Topological organization of RNA polymerase III-dependent genes in colorectal cancer cells

Yong-Jin Kim

Department of Medical Science

The Graduate School, Yonsei University

Topological organization of RNA polymerase III-dependent genes in colorectal cancer cells

Directed by Professor Hyoung-Pyo Kim

The Master's Thesis
submitted to the Department of Medical Science,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Master of Medical Science

Yong-Jin Kim

June 2021

This certifies that the Master's Thesis of

Yong-Jin Kim is approved.

_____

Thesis Supervisor : Hyoung-Pyo Kim

_____

Thesis Committee Member#1 : Hyun Seok Kim

_____

Thesis Committee Member#2 : Heon Yung Gee

The Graduate School

Yonsei University

June 2021

ACKNOWLEDGEMENTS

처음 대학원에 진학한 뒤 2년이 넘는 시간 동안의 석사 학위 과정을 마치고 어느새 졸업을 앞두게 되었습니다. 학위논문이 나오기까지 많은 도움을 주신 분들께 감사의 인사를 드리고자 합니다.

먼저 제 지도 교수님이신 김형표 교수님께 가슴 깊이 감사드립니다. 제가 학부 인턴으로 아직 대학원 입학도 정해지지 않았던 때, 생물정보학 교육과정에 참여할 수 있도록 애써 주셨던 날이 아직도 기억납니다. 덕분에 제 앞으로의 진로를 확실히 정할 수 있게 되었고 학위 과정 동안 후성유전학과 생물정보학 이라는 새로운 분야를 접하고 연구할 수 있는 기회를 주신 것에 감사 드립니다. 교수님의 많은 연구 지도 덕분에 저 또한 많은 성장을 할 수 있었던 것 같습니다. 또한 논문 작성에 있어서 관심 가져주시고 조언 해주신 두 분의 심사위원 김현석 교수님, 지헌영 교수님께도 감사 드립니다.

2년 넘는 시간 동안 같이 함께 해준 실험실 식구들 모두에게도 감사하다는 말 전하고 싶습니다. 부족한 저에게 때로는 짓궂은 장난도 치지만, 때로는 진심 어린 조언을 해주던 실험실 식구들이 있었기 때문에 학위 과정 동안 즐겁게 생활 할 수 있었습니다. 힘든 일 있을 때마다 공감 해주고 얘기 나눠주시는 민지 누나, 호쾌한 성격으로 잘 챙겨 주시는 미경 누나, 질문이 있으면 항상 친절하게 알려주시고 장난도 잘 받아주시는 웅재 형, 연구 관련해서 조언도 많이 해주시고 많은 일을 도맡아 하고 계신 은총이 형, 실험적으로나 실험실 생활로나 가장 많은 지원 해주신 보배 누나, 짓궂은 동생 장난도 잘 받아주시고 연구자로서 많이 본 받고 싶었던 정식이 형,

당차던 첫 후배 수경이, 모든 일에 관심 가져주던 경우, 입학 동기이자 항상 내 자신감을 북돋아 주던 무구, 입학 초기에 장난으로 긴장을 풀어 주시던 예은 누나, 어린 학생 의견도 경청해주시고 존중해주시던 양철민 박사님, 대학교부터 대학원까지 같이 진학하면서 고생했던 재우까지 즐거운 실험실 생활을 도와주셨던 분들 모두 감사드립니다.

마지막으로 짧지 않은 기간 동안 묵묵히 뒤에서 응원해주신 사랑하는 아버지, 어머니께 정말 감사합니다. 또 항상 부족한 건 없는지 챙겨주는 누나까지, 가족의 무한한 응원이 있었기 때문에 제가 온전히 연구에 집중해서 잘 마칠 수 있었던 것 같습니다.

지난 학위 과정 동안의 경험은 제게 더 없이 소중한 추억과 자산이 되었습니다. 많은 분들께서 도움 주셨을 때 느꼈던 고마움을 가지고 앞으로 더욱 가치 있는 사람이 되어 저도 누군가에게 도움이 될 수 있는 사람이 되도록 노력하겠습니다.
감사합니다.

2021년 6월
김용진 올림

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Topological organization of RNA polymerase III-dependent genes in colorectal cancer cells

Yong-Jin Kim

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Hyoung-Pyo Kim)

In eukaryotic cells, there are three major classes of RNA Polymerases. Among them, RNA Polymerase III is important for translation system and transcribes many different types of RNA molecules. Pol. III transcription is regulated in many different biological processes such as cell cycle, differentiation, regeneration, cellular stress. Deregulation of Pol. III transcription has been found to cause or to be linked to various diseases. Therefore, it is important to understand the mechanisms for regulation of Pol. III genes. The spatial organization of the genome into compartments and topologically associated domains can have an important role in the regulation of gene expression. However, it is not well known that how Pol. III genes are organized topologically and what extent they are involved in multi-level chromatin

structures. In recent years, many different high-throughput sequencing methods have been developed to study the three-dimensional organization of the genome and to measure genome-wide gene transcription. In this study, various types of data were analyzed which had been produced by several sequencing methods that are widely used in epigenetic researches such as ChIP-seq, ATAC-seq, *in situ* Hi-C and HiChIP to investigate Pol. III mediated chromatin landscape and PRO-seq to capture nascent RNA transcription level. As a result, integrative analysis using high-throughput sequencing methods could identify actively transcribed Pol. III genes and found that Pol. III genes are clustered in highly connected chromatin structure.

Key words : RNA polymerase III, 3D chromatin structure, tRNA

Topological organization of RNA polymerase III-dependent genes in colorectal cancer cells

Yong-Jin Kim

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Hyoung-Pyo Kim)

I. INTRODUCTION

In eukaryotes, there are three main RNA Polymerases(RNA Pol. I, Pol. II, Pol. III) which are highly specialized in their target gene transcription[1-4]. RNA Pol. I transcribes ribosomal RNA genes, Pol. II transcribes mRNA, miRNA, snRNA, and snoRNA genes, and Pol. III transcribes a variety of non-coding genes, for example, tRNA, vault RNA, 7SL, 7SK, U6, and other small RNA genes[1,4]. These Pol. III products have well-defined functions in translation and involved in other biological processes. In addition, many genomic loci contain the structure of

Pol. III gene promoter, and these genes are transcribed in a tissue-specific manner or specific malignant states. However, the functions of these RNAs are still understandable and new features continue to be revealed even as well-known Pol. III products[4].

Regulation of Pol. III transcription is involved in different biological processes such as the cell cycle[5], differentiation[6], development[7], regeneration[8], and cellular stress[9]. Not surprisingly, defects in Pol. III transcription have been found to be associated with or cause various diseases[4]. Although Pol. III transcription is associated with a variety of biological processes, recent studies are mainly focused on protein-coding genes transcribed by Pol. II and their regulation.

Gene expression is controlled by complex and various mechanisms, including the interaction of promoter with distal enhancers, which are short genomic elements often positioned several kilo-bases away from their target genes[10].

Regulatory mechanisms of gene expression are linked to the hierarchical organization of the genome. In mammalian cells, each chromosome is organized into hundreds of megabase-sized, local chromatin interaction domains termed topologically associated domains (TADs)[11].

The spatial organization of the human genome is known to play an important role in regulating gene expression[12-14]. tRNA genes(tDNA) are known to be enriched in the TAD boundary and are reported to play an important role in the chromosomal organization of eukaryotic cells in certain contexts[6]. However, it is not known to what extent the tRNA gene is involved in long-range interactions and 3D organization contributes to tRNA gene expression.

In recent years, some genome-wide studies have found the presence of Pol. III around active Pol. II gene promoters and Pol. II has also found near most of the loci that Pol. III binds[15]. Association of the two Polymerases with common loci harboring the histone modifications which mark transcriptionally active chromatin suggests that both polymerases may engage in a cross-talk[15]. And a research reported that a transcriptional interference by Pol. II could be confirmed as a direct

cause of the Pol. III gene repression[16]. These findings show that, even with gene-specific results, Pol. II contributes to Pol. III gene expression and vice versa, Pol. III may contribute to regulating Pol. II gene expression. In this study, various types of sequencing data were analyzed, which are commonly used in epigenetic studies such as ChIP-seq, ATAC-seq, HiChIP and PRO-seq, to investigate how Pol. III and its target genes are topologically organized and whether there are Pol. III mediated chromatin structures or other features.

ChIP-seq(Chromatin immunoprecipitation followed by sequencing) method is widely used to map DNA binding protein and histone modification throughout the genome[17]. Several ChIP-seq data for various target proteins such as transcription factor and histone protein were processed by using an in-house pipeline. In this study, RNA Pol. III, RNA Pol. II, which are major polymerases in human cell, CTCF, SMC1, which are well-known proteins for master regulator to organize genome topology[18], BRD4 which is a member of the bromo-domain protein family and interact with hyper-acetylated histone regions along the chromatin, accumulating on transcriptionally active regulatory elements[19] were sequenced and processed. Several histone modifications were also mapped such as H3K27Ac, H3K4me1, H3K4me3, H3K27me3 which are commonly used as markers for chromatin transcriptional states[20,21]. ATAC-seq(Assay for transposase-accessible chromatin) is used to profile the genome-wide accessible chromatin region and the composition of accessible chromatin throughout the genome reflects a network of physical interactions in which regulatory elements and transcription factors collaboratively regulate gene expression[22]. HiChIP (Hi-C chromatin immunoprecipitation) method is a recently developed sequencing technique that allows for the identification of interacting regulatory elements centric to specific target proteins[23].

By integrating and analyzing these data, genome-wide Pol. III occupancy and chromatin states were characterized in HCT116 cells. Active histone modification and chromatin accessibility have close relationship with Pol. III binding sites more

than transcription. Moreover, long-range interaction which is centric to Pol. III was identified by HiChIP. Chromatin interactions were observed throughout the genome, but were more frequent in chromosome 6. By performing 3D network analysis, surprisingly, it is found that a large number of tRNA genes are linked to other tRNA regions. These results suggest that these novel features of topological organization of Pol. III genes may have other unique gene regulation system in the 3D chromatin interacting domain.

II. MATERIALS AND METHODS

1. Dataset

All sequencing libraries were prepared in our laboratory and sequenced.

2. Annotation

A. Human Genome Sequence

Reference Human genome sequence file(v19, GRCh37.p13) and corresponding gene annotation GTF file were downloaded from gencode website[24]. (https://www.gencodegenes.org/). Genes annotated as 'level 3' which means automatically annotated loci in GTF file were excluded.

B. tRNA

The information for tRNA genes in the human genome were obtained from the Genomic tRNA Database[25] (GtRNAdb, http://gtrnadb.ucsc.edu/).

C. Repeat Sequence

The information for repeat sequence in the human genome were retrieved from UCSC genome browser database using table browser program.

3. Definition of Regulatory Element

A. Promoter

Promoter is defined as TSS ±2Kb flanking region. 'level 3' annotated genes

were excluded from GTF file.

B. Enhancer

H3K27Ac ChIP-seq peaks called by MACS2 software were determined as enhancers. (see method below).

4. Sequencing data processing

A. ChIP-seq

Paired-end sequencing libraries were constructed for all types of sequencing methods and sequenced on the Illumina platform. ChIP-seq reads were trimmed using trim_galore(v0.6.4) to remove adapter and low quality reads. Trimmed reads were mapped to hg19 human reference genome using bwa[26](v0.7.17) with default parameters. primary mapped reads were filtered using samtools[27](v1.9) with parameters -q 30 -F 1804. Duplicated reads were marked using picard(v2.18.23) default parameters. Reads marked as duplicated and mapped to mitochondrial chromosome were filtered using samtools. To visualize mapped reads on IGV or to use these reads in downstream analysis, bigwig files were generated using deepTools[28] function bamCoverage**(**v3.4.3**)** with parameters --normalizeUsing CPM --binSize 1.

ChIP-seq Peaks were identified from bam file using MACS2[29](v2.2.7.1) with parameters -f BAMPE --nomodel. Input DNA was sequenced and applied to peak-calling process to eliminate bias in fragmentation and sequencing. Peaks under q-value cutoff < 1e-2 were filtered out to optain significant binding sites. Consensus peak sets were identified by intersecting peaks from each replicate using bedtools[30] function intersect(v2.29.2).

B. ATAC-seq

ATAC-seq reads were trimmed using trim_galore with default parameters. Trimmed reads were mapped to reference genome using bowtie2[31] with parameters --end-to-end --very-sensitive -X 2000. mapped reads were filtered using samtools with parameters -q 30 -F 1804. Duplicated reads were marked using picard with default parameters. Marked reads and mitochondrial reads were removed using samtools. Fragment size distribution plot was generated after mapping process using Picard function CollectInsertSizeMetrics(v2.18.23). all uniquely mapped reads were shifted by +4bp on the positive strand and -5bp on the negative strand using deepTools function alignmentSieve to minimize Tn5 transposase binding biases. Bigwig files were generated using deepTools function bamCoverage with parameters --normalizeUsing CPM --binSize 1. ATAC-seq peaks were called using MACS2 with parameters -f BAMPE --nomodel --min-length 100. Significant and Consensus peak sets were identified as described above.

C. HiC & HiChIP

Hi-C reads of each replicates were pooled together for downstream analysis because they demonstrated high correlations. Raw reads were mapped and processed using HiC-Pro pipeline[32](v2.11.4). Configuration file for HiC-Pro was modified with parameters MIN_MAPQ=30, REFERENCE _GENOME=hg19, LIGATION_SITE=GATCGATC for DpnII restriction enzyme. All valid pair files from HiC-Pro were converted to .hic format file using hicpro2juicebox.sh in HiC-Pro utilities. .hic files were transformed to .cool file format using HiCExplorer[33] function hicConvertFormat(v3.4.3) to use in downstream analysis. Normalization was performed to minimize sequencing depth biases using HiCexplorer function hicNormalize with parameters --normalize smallest. Given matrices were normalized to the lowest read count matrix. Normalized count matrices were corrected using

HiCExplorer function hicCorrectMatrix with parameters --correctionMethod KR. Hi-C contact maps normalized and KR-corrected matrix were visualized by HiCExplorer function hicPlotMatrix.

FitHiChIP[23](v.8.0) was used to call loops with 10kb bin, corresponding ChIP-seq peaks, allValidPairs, and q-value < 1e-2 threshold was used to determine significant loop.

D. PRO-seq

PRO-seq raw sequence data were trimmed using trim_galore. Trimmed reads were aligned to human reference genome using bwa. Duplicated reads were marked using picard. Duplicated, low quality and mitochondrial reads were filtered out using samtools. Bigwig files were generated from bam files using deepTools function bamCoverage.

5. Topologically Associated Domain (TAD) boundary

Insulation score quantifies the interactions passing across each genomic bin and allows defining boundaries by identifying local minima. To calculate Insulation score, Sparse format ICE-balanced matrix at 10kb resolution in HiC-Pro output directory was converted to dense format matrix using sparseTodense.py python script contained in HiC-Pro utilities with parameters --ins -g hg19 --perchr. Each of the dense format matrices was compressed to be used as input for insulation score calculation. Insulation scores were calculated using matrix2insulation.pl perl script with default parameters and this script provides insulation score and TAD boundary region. Aggregate TAD plot was generated using FAN-C software[34](v.0.9.10).

6. Chromosomal A/B compartment

Genomic compartments were identified using eigenvector function of juicer software[35](v1.22.01) to calculate the first principal component (PC1) value of the Pearson's matrix at 100kb resolution with parameters -p KR BP 100000. The signs of PC1 values were manually adjusted on the basis of genome-wide ATAC-seq read density indicating open and active chromatin region. Saddle plot was generated at 100kb resolution using cooltools (v.0.3.2), and strength of compartmentalization was defined as the ratio of (AA+BB)/(AB+BA) interactions.

7. 3D connected clique

3D connected clique analysis was performed as previously described[36]. An undirected graph of regulatory interactions was constructed for Pol. III HiChIP loops where each vertex was a Loop anchor and each edge was significant interaction identified by FitHiChIP described as above. 3D Cliques were defined by spectral clustering of the regulatory graph interactions using cluster_louvain function in igraph R package with defulat parameters. A 3D clique connectivity was defined as the number of edges(loops) connecting vertices within the clique. the connectivity of cliques was ranked in ascending order and plotted against the rank[36].

III. RESULTS

1. Profiling genome-wide RNA Pol. III occupancy and Pol. III gene transcription in HCT116 cells

To find out how RNA Pol. III is distributed in the genome and what characteristics it has, an analysis was performed on ChIP-seq data (Table 1). ChIP-seq is a sequencing method that can identify the genome-wide binding sites of a specific target protein. Before the analysis, RNA Pol. III ChIP-seq profile was confirmed directly on genomic viewer to verify that it can reflect Pol. III specific binding profile. It shows sharp and specific binding patterns on its target gene sites (Fig. 1A). Next, the number of binding site for each chromosome was calculated to see how RNA Pol. III is distributed in the genome. The chromosome 6 was found to have 80 peaks out of total 330 peaks except for the peaks located in the blacklist region (Fig. 1B). Pol. III is known primarily for transcribing tRNA genes and transposable elements. The location informatoin of tRNA genes and transposable elements were downloaded from GtRNAdb and UCSC to find out what kind of elements Pol. III is binding. 257 tRNA elements overlap the most, followed by 87 SINE and 23 LINE sequences (Fig. 1C). To identify what types of elements are located in Pol. III binding sites according to its binding density, a boxplot was generated. Pol. III peaks are divided into tRNA, TE, and not annotated region and ChIP-seq signals for each peaks were quantified. It shows that most of the regions with high binding density were tRNA genes and signals at transposable elements were relatively low (Fig. 1D). RNA Pol. III is known to transcribe its target genes but Pol. III gene specific sequencing problems made it difficult to quantify Pol. III gene expression using standard RNA-seq method. Therefore, PRO-seq was used as an alternative method to quantify nascent RNA transcripts. Scatterplot shows Pol. III enrichment and PRO-seq signals are highly

correlated (Fig. 1E). and PRO-seq signals on tRNA genes which Pol. III is binding were higher than tRNA genes which Pol. III is not binding (Fig. 1F).

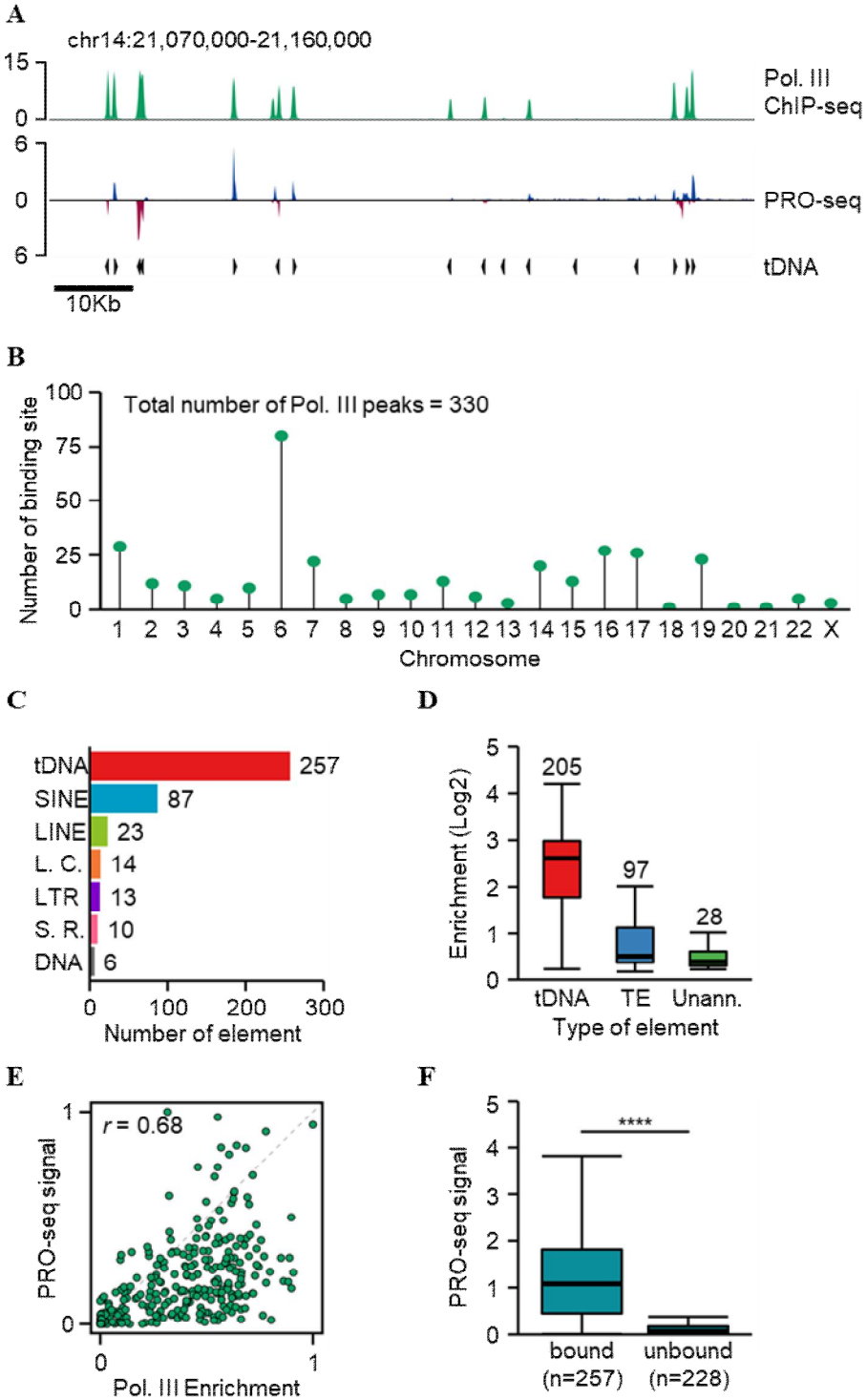

Figure 1. Genome-wide distribution of RNA Pol. III-dependent transcription. (A) Genomic snapshot showing enrichment of RNA Pol. III and nascent RNA transcription. Arrows in the bottom indicate tDNA genes with direction of transcription. (B) Number of Pol. III peaks across whole chromosomes. (C) Distribution of tRNA genes and transposable elements which are annotated to Pol. III binding sites (*L.C. : Low complexity, *S. R. : Simple Repeat). (D) Box plot showing the enrichment of RNA Pol. III in each type of elements. (E) Scatter plot indicating the relationship between Pol. III ChIP-seq signal and PRO-seq signal in tRNA gene body. Spearman's correlation coefficient is represented in the plot. (F) Box plot showing PRO-seq signal in tRNA genes with or without binding of Pol. III. Significance was determined by t-test (****, p-value < 0.0001).

Table 1. ChIP-seq, PRO-seq, ATAC-seq data mapping result summary

Data Type	Antibody	Replicate	Aligned Reads	Duplication Rate	Used Reads in Analysis
ChIP-seq	RNA Pol. III(POLR3A)	1	67,710,680	14%	52,012,220
ChIP-seq	RNA Pol. III(POLR3A)	2	67,499,464	16%	50,757,504
ChIP-seq	RNA Pol. II(Rpb1)	1	62,428,738	11%	49,813,613
ChIP-seq	RNA Pol. II(Rpb1)	2	69,879,350	12%	55,328,266
ChIP-seq	CTCF	1	85,191,271	13%	66,783,558
ChIP-seq	CTCF	2	65,903,470	16%	50,400,311
ChIP-seq	SMC1	1	62,635,502	10%	50,537,254
ChIP-seq	SMC1	2	65,052,530	11%	52,193,647
ChIP-seq	BRD4	1	69,782,482	11%	56,019,981
ChIP-seq	BRD4	2	65,829,391	10%	53,244,853
ChIP-seq	H3K27Ac	1	64,518,334	11%	52,851,091
ChIP-seq	H3K27Ac	2	59,799,588	10%	49,527,169
ChIP-seq	H3K4me1	1	123,980,223	14%	98,093,997
ChIP-seq	H3K4me1	2	110,126,280	17%	84,314,628
ChIP-seq	H3K4me3	1	72,417,946	18%	54,503,045
ChIP-seq	H3K4me3	2	65,714,455	19%	48,682,973
ChIP-seq	H3K27me3	1	151,796,102	9%	127,782,426
ChIP-seq	H3K27me3	2	107,699,337	11%	87,952,518
ChIP-seq	Input	1	23,157,027	7%	19,050,509
PRO-seq	-	1	52,829,834	16%	37,107,357
PRO-seq	-	2	56,033,738	19%	37,969,283
ATAC-seq	-	1	105,447,781	33%	58,044,932
ATAC-seq	-	2	111,706,944	36%	58,514,438

2. tRNA anticodon and transcription

There are several copies of tRNA genes in the human genome encoding anticodon sequences for specific amino acid. To compare the tRNA gene count and Pol. III enrichment, PRO-seq signals in these tRNA genes, gene counts for each anticodon family and read signals were quantified. $tRNA^{Asn\text{-}GTT}$, $tRNA^{Cys\text{-}GCA}$, $tRNA^{Ala\text{-}AGC}$ are most abundant anticodon family in human genome, each gene has more than 30 copies. However, Pol. III enrichment and PRO-seq signals are not higher as much as than other anticodon families. $tRNA^{Ile\text{-}GAT}$, $tRNA^{Asn\text{-}ATT}$, $tRNA^{Tyr\text{-}ATA}$, $tRNA^{Sup\text{-}TTA}$, $tRNA^{Pro\text{-}GGG}$, $tRNA^{Cys\text{-}ACA}$ exist less than 10 copies and ChIP-seq, PRO-seq signals are rarely observed. These results show that tRNA genes exist in mutl copies thought the genome, but this is not proportional to the Pol. III enrichment or transcription level.

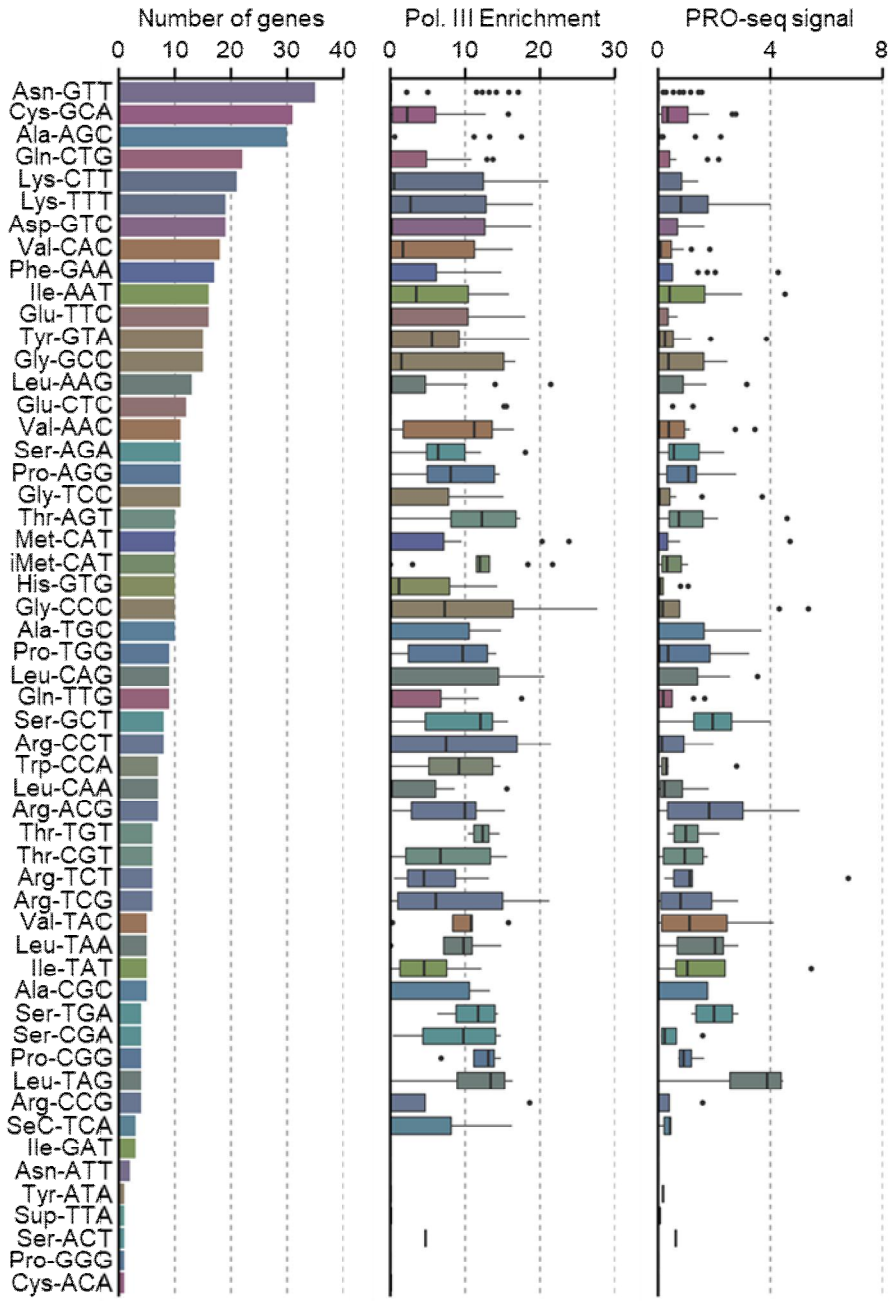

Figure 2. Number of tRNA gene(left), occupancy of RNA Pol. III(middle), and nascent RNA expression(right) in each tRNA gene classified into gene families according to their anticodon.

3. Association of Pol. III-dependet transcription with chromatin accessibility

Chromatin accessibility which means the degree to which nuclear molecules are able to physically contact DNA was identified using ATAC-seq. To see if Pol. III occupancy and transcription level are correlated with chromatin accessibility, the signals of the entire Pol. III occupied region were compared (Fig. 2A). As a result of the spearman correlation test, both Pol. III enrichment and PRO-seq signal were highly correlated with chromatin accessibility. Next, to determine whether these accessible chromatin features are dependent on Pol. III occupancy, tRNA genes are classified as genes with and genes without Pol. III, and ATAC-seq signals flanking 2Kb regions from the center of the genes are compared (Fig. 3B). The average signal difference is statistically significant (Fig. 3C). In order to verify that Pol. III dependent ATAC-seq signals are common patterns across the entire binding site rather than some specific regions, a heatmap was generated. As a result, most of regions have co-enriched Pol. III ChIP-seq, PRO-seq, and ATAC-seq signals (Fig. 3D). The co-enriched pattern of Pol. III ChIP-seq, ATAC-seq, and PRO-seq signals in tRNA genes is also observed in genomic tracks (Fig. 3E).

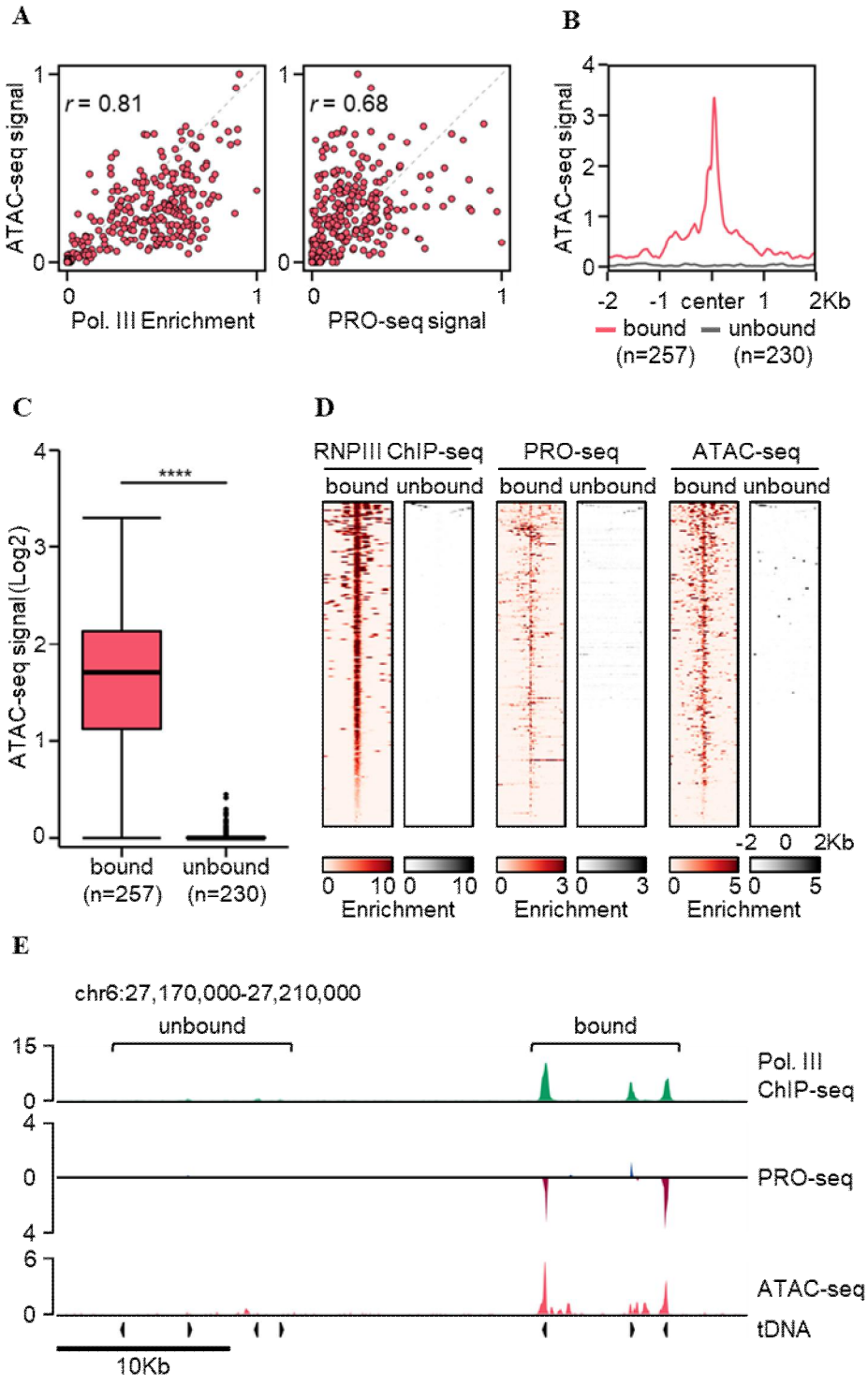

Figure 3. RNA Pol. III-dependent transcription is positively associated with chromatin accessibility. (A) Scatter plot showing the relationship of chromatin accessibility with either Pol. III enrichment (left) or PRO-seq signal in tRNA genes (right). Spearman's correlation coefficient is represented in the plot. (B) Average enrichment of ATAC-seq signal in tRNA genes with (pink) or without (grey) Pol. III binding. All signal was quantified from the center of tRNA gene to ±2Kb flanking region. (C) Box plot showing the distribution of ATAC-seq signal in tRNA genes with (pink) or without (grey) Pol. III binding. Significance was determined by t-test (****, p-value < 0.0001). (D) Heatmap of ChIP-seq (green), PRO-seq (blue), ATAC-seq (red) signals of tRNA genes with (left) or without (right) Pol. III binding. (E) Snapshot showing signal tracks for Pol. III ChIP-seq, PRO-seq, and ATAC-seq in representative tRNA genes.

4. Distinctive patterns of histone modification in actively transcribed RNA Pol. III genes

Several types of histone protein ChIP-seq data, such as H3K27Ac, H3K4me3, H3K4me1, H3K27me3, were used to identify enviornmental chromatin states around Pol. III binding sites. First, ChIP-seq signals in tRNA genes with and without Pol. III were quantified and compared (Fig. 4A-D). ChIP-seq signals of histone proteins which mark active epigenetic chromatin states, such as H3K27Ac, H3K4me3, H3K4me1, were more enriched in tRNA genes with Pol. III rather than tRNA genes without Pol. III. To determine whether there is a correlation between tRNA gene expression and these active chromatin states, spearman correlation test was performed, but ChIP-seq and PRO-seq signal show weak correlation (Spearman correlation coefficient(r), H3K27Ac = 0.22, H3K4me3 = 0.34, H3K4me1 = 0.14) (Fig. 4A-C). On the other hand, H3K27me3 which marks repressive chromatin state was more enriched in tRNA genes without Pol. III than genes with Pol. III and also shows a weak correlation with PRO-seq signals (Fig. 4D). signal tracks show distinctive ChIP-seq signal patterns of active histone modification nearby tRNA genes with Pol. III (Fig. 4E).

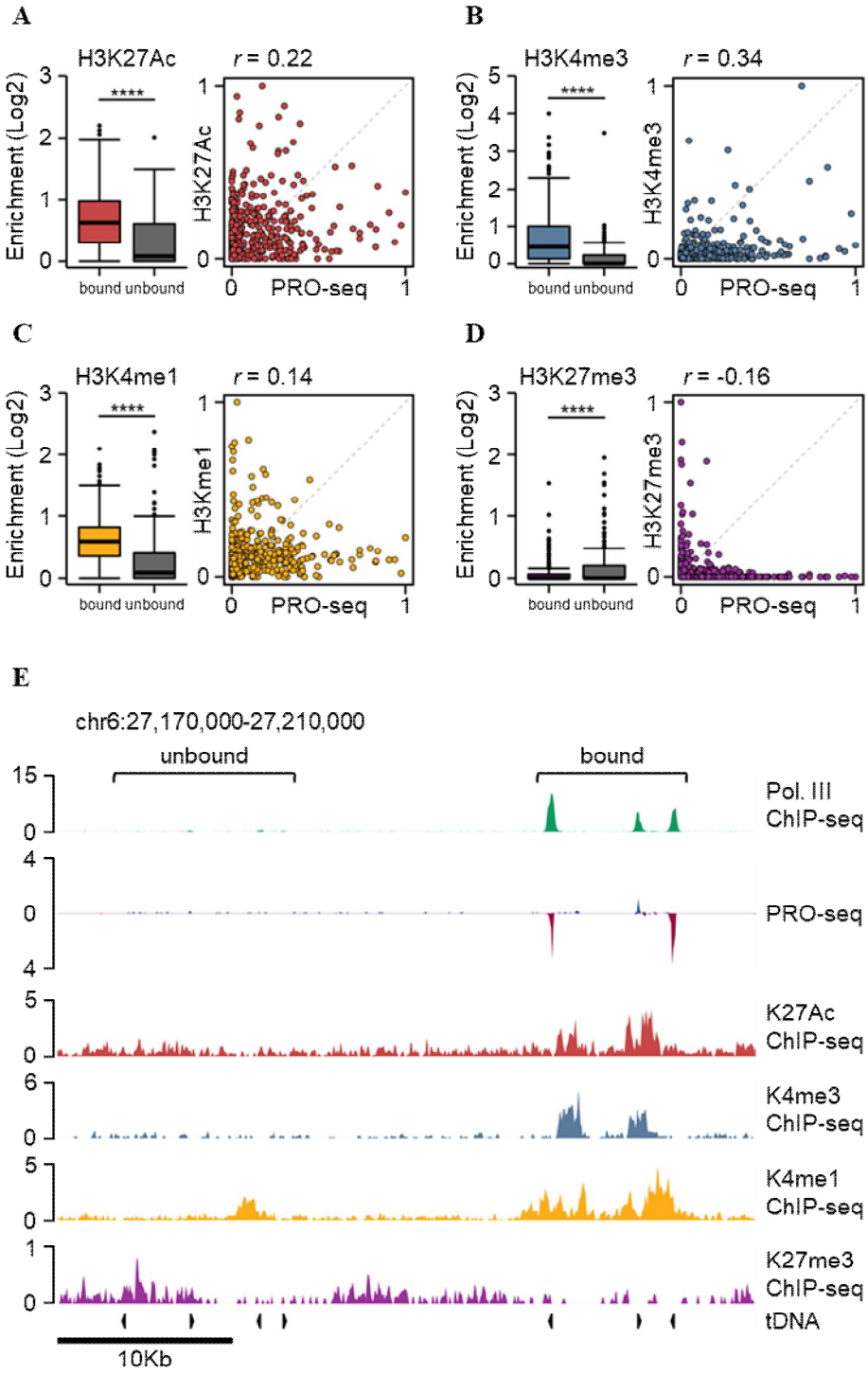

Figure 4. Distinctive patterns of histone modification in actively transcribed RNA Pol. III genes. (A-D) Box plots (left) showing the distribution of either H3K27ac (A), H34K4me3 (B), H3K4me1 (C), or H3K27me3 (D) signals in tRNA genes with (pink) or without (grey) Pol. III binding. Significance was determined by t-test (****, p-value < 0.0001). Scatter plots (right) showing the relationship of PRO-seq signal with either H3K27ac (A), H34K4me3 (B), H3K4me1 (C), or H3K27me3 (D) signals in tRNA genes. Spearman's correlation coefficient is represented in the plot. (E) Snapshot showing signal tracks for Pol. III ChIP-seq, PRO-seq, and histone modifications in representative tRNA genes.

5. RNA Pol. II-dependent transcription machinery in actively transcribed RNA Pol. III genes

Several types of transcription factor ChIP-seq data, such as Pol. II, BRD4, CTCF, SMC1 were used to identify whether Pol. III genes have a relationship with Pol. II related transcription machinery factors or chromatin structural proteins. As described in analysis with histone protein ChIP-seq data above, ChIP-seq signals in tRNA genes with and without Pol. III were quantified and compared (Fig. 5A-D). All transcription factors are more enriched in tRNA genes with Pol. III than genes without Pol. III. Pol. II and BRD4 signals are positively correlated with tRNA gene expression. CTCF and SMC1 also have positive correlation with tRNA gene expression, but not strong (Spearman correlation coefficient(r), Pol. II = 0.44, BRD4 = 0.58, CTCF = 0.17, SMC1 = 0.58) (Fig. 5A-D). These transcription factors binding pattern that Pol. II-dependent transcription machinery have positive correlation with tRNA gene expression but structural proteins are not well correlated is well illustrated in genomic viewer (Fig. 5E).

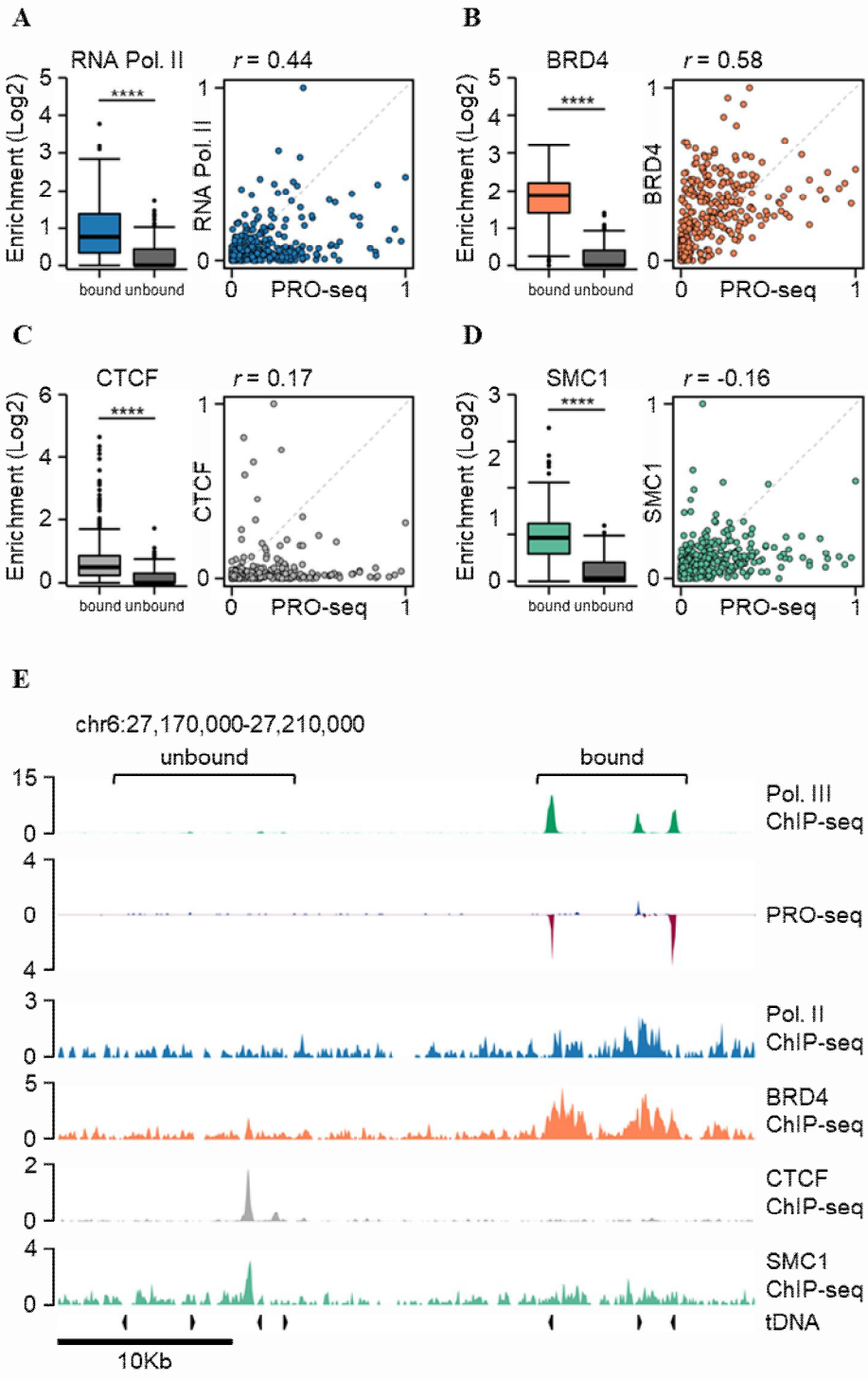

Figure 5. RNA pol. II-dependent transcription machinery components are enriched in actively transcribed RNA Pol. III genes. (A-D) Box plots (left) showing the distribution of either RNA Pol. II (A), BRD4 (B), CTCF (C), or SMC1 (D) signals in tRNA genes with (pink) or without (grey) Pol. III binding. Significance was determined by t-test (****, p-value < 0.0001). Scatter plots (right) showing the relationship of PRO-seq signal with either RNA Pol. II (A), BRD4 (B), CTCF (C), or SMC1 (D) signals in tRNA genes. Spearman's correlation coefficient is represented in the plot. (E) Snapshot showing signal tracks for enrichment of Pol. III, PRO-seq, Pol. II, BRD4, CTCF, and SMC1 in representative tRNA genes.

6. tRNA genes clustered in topologically associated domain

3D genome organization can be characterized at different scales. Topologically associated domain(TAD) is a fundamental unit of three dimensional nuclear organization. TADs are megabase-scale genomic regions and modulate gene regulation by limiting interactions of cis-regulatory sequences to target genes[37]. tRNA genes are dispersed throughout the genome and some genes are linearly organized into genomic clusters with other tDNAs. However, it is not well known how TADs contribute to tRNA gene organization. To determine the extent of the contribution of TADs in the tDNA organization, *in situ* Hi-C was performed (Table 2) and the structural units of the genome in HCT116 were defined (Fig. 6A). At large scales, chromosomes segregate into regions of preferential long-range interactions that form two mutually excluded types of chromatin, referred to as "A" and "B" compartments[38]. Eigenvector 1 values indicate these segregated compartment regions (Fig. 6B). Preferential homotypic interaction patterns (A-A, B-B) are well illustrated in saddle plot with compartment strength value 4.62 (Fig. 6C). Next, TAD boundaries were called and the total number of TAD boundaries was 4837 (Fig. 6D). Aggregated insulation scores from the center of TAD boundaries have confirmed that chromatin interactions are well insulated from the boundary (Fig. 6E). The average size of TADs is about 670 Kb, with IQR ranging from 390 Kb to 810 Kb (Fig. 6F). Aggregated TAD interactions show insulated interaction patterns from the TAD boundary (Fig. 6G). To find out the genome-wide distribution of tRNA genes at the TAD level, TADs containing one or more tRNA genes were classified and then the distribution of the number of tDNAs was conformed (Fig. 7A). Total number of transcribed tDNAs in TAD was 251 and total number of TADs containing one or more tDNAs was 81. 129 tRNA genes (51.4%) were located in 7 TADs. To confirm that whether there is a correlation between the number of tRNA genes contained in the TADs and transcription, TADs were

categorized into 5 groups according to the number of clustered tRNA genes and transcription level was compared. transcription level was higher in TADs where only one tRNA gene was located than TAD group which contains over 11 tRNA genes (***, p-value = 0.001). Chromatin interaction and organization of clustered tRNA genes around the TAD region which contains 49 tRNA genes, the largest number of clustered tRNA genes in TAD scale, is well illustrated (Fig. 7C). These results show that a large number of transcribed tRNA genes are not only positionally close, but are also clustered within certain TAD structures.

Table 2. Summarized mapping results of *in situ* Hi-C data

	Hi-C		
	Replicate 1	**Replicate 2**	**combined**
Total_pairs_processed	233,312,045	215,233,315	448,545,360
Unmapped_pairs	3,950,879	4,616,374	8,567,253
Low_qual_pairs	83,170,416	71,638,727	154,809,143
Unique_paired_alignments	130,126,225	122,356,282	252,482,507
Multiple_pairs_alignments	-	-	-
Pairs_with_singleton	16,064,525	16,621,932	32,686,457
Low_qual_singleton	-	-	-
Unique_singleton_alignments	-	-	-
Multiple_singleton_alignments	-	-	-
Reported_pairs	130,126,225	122,356,282	252,482,507
valid_interaction	113,505,715	102,346,638	215,852,353
valid_interaction_rmdup	98,001,210	84,646,092	182,645,596
trans_interaction	16,916,660	15,634,461	32,551,103
cis_interaction	81,084,550	69,011,631	150,094,493
cis_shortRange	18,237,716	15,185,625	33,421,942
cis_longRange	62,846,834	53,826,006	116,672,551

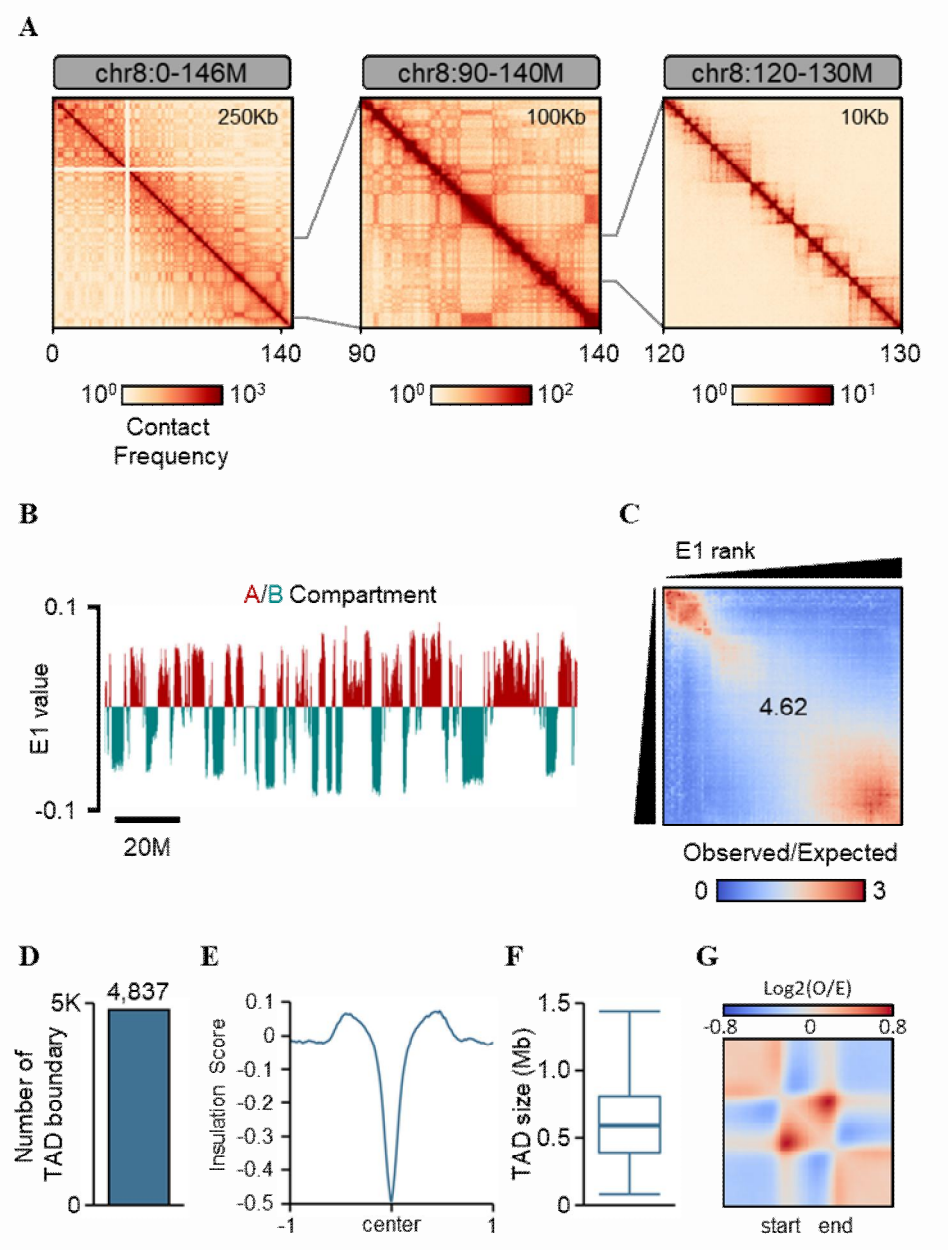

Figure 6. Three dimensional genome structure of HCT116 cells in terms of compartment and TAD. (A) Hi-C contact maps generated by HiCExplorer at 250-kb, 100-kb and 10-kb resolution (B) Distributions of cis Eigenvector 1 values across entire chromosome 8. (C) Saddle plot representing compartmentalization strength. (D) Number of TAD boundaries obtained with Hi-C data. (E) Genome-wide averaged insulation score plotted versus distance around insulation center at TAD boundaries. (F) Boxplot shows the size of TAD. (G) Heatmap shows the average observed/expected Hi-C interactions in the TAD regions.

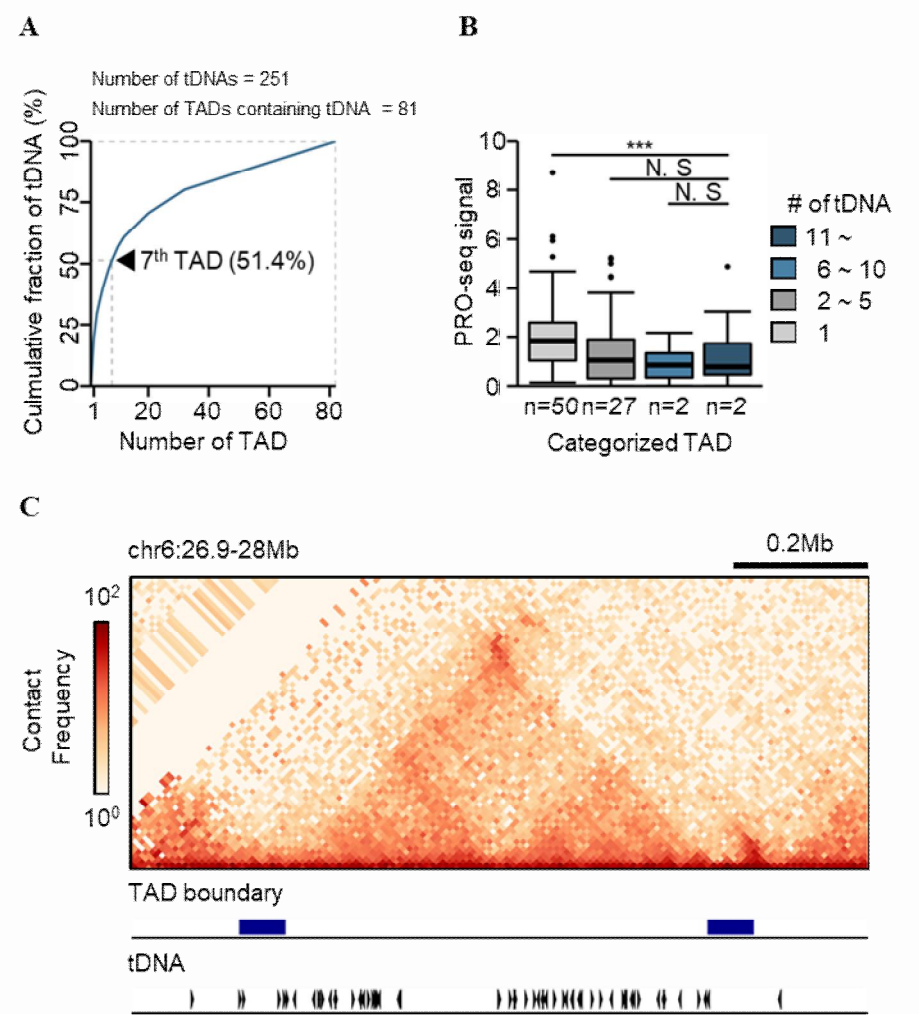

Figure 7. Distribution and expression of tRNA genes clustered within each TAD. (A and B) Cumulative curve showing the distribution of the number of clustered tDNA within TAD (A) and box plot showing the distribution of tDNA expression clustered within each categorized TAD according to the number of tDNA (B). Significance was determined by t-test (***, p-value < 0.001, N. S, p-value = 1). (C) Snapshot of *in situ* Hi-C contact maps with location of TAD boundaries and tRNA genes.

7. Identification of RNA Pol. III-mediated chromatin interactions

To identify the chromatin structure directly mediated by RNA Pol. III, HiChIP was performed using antibody targeting for Pol. III subunit(POLR3A) (Table 3). Total 810 of significantly interacting loops were identified from HiChIP data using FitHiChIP software (q-value < 0.01). A loop consists of two anchors connected. In order to clarify the characteristics of Pol. III mediated loops, loops were classified according to Pol. III occupancy in loop anchors and counted. 661 loops were bound by Pol. III in one of the two anchors, and the remaining 152 loops were bound by Pol. III in both anchors (Fig 8A). Since Pol. III mediated loops are calculated based on Pol. III ChIP-seq peaks, the aggregated chromatin contact of Hi-C and HiChIP data were compared in the interacting regions to confirm if it is a biased effect by an antibody specific to Pol. III. In 813 Pol. III loops, both data represent an enriched pattern (enrichment score, HiChIP = 5.19, Hi-C = 4.86) (Fig. 8B). The average distance between two genomic anchors is about 235 Kb, with IQR ranging from 90 Kb to 320 Kb (Fig. 8C). Next, all loops were classified as intra-TAD interaction and inter-TAD interaction to clarify Pol. III loop organization in TAD scale. As a result, 744 loops (91.05%) were formed within the TAD and 62 loops (7.6%) were formed beyond the TAD boundary (Fig. 8D). These results suggest that Pol. III loop interaction is mostly within the fundamental structural unit, TAD, and the Pol. III binding region may also interact with other types of genomic regions via looping structure (Fig. 8E). To determine which types of regions are connected by Pol. III mediated loops, each loop is classified according to the type of both anchors (Fig. 9). As a result, tDNA-tDNA interaction accounted for the largest number of population(148, 18.2%), followed by tDNA-Pol. II gene promoter(132, 16.2%), tDNA-CTCF(92, 11.3%). These results indicate that many Pol. III genes do not exist independently and are interacting with other Pol. III genes or other regions located at long distances.

Table 3. Summarized mapping results of RNA Pol. III HiChIP data

RNA Pol. III HiChIP			
	Replicate 1	Replicate 2	combined
Total_pairs_processed	227,740,526	220,471,054	448,211,580
Unmapped_pairs	3,646,627	4,624,434	8,271,061
Low_qual_pairs	74,328,709	76,762,436	151,091,145
Unique_paired_alignments	135,471,149	125,462,008	260,933,157
Multiple_pairs_alignments	-	-	-
Pairs_with_singleton	14,294,041	13,622,176	27,916,217
Low_qual_singleton	-	-	-
Unique_singleton_alignments	-	-	-
Multiple_singleton_alignments	-	-	-
Reported_pairs	135,471,149	125,462,008	260,933,157
valid_interaction	109,102,144	99,142,911	208,245,055
valid_interaction_rmdup	87,323,879	76,926,177	164,245,666
trans_interaction	15,868,241	12,803,483	28,671,703
cis_interaction	71,455,638	64,122,694	135,573,963
cis_shortRange	21,243,653	20,624,871	41,864,493
cis_longRange	50,211,985	43,497,823	93,709,470

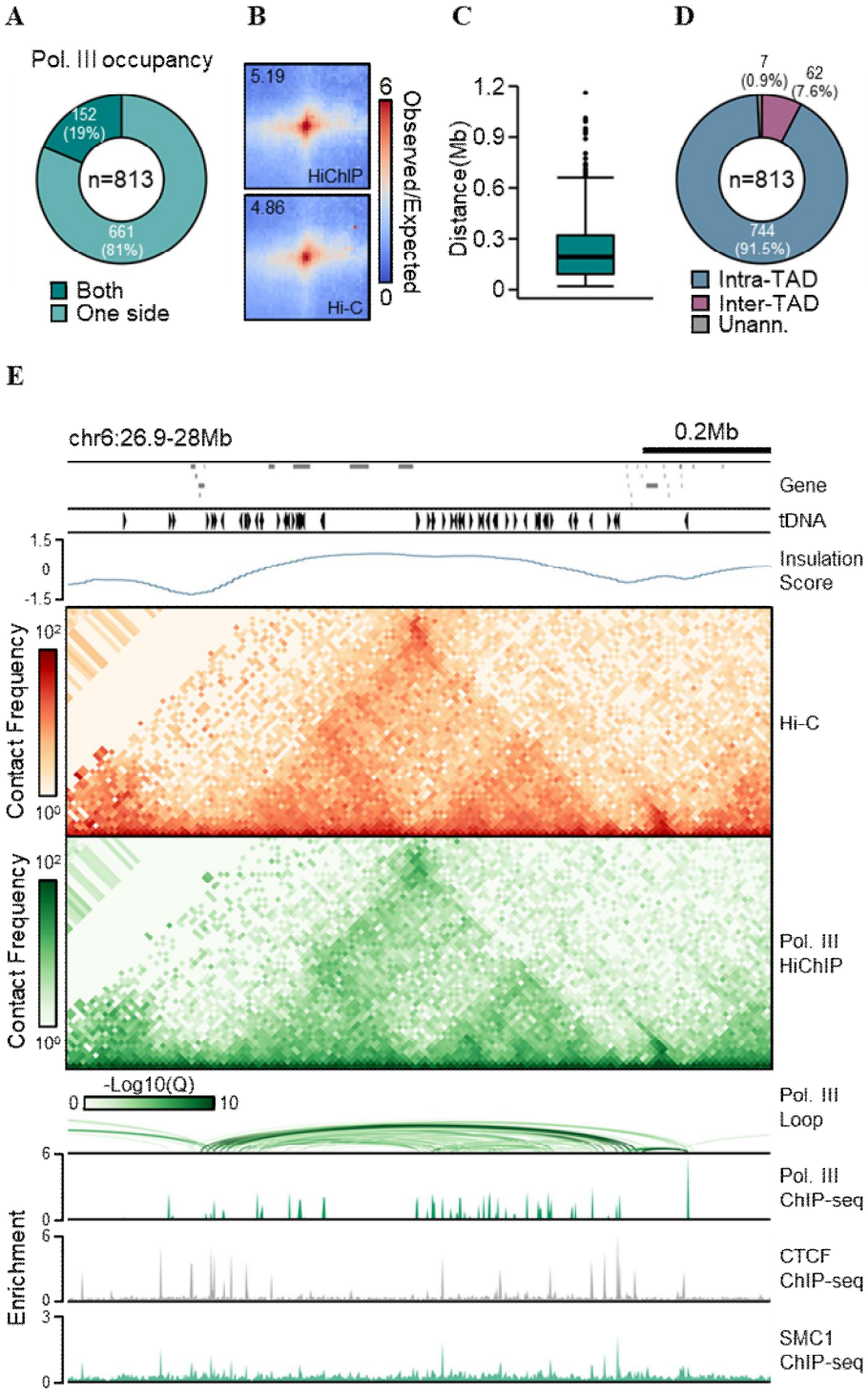

Figure 8. Identification of RNA Pol. III-mediated chromatin interactions. (A) Doughnut plot shows the distribution of Pol. III loop, classified by the presence of Pol. III peaks on loop anchor. Total number of Pol. III loops is indicated in the center of plot. (B) Heatmaps show the aggregated contact frequency of HiChIP(top) and Hi-C(bottom) in Pol. III loops. (C) Box plot shows the distribution of distances between two genomic anchors connected by the Pol. III. (D) Doughnut plot showing the number of Pol. III loops classified by intra-TAD or inter-TAD interaction. (E) Snapshot of insulation score curve, *in situ* Hi-C contact map, RNA Pol. III HiChIP contact map, Pol III HiChIP loops, and ChIP-seq signal tracks for of Pol III, CTCF and SMC1. The location of Pol. II genes and tDNA genes are shown on top of the snapshot.

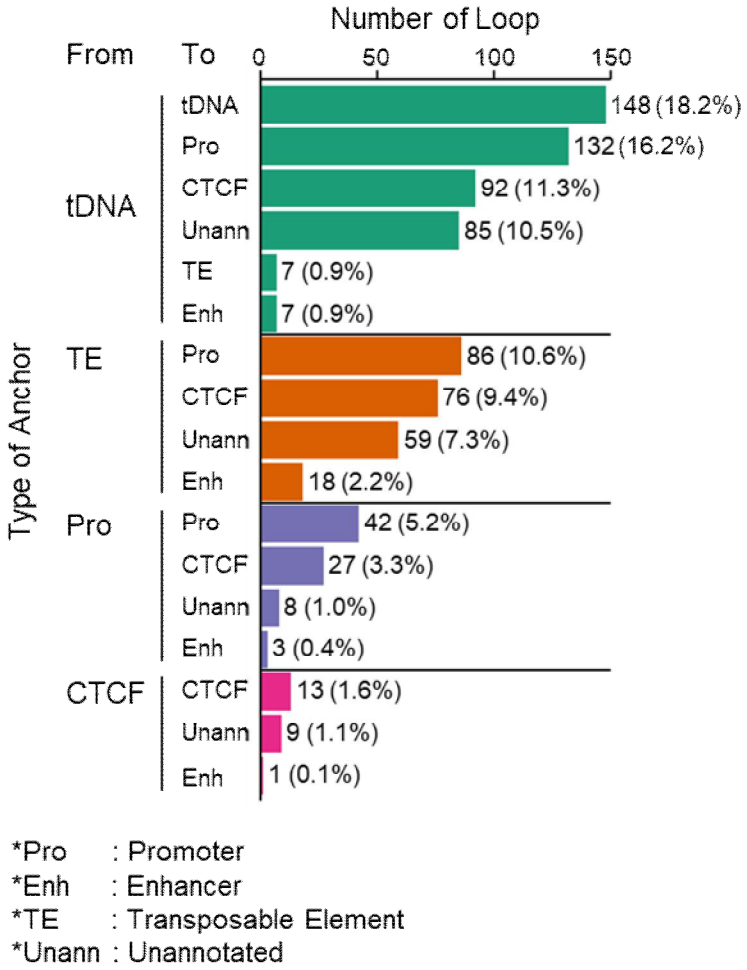

Figure 9. Classification of Pol. III loops categorized by their types of anchor.

8. Pol. III mediated chromatin interaction network

A total of 796 unique anchors form Pol. III mediated loops. Pol. III loop interaction patterns have very complicated patterns as illustrated in Fig. 8E. This means that one anchor is interacting complexly with several other anchors. According to the distribution of connections by anchor, most of the anchors (573, 72%) are interactive with one anchor, while the remaining (223, 28%) are connected with two or more anchors, with up to 21 interactions (Fig. 10A). These anchors were annotated to determine whether a particular type of region was related to a connection, and surprisingly, anchors with a large number of connections was found to involve a large number of tDNAs. In addition, the higher number of connections anchors have, the higher rate of tDNAs are involved (Fig. 10B). Of the anchors with more than 5 connections, 62% were tDNA anchors. It means that the Pol. III loop interactions are concentrated around the tRNA genes. Next, the distribution of tRNA gene expression level included by anchor groups according to the number of connections to see if there were any differences, but there were no significant differences (Fig. 10C). Furthermore, there were not much differences in expression between tRNA genes(200) related to Pol. III loop interaction and genes(57) which are not involved in chromatin interactions (Fig. 10D). In order to define a complicated Pol. III loop structure, network analysis was performed. The clique, a structure of two or more anchors connected, simplifies this complex relationship. From 813 Pol. III loops, 145 interaction network units, 3D cliques, were defined, and the number of connections included in each clique was calculated and the distribution was confirmed (Fig. 11A). The clique with the most connections has 196 chromatin interactions, followed by the clique with 37 interactions. 75% of cliques have no more than 5 interactions. It was confirmed above that many of the loop anchors have a high proportion of tDNA. Cumulative curve shows the distribution of tDNA counts by each clique (Fig. 11B). The clique

containing one or more tDNAs is 57 out of 145 and the clique containing the largest number of tDNAs has 72 tDNAs. Moreover, about half of the 200 tDNAs which are connected by Pol. III loops are located in 4 clique structure. A visualization of the top 3 connectivity and 10 randomly extracted 3D clique structures illustrates the relationship between tDNA anchors and anchors of different types within the network structure (Fig. 11C). The core of the rank 1 clique shows that several tDNA anchors are complex network and are also connected to the Pol. II gene promoter anchor around them. On the other hand, cliques with low connectivity are mainly composed of anchors that seem less relevant to tDNA or are not annotated. These results show the characteristics of chromatin structures specific to Pol. III biology.

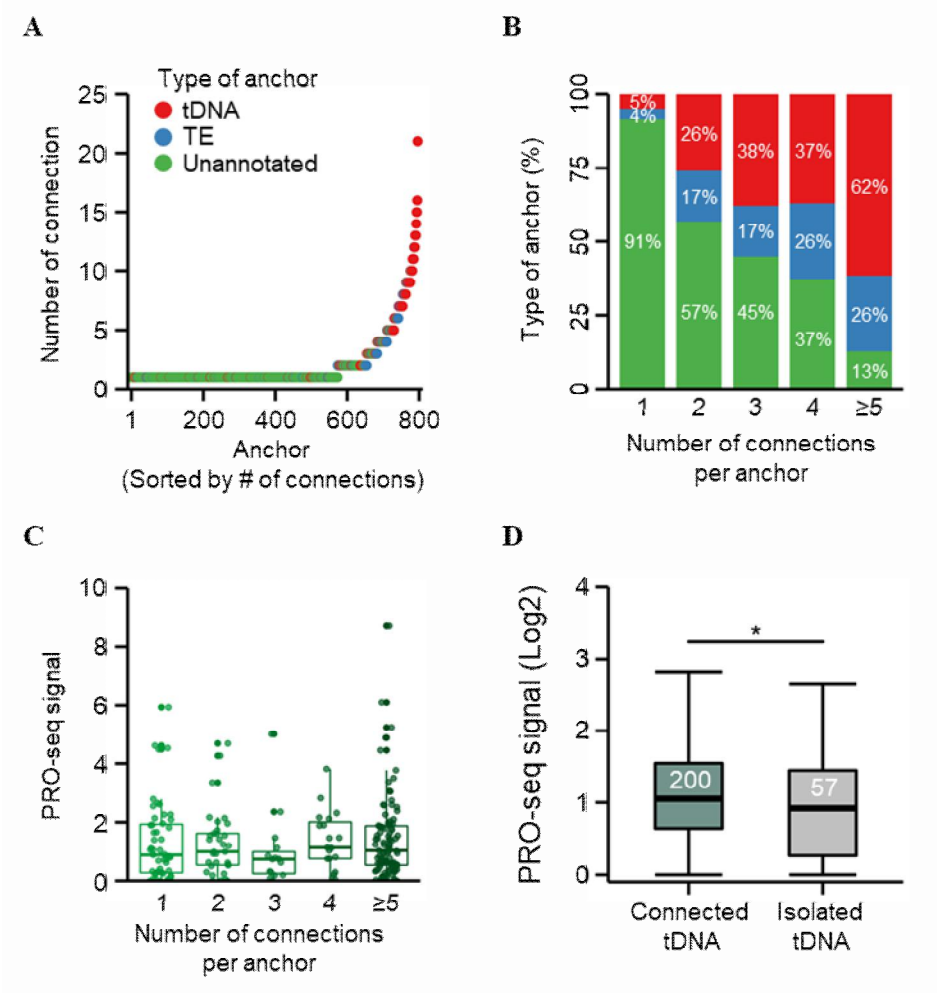

Figure 10. Genes encoding tRNA are highly connected by Pol. III-mediated chromatin interactions. (A) Number of connections categorized by type of anchor. (B) Proportion of anchors classified as the type of anchor. Anchors are sorted by the number of connections per anchor. (C) Box plot showing PRO-seq signals of tDNA genes categorized by the number of connections per anchor. (D) Box plot showing PRO-seq signals of tDNA genes either connected by Pol. III loops or isolated. Significance was determined by t-test (*, p-value < 0.05).

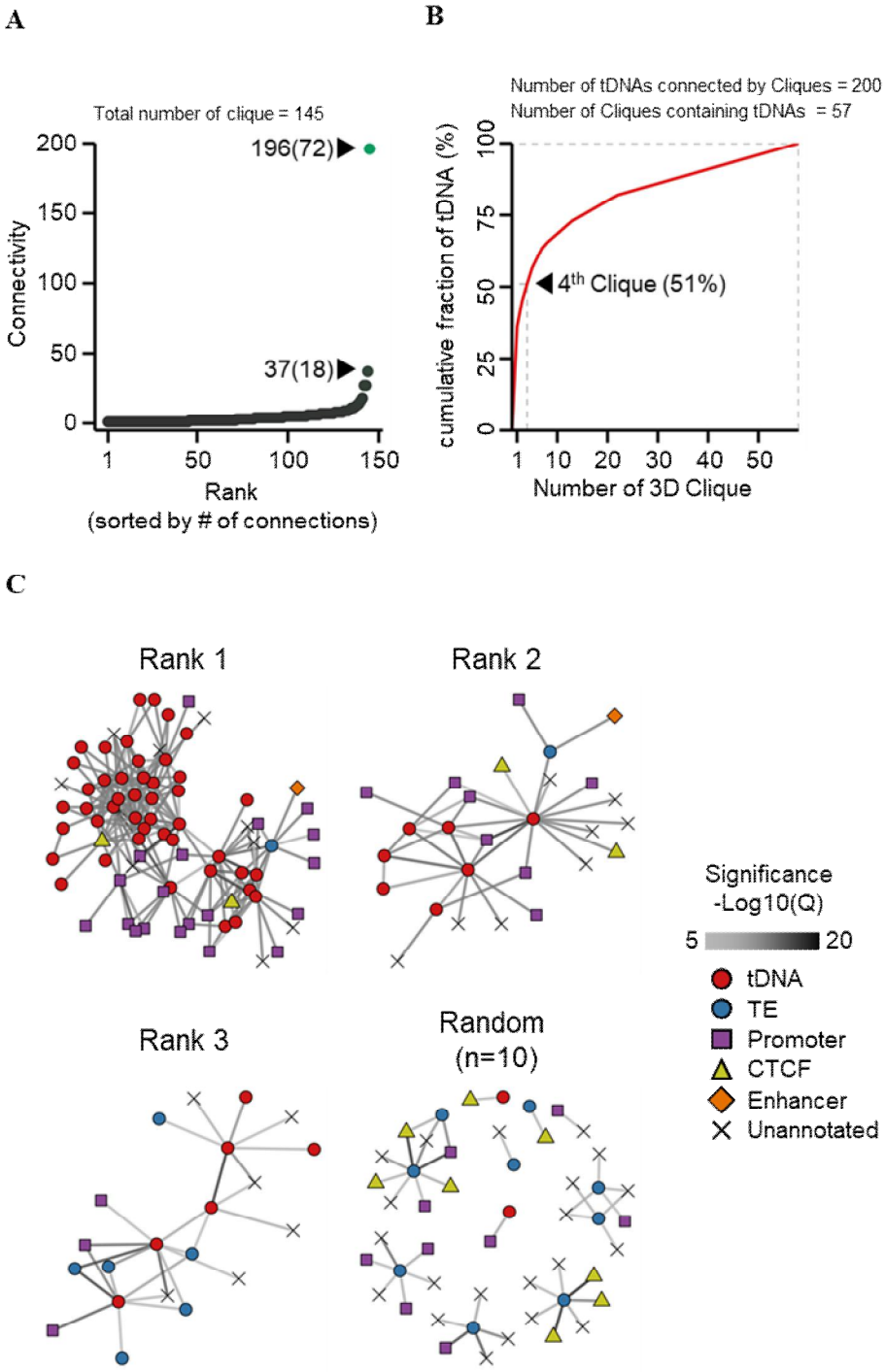

Figure 11. Pol III-mediated hyperconnected 3D cliques. (A) Distribution of connection in 3D cliques based on Pol III HiChIP loops. The number of tDNAs in top 2 hyperconnected cliques is provided in parenthesis. (B) Cumulative curve showing the distribution of the number of tDNA in each 3D clique. (C) 3D cliques where each edge represented a significant Pol III-mediated chromatin interaction. The color of each edge indicates the loop strength (-Log10(Q)). The color of each node represents tDNA, TE, promoter, CTCF, enhancer, or unannotated region.

IV. DISCUSSION

RNA Polymerase III is a specialized enzyme responsible for transcription of tRNA, 5S rRNA, 7SL RNA and other short non-coding RNAs and is involved in translational system. Despite detailed understanding of protein-coding gene transcription by Pol. II, the transcription of non-coding RNA genes by Pol. III is less well characterized. Chromatin structures are known to be closely related to the regulation of gene expression. Several omics data have been analyzed to investigate the chromatin structure that may regulate Pol. III gene transcription.

ChIP-seq was performed on two independent biological replicates of HCT116 cells, using antibodies against several target proteins involved in transcriptional chromatin states and Pol. III gene transcription. Pol. III binding sites throughout the genome were profiled by Pol. III ChIP-seq. Consensus Pol. III binding sites are 330. 80(24.2%) peaks of total 330 are located in chromosome 6 (Fig. 1B). tRNA genes which are known to be primarily transcribed by Pol. III is distributed genome-wide and, in many cases, linearly organized into genomic clusters[6]. This large number of Pol. III binding sites observed in chr6 suggests that clustered tRNA genes may be located in chr6. In order to find out what kind of regions Pol. III binds the most, tRNA genes and Repeat sequences downloaded from GtRNAdb and UCSC were annotated to Pol. III peaks. the result shows that Pol. III is binding to 257 tRNA genes, followed by 87 to SINE(Short Interspersed Nuclear eElement), 23 to LINE(Long Interspersed Nuclear Element), and many other annotated repeat elements and indicates that the peak calculated from Pol. III When box plots were created to identify whether there was an element type according to Pol. III binding density, Most of the regions with strong Pol. III binding signal seemed to be tRNA. Of course, there were some places where the Pol. III signal was strong among TE and unannotated regions which is excluded when the tRNA gene and the TE annotation (Fig. 1D).

RNA Pol. III binding tRNA region indicate a pattern of histone modifications

associated with actively transcribed Pol. II genes such as H3K27Ac, H3K4me1, H3K4me3 (Fig. 4A-C), but weakly correlated with tRNA gene transcription at Pol. III dependent genes. It is a result that reflects the epigenetic state of the surrounding environment, rather than precisely indicating the state of the binding site. Transcription factors such as CTCF, SMC1, BRD4, Pol. II signals have also been enriched at tRNA genes with Pol. III (Fig. 5A-C). Among these factors, Pol. II and BRD4 were highly correlated with PRO-seq signals. However, the H3K27me3 signal was higher in the tRNA region where Pol. III was not binding (Fig. 4D). These shared pattern between Pol. III and Pol. II genes are also shown in other studies[15,39,40]. These data reveal that certain features of chromatin are shared between genes transcribed by Pol. III and Pol. II on a genome-wide scale.

Quantifying Pol. III gene expression level has been considered challenging due to its unique compexity. Cellular tRNA becomes highly modified after transcription and consequently difficult to reverse transcribe during library preparation[6]. Mapping RNA Pol. III as an alternative also shows inappropriate measurement. To overcome this challanges, PRO-seq data was used, which can quantify nascent RNA transcripts. Correlation test also shows that there is a positive correlation between Pol. III binding density and expression (Fig. 1E). To verify whether tRNA expression is dependent on Pol. III binding, tRNA regions were divided into the genes with and without, and PRO-seq signals were compared. The result shows that the PRO-seq signal was much higher in the region with Pol. III (Fig. 1F). This result means that Pol. III genes bound by Pol. III are actively transcribed.

tRNA genes have been reported to play a role in the organization of eukaryotic chromosomes in certain case[41,42]. However, Little has been studied how tRNA genes are related to chromatin structures and how chromatin structures play a role in regulating these tRNA expressions. To identify chromatin interaction directly mediated by Pol. III, loops were called from HiChIP data. 813 Pol. III mediated loops were identified (Fig. 8A), and mainly formed within TAD (Fig. 8D). In addition, many loops connect tRNA gene to tRNA gene region, followed by Pol. II

gene promoter, and CTCF binding sites (Fig. 9). These results indicate that Pol. III is participating in long-range chromatin interactions, and also has well organized chromatin structure. Network anlaysis identified the 3D hyperconnected clique structures and revealed that many of tRNA genes are involved in a few cliques in restricted regions of the genome (Fig. 11A, B). A research previously reported that the transcription of tRNA gene which are linearly organized into cluster decreases in developing macrophages[6]. In summary, these results suggest that these Pol. III-centric chromatin structures may contribute co-regulation of Pol. III genes in responding to internal or external stimuli.

V. CONCLUSION

Eukaryotic gene expression process is very complicated and precisely regulated by coordinated systems. The relationship between Pol. III gene transcription and 3D chromatin structure is one of the important questions to be explored. In this study, several sequencing methods, such as ChIP-seq, ATAC-seq, PRO-seq, Hi-C, HiChIP, were analyzed to reveal a novel feature of the Pol. III centric 3D chromatin structures which may contribute Pol. III gene regulation in HCT116 colorectal cancer cells. As a result, actively transcribed Pol. III genes were identified by measuring Pol. III occupancy and nascent RNA expression. Accessible chromatin structure and active histone modification states exhibited positive correlations with Pol. III-dependent gene expression. In point of view of the genome architecture, higher-order chromatin structure confined most of tRNA genes into only a few topologically associated domains. And 80% of actively transcribed tRNA genes were physically connected with each other via long-range chromatin interactions. Furthermore, half of tRNA genes were transcribed within as few as 4 hyperconnected spatial clusters. This study suggests that highly interconnected Pol. III-mediated chromatin interactions may contribute to the coordinated transcriptional regulation of tRNA genes in response to intrinsic and eternal stimuli.

REFERENCES

1. Carter R, Drouin G. Structural differentiation of the three eukaryotic RNA polymerases. Genomics 2009;94:388-96.
2. Policarpi C, Crepaldi L, Brookes E, Nitarska J, French SM, Coatti A, et al. Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. Cell Rep 2017;21:2879-94.
3. White RJ. Transcription by RNA polymerase III: more complex than we thought. Nat Rev Genet 2011;12:459-63.
4. Yeganeh M, Hernandez N. RNA polymerase III transcription as a disease factor. Genes Dev 2020;34:865-82.
5. White RJ, Gottlieb TM, Downes CS, Jackson SP. Cell cycle regulation of RNA polymerase III transcription. Mol Cell Biol 1995;15:6653-62.
6. Van Bortle K, Phanstiel DH, Snyder MP. Topological organization and dynamic regulation of human tRNA genes during macrophage differentiation. Genome Biol 2017;18:180.
7. Schmitt BM, Rudolph KL, Karagianni P, Fonseca NA, White RJ, Talianidis I, et al. High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. Genome Res 2014;24:1797-807.
8. Yeganeh M, Praz V, Carmeli C, Villeneuve D, Rib L, Guex N, et al. Differential regulation of RNA polymerase III genes during liver regeneration. Nucleic Acids Res 2019;47:1786-96.
9. Upadhya R, Lee J, Willis IM. Maf1 is an essential mediator of diverse signals that repress RNA polymerase III transcription. Mol Cell 2002;10:1489-94.
10. Kolovos P, Knoch TA, Grosveld FG, Cook PR, Papantonis A. Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin 2012;5:1.

11. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012;485:376-80.
12. Bickmore WA. The spatial organization of the human genome. Annu Rev Genomics Hum Genet 2013;14:67-84.
13. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2001;2:292-301.
14. Sexton T, Schober H, Fraser P, Gasser SM. Gene regulation through nuclear organization. Nat Struct Mol Biol 2007;14:1049-55.
15. Bhargava P. Regulatory networking of the three RNA polymerases helps the eukaryotic cells cope with environmental stress. Curr Genet 2021.
16. Gerber A, Ito K, Chu CS, Roeder RG. Gene-Specific Control of tRNA Expression by RNA Polymerase II. Mol Cell 2020;78:765-78 e7.
17. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 2009;10:669-80.
18. Barrington C, Georgopoulou D, Pezic D, Varsally W, Herrero J, Hadjur S. Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. Nat Commun 2019;10:2908.
19. Donati B, Lorenzini E, Ciarrocchi A. BRD4 and Cancer: going beyond transcriptional regulation. Mol Cancer 2018;17:164.
20. Zhang T, Zhang Z, Dong Q, Xiong J, Zhu B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biol 2020;21:45.
21. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 2005;120:169-81.

22. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 2019;20:207-20.
23. Bhattacharyya S, Chandra V, Vijayanand P, Ay F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. Nat Commun 2019;10:4221.
24. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 2019;47:D766-D73.
25. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res 2016;44:D184-9.
26. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589-95.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
28. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187-91.
29. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc 2012;7:1728-40.
30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-2.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-9.
32. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 2015;16:259.

33. Ramirez F, Bhardwaj V, Arrigoni L, Lam KC, Gruning BA, Villaveces J, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun 2018;9:189.
34. Kruse K, Hug CB, Vaquerizas JM. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. Genome Biol 2020;21:303.
35. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 2016;3:95-8.
36. Chiang MY, Radojcic V, Maillard I. Oncogenic Notch signaling in T-cell and B-cell lymphoproliferative disorders. Curr Opin Hematol 2016;23:362-70.
37. McArthur E, Capra JA. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am J Hum Genet 2021;108:269-83.
38. Liang Z, Fu XD. 3D genome encoded by LINE and SINE repeats. Cell Res 2021;31:603-4.
39. Moqtaderi Z, Wang J, Raha D, White RJ, Snyder M, Weng Z, et al. Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. Nat Struct Mol Biol 2010;17:635-40.
40. Barski A, Chepelev I, Liko D, Cuddapah S, Fleming AB, Birch J, et al. Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. Nat Struct Mol Biol 2010;17:629-34.
41. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, et al. Human tRNA genes function as chromatin insulators. EMBO J 2012;31:330-50.

42. Snider CE, Stephens AD, Kirkland JG, Hamdani O, Kamakaka RT, Bloom K. Dyskerin, tRNA genes, and condensin tether pericentric chromatin to the spindle axis in mitosis. J Cell Biol 2014;207:189-99.

ABSTRACT(IN KOREAN)

대장암세포에서 RNA 중합효소 III 의존적 유전자들의 위상학적 구조

< 지도교수 김 형 표 >

연세대학교 대학원 의과학과

김 용 진

진핵 세포에는 세 가지 주요한 RNA 중합효소가 존재한다. 그 중 RNA 중합효소 III는 단백질 번역 시스템이 있어서 중요하며 여러가지 다양한 non-coding RNA를 전사하는 것으로 알려져 있다. RNA 중합효소 III의 전사는 세포 주기, 분화, 재생, 세포 스트레스와 같은 여러 가지 다양한 생명현상 안에서 조절된다. 이러한 RNA 중합효소 III 전사과정에서의 결함은 여러 가지 질병을 야기하거나 관련되어 있다고 알려져 있다. 그러므로 RNA 중합효소 III가 전사하는 유전자들의 조절 메커니즘을 이해하는 것은 중요하다.

Compartment 또는 TAD 로 나누어지는 유전체의 위상적 구조는

유전자 조절에 있어서 중요한 역할을 할 수 있다. 그러나 RNA 중합효소 III 의존적 유전자들이 위상학적으로 어떻게 조직되어 있고 여러 계층의 크로마틴 구조와 어느 정도로 연관되어 있는지는 잘 알려져 있지 않다.

최근 몇 년 동안 유전체의 3차원 크로마틴 구조를 연구하고 유전자 발현을 측정하기 위한 많고 다양한 high-throughput sequencing 방법들이 개발되어 왔다. 본 연구에서는 RNA 중합효소 III에 의해 매개되는 크로마틴 구조를 연구하기 위해 ChIP-seq, ATAC-seq, *in situ* Hi-C, HiChIP과 같이 후성유전학 연구에서 널리 사용되는 다양한 시퀀싱 데이터와 nascent RNA를 정량 할 수 있는 PRO-seq 데이터를 분석하였다. 그 결과 이러한 NGS 데이터의 통합 분석 결과로부터 활발히 전사되는 RNA 중합효소 의존적 유전자들을 규명할 수 있었고, 또한 많은 수의 RNA 중합효소 III 의존적 유전자들은 크로마틴 구조를 통해 위상적으로 가까이 모여있는 것을 확인하였다.

핵심되는 말 : RNA 중합효소 III, 3차원 크로마틴 구조, tRNA