8). Here we showed one down-regulated case (TRMT1, Figure 8), to exemplify.

Here, we are particularly interested in these 522 down-regulated genes after Cohesin depletion as identified by TT-Seq, and we further performed a functional enrichment analysis on these genes. The functional enrichment analysis showed that these genes were highly enriched in ribosome biogenesis genes (Figure 9a); This result compels us to question what co-factors may work together. With Cohesin to regulate these genes here, we performed a Cistrome GIGGLE analysis with these down-regulated genes promoter regions, seeking potential transcription factors that are highly occupied at these genes' promoters (Figure 9b). The giggle scores showed that oncogene MYC is the top TF that binds at these regions, suggesting a novel link between MYC and Cohesin complex in transcriptional regulation at promoters. To further

demonstrate MYC directly targeting these Cohesin down-regulated genes, we took advantage of a public valuable dataset that is SLAM-Seq (another nascent transcriptome method) of MYC-mAID HCT-116 cell line (45). Re-analysis of SLAM-Seq data after acute depleting MYC showed that there are 39 genes out of 95 down-regulated genes by MYC depletion that are also down-regulated genes after RAD21 depletion, suggesting overlapped functions of MYC and Cohesin complex in transcription. Further combination analysis of Rad21-mAID TT-Seq and MYC-mAID SLAM-Seq demonstrated that a significant portion of genes that are responsive to both MYC and Cohesin depletion (Figure 10), emphasizing the conclusion that MYC and Cohesin directly co-target these genes and they shared a large portion of targets.

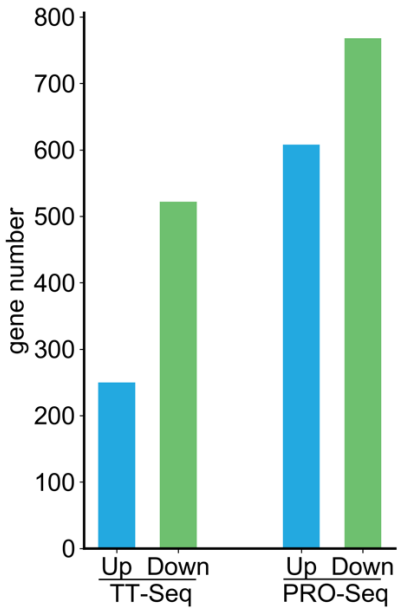

Figure 7, The up-regulated and down-regulated genes after Cohesin depletion

The bar-plot of gene numbers that are up-regulated or down-regulated after Cohesin depletion in both TT-Seq and PRO-Seq data.

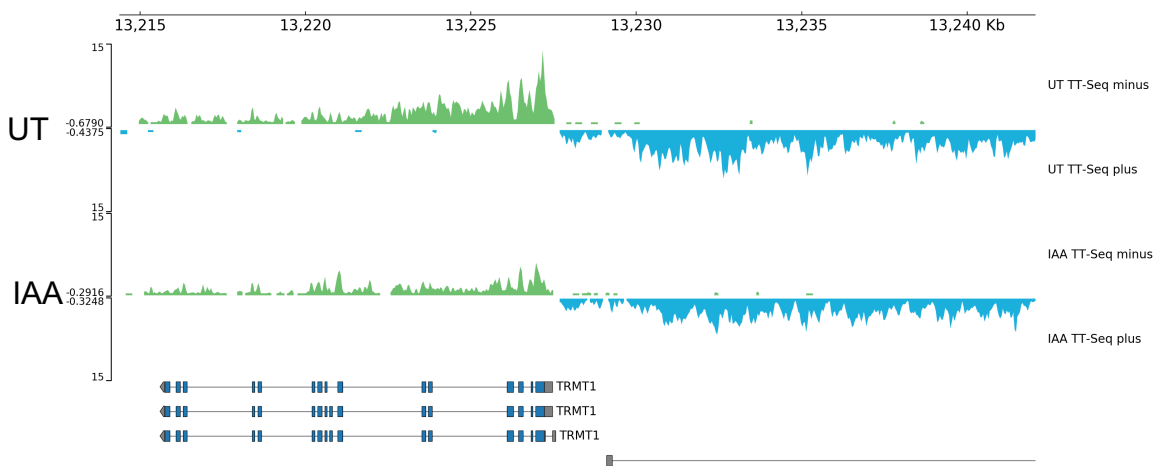

Figure 8, A case to exemplify the Cohesin depletion's effects on gene transcription.

The genome-browser snapshots of TT-Seq to exemplify the Cohesin depletion effects.

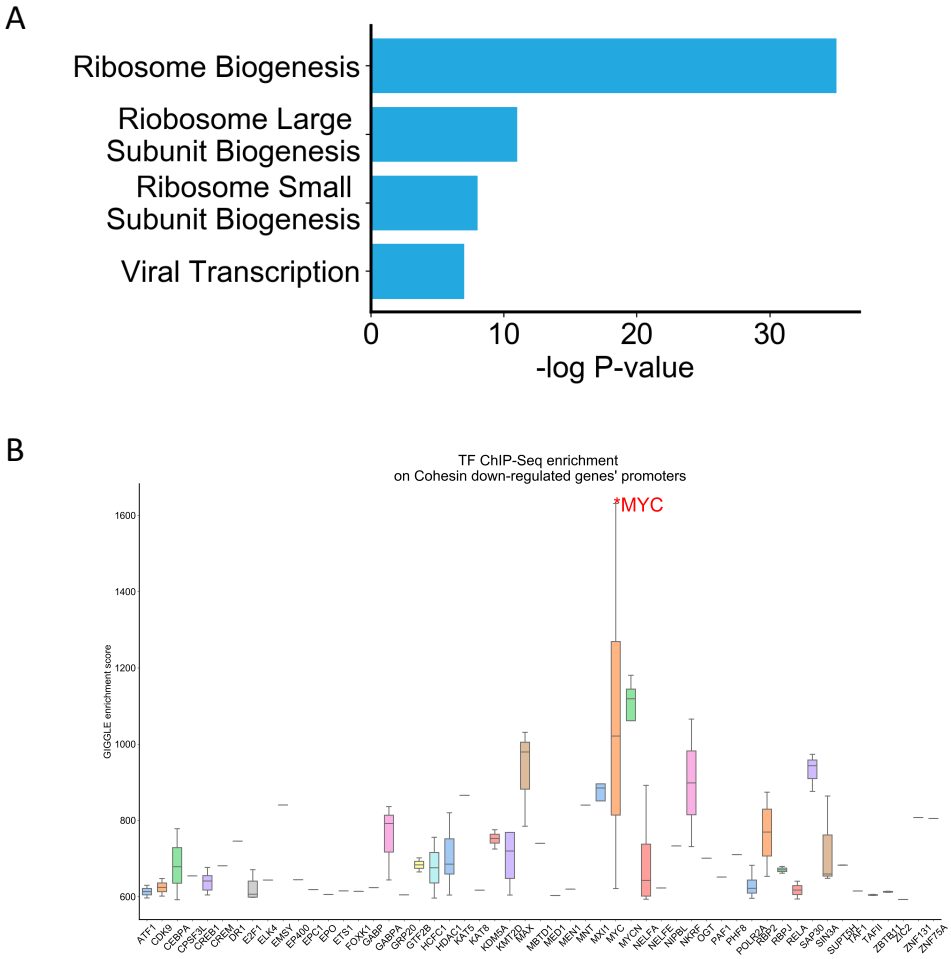

Figure 9, The functional enrichment analysis of Cohesin depletion down-regulated genes

A) Functional enrichment of Cohesin down-regulated genes showed that ribosome biogenesis genes are highly enriched; B) The Cistrome enrichment analysis showed that MYC is the highly enriched TF that occupied these Cohesin-down regulated genes promoters.

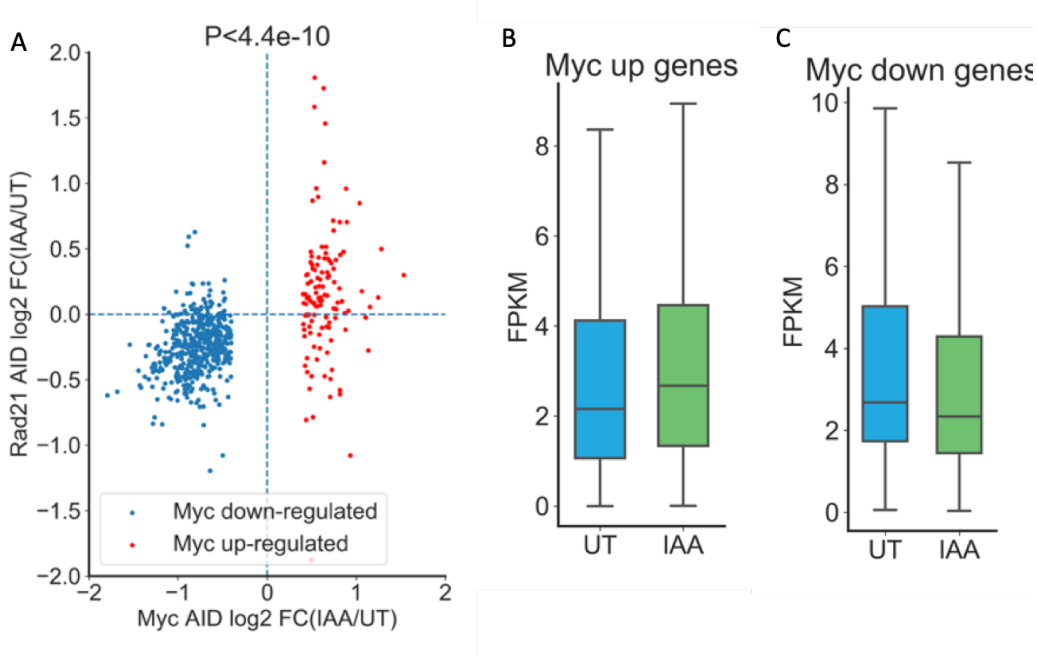

Figure 10, Cohesin complex and MYC are directly co-targeting a large portion of MYC-dependent genes

The Cohesin and MYC are directly co-targeting a large portion of MYC dependent genes: A) The scatter plot of log2 FC (UT/IAA) FPKM of TT-Seq in Rad21-mAID HCT-116 cells, and log2 FC(UT/AA) FPKM of SLAM-Seq in MYC-AID HCT-116 cells; B) The boxplot of TT-Seq FPKMs in Rad21-mAID cells of identified MYC-upregulated genes; C) The boxplot of TT-Seq FPKMs in Rad21-mAID cells of identified MYC-downregulated genes.

Since MYC was previously demonstrated to regulate transcription pause-release and elongation (46) and given the highly-correlated direct transcriptional effects between MYC and Cohesin, we sought to test whether Cohesin depletion has effects on transcription pause-release by checking the TT-Seq and PRO-Seq pattern around TSS regions after depleting Cohesin. Our metagene analysis of TT-Seq and PRO-Seq reads densities around TSS regions clearly showed that Cohesin depletion caused TT-Seq signal reducing at TSS regions (Figure 11a), while TSS's PRO-Seq signal increased after Cohesin depletion in contrast (Figure 11b). Since PRO-Seq measures all chromatin associated Pol II activities and TT-Seq is measuring only actively ongoing Pol II activities, we reasoned that decreasing of TT-Seq around TSS and increasing of PRO-Seq around TSS regions consistently suggest that Cohesin depletion increased the pausing of Pol II, while reduced the releasing of Pol II into gene bodies (at least the very starting regions of gene bodies). Given the fact that we didn't observe a strong transcriptional effect genome-widely by depleting Cohesin, we actually were surprised to observe a potential pause-release effect on Pol II by Cohesin depletion. To further confirm this finding, we quantify the TT-Seq reads only for TSS regions and compared with whole gene body quantification. We found that there is a more obvious change on TSS regions' reads rather than whole gene bodies' reads (Figure 12).

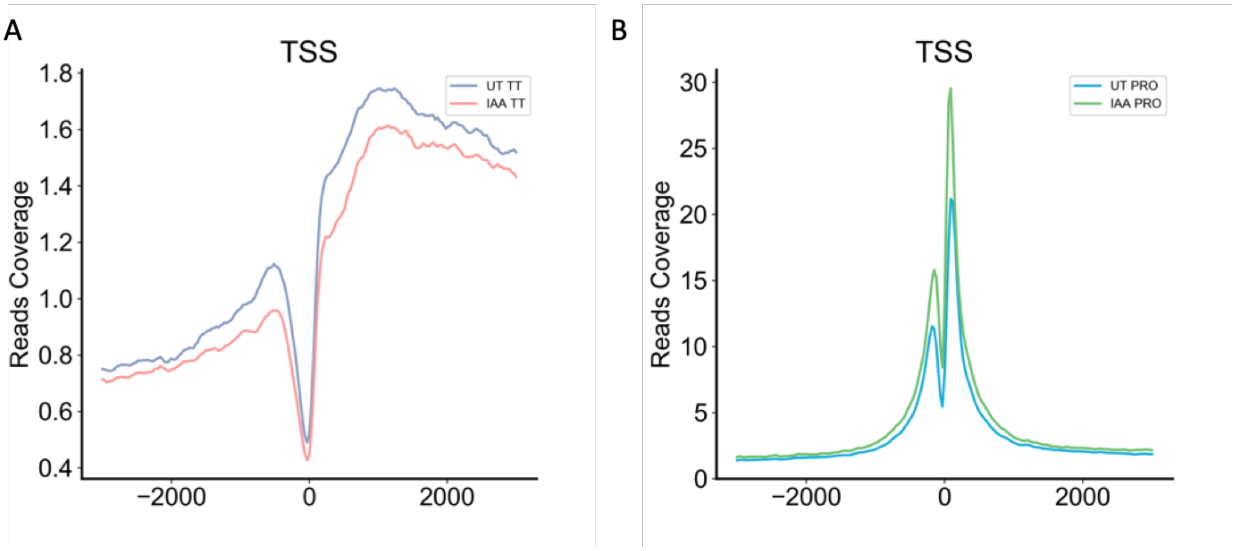

Figure 11, The Cohesin depletion's effect on TSS region's transcription

The Cohesin depletion's effect on TSS regions: A) The metagene plot of TT-Seq reads on TSS regions showed that Cohesin depletion decreased TT-Seq signal on TSSs; B) The metagene plot of PRO-Seq reads regions showed that Cohesin depletion increased PRO-Seq signal on TSSs.

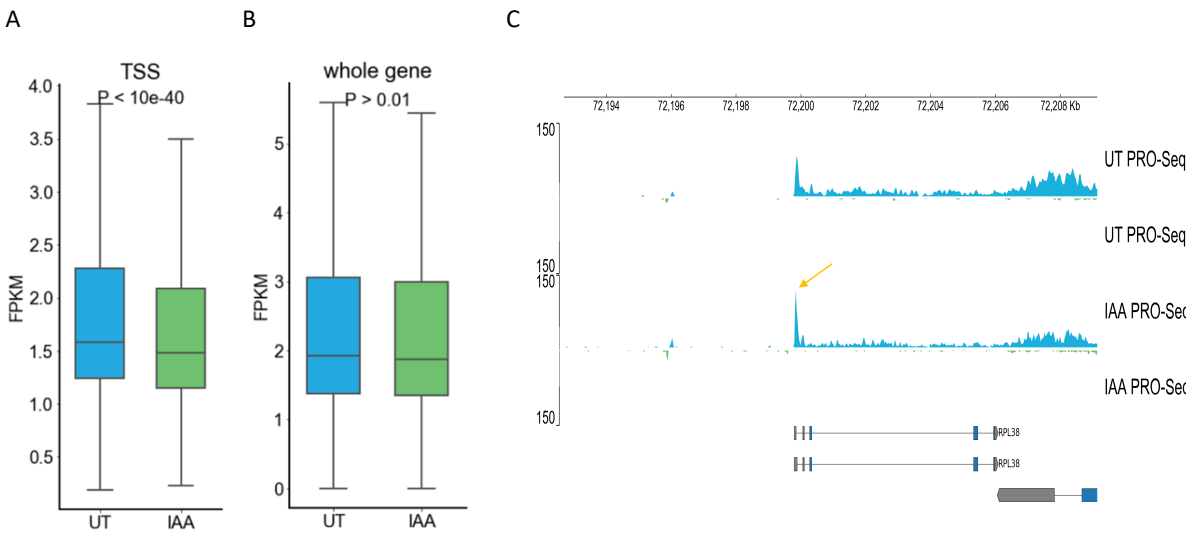

Figure 12, The comparison of TT-Seq reads quantifications on TSS regions and whole gene bodies

The comparison of TT-Seq reads quantifications on TSS regions and whole gene bodies: A). The boxplot of TT-Seq quantifications on TSSs showed that TT-Seq reads densities on TSS regions are significantly reduced by Cohesin depletion; B) The box-plot of TT-Seq quantifications on whole gene bodies showed no significant differences (P-value greater than 0.01). The P-value here was calculated by the Mann-Whitney U test. C). The genome-browser track snapshot of UT and IAA PRO-Seq at one locus showed RAD21 depletion regulate Pol II pause-release. The orange arrow indicated the paused promoter.

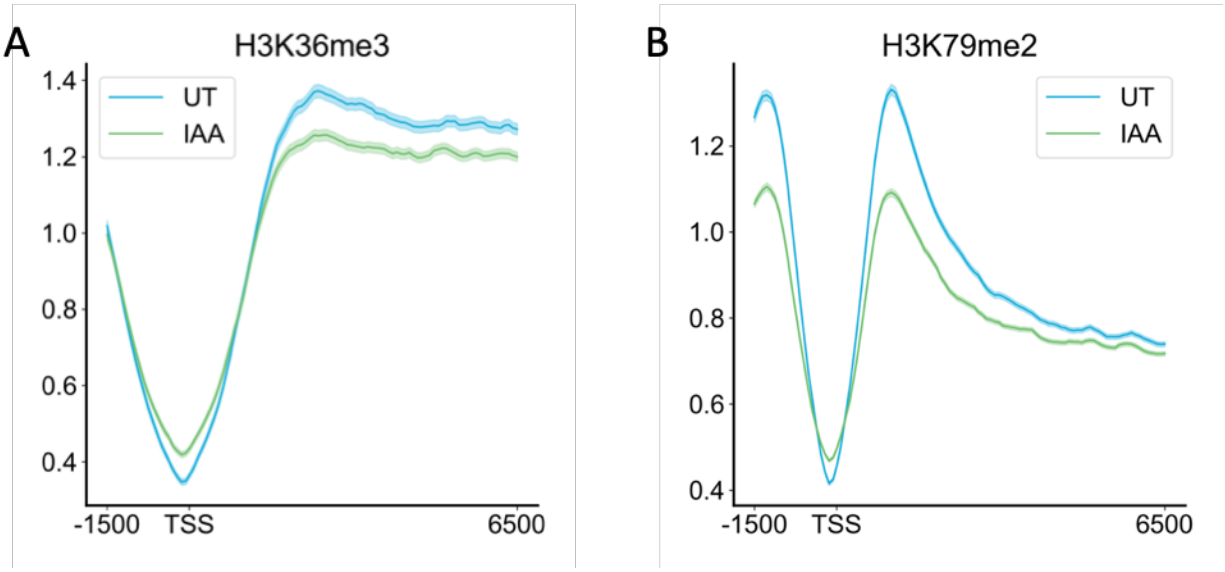

Figure 13, The transcription elongation associated histone modification ChIP-Seq metagene patterns after Cohesin depletion

The transcription elongation associated histone modification ChIP-Seq patterns after Cohesin depletion: A) The metagene plot of H3K36me3 ChIP-Seq after Cohesin depletion showed decreasing of H3K36me3; B) The metagene plot of H3K79me2 ChIP-Seq showed decreasing H3K79me2 after Cohesin depletion.

To further understand the underlying mechanism of Cohesin's effect on transcription, we analyzed published ChIP-Seq of several histone marks in the same Cohesin depletion cell model (47). Unexpectedly, we found two transcription elongation related histone marks (H3K36me3, H3K79me2) showed dramatic decreasing effects after Cohesin depletion (Figure 13), further suggesting a Cohesin' role in regulating transcription elongation possibly through affecting chromatin histone modifications.

Here, our nascent transcriptome data after Cohesin depletion have revealed that Cohesin depletion will affect the transcription of a large portion of MYC-targeted genes and will also impact the nascent transcription around TSS regions, possibly through RNA Pol II pause-release mechanisms. Although we uncovered an unprecedented link between MYC and Cohesin in transcription regulation here, we still don't understand how they are affecting each other and in what mechanism they are regulating the same targets' transcription. ChIP-Seq of MYC in Cohesin-depletion cells and Cohesin ChIP-Seq in MYC-depletion cells would be very helpful to understand MYC and Cohesin's dependencies on each other. Further experiments also need to be performed to consolidate the finding that Cohesin complex is regulating the pause-release at promoter regions in human cells. Study of different modified version of RNA Pol II (e.g. Pol II Ser2-phos, Pol II Ser5-phos, and total Pol II) residence on promoter regions after Cohesin depletion would be important to understand how Cohesin regulates Pol II transcription. Besides promoter regions, how Pol II transcription activities on distal enhancer regions are changed by Cohesin depletion is another important aspect to further study, especially given the recent evidences that enhancer RNAs per

se can regulate genes' transcription and chromatin structures. It is tempting to infer that Cohesin exerts its regulatory role through enhancer RNA controls.

2.4 Result and Conclusion II: Studying the Cohesin's function on signal-stimulated transcription regulation

After studying the Cohesin's direct effect on basal transcription, we next sought to ask whether Cohesin plays a role in the regulation of signal-stimulated transcription. Here, we utilized two well-characterized transcription stimulating signals- heat-shock (HS) and tumor necrosis factor-alpha (TNF-alpha). The basic principle here is that we treated two different signals in both normal cells (UT: Untreated) and Cohesin-depleted cells (IAA treated) and further measured their nascent transcriptome changes by performing PRO-Seq. For heat-shock treatment (43 degree for only 90 mins), we observed a dramatic change on nascent transcriptome, which includes 5483 out of 21862 genes down-regulated and 1612 genes up-regulated (Figure 14) (fold-change greater than 2). Next, we examined the Cohesin's effect on these HS-stimulated transcriptional changes. It turned out that Cohesin depletion has almost no effect on both HS up-regulated genes and HS down-regulated genes genome-widely (Figure 15a and Figure 15b), suggesting that Cohesin is generally not required for HS-stimulated transcriptional regulation.

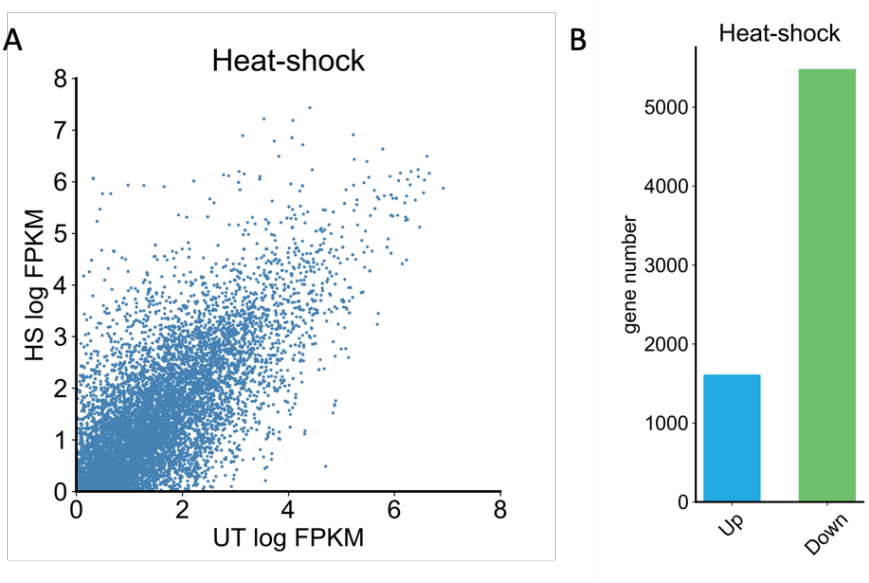

Figure 14, The nascent transcriptome alterations after Heat-shock stimulation

The nascent transcriptome changes after Heat-shock stimulation. A). The scatter plot of each genes' HS FPKM versus UT FPKMs; B) The bar-plot of HS up-regulated gene numbers and down-regulated gene numbers.

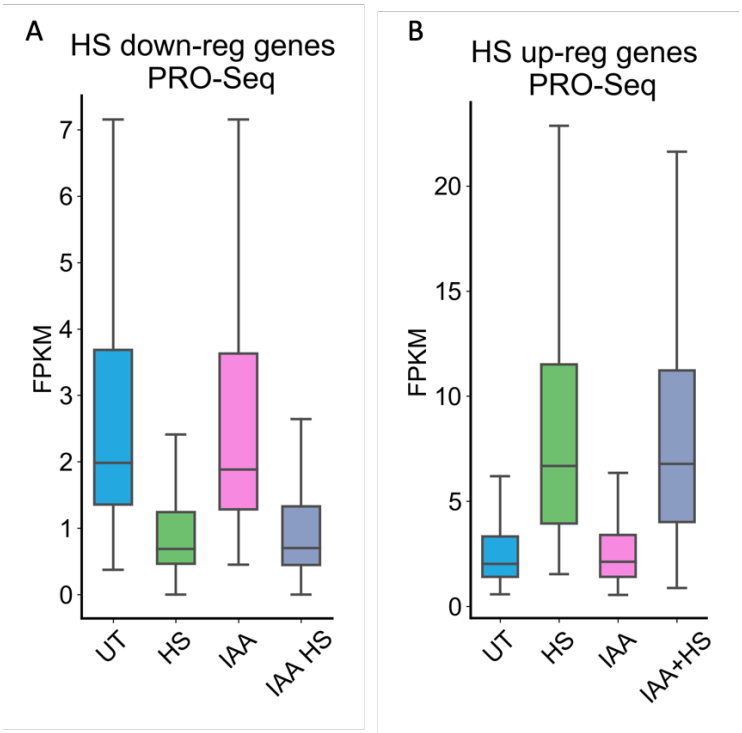

Figure 15, The transcriptional effects of Cohesin depletion on HS-stimulated transcriptional changes

The transcriptional effects of Cohesin depletion on HS-stimulated transcriptional changes: A). The boxplots of HS down-regulated genes FPKMs showed that Cohesin has almost no-effect on HS down-regulated genes transcriptional changes; B) The boxplots of HS up-regulated genes FPKMs showed that Cohesin has no-effect on HS up-regulated genes transcriptional changes.

In contrast to HS(Heat-shock), we observed a much more modest change of transcription after TNF-alpha stimulation. Overall, there are 226 genes that are up-regulated by TNF-alpha, and 74 genes that are down-regulated by TNF-alpha (Figure 16) (with a fold-change greater than 2). And combining with Cohesin's depletion effect, we found that Cohesin's depletion has very little effect on TNF-alpha up-regulated genes (Figure 17a), suggesting Cohesin is not required for TNF-alpha induced gene activation. However, we found that Cohesin's degradation has a stronger effect on reducing TNF-alpha down-regulated genes' reduction by TNF-alpha (Figure 17b), indicating that Cohesin may contribute to TNF-alpha's transcription inhibitory functions.

Although we observed that there are some effects of Cohesin depletion on small set of genes' responses to TNF-alpha (TNF-alpha inhibiting genes), the altered genes' number is limited. Thus, based on our HS and TNF-alpha stimulated transcription study, we concluded that Cohesin is largely not required for signal-stimulated transcription regulation, at least for two signals studied here- Heat-shock and TNF-alpha. Whether TNF-alpha and HS could represent another transcriptional-stimulus is unclear, it would be of great significance to extend our study to other pathways' signals, like TGF-beta and LPS, which are highly connected with disease and development.

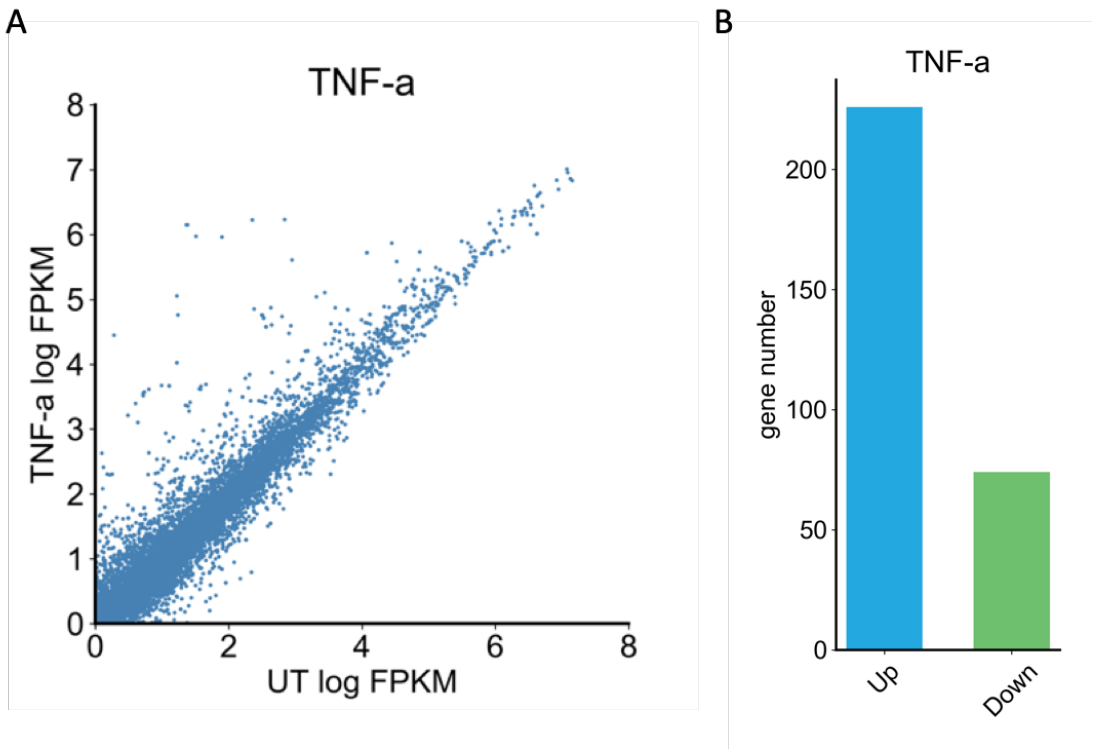

Figure 16, The modest transcriptional changes stimulated by TNF-alpha treatment

The modest transcriptional changes stimulated by TNF-alpha treatment. A). The scatterplot of each genes' TNF-alpha PRO-Seq FPKMs versus UT PRO-Seq FPKMs; B). The bar-plots of numbers of genes that are up-regulated by TNF-alpha, and down-regulated by TNF-alpha.

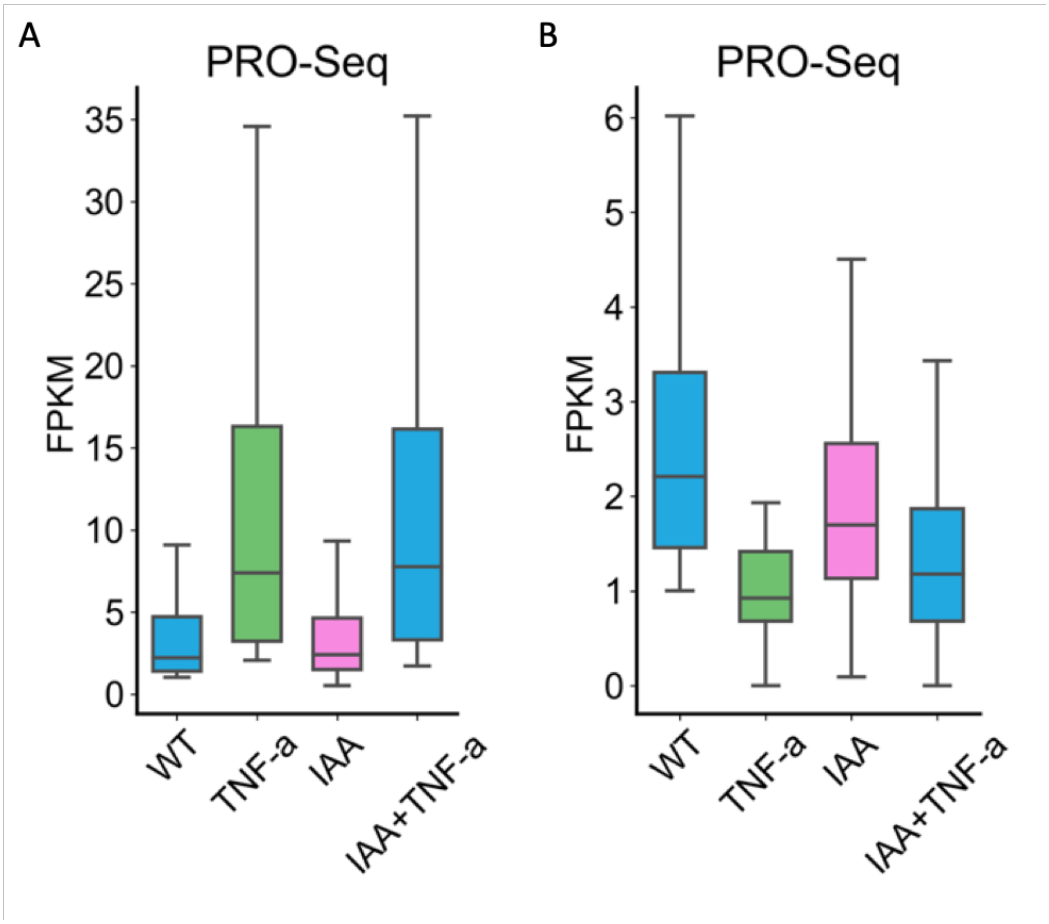

Figure 17, The transcriptional effects of Cohesin on TNF-alpha stimulated transcriptional changes

The transcriptional effects of Cohesin on TNF-alpha-stimulated transcriptional changes: A). The boxplots of TNF-alpha up-regulated genes FPKMs showed that Cohesin has almost no-effect on HS down-regulated genes transcriptional changes; B) The boxplots of TNF-alpha down-regulated genes FPKMs showed that Cohesin depletion would reduce the TNF-alpha down-regulated genes inhibition of transcription.

3. Dissecting Cohesin's role in 3D genome organization

3.1 Introduction to the Hi-C/HiChIP methods

Chromosome conformation capture (3C) was originally developed in 2002 (48), which has become the most powerful method to study chromatin structure since its development. The essence of 3C related methods is the proximity ligation of two spatially closed chromatin fragments. In a 3C or 3C derived assay, cells were first crosslinked to preserve the spatial information of chromatin contacts, and then the chromatin of cells were digested with certain restriction enzymes to kb-level fragmented chromatin complexes. The two crosslinked interacting chromatin fragments inside one chromatin complexes were further ligated into one fragment by T4 DNA ligase. Then this ligated chimeric chromatin fragment is ready to PCR by specific primers for detection purpose. Based on this key idea of 3C, there have been many derivatives emerged. With the development of next-generation sequencing, 4C (one-to-all) was developed to systemically quantify one specific regions' interaction frequencies with all other regions (49). Almost at the same time. 5C (many-to-many) method was developed to loop at all the interactions between chromatin fragments in a specific window of chromatin regions (50).

However, we were not able to map the chromatin interaction genome-widely until the development of Hi-C (High through-put Chromosome conformation capture) (15), an all-to-all 3C assay. The revolutionary Hi-C method revealed multiple layers of the hierarchical 3D genome structure, namely TADs, A/B compartments, and domain loops. Concordantly, another distinctive genome-wide 3C-derived method, named ChIA-PET (51), was devised to chart the protein-mediated chromatin interactome.

ChIA-PET data can provide a high-resolution chromatin interaction map than Hi-C, and study specific proteins' function in mediating chromatin interactions. However, ChIA-PET required a huge amount of cells (100 millions cells) as starting material and protocol itself is sophisticated, thus not robust enough to apply widely. Lately, a more effective and robust derivative of ChIA-PET called HiChIP was developed, enabling the study of enhancer-promoter connectome achievable in rare cells (52, 53). And more recently, Hi-C related methods were further extended to single cell level, which can study the chromatin contacts' heterogeneity in an unprecedented level (54–57). Besides all these 3C-derived methods, imaging-based methods like FISH combined with super resolution microscopy could serve as a complementary approach to validate the conclusion drawn from genome-wide 3C-related methods, or provide a new perspective of 3D genome organization.

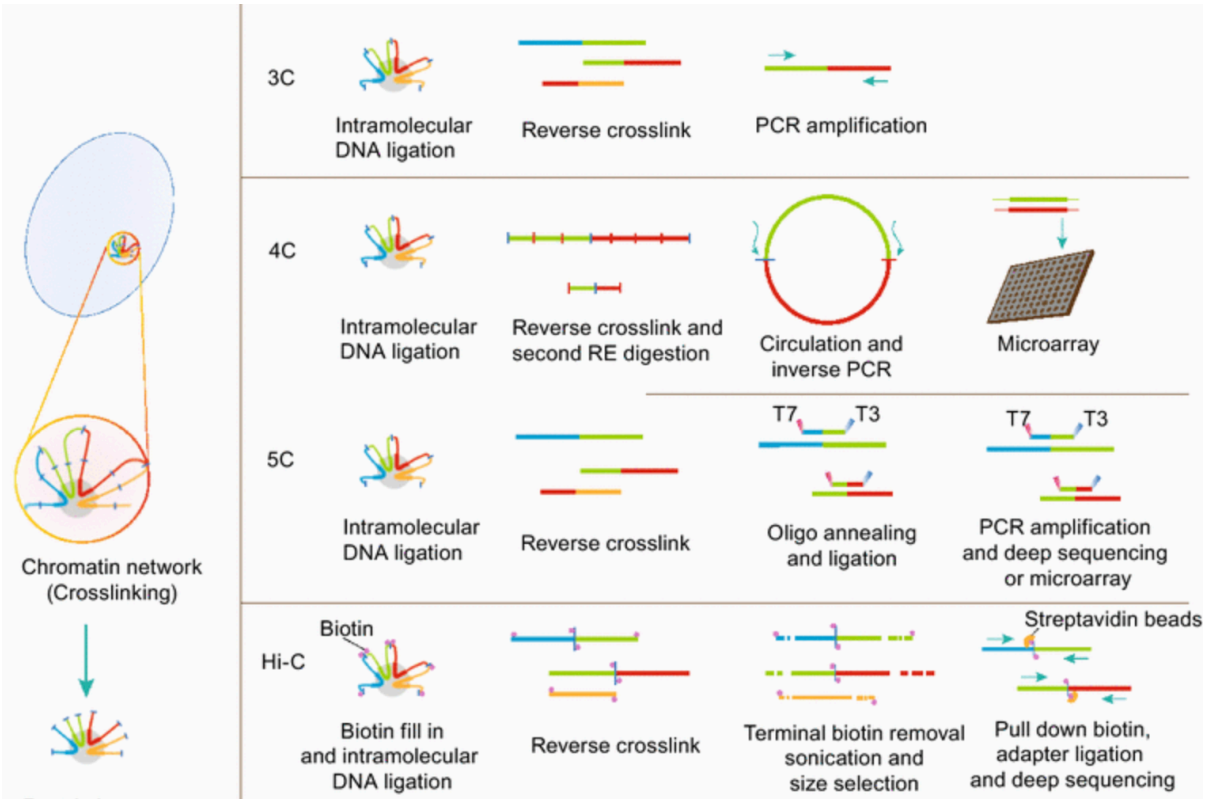

Figure 18, The brief summary of major 3C-related methods' diagram

The brief summary of major 3C-related methods' diagram. * This figure is adapted from (58), with the permission of authors.

3.2 Materials and Methods

Cell Culture and Treatment:

Rad21-mAID HCT-116 (#555) cells were obtained from Dr. Masato T. Kanemaki lab at NIG, Japan. And these cells were cultured in McCoy's 5A medium supplemented with 10% FBS in a 5 % CO2 humidified cell culture incubator. The auxin (IAA, Sigma) were treated at a final concentration of 500 uM for 6 hours. The TNF-alpha were treated at a final concentration of 25 ng/ul of 1 hour. The heat-shock of cells were performed at 43 degree for 90 mins in a 5% CO2 cell culture incubator.

Hi-C and HiChIP:

The starting material of Hi-C and HiChIP is around 10 million cells per sample. The cells were fixed and washed as routinely in ChIP-Seq and Hi-C Protocols (16). The fixed cells were resuspended in cold Hi-C Lysis buffer and rotate for 30 mins at 4 degrees. Then cells were pelleted down, washed with Hi-C Lysis buffer once, and resuspend with 100 ul 0.5% SDS. The samples were further incubated at 62 degree for 10 mins and quenched with 260 ul water and 50 ul 10% Triton X-100 at 37 degree for 15 mins. The enzymatic digestions were performed with 50 ul of 10X NEB Buffer 2 and 400 Units of MboI (NEB, R0147M) at 37 degree overnight. After digestion, the MboI was inactivated at 62 degree for 20 mins. The sticky ends of digested chromatins were filled-in and marked with biotin by treating DNA Klenow with biotin-dATP at 37 degree for 1hr. Then, we performed proximity ligation using NEB T4 DNA ligase with BSA and 10% Triton X-100 at RT for 4 hours. After ligation, the nuclei were pelleted-

down and lysis by Nuclear Lysis buffer. The nuclei lysis (chromatin) were sonicated with the parameter of 10/20 secs,25% Amp, 5 mins. The sonicated chromatins were further proceeded with ChIP or directly de-crosslink with ChIP elution buffer (containing Proteinase K). The de-crosslinked DNAs were purified by C1 beads to enrich biotin-labeled ligated DNAs. The incubated C1 beads were washed with TWB buffer for twice and proceeded to KAPA Hyper-sensitivity kit for library constructions.

Hi-C and HiChIP data analysis:

Modified HiC-Pro (59) was used to process Hi-C/HiChIP raw data first. Briefly, sequencing reads of each end of Hi-C and H3K27ac HiChIP libraries were independently mapped to reference genome version hg19 with SNAP-aligner. Then we trimmed the double MboI sites containing reads from all the unmapped reads and further mapped these trimmed reads to the reference genome again with SNAP-aligner. The two rounds of unique mapped reads were combined, paired, and assigned to each digested chromatin fragment in the genome. The reads pairs with dangling-end and self-ligated reads were considered as un-valid pairs and removed. All valid pairs were merged to build binned contact matrices. For Hi-C, the binned contact matrices were further normalized by using iterative correction method, and for HiChIP binned matrices were not balanced. All Hi-C and HiChIP matrices were further transferred to mcool format for visualization purpose in HiGlass (60). The insulation scores of potential boundaries were calculated as previously described (61). The APA (Aggregation Peak Analysis) was performed as previously described (62). The loop's coordinates are randomly shift 10 times, and resulted matrices are served as the normalization factors

for APA analysis. HiChIP loop calling was performed by CID (Chromatin Interaction Discovery) (63). The differential loop quantities analyses were performed with custom scripts. The P(s) analysis was utilized to depict the relationship between chromatin interaction probability and linear distance between two DNA fragments as previously described(64).

3.3 Results and Discussion I: H3K27ac HiChIP in Cohesin-depleted cells revealed regulatory loops persisted after acute Cohesin depletion.

Previous study showed that Cohesin depletion could almost eliminate all the loops identified by in situ Hi-C (47). Given the results that Cohesin depletion has nuanced effects on protein-coding genes' transcription genome-widely acutely, it is hard to imagine that our genome's transcriptional regulation is largely independent on chromatin loops. We hypothesized that there are still some chromatin loops that are independent of Cohesin depletion and are regulating gene transcriptions, termed regulatory loops. To testify this hypothesis, we sought to perform H3K27ac HiChIP to identify chromatin loops that are connecting transcriptionally-active regions in Cohesin depleting cells. We successfully obtained H3K27a HiChIP maps in both UT and IAA cells. The snapshot of our HiChIP data, combining with in situ Hi-C data in the same cell line, clearly showed that there are chromatin loops that are not well detected in situ Hi-C map (Figure 19). Moreover, these under-detected chromatin loops can persist after Cohesin acute depletion (Figure 19). To ask whether this is a genome-wide phenomenon, we performed APA (Aggregate Peak Analysis) analysis on all putative enhancer-promoter loops in HCT-116 cells. The APA results unexpectedly showed that these putative enhancer-promoter loops persisted in total after Cohesin depletion, albeit their intensities decreased (Figure 20). We also performed APA analysis on another set of loops-domain loops (usually CTCF loops, identified by in situ Hi-C), and the results clearly showed that domain loops were abolished by Cohesin depletion, serving as a control for regulatory loops described here (Figure 20).

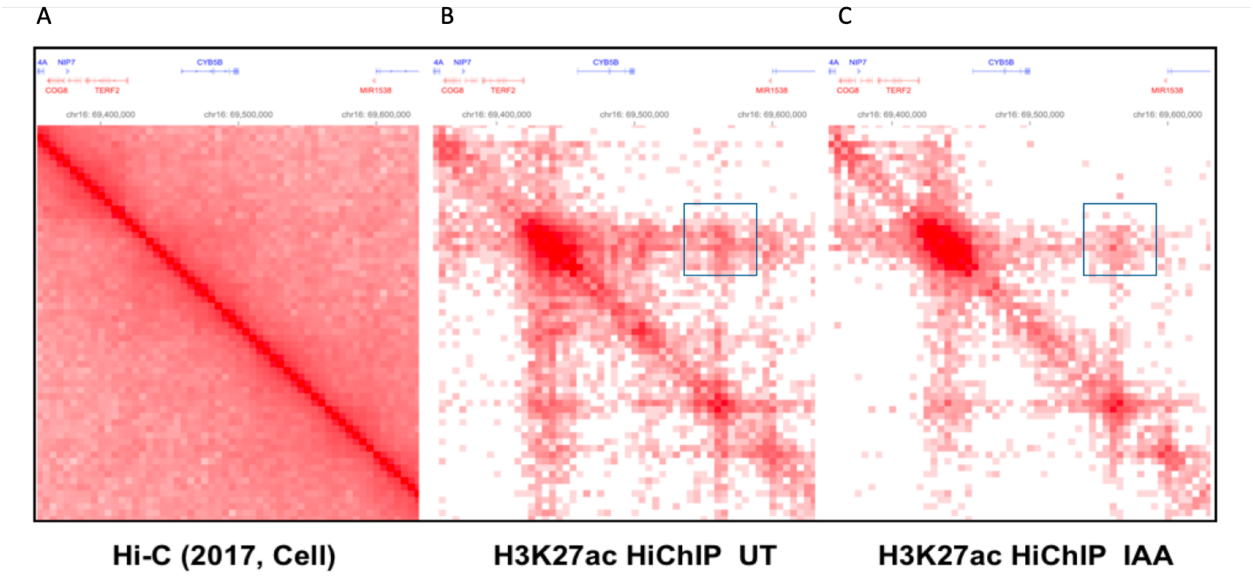

Figure 19, The snapshot of Hi-C and H3K27ac HiChIP matrices

The snapshot of Hi-C and H3K27ac HiChIP matrices: A). The published in situ Hi-C map in HCT-116; B). H3K27ac HiChIP map in UT cells; C). H3K27ac HiChIP map in IAA cells. The blue box here indicates a putative enhancer-promoter loop (TERF2 gene) which are persisted after Cohesin depletion.

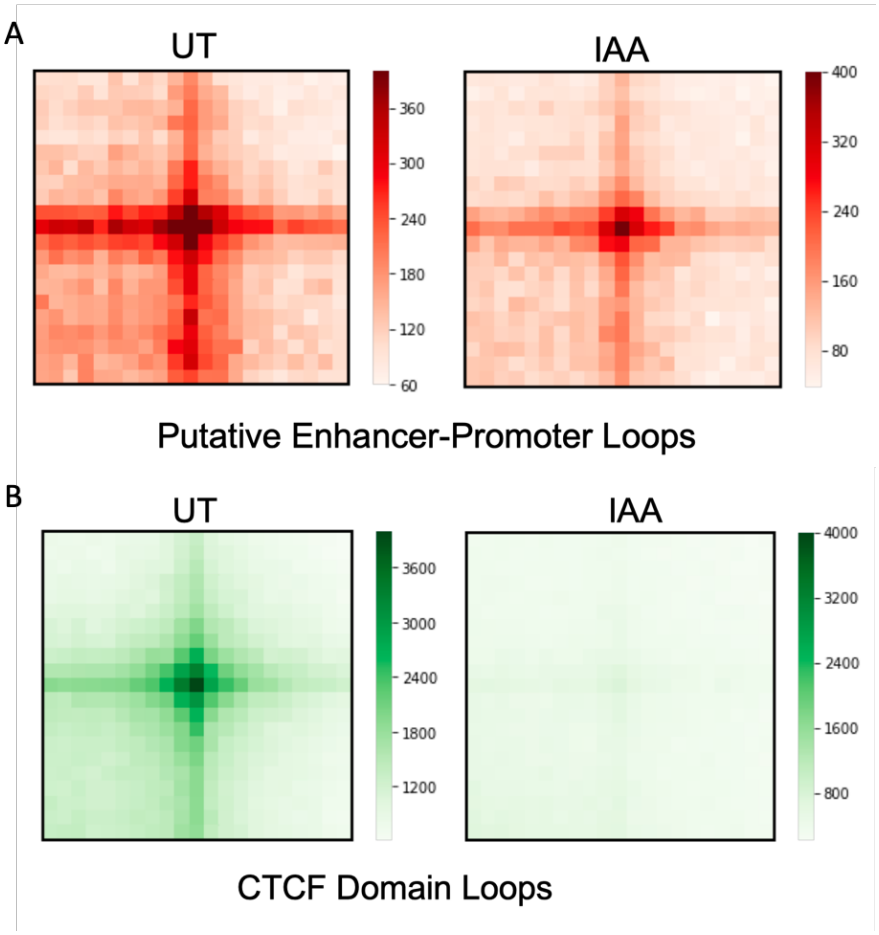

Figure 20, The APA analysis of putative enhancer-promoter loops in H3K27ac HiChIP data

The APA analysis plot of putative enhancer-promoter loops (A) and CTCF domain loops (B) showed that domain loops were totally abolished while regulatory loops persisted genome-widely.

To further identify these persisted regulatory loops in a non-biased manner, we performed HiChIP loop calling using a newly developed loop-calling algorithm named CID (Chromatin Interaction Discovery) (63). Genome-widely, we have identified 129,937 chromatin loops based on our H3K27ac HiChIP data, and we further compared these chromatin loops intensities in between UT and IAA two conditions by normalizing the each identified loop's intensity with its anchors' 1D reads depths. After calculating a multiple-testing corrected Binomial FDR, we identified 65,432 chromatin loops were significantly reduced after acute depletion of Cohesin, suggesting that there are around half of H3K27ac associated chromatin loops were largely Cohesin independent (Figure 21A). Interestingly, these persisted loops' lengths were significantly shorter than those altered loops (Figure 21B).

How these large portions of chromatin loops detected from HiChIP persisted after Cohesin depletion, and what other mechanisms are mediating these Cohesin-independent loops are still unknown. A large-scale combined analysis with different transcription factors ChIP-Seq in these loops' anchors would be helpful to identify potential protein factors that are mediating these loops. Also, enhancer transcription products-eRNAs could also facilitate these loops' formation independent of Cohesin. This RNA mediating hypothesis needs to be tested in a more comprehensive way than previous studies.

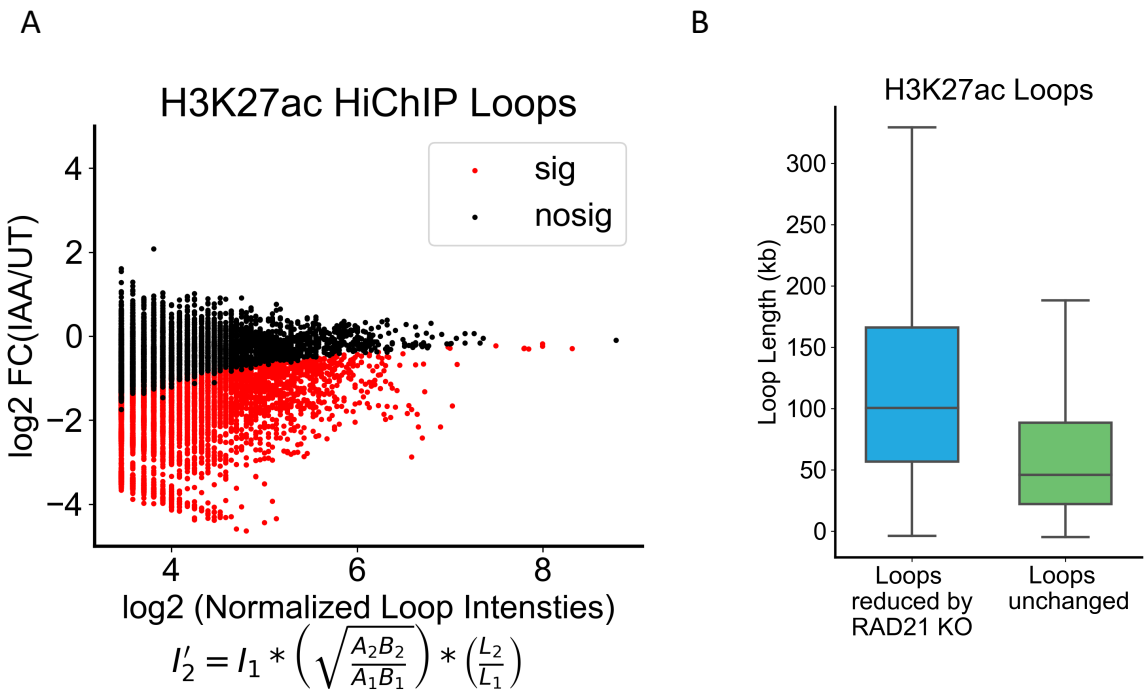

Figure 21,The comparison of H3K27ac associated chromatin loops' normalized intensities in UT and IAA cells

A). The comparison of H3K27ac associated chromatin loops' normalized intensities in UT and IAA cells. The scatter plot of normalized loop intensities in UT and IAA cells, each dot here represents a single chromatin loop that is called from H3K27ac HiChIP datasets. The formula here showed how the expected interaction was computed. The red dot indicates a significant loop that has a corrected FDR < 0.05. B). The boxplots of loop lengths of RAD21 KO persisted loops and altered loops. P-value <0.001, (Mann-Whitney test)

3.4 Results and Discussion II: Underlying mechanisms of Cohesin antagonized super-enhancer loops

As previously reported, we also detected super long-distance loops between super-enhancers were enhanced by Cohesin depletion from our H3K27ac HiChIP data (Figure 22). However, these super-enhancer loops' formation and function are not understood. Thus, we sought to study these super-enhancer loops' underlying mechanisms with our model systems and H3K27ac HiChIP technique. Since these super-loops are formatted between super-enhancer regions, we sought to target proteins that are highly enriched on super-enhancers (Higher H3K27ac). BRD4 is considered as our first target, not only because BRD4 is H3K27ac reader, but also there is a potent BRD4 inhibitor-JQ1 available. To specifically target BRD4 here, we treated JQ1 with both UT and IAA-treated cells and performed H3K27ac HiChIP. The APA analysis of extra-long distance (greater than 2Mb) SE (Super-enhancer) loops in UT and IAA samples confirmed that SE-loops were enhanced genome-widely after depleting Cohesin here (Figure 23a). Surprisingly, JQ1 treatment greatly diminished the enhancement of Cohesin depletion on these SE-loops, and has minor reduction effect on basal SE-loop intensities (Figure 23a and Figure 23b). These analyses suggest that BRD4 is crucial for enhancement of extra long-distance SE loops. It is unlikely for a distance greater than 2Mb chromatin loop to be formed through loop extrusion. Thus, these SE loops must be formed by other biophysical models rather than loop extrusion. Considering previous study showed that BRD4 is a highly intrinsic disordered protein and have strong propensities to form phase separation condensates. It is tempting to speculate that these SE-loops are formed by BRD4 phase separation.

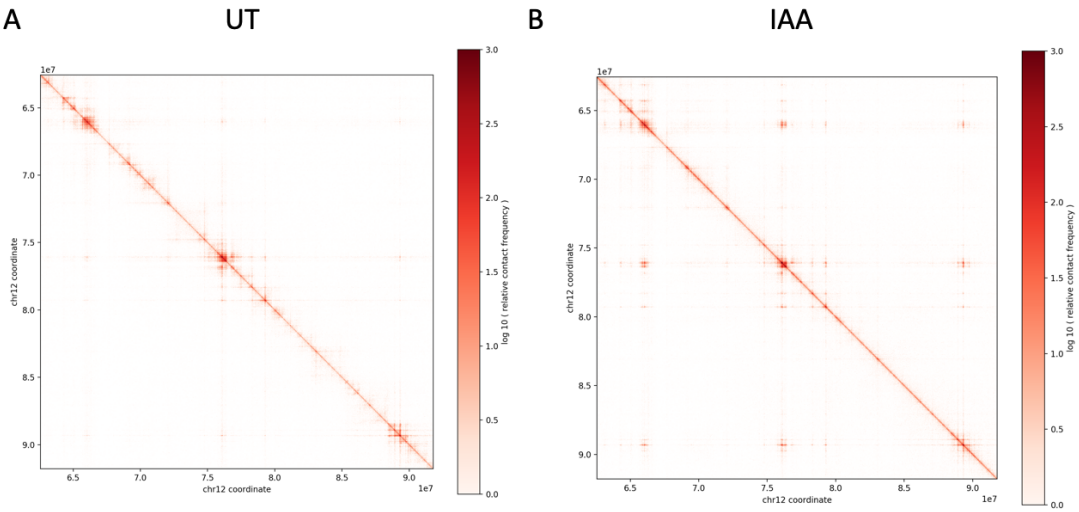

Figure 22,The dramatic enhancement of SE (Super-enhancer) loops after Cohesin depletion in H3K27ac HiChIP dataset

The dramatic enhancement of SE (Super-enhancer) loops after Cohesin depletion in H3K27ac HiChIP dataset. A) The H3K27ac HiChIP snapshot of KITLG gene locus in UT dataset. B) The H3K27ac HiChIP snapshot of KITLG gene locus in IAA dataset. The color bar here indicates the log relative interaction frequencies.

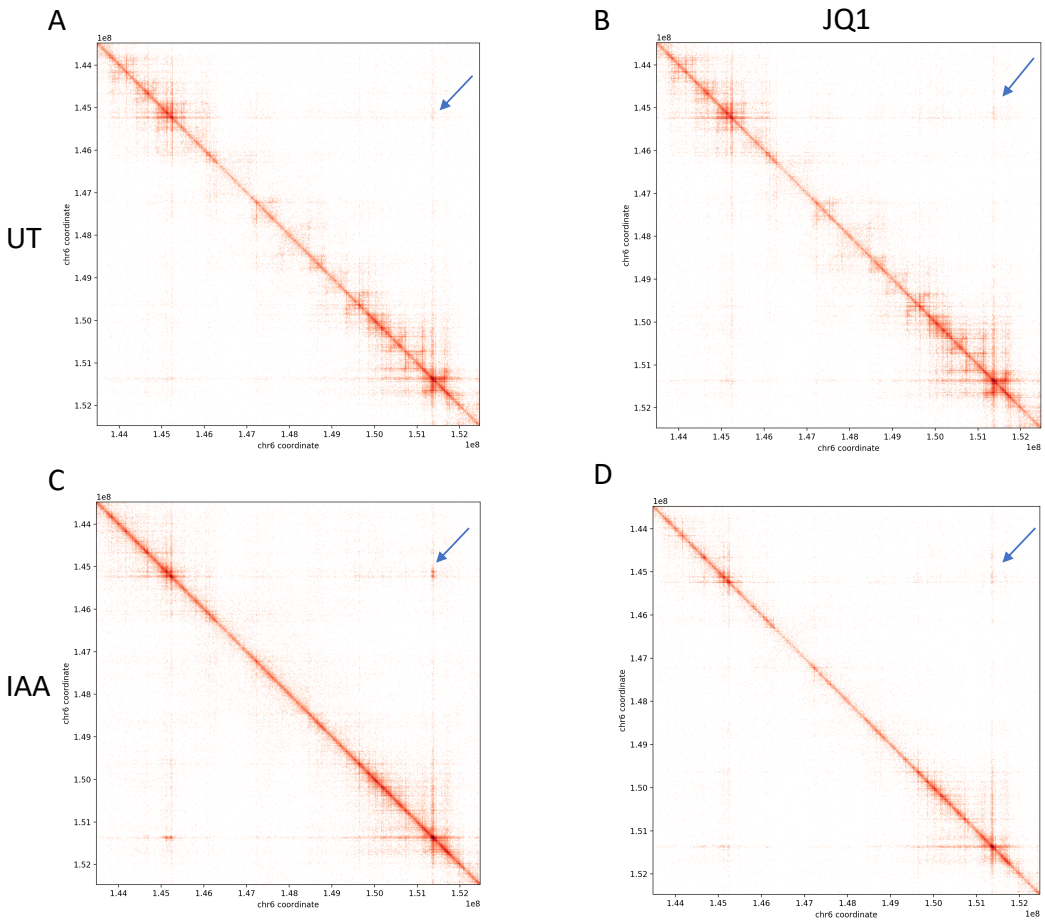

Figure 23a, The Cohesin depletion enhanced extra-long-distance SE loops were diminished by BRD4 inhibitor-JQ1

The Cohesin depletion enhanced extra-long distance SE loops were diminished by BRD4 inhibitor-JQ1: The H3K27ac HiChIP snapshot at a specific region (chr6: 144mb – 152mb) in four different conditions, UT (Untreated, A), JQ1 (B), IAA (C), and JQ1+IAA (D), showed that IAA induced SE loops were largely reduced by JQ1 treatment.

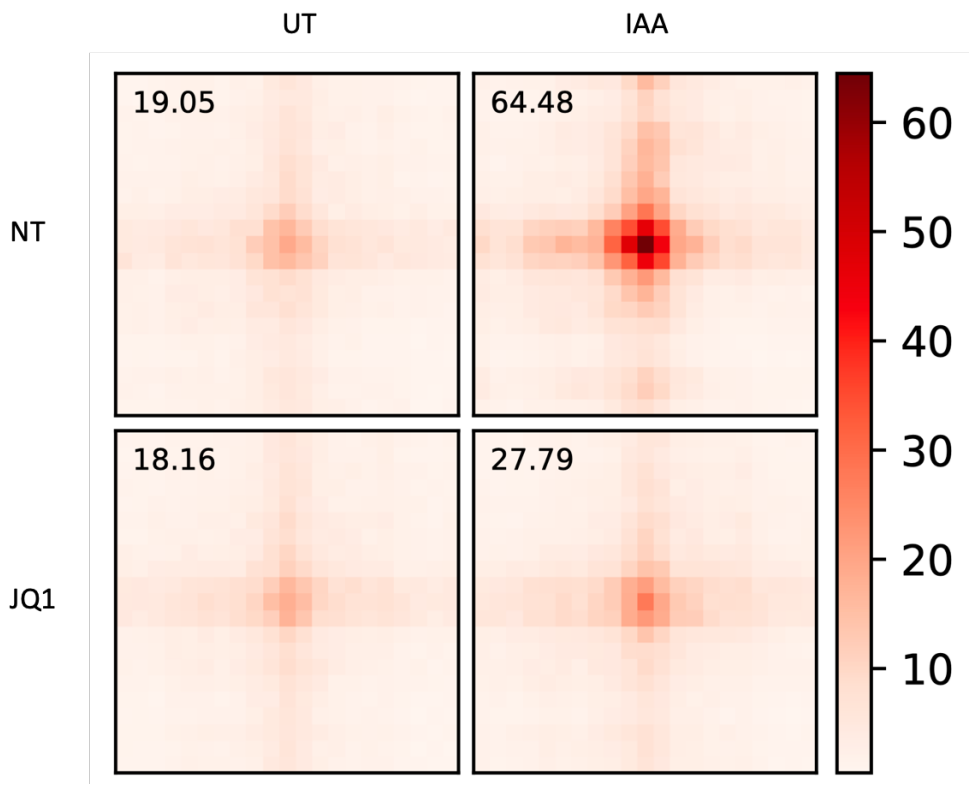

Figure 23b, The APA Analysis of extra-long-distance SE loops after JQ1 and IAA treatment

The APA Analysis of extra long-distance SE loops after JQ1 and IAA treatment; The window size here is 400kb, the left-up numbers indicate the APA enrichment score.

To validate whether these SE-loops are indeed phase-separated loops, we sought to perturb these SE-loops with two general phase separation perturbation chemicals 1,6HD (perturbing hydrophobic interactions) and NH4OAC (perturbing electrostatic interactions), and performed H3K27ac HiChIP. Surprisingly, our HiChIP analysis clearly showed that these enhanced SE loops were diminished by both of these drugs' 10 mins treatment, both in Hi-C map and APA analysis (Figure 24 and Figure 25), demonstrating these SE loops were formed, at least facilitated, by BRD4 phase separation (Figure 22). To further confirm this phenomenon, we performed in situ Hi-C, which is not biased towards active regions, after these two drugs treatment. Concordantly, our Hi-C APA analysis of these SE-loops were consistent with our H3K27ac HiChIP results, further consolidating the conclusion that extra long-distance SE loops are formed by phase separation (Figure 26). In summary, our results here revealed that Cohesin antagonized super enhancer loops are formed through a novel mechanism, LLPS of BRD4, which is distinctive from the loop extrusion model.

Based on this intriguing data, it is tempting to hypothesize that Cohesin complex is a general antagonizing factor for LLPS events inside nuclei. Since a large portion of nuclear proteins (e.g., transcription factors) have the strong propensities to form LLPS mediated condensates, thus cells must devise multiple ways to control and regulate all these LLPS events inside nuclei. Cohesin complex, and even TAD structures, could be one possible mechanism to control the LLPS of genomic regions, like SEs (Super-enhancers). Further biochemical study and imaging-based research about Cohesin's role in nuclear LLPS events are required to test this hypothesis.

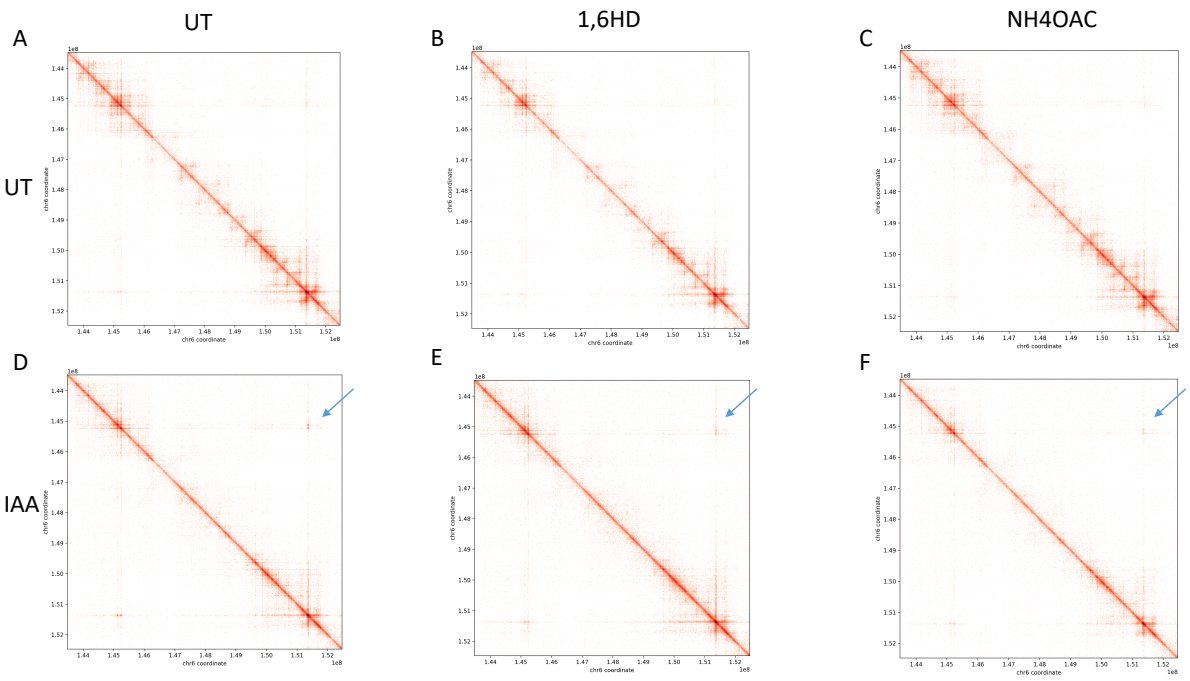

Figure 24, The extra-long-distance Super enhancer loops are perturbed by 1,6 HD and NH4OAC

The H3K27ac HiChIP snapshot of chr6: 144Mb – 152 Mb regions in six different conditions: UT (Untreated) (A), 1,6HD (B), NH4OAC(C), IAA (D), IAA+1,6HD (E), IAA + NH4OAC (F). The blue arrow here indicates the specific extra-long-distance SE loop, which is perturbed by 1,6 HD and NH4OAC, two general phase separation perturbation drug.

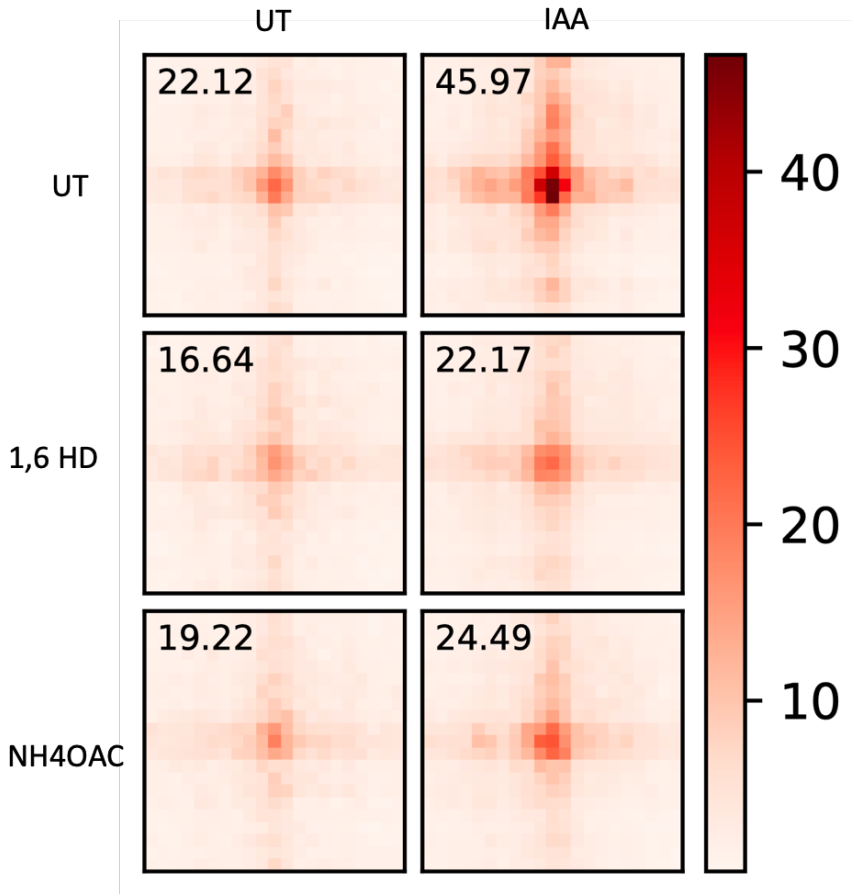

Figure 25, H3K27ac HiChIP APA analysis of extra-long-distance SE loops after 1,6HD, NH4OAC, and IAA treatment

APA Analysis of extra long-distance SE loops (H3K27ac HiChIP) after 1,6 HD, NH4OAC and IAA treatment; The window size here is 400kb, the left-up numbers indicate the APA enrichment score.

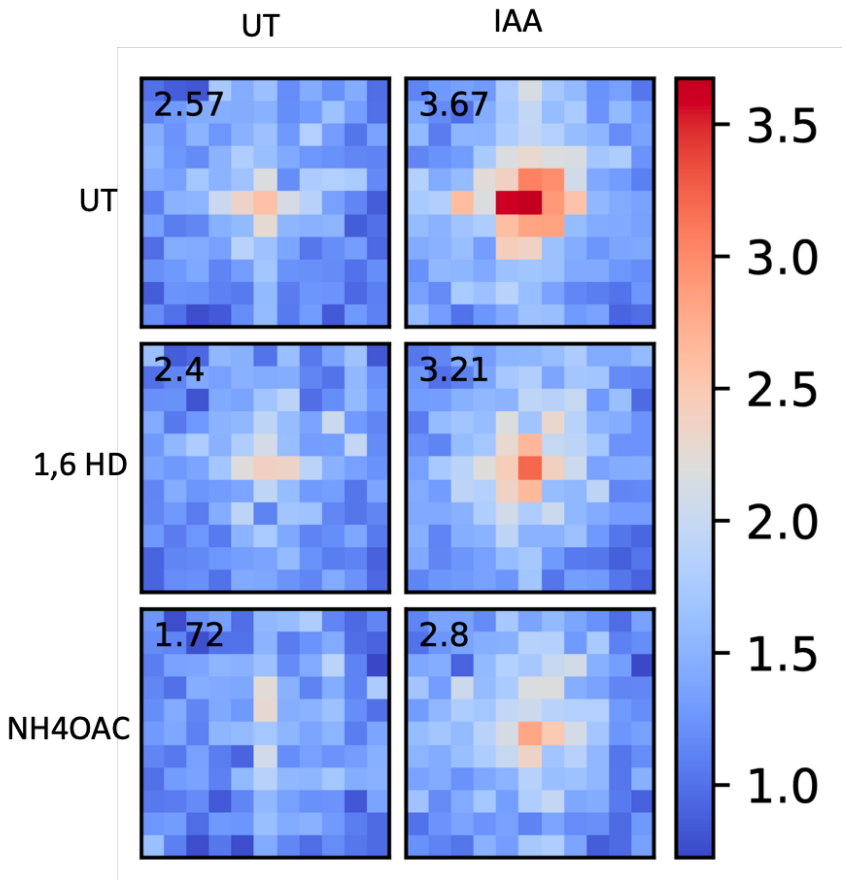

Figure 26, Hi-C APA analysis of extra-long-distance SE loops after 1,6HD, NH4OAC, and IAA treatment

APA Analysis of extra long-distance SE loops (Hi-C) after 1,6 HD, NH4OAC and IAA treatment; The window size here is 400kb, the left-up numbers indicate the APA enrichment score.

4. Condensin I ‘s role in transcription regulation and 3D genome organization

4.1 Materials and Methods:

The major techniques and data analysis methods used in this part has been described and discussed in section 2 and section 3.

4.2 Results and Discussion I: Transcriptional effect after acutely depleting Condensin I complex

Multiple previous studies has suggested Condensin complexes also play an important role in transcription regulation through organizing the chromatin organization (36). However, all previous studies were using siRNA/shRNA-based knockdown which usually last more than few days and could involve any secondary effect, and no research has been performed to assess the direct effect on transcription by acute depletion of Condensin complex. Here, starting from Condensin I complex, we utilized a previous established cell model NCAPH-mAID HCT-116 cell (65) to quickly degrade Condensin I complex in short time (6 hour). We have used both microscopy-based approach and immune-blot to validate the complete degradation of NCAPH in our models (Figure 27).

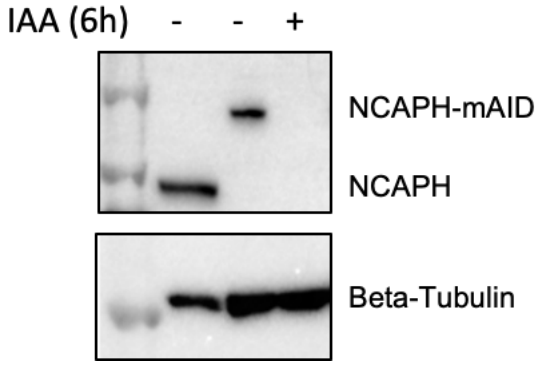

Figure 27, The validation of Condensin I depletion in HCT-116

The validation of Condensin I depletion in HCT-116. The lower band of NCAPH in first lane indicate the untagged endogenous NCAPH in another HCT-116 cell line that does not contain any NCAPH-mAID allele, the higher band of NCAPH in other two lanes indicate the mAID tagged NCAPH.

After confirming the complete removal of Condensin I in 6 hours, we performed PRO-Seq to directly measure the nascent transcriptome by depleting NCAPH. Unexpectedly, the nascent transcriptome of wild-type cells (abbreviated as UT) and NCAPH depleted cells (abbreviated as IAA) is highly consistent (Figure 28a). By counting the FPKM value of all protein-coding genes along the gene body, we identified 23,344 transcribing RefSeq annotated genes with a cutoff of genes' average FPKM is greater than 1. Among these transcribed genes, there are only 4 genes that are down-regulated with a fold change of 2, and none of them are up-regulated. If we loosen out cutoff to 1.3-fold, we can identify 94 down-regulated genes and 309 up-regulated genes, which have much less down-regulated genes than Cohesin depletion (Figure 28b). Also, we performed functional enrichment analysis of these down-regulated genes, and we couldn't find a strong enrichment of any functional annotation terms. These analyses suggest that Condensin I depletion has a very minor effect on basal transcription.

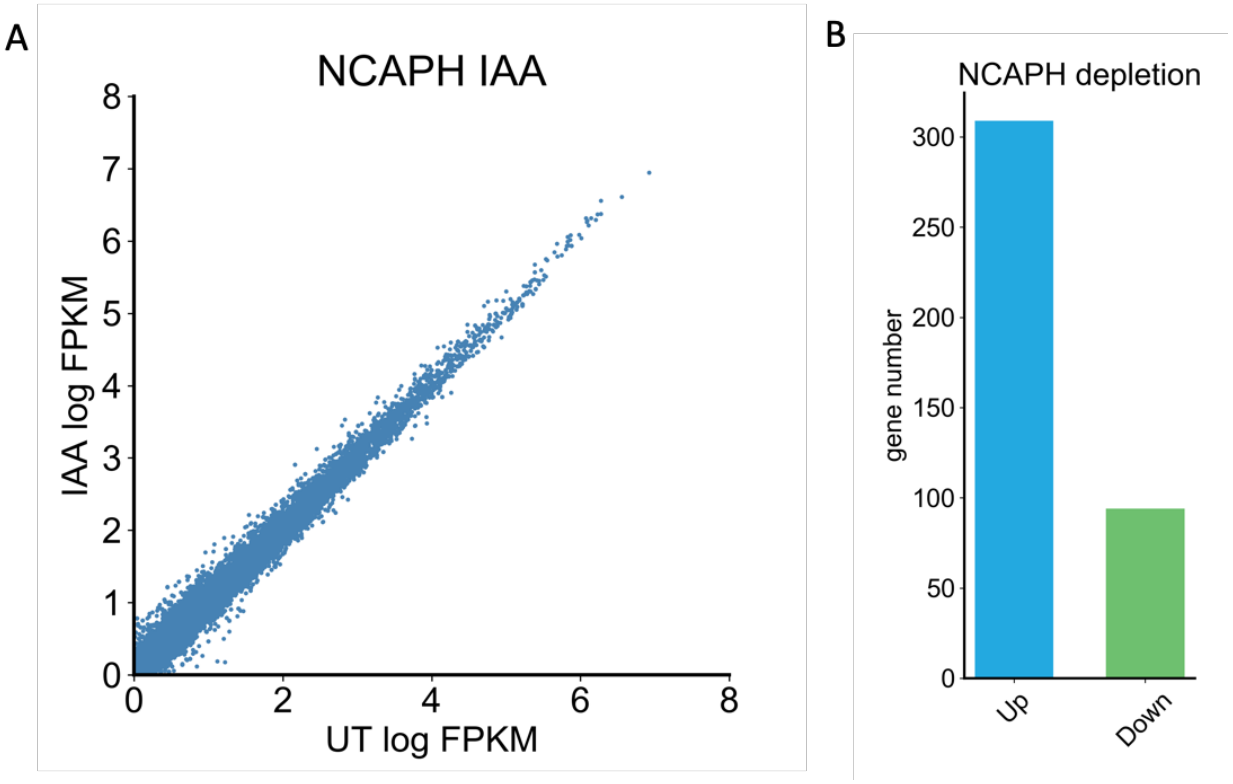

Figure 28, The nascent transcriptome after acute depletion of Condensin I complex

The nascent transcriptome after acute depletion of Condensin-I complex subunit NCAPH: A). The scatterplot of each gene's PRO-Seq FPKM values in both UT and IAA cells. B). The bar-plot of up-regulated gene number and down-regulated gene number in PRO-Seq. UT here denoted as untreated condition, IAA here denoted as 6 hours auxin (IAA) treated condition.

Since previous work demonstrated that Condensin I complex could be recruited to interphase chromatin highly in an estrogen (E2) -inducible manner, we hypothesize that Condensin I could play a regulating role in signal-stimulated transcription, rather than just basal transcription. To test this hypothesis, we also applied TNF-alpha signaling here as a stimulus agent and performed PRO-Seq on that. After counting the FPKM value, it is consistent with Rad21-mAID HCT-116 cells, that TNF-alpha would increase around 200 genes' transcription in 1 hour. However, in contrast to the previous conclusion that Cohesin depletion almost had no effect on TNF-alpha induced genes' transcription, we observed a strong increase of TNF-alpha induced genes' transcription level after acutely depleting Condensin I, genome-widely ($P<0.001$) (Figure 29 a,b). And for TNF-alpha inhibited genes, we also observed a significant effect that Condensin I depletion reduced the TNF-alpha inhibitory effects ($P< 0.001$) (Figure 29 c). These results suggest a role of Condensin I in antagonizing TNF-alpha mediated transcription responses. And in this TNF-alpha case, it is quite different from estrogen-mediated transcription effects, as previous work suggest Condensin I is facilitating for E2-stimulated transcription. Nevertheless, our work here revealed the Condensin I is not required for basal transcription regulation, but is antagonizing TNF-alpha stimulated transcription regulation.

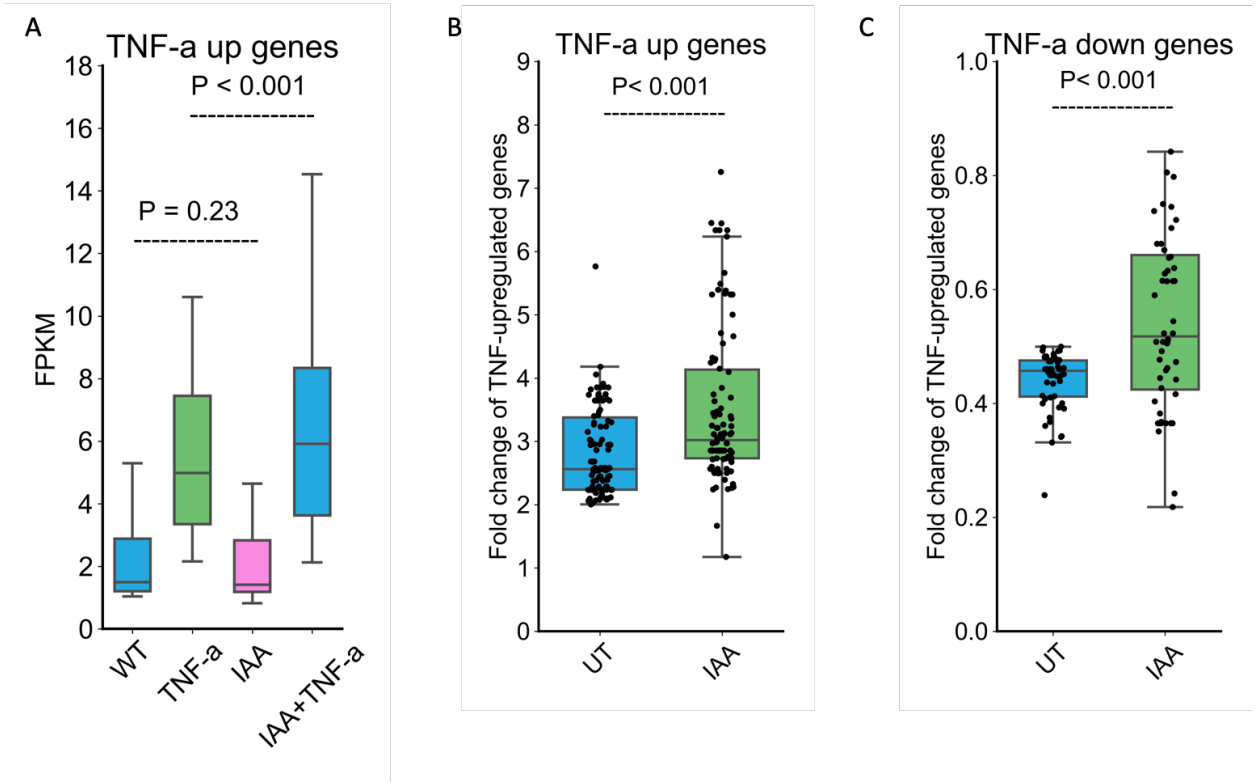

Figure 29, Condnesin I depletion's effect on TNF-alpha stimulated transcription response.

Condensin I depletion's effect on TNF-alpha stimulated transcription regulation. A) The boxplot of FPKM values' TNF-alpha upregulated genes in four different PRO-Seq datasets (WT, TNF-alpha, IAA, IAA+ TNF-alpha); B) The boxplot of fold change of TNF-alpha up-regulated genes' FPKM in both UT and IAA conditions. C) The boxplot of fold change of TNF-alpha down-regulate genes' FPKM in both UT and IAA conditions. The P-vaules here are calculated with Mann-Whitney U test.

4.3 Results and Discussion II: The Condensin I's direct role in 3D genome organization

The SMC complex Cohesin, has been widely demonstrated to be a bona fide key organizer in 3D genome structure. And a major model of how 3D genome structure, specifically chromatin loops, is loop extrusion through Cohesin complex. Given the fact that the similar ring structure that Cohesin and Condensin share, people hypothesized that Condensin can also have the loop extrusion phenomenon and further drive the formation of specific chromatin structures. And recent efforts have been made to show directly that Condensin complex can extrude on the DNA for the first time (66), further support the hypothesis that Condensin is a chromatin organizer. In several mammalian systems including embryonic stem cells, Condensin complex has been showed to reside on enhancer regions, and potentially regulate enhancer-promoter loops and TADs (35–37). However, there is not genome-wide evidence yet to directly demonstrate Condensin complex regulate 3D genome structure in mammals. Here, to tackle this knowledge gap, we took the advantage of this NCAPH-mAID cell model to systematically study the 3D genome organization after acute depletion of Condensin I. We utilized Hi-C to map the genome-wide chromatin interactions after Condensin I depletion in NCAPH-mAID HCT-116 cells. Overall, we can get 200~300 million effective read pairs, suggest a high quality of our Hi-C data and deep depth. At the scale of 20Mb, we didn't observe strong differences between NCAPH-UT Hi-C maps and NCAPH-IAA Hi-C maps (Figure 30). Specifically, the triangle shaped Hi-C map pattern indicating the TADs structure are distinctive in both maps, and the plaid-

like Hi-C map pattern indicating the A/B compartment are also well-observed in both maps. In a quantitative manner, the UT/IAA comparative Hi-C map also showed that the Hi-C interaction differences between UT and IAA majorly come from the region outside of TAD structures, indicating that Condensin I depletion doesn't affect chromatin organization at TAD level. We further validate these conclusions at a genome-wide scale by calculating the insulation scores of all potential TAD boundaries. The insulation scores scatterplot also suggest that TADs' were not much affected by Condensin I depletion genome-widely (Figure 31).

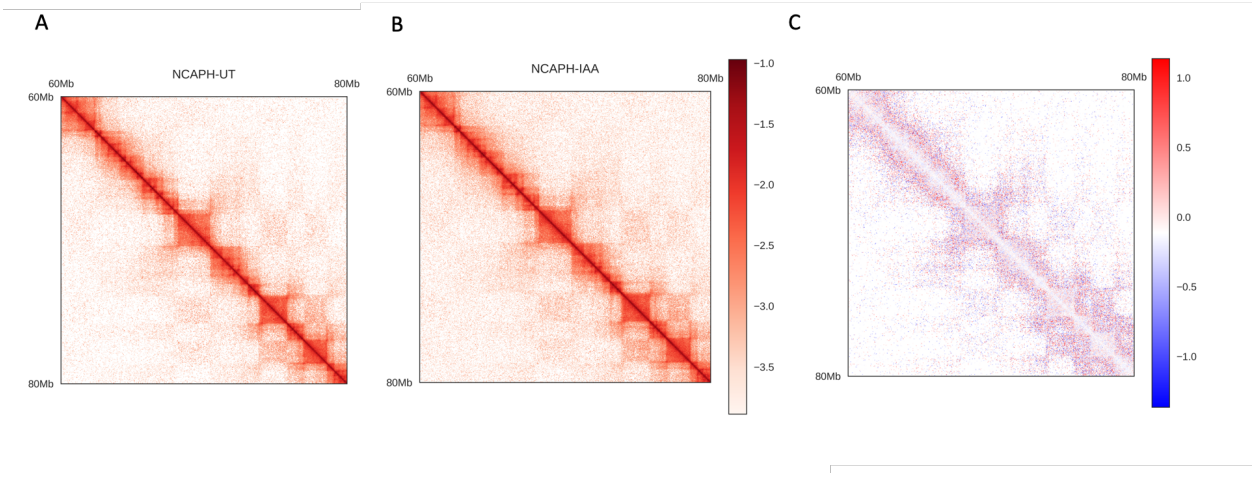

Figure 30, The snapshot of Hi-C matrices in Condeinsin I depleted cells

The snapshot of chromatin interaction map (Hi-C) in NCAPH-mAID cells. A). The Hi-C map from chr1:60Mb to 80Mb in NCAPH UT cells, the color scale here indicates the log10 value of normalized interaction contacts; B) The Hi-C map from chr1:60 to 80 Mb in NCAPH IAA cells, the color scale here indicates the log10 value of normalized interaction contacts. C) The comparative Hi-C map of UT/IAA from chr1: 60Mb- 80 Mb, the color scale here indicates the log10 ratio (UT/IAA).

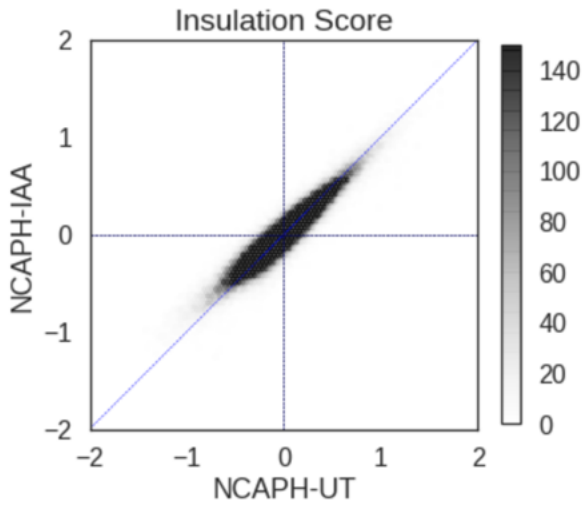

Figure 31, The insulation score analysis in Condensin-I depleted cells

The insulation score analysis of Hi-C map in both NCAPH-UT and NCAPH-IAA samples. Each dot in the scatter plot stands for on potential TAD boundary, the insulation score represents the potential of that region to shape a TAD in Hi-C map.

Next, we asked whether there is any change after depleting Condensin I at the domain loops level. Since our current sequencing depth is still not deep enough for a precise de novo loop calling using HICCUPS algorithms, we performed APA (Aggregate Peak Analysis) on previous published domain loops (~3000 loops) by using our Hi-C data. Based on APA results (Figure 32), we can conclude that Condensin I depletion caused a minor reduction of domain loops enrichment genome-widely, but this effect is not comparable to Cohesin's effect on domain loops. Furthermore, to obtain a global view of the chromatin contacts at different genomic distances, we computed P(s) curve using these two Hi-C datasets. From P(s) curve, we can clearly observe that there are two major changes between UT and IAA conditions (Figure 33). The first change of Hi-C interaction frequencies is at the distance ranging from 40kb to 500 kb, which usually resemble the chromatin contacts inside TADs. This is consistent with the conclusion that domains loops have a minor reduction after depleting Condensin I. The other major change of Hi-C interaction frequencies is at the distance ranging from 2Mb to 25Mb, representing the chromatin interaction changes beyond TADs that we observed on Hi-C heatmap (Figure 30).

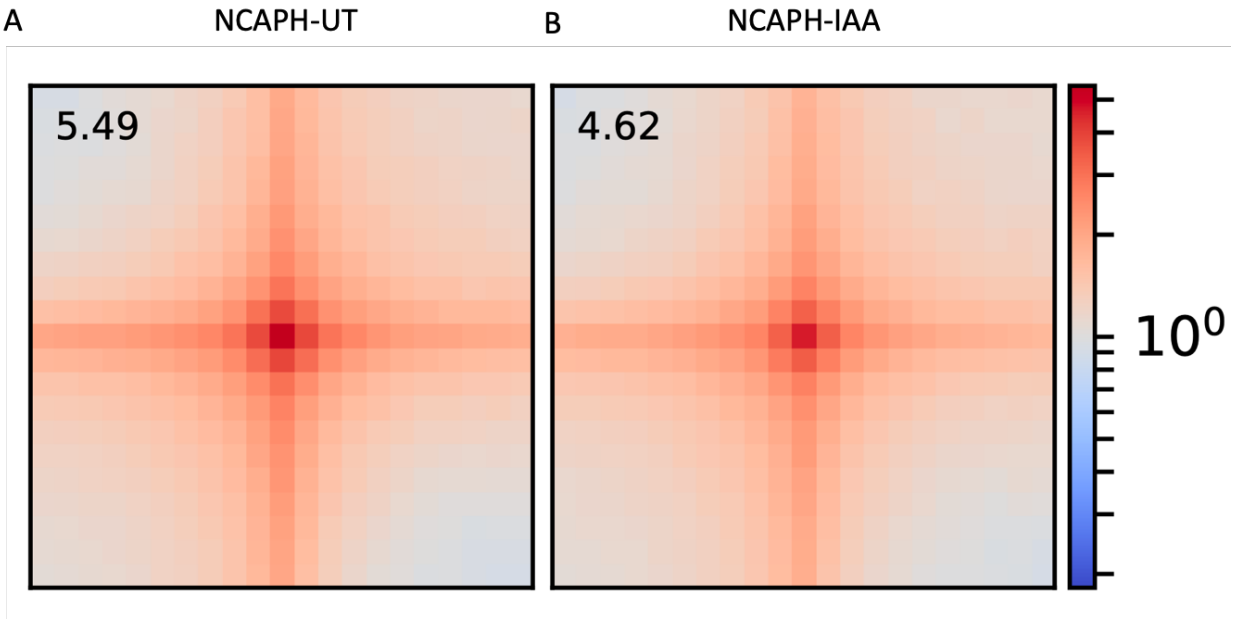

Figure 32, The APA analysis of domain loops in Condensin I depleted cells

The APA analysis of NCAPH-UT (A) and NCAPH-IAA (B) Hi-C matrices on domain loops. The value on the left-up corner indicates the APA enrichment score.

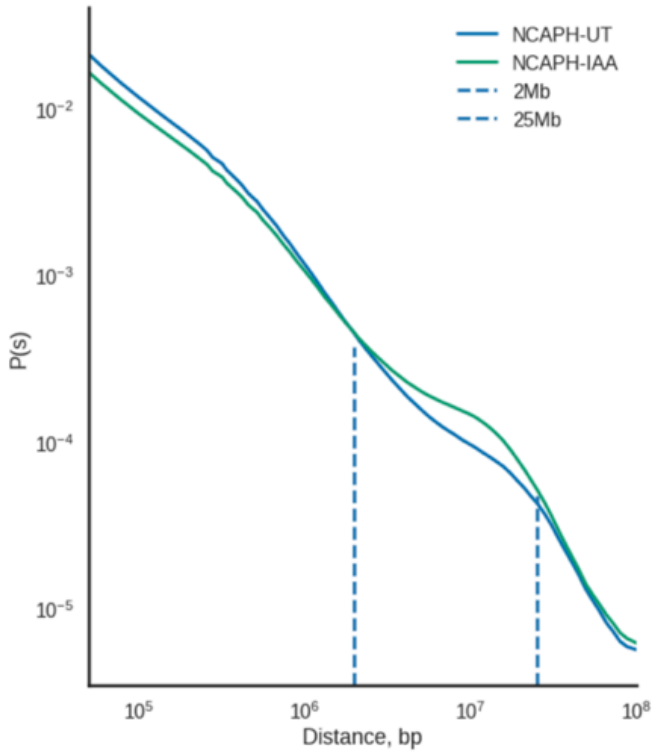

Figure 33, The P(s) curve in NCAPH-IAA and NCAPH-UT Hi-C matrices

The P(s) curve of NCAPH-UT and NCPAH-IAA Hi-C. This curve depicts the relationship between any possible chromatin contacts' distance and interaction probabilities measured by Hi-C.

We reasoned that the increasing of long-distance (2Mb- 25Mb) chromatin contacts are changes that happened at the A/B compartment level. To validate that, we performed A/B compartment analysis by calculating the first component of PCA values of Hi-C matrices. Unexpectedly, we didn't observe a huge increase of A/B compartment score, even at some locus we observed a decrease of A/B compartment score (Figure 34). We further calculated the interaction frequencies of intra-compartment and inter-compartments and found that the ratio of intra versus inter compartment decreased a lot after degradation of Condensin I (Figure 35). This effect can be also observed locally at specific locus (Figure 36). These results suggest that Condensin I depletion will increase the inter-compartmental interaction at a large scale.

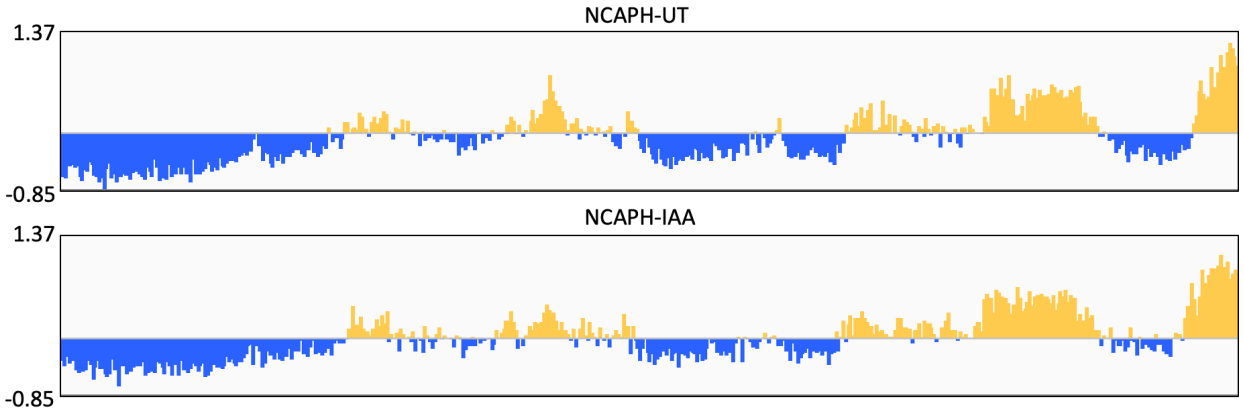

Figure 34, The A/B compartmental scores are decreased in NCAPH-IAA, at ACADM locus

The PCA value (A/B compartment score) at specific loci (ACADM gene) showed reduction to some extent after Condensin I depletion.

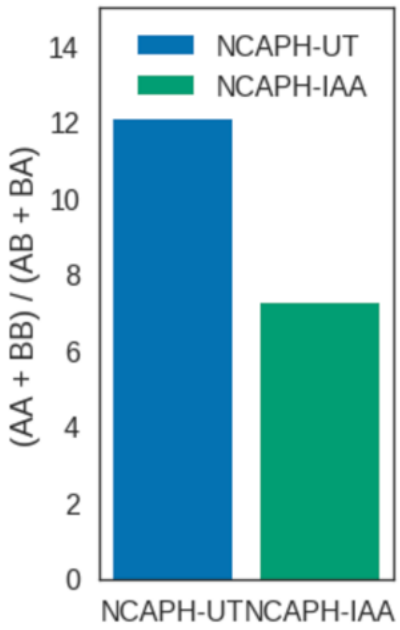

Figure 35, the inter-compartmental interaction increased after Condensin I depletion

The ratio of intra-compartment contacts versus inter-compartment contacts decreased after Condensin I depletion. The bar-plots indicate the ratio of (AA+BB)/ (AB +BA), where AA means the interaction between A compartment, AB means the interaction between A compartment and B compartment, BB means the interaction between B compartment. This figure suggests that NCAPH loss decreased the chance of inter-actions to happened between the same compartment but increased the chance of inter-mingling between A and B compartments.

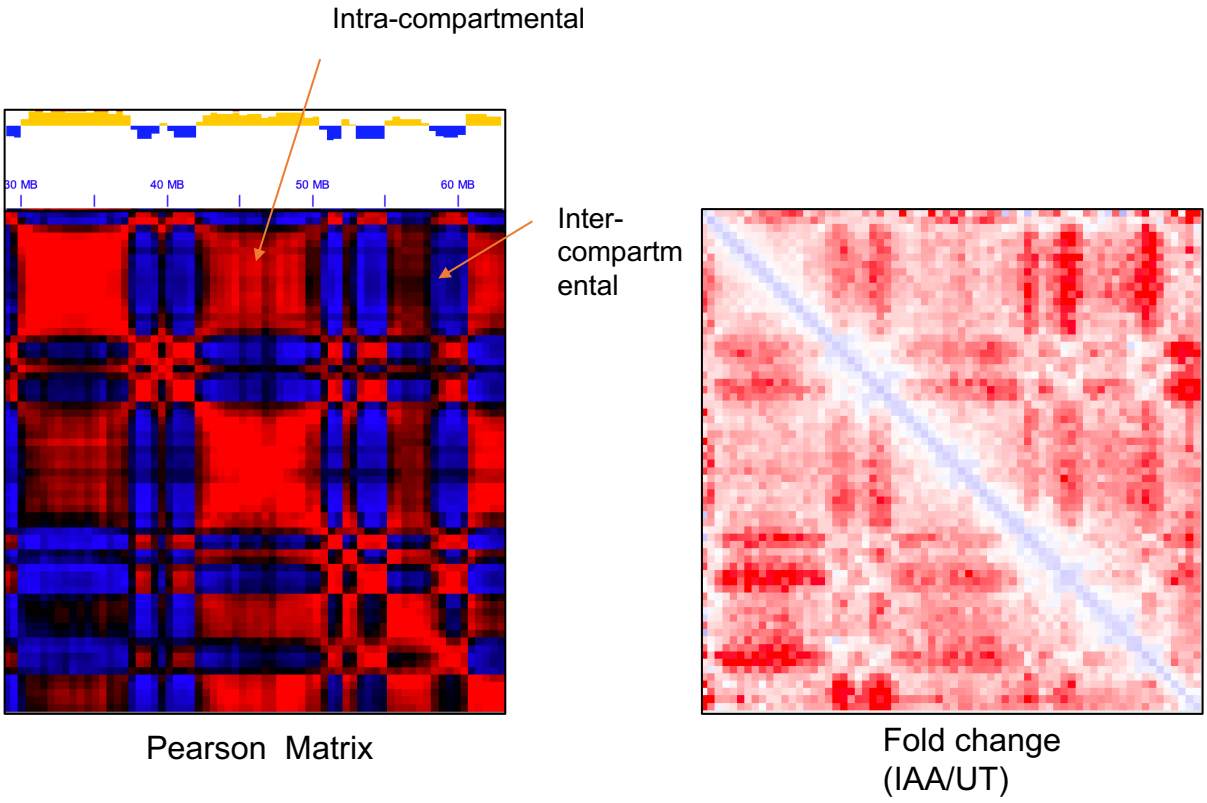

Figure 36, A snapshot of Hi-C matrices showed the increasing inter-compartmental interactions.

A snapshot of Hi-C Pearson Correlation map (left) and fold change (IAA/UT) map suggest that the inter-compartment interaction has been increased by Condensin I depletion. The arrow indicates the specific region that is decreased or increased.

Here, our comprehensive analysis of Condensin I depletion Hi-C data showed that Condensin I is regulating the large-scale chromatin organization at A/B compartment level, rather than at TAD level. Namely, Condensin I is counteracting the A/B inter-compartment interactions, which is in contrast to the fact that Cohesin is counteracting the A/B intra-compartment (47). However, the underlying mechanism of how Condensin I regulate this interphase chromatin compartmental interaction is still not clear. We reasoned that it is unlikely through loop extrusion of Condensin I since it is not possible for Condensin I to extrude such a long distance (Compartmental interaction usually is at the level of 10Mb) in such a short time (6 hours). As other people's hypothesis and our current data in Cohesin part of this paper (part 3), the compartmental interactions could be formed by a novel mechanism, namely liquid-liquid phase separation (LLPS). But there is not any evidence to our knowledge that have demonstrated that Condensin I is involved in LLPS. We also need to consider the fact that Condensin I is a ring-like structure, thus is not intrinsically disordered. Besides the loop extrusion model and LLPS model, a new mechanism may exist to explain the role of Condensin I in 3D genome organization.

Reference

1. Stadhouders, R., G. J. Filion, and T. Graf. 2019. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* 569: 345–354.

2. Jacob, F., and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3: 318–356.

3. Haberle, V., and A. Stark. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Bio* 19: 1–17.

4. Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27: 299–308.

5. Lagha, M., J. P. Bothma, and M. Levine. 2012. Mechanisms of transcriptional precision in animal development. *Trends Genet* 28: 409–416.

6. Shi, J., W. A. Whyte, C. J. Zepeda-Mendoza, J. P. Milazzo, C. Shen, J.-S. Roe, J. L. Minder, F. Mercan, E. Wang, M. A. Eckersley-Maslin, A. E. Campbell, S. Kawaoka, S. Shareef, Z. Zhu, J. Kendall, M. Muhar, C. Haslinger, M. Yu, R. G. Roeder, M. H. Wigler, G. A. Blobel, J. Zuber, D. L. Spector, R. A. Young, and C. R. Vakoc. 2013.

Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Gene Dev* 27: 2648–2662.

7. Li, W., D. Notani, and M. G. Rosenfeld. 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* 17: nrg.2016.4.

8. Lam, M., W. Li, M. G. Rosenfeld, and C. K. Glass. 2014. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39: 170–182.

9. Engreitz, J. M., J. E. Haines, E. M. Perez, G. Munson, J. Chen, M. Kane, P. E. nel, M. Guttman, and E. S. Lander. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539: 452.

10. Hnisz, D., B. J. Abraham, T. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, and R. A. Young. 2013. Super-Enhancers in the Control of Cell Identity and Disease. *Cell* 155: 934–947.

11. Levine, M., C. Cattoglio, and R. Tjian. 2014. Looping Back to Leap Forward: Transcription Enters a New Era. *Cell* 157: 13–25.

12. Consortium, T. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57.

13. Schoenfelder, S., and P. Fraser. 2019. Long-range enhancer–promoter contacts in gene expression control. *Nat Rev Genet* 1–19.

14. Deng, W., J. Lee, H. Wang, J. Miller, A. Reik, P. D. Gregory, A. Dean, and G. A. Blobel. 2012. Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell* 149: 1233–1244.

15. Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326: 289–293.

16. Rao, S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. Aiden. 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159: 1665–1680.

17. Dixon, J. R., I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget, A. Lee, Z. Ye, A. Kim, N. Rajagopal, W. Xie, Y. Diao, J. Liang, H. Zhao, V. V. Lobanenkov, J. R. Ecker, J. A. Thomson, and B. Ren. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* 518: 331.

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

19. Merkenschlager, M., and E. P. Nora. 2015. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genom Hum G* 17: 1–27.

20. Lupiáñez, D. G., M. Spielmann, and S. Mundlos. 2016. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* 32: 225–237.

21. Flavahan, W. A., Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suvà, and B. E. Bernstein. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529: 110.

22. Bonev, B., N. Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J.-P. Hugnot, A. Tanay, and G. Cavalli. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171: 557-572.e24.

23. Dekker, J., and E. Heard. 2015. Structural and functional diversity of Topologically Associating Domains. *Febs Lett* 589: 2877–2884.

24. van Steensel, B., and A. S. Belmont. 2017. Lamina-Associated Domains: Links

with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169: 780–791.

25. Javierre, B. M., O. S. Burren, S. P. Wilder, R. Kreuzhuber, S. M. Hill, S. Sewitz, J. Cairns, S. W. Wingett, C. Várnai, M. J. Thiecke, F. Burden, S. Farrow, A. J. Cutler, K. Rehnström, K. Downes, L. Grassi, M. Kostadima, P. Freire-Pritchett, F. Wang, T. Consortium, J. H. Martens, B. Kim, N. Sharifi, E. M. Janssen-Megens, M.-L. Yaspo, M. Linser, A. Kovacsovics, L. Clarke, D. Richardson, A. Datta, P. Flicek, H. G. Stunnenberg, J. A. Todd, D. R. Zerbino, O. Stegle, W. H. Ouwehand, M. Frontini, C. Wallace, M. Spivakov, and P. Fraser. 2016. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167: 1369-1384.e19.

26. Yu, M., and B. Ren. 2016. The Three-Dimensional Organization of Mammalian Genomes. *Annu Rev Cell Dev Bi* 33: 1–25.

27. Uhlmann, F. 2016. SMC complexes: from DNA to chromosomes. *Nat Rev Mol Cell Bio* 17: 399–412.

28. Niki, H., A. Jaffé, R. Imamura, T. Ogura, and S. Hiraga. 1991. The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of E. coli. *Embo J* 10: 183–193.

29. Skibbens, R. V. 2008. Chapter 5 Mechanisms of Sister Chromatid Pairing. *Int Rev Cel Mol Bio* 269: 283–339.

30. Hirano, T. 2016. Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell* 164: 847–857.

31. Anderson, D. E., A. Losada, H. P. Erickson, and T. Hirano. 2002. Condensin and cohesin display different arm conformations with characteristic hinge angles. *J Cell Biology* 156: 419–424.

32. Fudenberg, G., N. Abdennur, M. Imakaev, A. Goloborodko, and L. Mirny. 2018. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Biorxiv* 264648.

33. Hassler, M., I. A. Shaltiel, and C. H. Haering. 2018. Towards a Unified Model of SMC Complex Function. *Curr Biol* 28: R1266–R1281.

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

35. Dowen, J. M., S. Bilodeau, D. A. Orlando, M. R. Hübner, B. J. Abraham, D. L. Spector, and R. A. Young. 2013. Multiple Structural Maintenance of Chromosome

Complexes at Transcriptional Regulatory Elements. *Stem Cell Rep* 1: 371–378.

36. Li, W., Y. Hu, S. Oh, Q. Ma, D. Merkurjev, X. Song, X. Zhou, Z. Liu, B. Tanasa, X. He, A. Chen, K. Ohgi, J. Zhang, W. Liu, and M. G. Rosenfeld. 2015. Condensin I and II Complexes License Full Estrogen Receptor α-Dependent Enhancer Activation. *Mol Cell* 59: 188–202.

37. Yuen, K. C., B. D. Slaughter, and J. L. Gerton. 2017. Condensin II is anchored by TFIIIC and H3K4me3 in the mammalian genome and supports the expression of active dense gene clusters. *Sci Adv* 3: e1700191.

38. Gligoris, T. G. 2018. Chromosome Biology: The Smc–Kleisin Enzymology Finally Comes of Age. *Curr Biol* 28: R612–R614.

39. Core, L. J., J. J. Waterfall, and J. T. Lis. 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* 322: 1845–1848.

40. Schwalb, B., M. Michel, B. Zacher, K. Frühauf, C. Demel, A. Tresch, J. Gagneur, and P. Cramer. 2016. TT-seq maps the human transient transcriptome. *Science* 352: 1225–1228.

41. Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut,

M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.

42. Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* 38: 576–589.

43. Zhou, Y., B. Zhou, L. Pache, M. Chang, A. Khodabakhshi, O. Tanaseichuk, C. Benner, and S. K. Chanda. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10: 1523.

44. Natsume, T., T. Kiyomitsu, Y. Saga, and M. T. Kanemaki. 2016. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Reports* 15: 210–218.

45. Muhar, M., A. Ebert, T. Neumann, C. Umkehrer, J. Jude, C. Wieshofer, P. Rescheneder, J. J. Lipp, V. A. Herzog, B. Reichholf, D. A. Cisneros, T. Hoffmann, M. F. Schlapansky, P. Bhat, A. von Haeseler, T. Köcher, A. C. Obenauf, J. Popow, S. L. Ameres, and J. Zuber. 2018. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* 360: eaao2793.

46. Rahl, P. B., C. Y. Lin, A. C. Seila, R. A. Flynn, S. McCuine, C. B. Burge, P. A.

Sharp, and R. A. Young. 2010. c-Myc Regulates Transcriptional Pause Release. *Cell* 141: 432–445.

47. Rao, S., S.-C. Huang, B. Hilaire, J. M. Engreitz, E. M. Perez, K.-R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, and E. Aiden. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell* 171: 305-320.e24.

48. Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. Capturing Chromosome Conformation. *Science* 295: 1306–1311.

49. Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, and W. de Laat. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat Genet* 38: ng1896.

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

51. Fullwood, M. J., M. Liu, Y. Pan, J. Liu, H. Xu, Y. Mohamed, Y. L. Orlov, S. Velkov, A. Ho, P. Mei, E. G. Chew, P. Huang, W.-J. Welboren, Y. Han, H. Ooi, P. N. Ariyaratne, V. B. Vega, Y. Luo, P. Tan, P. Choy, S. K. Wansa, B. Zhao, K. Lim, S. Leow, J. Yow, R. Joseph, H. Li, K. V. Desai, J. S. Thomsen, Y. Lee, K. R. Karuturi, T. Herve, G. Bourque, H. G. Stunnenberg, X. Ruan, V. Cacheux-Rataboul, W.-K. Sung, E. T. Liu, C.-L. Wei, E. Cheung, and Y. Ruan. 2009. An oestrogen-receptor-α-bound human chromatin interactome. *Nature* 462: 58.

52. Mumbach, M. R., A. J. Rubin, R. A. Flynn, C. Dai, P. A. Khavari, W. J. Greenleaf, and H. Y. Chang. 2016. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 13: 919–922.

53. Mumbach, M. R., A. T. Satpathy, E. A. Boyle, C. Dai, B. G. Gowen, S. Cho, M. L. Nguyen, A. J. Rubin, J. M. Granja, K. R. Kazane, Y. Wei, T. Nguyen, P. G. Greenside, R. M. Corces, J. Tycko, D. R. Simeonov, N. Suliman, R. Li, J. Xu, R. A. Flynn, A. Kundaje, P. A. Khavari, A. Marson, J. E. Corn, T. Quertermous, W. J. Greenleaf, and H. Y. Chang. 2017. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* 49: ng.3963.

54. Ramani, V., X. Deng, R. Qiu, K. L. Gunderson, F. J. Steemers, C. steche, W. S. Noble, Z. Duan, and J. Shendure. 2017. Massively multiplex single-cell Hi-C. *Nat Methods* 14: 263–266.

55. Nagano, T., Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, and P. Fraser. 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502: 59.

56. Stevens, T. J., D. Lando, njan Basu, L. P. Atkinson, Y. Cao, S. F. Lee, M. Leeb, K. J. Wohlfahrt, W. Boucher, A. O'Shaughnessy-Kirwan, J. Cramard, A. J. Faure, M. Ralser, E. Blanco, L. Morey, M. Sansó, M. G. Palayret, B. Lehner, L. Croce, A. Wutz, B. Hendrich, D. Klenerman, and E. D. Laue. 2017. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544: 59.

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

58. Cao, J., Z. Luo, Q. Cheng, Q. Xu, Y. Zhang, F. Wang, Y. Wu, and X. Song. 2015. Three-dimensional regulation of transcription. *Protein Cell* 6: 241–253.

59. Servant, N., N. Varoquaux, B. R. Lajoie, E. Viara, C.-J. Chen, J.-P. Vert, E. Heard, J. Dekker, and E. Barillot. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 16: 259.

60. Kerpedjiev, P., N. Abdennur, F. Lekschas, C. McCallum, K. Dinkla, H. Strobelt, J.

M. Luber, S. B. Ouellette, A. Azhir, N. Kumar, J. Hwang, S. Lee, B. H. Alver, H. Pfister, L. A. Mirny, P. J. Park, and N. Gehlenborg. 2018. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol* 19: 125.

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

62. Flyamer, I. M., R. S. Illingworth, and W. A. Bickmore. 2019. Coolpup.py – a versatile tool to perform pile-up analysis of Hi-C data. *Biorxiv* 586537.

63. Guo, Y., K. Krismer, M. Closser, H. Wichterle, and D. K. Gifford. 2019. High resolution discovery of chromatin interactions. *Nucleic Acids Res* 47: gkz051-.

64. Schwarzer, W., N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. H. Haering, L. Mirny, and F. Spitz. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551: 51.

65. Takagi, M., T. Ono, T. Natsume, C. Sakamoto, M. Nakao, N. Saitoh, M. T. Kanemaki, T. Hirano, and N. Imamoto. 2018. Ki-67 and condensins support the integrity of mitotic chromosomes through distinct mechanisms. *J Cell Sci* 131: jcs212092.

66. Ganji, M., I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, and C. Dekker. 2018. Real-time imaging of DNA loop extrusion by condensin. *Science* 360: eaar7831.

Vita

Ruoyu Wang was born in Chuzhou, China, the son of Yubao Wang and Hong Zhao. After completing his middle school education at Chuzhou No. 1 High School, he entered University of Science and Technology of China (USTC) at Hefei, China. He received the degree of Bachelor of Bioscience from USTC in June 1st, 2016. In August of 2017, He joined The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences (UT GSBS).