

博士論文

**A Computational Analysis of the Impact of 3D  
Chromosome Organization on its Gene  
Expression Regulation**

(染色体 3 次元構造がその遺伝子発現制御に及ぼ  
す影響の情報解析)

ナガイ ルイス アウグスト エイジ

THE UNIVERSITY OF TOKYO

DOCTORAL THESIS

---

**A Computational Analysis of the  
Impact of 3D Chromosome  
Organization on its Gene  
Expression Regulation**

---

*Author:*  
Luis Augusto Eijy  
NAGAI

*Supervisor:*  
Dr. Kenta NAKAI

*A thesis submitted in fulfillment of the requirements  
for the degree of Doctor of Philosophy*

*in the*

Laboratory of Functional Analysis *in silico*  
Department of Computational Biology and Medical Sciences  
Graduate School of Frontier Sciences

July, 2018

# Abstract

Luís Augusto Eijy NAGAI

*A Computational Analysis of the Impact of 3D  
Chromosome Organization on its Gene Expression  
Regulation*

In eukaryotes, the long genomic DNA is divided into chromosomes and densely compacted within a small nuclear compartment. Consequently, each chromosome is non-randomly organized and the three-dimensional (3D) architecture has an important effect on cellular function. Using the information of genomic DNA, cells differentiate and exert their function mediated by a spatiotemporal control of gene expression. Although cell differentiation has been attributed to the fine orchestration of specific genes, the full scope of how gene expression is regulated by the 3D chromatin organization remains to be clarified. Moreover, how the differences in the chromatin structure of normal and abnormal cells impact the regulation is also an important and unanswered question.

Throughout the chromosomes, topologically associating domains (TADs) represent contiguous regions of chromatin with high self-interaction and provide means for distal regulatory elements to interact with their target

genes (Lieberman-aiden et al. 2009). The compartmentalization of specific chromosome territories has been disclosed by fluorescence in situ hybridization (FISH), and recent studies using genome-wide high-throughput chromatin conformation capture (Hi-C) have shown that the 3D organization of the mammalian genome and its chromatin modifications are conserved between cell-lineages and species.

In this manuscript, I have attempted to investigate the chromatin dynamics in normal and abnormal B cell lymphocytes and clarify its functional impact on the regulation of gene expression. By integrating publicly available multi-omics data, I dissected the 3D structure reorganization loci and observed a subtle coordination with gene expression level changes. Remarkably, the set of genes in dynamic changes are involved in B cell biological process. Moreover, the boundaries of topologically associating domains were significantly enriched with Prdm1 motif, which is a key factor for dysfunctional in aggressive lymphoma. Although chromatin reorganization appeared in less impact on the gene regulation, I have discovered an unknown mechanism that impacts the structure and function of chromosomes and cognates with genes in a specific manner. My findings suggest the presence of intricate cross-talk between the higher-order chromatin structure and cancer development.

Overall, I attempted to investigate: (i) how does the 3D genomic structure defined by repressive and permissive compartments domains could be reflecting the cell-specific gene regulation, (ii) whether TADs coordinate with chromatin compartments for gene regulation, and (iii) whether specific chromatin interactions would define the fine orchestration of transcription.



# Contents

<b>Abstract</b>	<b>1</b>
<b>Contents</b>	<b>3</b>
<b>List of Figures</b>	<b>6</b>
<b>List of Tables</b>	<b>8</b>
<b>1 Introduction</b>	<b>10</b>
1.1 Overview . . . . .	10
1.2 The Eukaryotic Genome in a Three-Dimensional Nuclear Organization . . . . .	11
1.2.1 The Nuclear Lamina . . . . .	13
1.2.2 The Nuclear Pore Complex . . . . .	13
1.2.3 The Nucleolus . . . . .	14
1.3 Visualization of the Genome Organization . . . . .	16
1.3.1 3C-based techniques . . . . .	16
1.4 Chromosome Organization in the Nuclear Space . . . . .	19
1.4.1 Chromosome Territories . . . . .	20
1.4.2 Compartments . . . . .	22
1.4.3 Topologically Associating Domains . . . . .	23
1.4.4 Chromatin Loops . . . . .	25

	4
1.5 Summary . . . . .	27
1.6 Goal of my research . . . . .	29
<b>2 Material and Methods</b>	<b>30</b>
2.1 Data preparation . . . . .	30
2.2 RNA-seq data analysis . . . . .	31
2.3 Hi-C matrix . . . . .	31
2.4 Compartment and TAD identification . . . . .	32
2.5 Compartment classification of mouse genes . . . . .	32
2.6 Gene ontology enrichment analysis . . . . .	33
2.7 Calculating normalized scores . . . . .	33
<b>3 Analyzing the 3D Chromatin Organization Coordinating with Gene Expression Regulation in B-cell Lymphoma</b>	<b>35</b>
3.1 Background . . . . .	35
3.2 Results . . . . .	38
3.2.1 Highly Conserved Folding Patterns on Chromatin Compartment Domains from Heterogeneous NGS Resources . . . . .	38
3.2.2 Extensive Reorganization of the Mouse Genomic Com- partments and the Impact on Gene Expression Levels	40
3.2.3 Influence of topologically associating domains on compartment reorganization . . . . .	43
3.2.4 TAD boundaries suggest gene regulation function in cancer . . . . .	45
3.2.5 Inter-chromosomal interactions in B-cell lymphoma	47
<b>4 Discussion</b>	<b>61</b>

	5
<b>5 Conclusion</b>	<b>64</b>
<b>6 Future Directions</b>	<b>65</b>
<b>Acknowledgements</b>	<b>66</b>
<b>7 Bibliography</b>	<b>68</b>
<b>A Supplementary Figures</b>	<b>78</b>
<b>B Supplementary Tables</b>	<b>81</b>

## List of Figures

1.1	Representation of chromosome territories during interphase inside the nuclear lamina . . . . .	12
1.2	The nuclear envelope is composed of two membranes that are fused at nuclear pore complexes . . . . .	14
1.3	Representation of the genome organization from FISH and C-techniques . . . . .	17
1.4	Visualization of single gene (Nanog) interaction points in the whole-genome . . . . .	19
1.5	Chromosome organization across different scales . . . . .	21
1.6	Influence of CTCF of TAD structures. . . . .	24
1.7	Regulation of gene expression by distal regulatory elements	26
2.1	Proposed integrative pipeline for multi-omics data analysis using Hi-C, RNA-seq and ChIP-seq . . . . .	34
3.1	Representation of switching regions from compartment identification analysis . . . . .	39
3.2	Chromatin organization at compartment level in ES cell, pro-B cell, and Lymphoma . . . . .	49
3.3	Compartment normalized score by chromosome size . . .	50

3.4	Profile of gene expression in compartment reorganization comparing ES cell, pro-B cell and lymphoma . . . . .	51
3.5	Profile of gene expression in compartment reorganization in comparison with random genes from stable profile . . .	52
3.6	Enriched GO biological process terms in set of genes from compartment reorganization . . . . .	53
3.7	GO terms enriched in switching regions. . . . .	54
3.8	Violin plot of topologically associating domain size . . . .	55
3.9	Chromatin organization at TAD level in ES cell, pro-B cell, and Lymphoma . . . . .	56
3.10	Normalized TAD number per chromosome size performed in ES cell, B cell, and B cell lymphoma . . . . .	57
3.11	Heatmap of set of some of genes known to be involved in B-cell fate and B-cell lymphoma . . . . .	58
3.12	Comparison of gene expression values based on the chro- matin structure disposition of TADs . . . . .	59
3.13	Significant intra- and inter-chromosomal interaction ratios in mouse cells (P-value < 0.001) . . . . .	59
3.14	Distribution of inter-chromosomal interaction pairs in B- cell lymphoma . . . . .	60
A.1	Principal Component Analysis of gene expression values .	78
A.2	Representation of the correlation between compartments and gene expression . . . . .	79
A.3	Heatmaps of the chromatin interaction maps of the whole genomes of Pro-B cell and Lymphoma . . . . .	80

## List of Tables

3.1	Gene expression values including compartment A and B information of set of genes known to be involved with B cell and lymphoma. . . . .	41
B.1	RNA-seq detailed data information . . . . .	81
B.2	Hi-C detailed data information . . . . .	82
B.3	Gene enrichment analysis in Pro-B cell (from B to A com- partment). GO terms, biological function and statistical test (FDR). . . . .	83
B.4	Gene enrichment analysis in B-cell lymphoma (from B to A compartment). GO terms, biological function and statis- tical test (FDR). . . . .	84

Dedicated to my mentor in  
life Dr. Daisaku Ikeda, my  
family, friends and my  
beloved future wife.

# Chapter 1

## Introduction

### 1.1 Overview

In this dissertation work, I dissected the eukaryotic 3D genome structure and its relationship with gene regulatory mechanism focusing on the differential structure of chromatin in specific contexts of cell and function. To tackle this problem, I extend a comprehensive analysis on chromosome conformation data in healthy/disease study using pro-B cell, B-cell derived lymphoma, and embryonic stem cell to delineate chromatin conformation principles involved in disease development. Prior to detailing my findings, this chapter provides an overview of: (i) the current knowledge of the nuclear organization interaction units and technologies applied, (ii) distinct levels of chromosomal organization inside the nucleus, and (iii) the main goals and questions leading this work.



## 1.2 The Eukaryotic Genome in a Three-Dimensional Nuclear Organization

The compaction of the eukaryotic genome within a small nucleus has been studied for over a century. At the molecular level, the genomic DNA binds and wraps around the surface of the histone octamer (eight histone protein cores) to form a nucleosome. Histones, represented by two copies of the four types H2A, H2B, H3, and H4, are the most abundant proteins bound to chromatin. The histone amino (N)-terminal tails are highly basic and protrude from each nucleosome interacting with neighbor nucleosomes thus affecting inter-nucleosomal interactions in such a way that also affects the overall chromatin architecture (Bannister and Kouzarides 2011). This structure is the fundamental unit of the eukaryotic chromatin and represents the first level in packaging, which effectively reduces the initial length by 7-fold (Berger 2007). However, not only reducing chromosome size function, but it is also recognized that the chromatin structure has crucial importance for genome functions, such as transcriptional control. For example, posttranslational modification (PTMs) of histone tails can control the interaction between nucleosomes and DNA, thereby controlling access to specific regions of the DNA in a spatiotemporal coordination manner. Inside an interphase nucleus, chromatin forms distinct microenvironments composed of at least two subtypes of chromatin: euchromatin and heterochromatin. Euchromatin is the type that decondenses showing a “open” or “active” chromatin, which also contains most of the coding genes; and heterochromatin, condensed structure also called “closed” or “inactive” chromatin, often

contains repetitive DNA and fewer genes (Grunstein 1997). These distinct environments affect transcriptional control by establishing physical connections among regulatory elements and target genes. However, in order to understand how this mechanism works, we need to know not only the information encoded in the sequence of DNA but also the ways this sequence is organized in chromosomes (Gibcus and Dekker 2013) as well as the positioning inside the nucleus. Prior to dissect the chromatin structure of the DNA, I briefly introduce the physical components of the nucleus and their influence in the chromatin conformation (Figure 1.1).

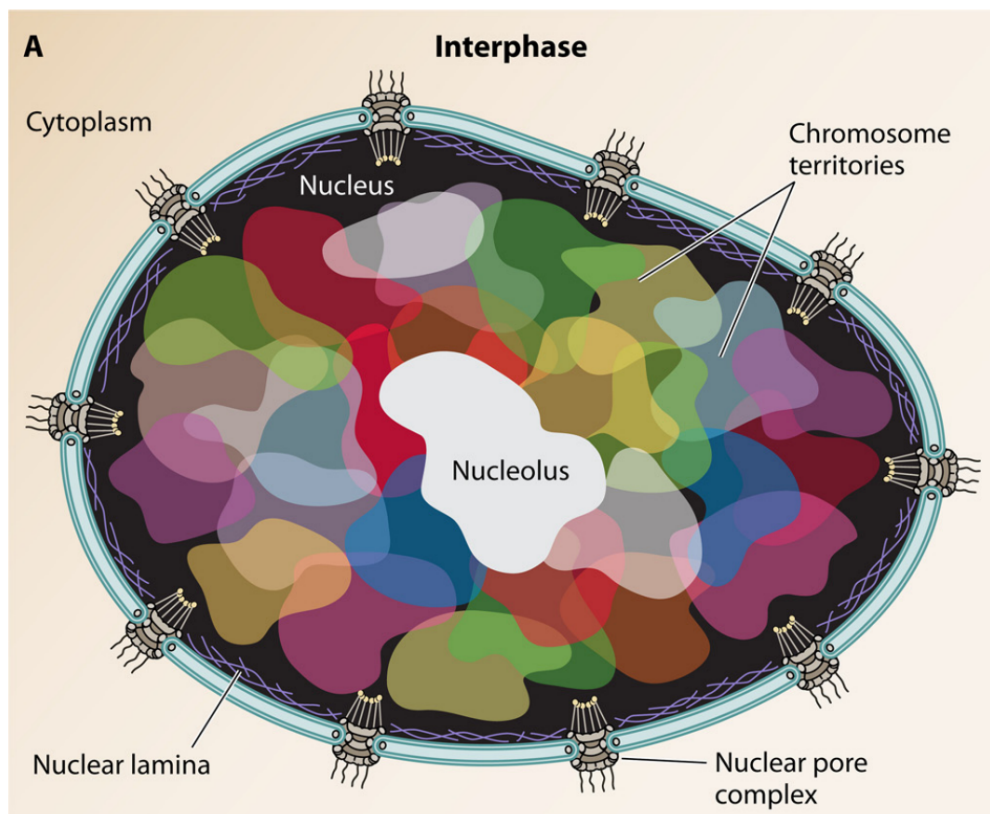


FIGURE 1.1: Representation of chromosome territories during interphase inside the nuclear lamina perforated by nuclear pore complexes. In the center, the nucleolus is shown in white (From (Fraser et al. 2015)).

### 1.2.1 The Nuclear Lamina

One of the main characteristics of eukaryotes is the presence of a barrier that organizes their genome inside a compartment, the nucleus, and separates it from the cytoplasm. The boundary of the nucleus is formed by a two-layer structure called nuclear envelope (NE). The outer nuclear membrane (ONM) and inner nuclear membrane (INM), which combined with nuclear envelope transmembrane proteins (NETs) and lamin proteins form the nuclear lamina (Figure 2). The nuclear lamina is essential for many activities such as replication, RNA transcription, nuclear and chromatin organization, cell cycle regulation, cell development and differentiation, nuclear migration, and apoptosis (Gruenbaum et al. 2003). Lamina interacts with chromatin through lamin-associated domains (LADs) and contain transcriptionally inactive heterochromatin (Gonzalez-Sandoval and Gasser 2016). Moreover, the peripheral layer of chromatin, also called perinuclear heterochromatin, is associated with gene silencing and regulates the position of chromosomes in the nucleus. Thus, studies suggest that just the localization of chromatin regions near LADs is often sufficient to decrease gene expression level (Finlan et al. 2008; Kumaran and Spector 2008; Stevens et al. 2017).

### 1.2.2 The Nuclear Pore Complex

The nuclear envelope (NE) contains more than 500 individual protein pores and form a nuclear pore complex (NPC). This structure has the function of controlling the transportation of macromolecules between the

cytoplasm and nucleus. In addition, NPC is also involved in the regulation of gene expression, however, whereas nuclear lamina is associated with heterochromatin and gene silencing, the NPCs are enriched for associations with active genes and euchromatin (Brown et al. 2008; Capelson et al. 2010). NPC creates a gap in the nuclear lamina, which does not interact with chromatin (Figure 1.2).

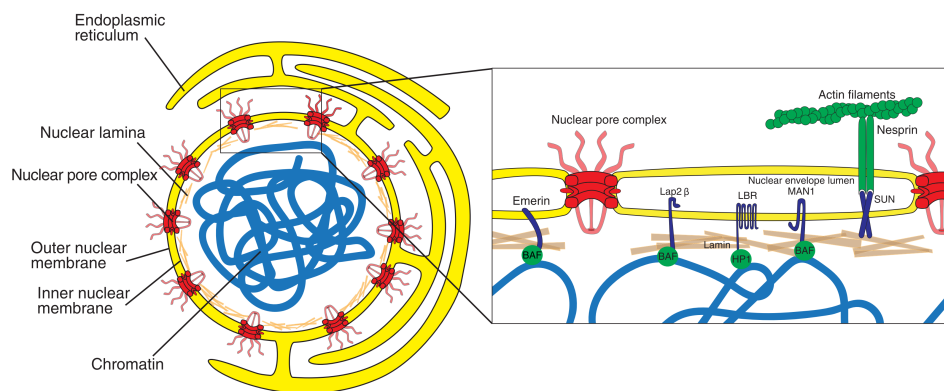


FIGURE 1.2: The nuclear envelope is composed of two membranes that are fused at nuclear pore complexes in red. Not only have the function of mediating molecular transportation, but also is responsible for chromatin interaction and gene regulation (From (Magistris and Antonin 2018)).

### 1.2.3 The Nucleolus

The Nucleolus is the largest structure within the nucleus and is also involved in genome organization. Nucleoli are dense structures dedicated to expression by RNA polymerase I where rRNA synthesis and assembly occurs. However, several RNA polymerase II genes were found around nucleolus when transcriptionally inactive (Koningsbruggen et al. 1992; Stevens et al. 2017). The nucleoli are formed around ribosomal-DNA

gene clusters located in specific chromosomes (Zink, Fischer, and Nickerson 2004). The size and shape of nucleoli typically reflect the rate of ribosome production; however, the prominent changes were also observed in many types of cancer. For example, increased nucleoli are important for Hodgkin's disease diagnostic, however in the opposite direction, inconspicuous nucleoli were used for small-cell anaplastic lung cancer diagnostic (Zink, Fischer, and Nickerson 2004). It has been proposed that changes in nucleoli observed in cancer cells do not always be related to cancer development and other functions other than ribosome synthesis might be related to the nucleoli dynamic sizes.

Interestingly, previous studies found similar characteristics between nucleolus-associated domains (NADs) and LADs. For example, GC- and gene-poor contacts, the size range and median sequence length in NADs (0.1-10 Mb; 749 kb) and in LADs (0.1-10 Mb; 553 kb) (Németh et al. 2010), and finally, both loci demonstrated heterochromatic regions which chromosomes formed subnuclear structures clustering for actively transcription with RNA pol I and pol III, and not with RNA pol II (Kind et al. 2013; Thomson et al. 2004).

## 1.3 Visualization of the Genome Organization

In order to investigate the genome organization guided by chromatin interactions within cis- and trans- chromosomes, two main approaches have been extensively used and derived (Figure 1.3). The first technique is microscopy-based techniques, including several fluorescence in situ hybridization (FISH) visualized by high-resolution microscopy. It creates a photography-type image in high-resolution and is used to directly measure distance between DNA segments. The second approach uses molecular techniques such as chromosome conformation capture (3C), which infers distance between DNA segments by quantifying the frequencies of contacts between two loci considered closed in vivo.

### 1.3.1 3C-based techniques

The development of 3C technologies allowed researchers to investigate the 3D chromosomal folding principles at high resolution by quantifying contact probability between two contact loci (Dekker et al. 2002). Later, several 3C-based techniques appeared, such as circularized chromosome conformation capture (4C), carbon copy chromosome conformation (5C), chromatin interaction analysis coupled to paired-end tagging (ChIA-PET), genome-wide 3C (Hi-C). The basic steps among them includes cell cross-linking using formaldehyde followed by chromatin sonication or restriction endonucleases, followed by ligation of cross-linked fragments that were spatial close. Then, junctions are detected, sequenced and quantified (Denker and De Laat 2016) (Figure ??). After, interaction between ligation is calculated based on the number of reads produced

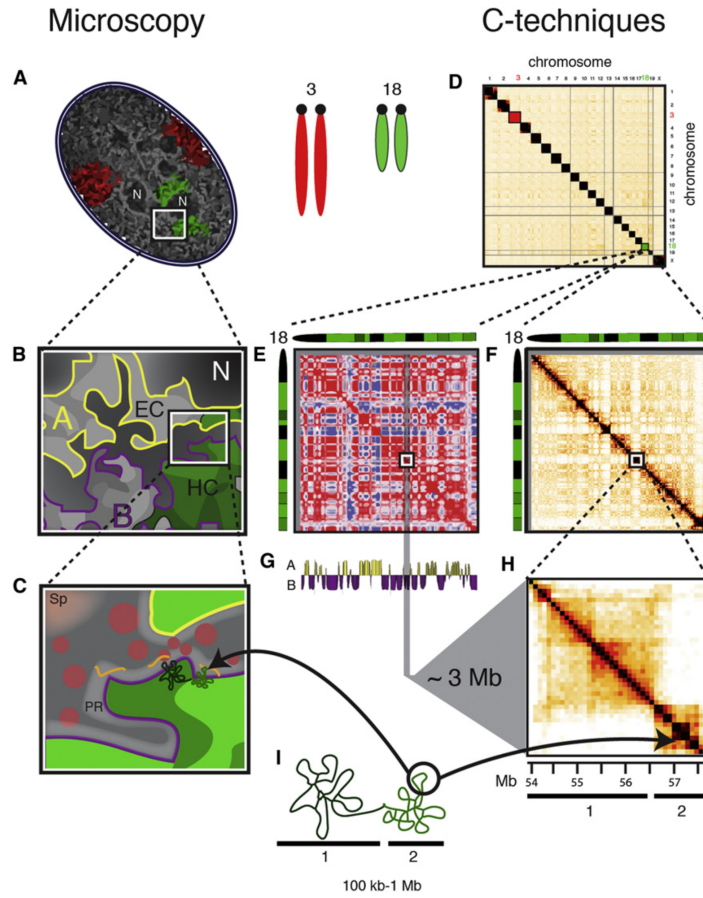


FIGURE 1.3: Representation of the genome organization from the two main approaches. On the left panels, a single cell can be visualized in high resolution microscopy (A); chromosomes territories and inter connections (B-C). On the right side, chromatin maps from interaction matrix are usually reported as heatmaps and describe the probability of two loci proximity. The high-resolution allows go to regions from kilobase of distance from each other (D-H) (From (Gibcus and Dekker 2013)).

and normalized. Among 3C techniques, Hi-C offers whole-genome contact maps in an “all-versu-all” resource (Lieberman-Aiden et al. 2009). In this project, I use Hi-C approach to interrogate genome-wide chromatin interactions reorganization loci.

More recently, 3C-derived approach integrated with computational modeling and image processing, have revolutionized the chromosome organization studies. By creating high-resolution maps ( $<1$  Kb), scientists could create virtual map of an entire cell and confirm genome-wide interactions between regulatory elements and single genes (Stevens et al. 2017) Figure 1.4). They observed that the interactions of a single gene are complex and difficult to predict, in fact, the interactions are variable and might involve many others biological process in a cell-specific manner. Moreover, recent advances have allowed analysis in very high-resolution, however, the scalability of the increasing resolution experiments and the corresponding size of Hi-C data-sets created a primary concern and bottleneck for Hi-C analysis making crucial the use of big servers and high performance machines (Sauria et al. 2015).



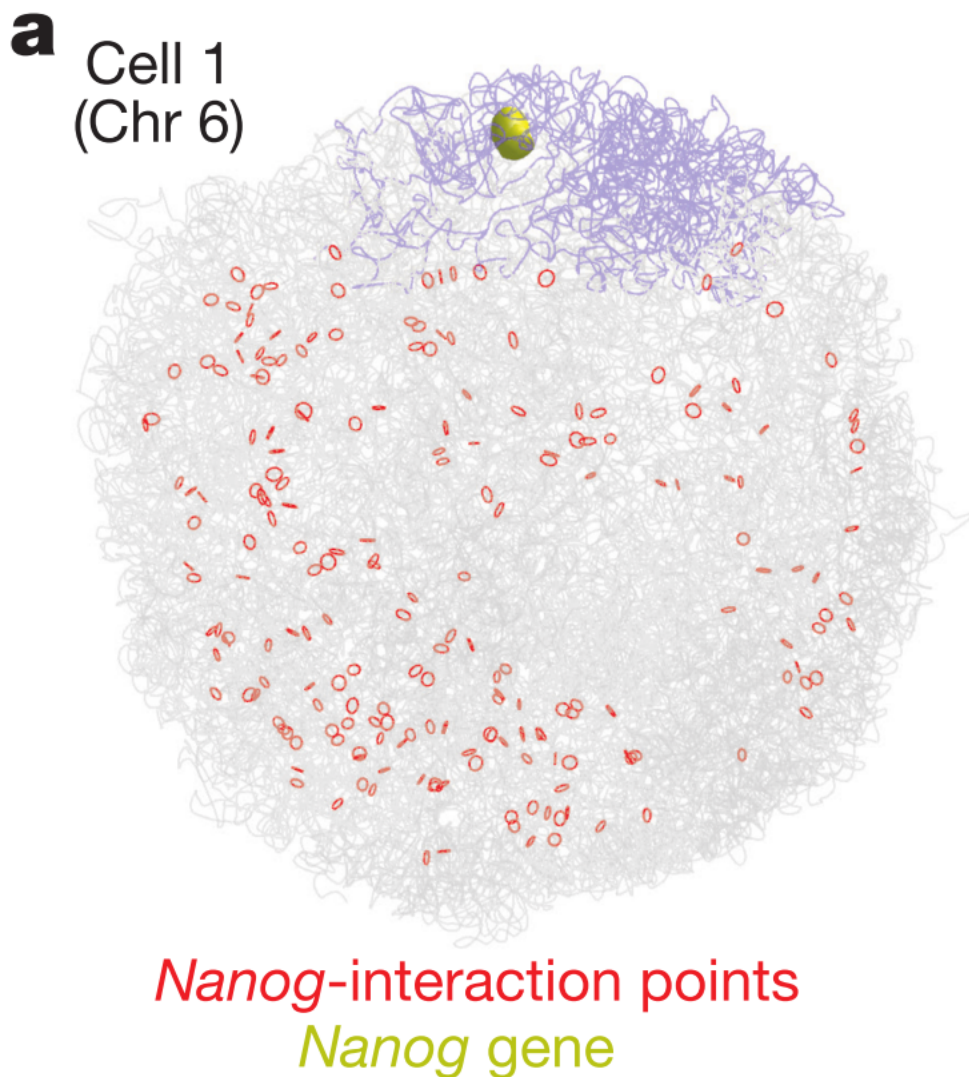


FIGURE 1.4: Visualization of single gene (Nanog) interaction points in the whole-genome (From (Stevens et al. 2017)).

## 1.4 Chromosome Organization in the Nuclear Space

From FISH studies in living cells, researchers first interpreted the nuclear content as lack of organization between cell types. Later studies using

higher resolution got improved imaging results and revealed several organization patterns in the whole nucleus (Boyle et al. 2011). (i) chromosomes in interphase not only do not mix with each other but instead tends to locate specific territories. (ii) from a single chromosome position, each chromosome can connect to one nearby and potentially forms functional interactions. (iii) RNA polymerase II and other transcription machinery tends to occur in subnuclear loci instead of randomly. (iv) lastly, transcriptionally active and inactive segments of the genome tend to form groups and interact with each other within in the same group rather than other type. These observations created the basis for a subdivision of the chromosome organization (Figure 1.5) into chromosome territories, compartment, topologically associating domains (TADs), and chromatin loops.

### 1.4.1 Chromosome Territories

Inside the nucleus, the first hierarchical the genome is organized into chromosomes territories (CTs). This chromosomal structure disclosure was verified by microscopic approaches at increased resolution and details (Schermmelleh, Heintzmann, and Leonhardt 2010). More recently, with the development of 3C technologies the chromosomal arrangement was also confirmed by Hi-C. They have observed that CT is not randomly located inside the nucleus. In fact, although each chromosome tends to interact more with itself, they also observed inter-chromosomal interaction between small, gene-rich loci in Hi-C human lymphocytes (Lieberman-Aiden et al. 2009). Moreover, the contact loci of those small

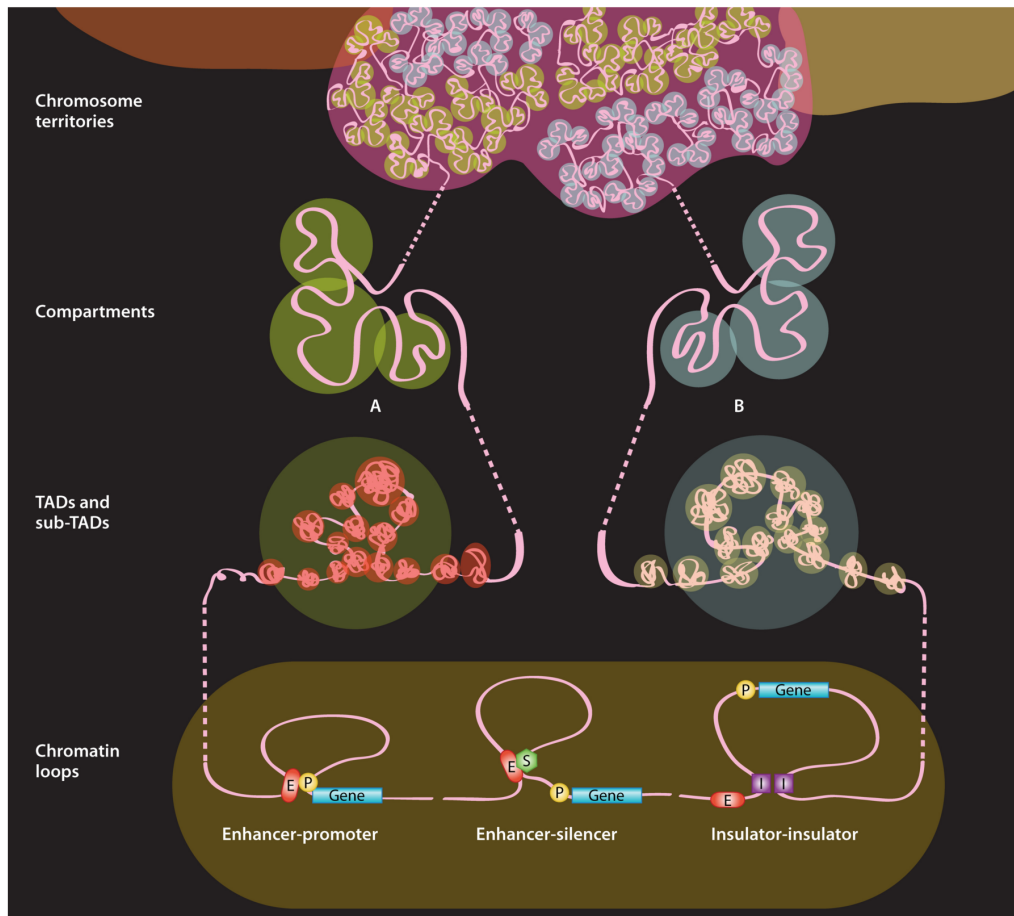


FIGURE 1.5: Chromosome organization across different scales (From (Fraser et al. 2015)).

loops were coincidentally enriched with open chromatin and active expression (Apostolou et al. 2013). These results suggest that chromosomes have some predefined position inside the nucleus in order to facilitate specific contact with other CTs, and this association influence gene expression control.

### 1.4.2 Compartments

The genome can be divided into two compartments, called A and B compartments. Compartments were established at megabase scale and represent constrained chromatin interactions between loci more likely interacting with regions belonging to the same compartment type than expected based on their size (Lieberman-Aiden et al. 2009). The current consensus method for compartment identification is based on the Principal Component Analysis (PCA) of the normalized interaction matrices from Hi-C experiment. This method was coined by (Lieberman-Aiden et al. 2009) which defined a normalized contact matrix and observed a plaid pattern of large blocks of enriched and depleted interactions. These plaid patterns can be decomposed in two sets of regions, which contacts within same set are enriched and contacts from the other set are depleted. Then, the principal component 1 (first eigenvector) represent the compartment score, and a genomic window with positive and negative scores define A and B compartments. A compartments are assigned to the regions more gene-rich, usually more transcriptionally active, enriched for the active histone H3K36me3 modification, and more accessible to DNase I than the B compartment (Eagen 2018). In contrast, B compartments show higher frequencies of chromatin contacts, enrichment for silencing histone H3K9me3 modification, stronger tendency for self-association, and closed chromatin. Another characteristic regards to the position is that B compartments tend to correlate with late replication timing, LADs, and more recently to NADs, suggesting proximity to peripheral regions but also around the hollow nucleoli (Stevens et al. 2017).

The partition of CTs into A and B compartments has been observed in several cell types and autosomes. In fact, the distribution of A and B compartments along chromosomes is considered highly conserved across cell types, however small changes were observed, suggesting important cell-type regulatory function (Lin et al. 2012; Phillips-Cremins et al. 2013).

### 1.4.3 Topologically Associating Domains

Topologically associating domains (TADs) has been considered one of the most important 3D chromosomal conformation feature and one of the most recent discovery (J R Dixon et al. 2012). These structures are known to have more interactions within themselves than with neighbor units (E P Nora et al. 2012). TADs are identified in Hi-C studies in whole genome, divides the genomes in 2000 units with average size of 880 kb (J R Dixon et al. 2012). It has been proposed that TADs are conserved among different cell types and tissues (Rowley et al. 2017; Weinreb and Raphael 2016). An important feature regarding TADs is the assumption of regulating gene activity by specific spatial chromatin folding. This structure are found to cover around 90% of the genomes in mammals (J R Dixon et al. 2012) and usually the separation between consecutive TADs along chromosomes is associated with binding sites for architectural CCCTC-factor (CTCF) and cohesin (Figure 1.6). Several studies have confirmed the presence of CTCFs at TAD boundaries (Dekker and Heard 2015; Narendra et al. 2015; Rodríguez-Carballo et al. 2017), however, remains unclear whether boundaries of TADs deplete or increase interaction frequencies between neighbor TADs (Elphège P. Nora et al. 2017). In fact, in a recent

study (Ruiz-Velasco and Zaugg 2017) found a robust evidence that 3D chromatin architecture has strong influence on DNA mutations around TAD and chromatin loop boundaries.

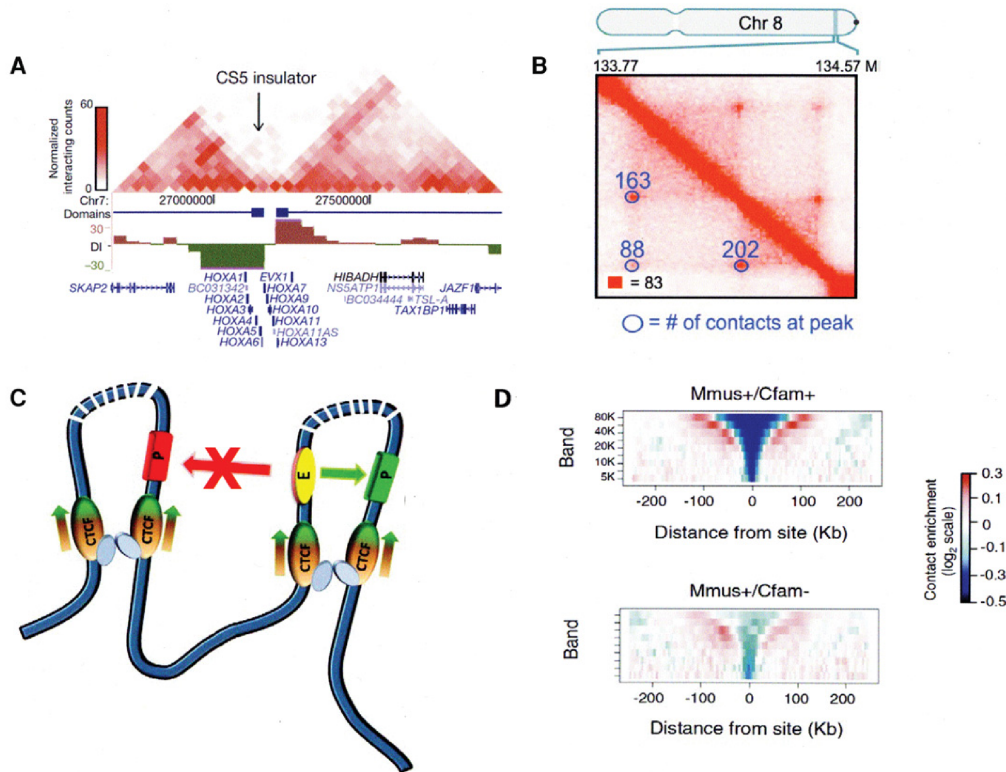


FIGURE 1.6: Influence of CTCF of TAD structures. TADs of HOXA locus delimited with CTCF insulator site (A). Contact peaks Dixon, et al 2012 (B), CTCF pairs interacting between enhancers and promoter in chromatin loops. Interaction from different loops are blocked by CTCFs. (D) reduced frequency of contacts across CTCFs boundaries sites conserved between mice and dogs (From (Ghirlando and Felsenfeld 2016))

### 1.4.4 Chromatin Loops

Inside TADs, chromatin loops are another class of chromatin interaction that also receive influence of CTCF and Cohesin. This interaction enable the connection of distal enhancers to their target promoter-genes (Shlyueva, Stampfel, and Stark 2014). Enhancer-promoter interactions are known to play important role in cell-specific gene expression programs, which underlie cellular identity. Enhancer interaction activity is typically restricted to its target gene within the same TAD (Lupiáñez et al. 2015). Moreover, epigenetics histone modifications can bind enhancer regions and regulate it (Figure 1.7). This suggests that TADs and chromatin loops have similar function in the chromatin organization.

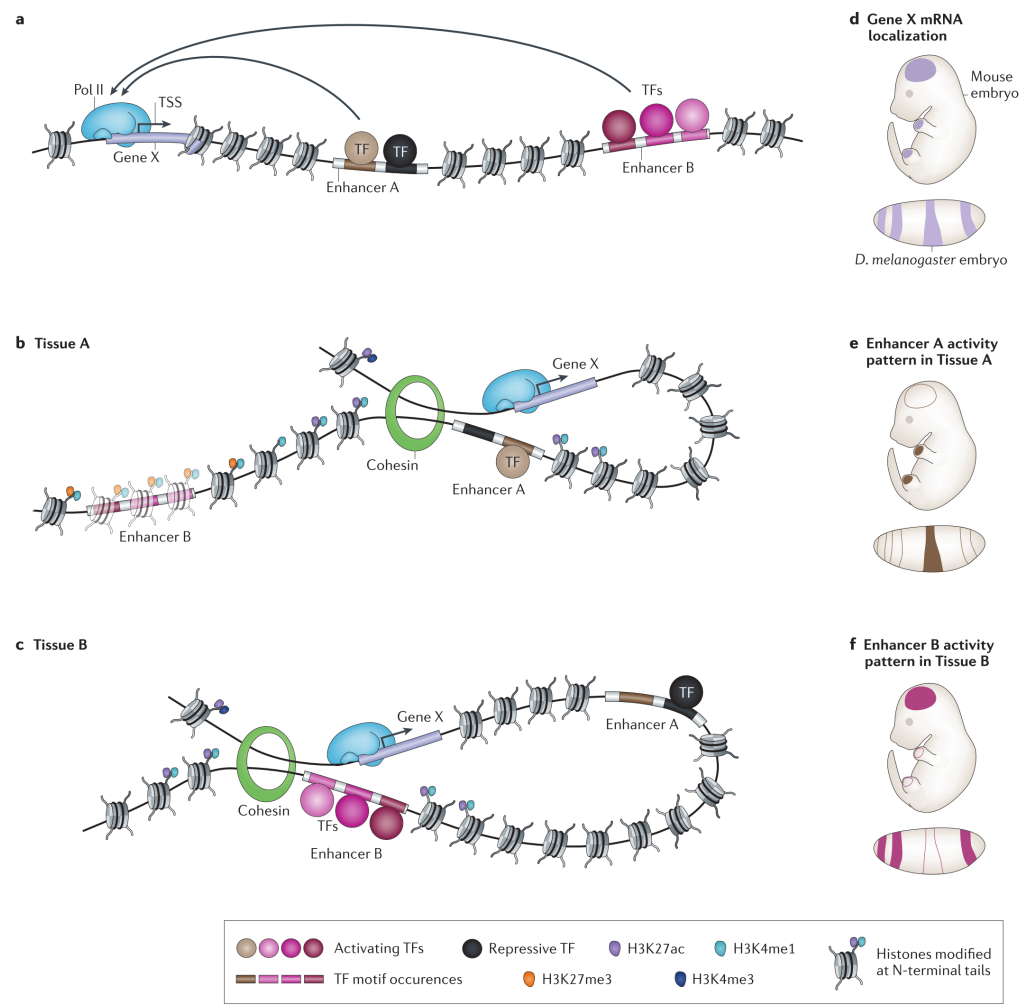


FIGURE 1.7: Regulation of gene expression by distal regulatory elements. Along the linear genomic DNA, enhancer can be located far away from its transcription start site and enhance gene expression in a cell/tissue-specific manner (From (Shlyueva, Stampfel, and Stark 2014)).



## 1.5 Summary

Several studies have revealed important aspects of the 3D genome architecture. (i) In the normal cell cycle, entering the mitotic process chromosomes begin to condense and can be visualized by microscope. After mitosis, chromosomes quickly establish a different conformation in which is proposed to start the formation of compartments, TADs and long-range interactions in a cell type-specific manner (Dekker 2014). Researchers have observed in different stages of interphase, that the A and B compartmentalization and TADs can change their formation in a manner that correlates with DNA replication (Nagano et al. 2017). (ii) Other studies have focused on the differentiation from the human embryonic state and the relationship of the chromatin organization. They found that TADs are important units to organize and facilitate intra- and inter- TAD interactions. This also correlated with transcription levels and epigenetic states (Jesse R. Dixon et al. 2015). Also in human cells, scientists have tracked immune system cells during differentiation from naïve B cells to germinal center B cells and revealed cell specific networks between promoters and enhancers (Bunting et al. 2016). (iii) During reprogramming, embryonic stem cell(ESC)-like structures were observed in somatic cells (Krijger et al. 2016). At a higher order chromatin structures, ESCs and induced pluripotent stem cells (iPSCs) have very similar compartment and TADs structures suggesting important role of the 3D organization to the cell function.

Despite several efforts to connect the chromatin conformation to gene

regulation in different biological scenarios, a contribution in disease reorganization of the 3D genome architecture remains necessary. Other studies using disease-associated variants to target genes using genome-wide association studies (GWAS) have clarified that target genes of distal regulatory elements are not always the closest one in linear manner (Dunham et al. 2012; Javierre et al. 2016). Another study have demonstrated that TADs disruption are associated with interdomain interactions between TAD neighbors contributing to different orchestration of gene expression and aberrant formation in mouse (Lupiáñez et al. 2015).

In conclusion, a great opportunity to study the complexity of 3D genome organization as well as the gene regulation mechanism involved in this process has become the focus of current state-of-the-art research in chromosomal conformation interaction and became the focus of this research dissertation.

## 1.6 Goal of my research

The main goal of this dissertation is to start tackling the complex challenge of merging epigenetics with transcriptomics by integrating the dynamics of differential chromatin structures with gene expression levels to clarify how do the changes in the 3D structure might be affecting gene expression regulation. More specifically, (i) I aim to investigate chromatin dynamics upon locus-specific and genome-wide manner by first investigating compartments reorganization in different cell contexts; (ii) to gain information on basic principles delineating three-dimensional chromatin organization function on its gene expression regulation; and (iii) defining higher-order chromatin interactions involved in the orchestration of gene transcription.

## Chapter 2

# Material and Methods

### 2.1 Data preparation

RNA-seq data were downloaded from Gene Expression Omnibus (GEO); (i) GSM2400249 and GSM2400250 for mouse embryonic stem cells (ES cells), cell type ES-E14, male, from total RNA-Seq. (ii) GSM1897405, GSM1897406, and GSM1897407 for mouse normal B cells, cell type C57BL/6 RAG<sup>-/-</sup> pro-B cell, pool of male and female individuals, from total RNA-Seq. (iii) GSM2072416 and GSM2072417 for mouse B-cell lymphoma from strain B10.H-2aH-4bp/Wts CH12.LX immortalized cell line, female, from total RNA-Seq. Complete information from samples can be found in Table B.1.

Hi-C data were downloaded from GEO; (i) GSM862720 and GSM862721 for mouse ES cells, cell type mESC line J1, male, from genomic DNA. (ii) GSM987818 for mouse pro-B cells, sex not informed, from genomic DNA. (iii) GSE63525 for mouse B-cell lymphoma, sex not informed, from genomic DNA. Complete information from samples can be found in Table B.2.

## 2.2 RNA-seq data analysis

Sequencing reads were first trimmed using Trimmomatic 0.36 (with the parameters: ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) (Bolger, Lohse, and Usadel 2014). Processed reads were counted and assigned to transcripts using Salmon (Patro et al. 2017). Gene expression levels were obtained by R/Bioconductor package tximport (Soneson, Love, and Robinson 2016).

## 2.3 Hi-C matrix

The paired-end reads were trimmed by Trimmomatic (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:25), and mapped separately to mm10 by BWA-MEM (Li and Durbin 2009) with gap extension penalty and clipping at 5' and 3' ends (-A 1 -B 4 -E 50 -L 0 -T 25 -t 10). The paired-end reads, which both ends were mapped into unique genomic locations, were further used. HiCExplorer 2.0 (Ramírez et al. 2018) built Hi-C matrices with read counts over the bins of unequal size considering restriction sites; HindIII (AAGCTT) for pro-B cell and ES cell, and MboI (GATC) for B cell lymphoma. Briefly, the values of rows and columns in a Hi-C matrix stand frequencies that any two bins were connected by any pairs of processed read. This process discards non-uniquely mapped reads, lower mapping score reads, duplicated, re-ligation and dangling ends. To avoid amplification biases, low count bins and higher outliers are filtered out by setting a threshold on bimodal distribution. Hi-C replicates of each sample were merged as

recommended in the HiCEXplorer manual. To avoid the sex dependent bias, I removed chromosome Y from Hi-C merged matrices. Then, iterative correction was performed as described in Imakaev et al. (Imakaev et al. 2012).

## 2.4 Compartment and TAD identification

HOMER (Heinz et al. 2010) performed principal component analysis (PCA) on normalized interaction matrices and integrated H3K36me3 peaks to assign positive values to A compartment and negative values to B compartment. I downloaded ChIP-seq BED files from ENCODE (Dunham et al. 2012); ENCSR000CGR for ES cells, ENCSR000CFY for B cells, and ENCSR000CFL for B-cell lymphoma. To identify TADs, I ran the program “hicFindTADs” of HiCEXplorer; it first transforms the Hi-C contact matrix into a z-score matrix considering all contacts at the same genomic distance. Then, separation score is computed for different values of window, and low scores are indicative of TAD boundaries. To compare submatrices values, Wilcoxon rank-sum test is applied and the highest of two p-values is kept. P-values are corrected using Bonferroni method and boundaries with  $p < 0.01$  are reported.

## 2.5 Compartment classification of mouse genes

To avoid multiple counting and assigning to both compartments, I considered only transcription start site (TSS) positions of genes. By using

the program “intersect” in bedtools (Quinlan and Hall 2010) with a parameter “-wo -F 1.0” , I prepared genes whose TSSs were overlapped either A or B compartments. I used the R language (<https://www.r-project.org/>) for visualizing. Representation of data integration in Figure 2.1.

## 2.6 Gene ontology enrichment analysis

I conducted gene ontology (GO) enrichment analysis using DAVID (Huang, Sherman, and Lempicki 2009). I first compartmentalized the genome as described above, and prepared gene sets that were located in different compartments in a pair of cells. I analyzed GO biological process terms for each gene set. I used 0.05 as a threshold of P-values and EASE score threshold (Maximum Probability) = 0.01. I exhibit additional GO terms in supplementary files using Bonferroni correction with threshold 0.05.

## 2.7 Calculating normalized scores

I calculated the normalized scores for compartments found samples; for each chromosome in a sample, A or B compartment count is divided by total number of compartments in the respective chromosome, and is divided by the chromosome size. I also calculated the normalized scores for TADs in the same manner.

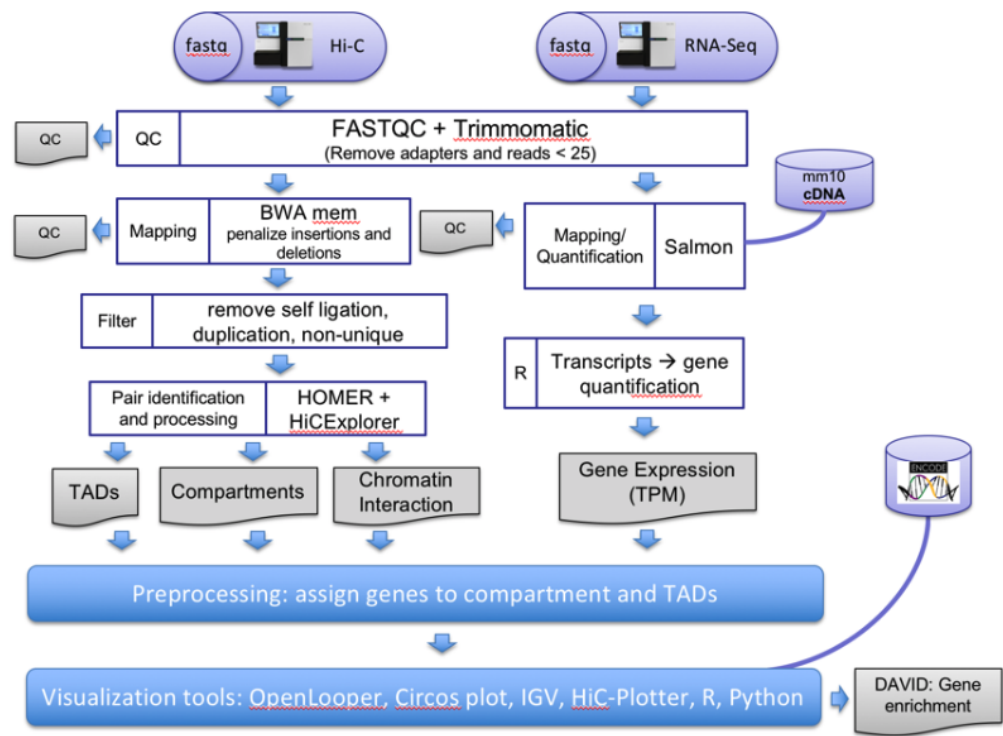


FIGURE 2.1: Proposed integrative pipeline for multi-omics data analysis using Hi-C, RNA-seq and ChIP-seq.



## **Chapter 3**

# **Analyzing the 3D Chromatin Organization Coordinating with Gene Expression Regulation in B-cell Lymphoma**

### **3.1 Background**

To define three-dimensional (3D) chromatin structures in eukaryotic nuclei, Chromosome Conformation Capture (3C) sequencing technologies, such as the genome-wide 3C version (Hi-C), have emerged as a promising strategy, and revealed that the 3D structures non-randomly compacted have a functional impact on gene expression (J R Dixon et al. 2012; Jin et al. 2013; Jung et al. 2017; Rudan et al. 2015; Yue et al. 2014). For example, in B cells (B lymphocytes), the nuclear lamina interacting directly and indirectly with DNA and chromatin are disrupted during early lymphocyte

development (Kosak et al. 2002). Another study (Lin et al. 2012) combining 3D fluorescence in situ and Hi-C analysis has shown that to establish a lineage specific signature, particular genome-wide structural transformations, i.e. chromatin compartment switches, are observed from loci of key developmental genes. In addition, the recent advances in 3C technologies have identified sub-compartment regions involved in B-cell fate determination (Bonev et al. 2017).

B cells are central in the humoral immune system and the abnormal gene regulation in the cells is associated with cancer development (Klein et al. 2001). Diffuse large B-cell lymphoma, one of the most common type of cancer in B cells, represents 30-40% of all non-Hodgkin lymphoma. Genetic translocations on the chromosome structure deregulate B Cell CLL/Lymphoma 6 (Bcl6) gene in germinal-center response in mouse giving rise to different types of lymphoma (Lenz and Staudt 2010). Moreover, a recent study (Xia et al. 2017) using gene expression profiling revealed that PRDM1/BLIMP-1, a master regulator of plasma-cell differentiation, is inactivated in lymphoma which loss of its expression correlates with tumor cell proliferation.

Here, I sought to identify chromatin dynamics involved in the gene regulation of B-cell lymphoma. I combined different scales of genome structures from Hi-C of published data (J R Dixon et al. 2012; Lin et al. 2012; Rao et al. 2014) with mouse gene expression profiles (RNA-seq). I observed that the higher-order chromatin organizations characterized as compartments and topologically associating domains (TADs) are highly conserved among cells. However, the compartment switch from repressive in B cells to permissive in lymphoma involves a specific gene set

( 8%) that exhibits increased gene expression levels in lymphoma. Interestingly, TAD boundaries are enriched with Prdm1 motif, suggesting coordination between the higher-order of chromatin structures and cancer development.

## 3.2 Results

### 3.2.1 Highly Conserved Folding Patterns on Chromatin Compartment Domains from Heterogeneous NGS Resources

Eukaryotic genomes are composed of sets of loci that are more likely to interact with one another than expected by random conformation of a chromosome. These sets show a plaid pattern that classifies each genomic locus into either A or B compartments (Lieberman-Aiden et al. 2009). Interaction maps from Hi-C data can provide information in multiple levels of genome organization hierarchy (Gorkin, Leung, and Ren 2014). The first level to examine chromatin interaction is the compartment domain. To examine the three-dimensional chromatin folding dynamics in B cell and lymphoma, I prepared public Hi-C data of pro-B cells (Lin et al. 2012), B-cell lymphoma (Rao et al. 2014), and ES cells (J R Dixon et al. 2012). Then, I performed PCA analysis at a higher resolution (100 kb) on the mouse genome.

Overall, my analysis classified 1.48 Gb of the mouse pro-B cell genome in B compartment, containing 4,900 genes, whereas 1.1 Gb was classified in A compartment, containing 14,600 genes. To compare the folding patterns among the cells, I compared the A and B coordinates as previously described (Boya et al. 2017) (Figure 3.1). I found that 84.5% and 83.8% of the genomic coordinates, in pro-B cell and lymphoma respectively, remained stable when compared them in ES cell. Furthermore, I found a higher similarity (88.4%) between pro-B cell and lymphoma compartment coordinates. These results are consistent with the observation in a study that

90.7% of compartments are conserved in pre-pro-B and pro-B mice cells (Boya et al. 2017). In human, a previous study found a similarity degree of 64% among ES cell and four derived lineage compartments (Jesse R. Dixon et al. 2015). While another study across 21 human cells and tissues, researchers observed 40.4% of conservation (Schmitt et al. 2016). These results suggest that my analysis achieves satisfactory chromatin organization structures by finding very similar chromatin compartment domains between pro-B cell and lymphoma using heterogeneous data resources.

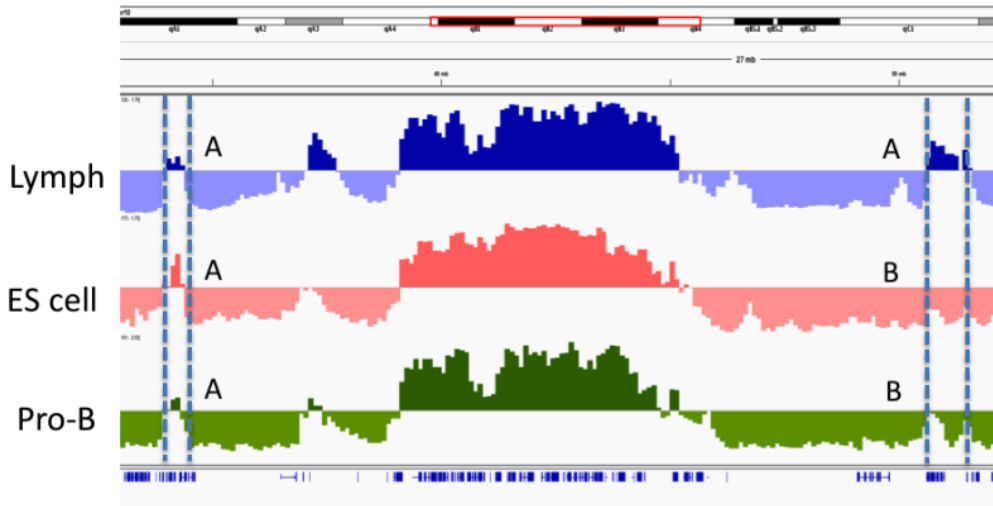


FIGURE 3.1: Representation of switching regions from compartment identification analysis. Positive values from PCA were assigned based on ChIP-Seq H3K36me in all samples. A compartment coordinates with active and gene rich region of the genome, whereas B compartment with inactive gene islands.

### 3.2.2 Extensive Reorganization of the Mouse Genomic Compartments and the Impact on Gene Expression Levels

It has been known that compartment reorganizations were associated with the disruption of normal gene expression program leading to breast cancer (Barutcu et al. 2015). In order to investigate whether this phenomenon is also observed in mouse B cell-derived lymphoma, I identified chromatin compartments at 100 Kb resolution from normalized chromatin interaction matrices and obtained gene expression values from RNA-seq paired-end sequences (Supplementary Figure A.1). I examined regions switching compartments across cell types where 8% switched from inactive B-cell compartment to active lymphoma compartment, 11% between pro-B cell and ES cell, and 12% between lymphoma and ES cells (Figure 3.2). This result is consistent with those of other studies; for example, the high gene expression level of *Myc* has changed from B compartment in ES cell to A compartment in B-cell lymphoma (Nguyen, Papenhausen, and Shao 2017). Another gene described in the literature is *Ebf1*, as an important regulator that orchestrates B-cell fate, changed from B to A compartments. Although it was possible to observe reported genes from literature switching compartments between cell-types, the overall correlation between gene expression and compartment classification was weak (Supplementary Figure A.2).

To investigate the influence of chromatin compartmentalization frequency in chromosomes, I calculated a normalization score by dividing the sum

TABLE 3.1: Gene expression values including compartment A and B information of set of genes known to be involved with B cell and lymphoma.

ID	Gene <sub>n</sub> name	Expression(TPM)			Compartment		
		EScell	Pro – Bcell	Lymphoma	EScell	Pro – Bcell	Lymphoma
ENSMUSG00000004698	<i>Hdac9</i>	0.160	0.361	0.63	B	A	B
ENSMUSG00000005583	<i>Mef2c</i>	0	1.683	4.39	B	A	A
ENSMUSG00000013523	<i>Bcas1</i>	1.544	0.617	0.075	A	B	B
ENSMUSG00000022346	<i>Myc</i>	12.073	8.008	47.999	B	A	A
ENSMUSG00000022508	<i>Bcl6</i>	1.056	1.338	0.248	B	A	A
ENSMUSG00000038151	<i>Prmd1</i>	1.635	0.484	6.121	A	A	A
ENSMUSG00000053175	<i>Bcl3</i>	22.805	2.049	15.269	A	A	A
ENSMUSG00000057098	<i>Ebf1</i>	2.076	209.817	35.367	B	A	A
ENSMUSG00000057329	<i>Bcl2</i>	0.339	0.852	6.766	A	A	A
ENSMUSG00000076617	<i>Ighm</i>	8.598	367.041	4023.223	NA	NA	NA
ENSMUSG00000095079	<i>Igha</i>	1.457	0.035	995.126	NA	NA	NA

of compartments in each chromosome by its chromosome size. The distribution of compartments throughout the genome was much more similar between pro-B cell and lymphoma than between those cells and ES cell (Figure 3.3). The genes located in the compartments switching from B to A tend to show increased expression levels, whereas the genes positioned in the A to B compartment change show the opposite tendency (Figure 3.4). I further compared the expression levels of switching-gene with those of random genes located in stable regions (Figure 3.5). The overall tendency of gene expression in compartment changes is subtle due in part to that a subset of genes are affected by the compartment changes (Barutcu et al. 2015; Jesse R. Dixon et al. 2015). I selected all the genes from changes of repressive B compartment to active A compartment for a functional enrichment analysis. Pro-B cell and lymphoma were enriched for similar GO terms related to B cell function, such as natural killer cell activation involved in immune response, humoral immune response, B cell proliferation, and immune response process (Figure 3.6). To identify specific GO terms from each of pro-B cell and lymphoma, I next filtered

out 846 common genes and found that pro-B cell were enriched with immune response terms, such as negative regulation of viral entry into host cell and proteolysis (Figure 3.7), whereas lymphoma was enriched with sensory perception of chemical stimulus, V(D)J recombination, negative regulation of T cell apoptotic process.

Taken together, in a global level, my results indicate that the genes identified in A and B compartments have a high specificity for the cell type when compared GO term enrichment. In addition, changes in compartment corresponded to changes in gene expression, indicating that A and B compartments might have contributions to the cell differentiation although it might not be determinant.



### 3.2.3 Influence of topologically associating domains on compartment reorganization

I next examined the sub-compartment structure organized into dense and contiguous self-interacting regions known as TADs (J R Dixon et al. 2012). Although TADs tend to be conserved across different cell types, chromatin interactions vary from cell to cell (Jesse R. Dixon et al. 2015). Here I raised the question about the possibility that TADs contribute to the gene expression programs between pro-B cell and lymphoma.

At 40 kb resolution, my TAD calling classified the genome structures into similar numbers; 2,829 in lymphoma, 2,807 TADs in pro-B cell, and 2,808 in ES cell (Figure 3.8). Among them, the majority of TADs identified in a cell was conserved in another cell; by applying an approach previously described (Wu et al. 2017), I identified >70% overlapped TADs in a pair of samples, resulting in 2,348 (83.6% of total TADs) in lymphoma and ES cell, 2,319 (82.6%) in pro-B cell and ES cell, 2,235 (79%) in lymphoma and pro-B cell (Figure 3.9 ). In addition, TADs were largely identified from chromosome 7 in all samples, whereas lymphoma forms notably larger number of TADs in chromosome 14 and smaller number of TADs in chromosome X (Figure 3.10). These numbers support the study that the prostate cancer cell exhibits an increasing number of TADs compared to in its normal state (Taberlay et al. 2016). I also identified 301 unique TADs in pro-B cell, 191 in ES cell, and 198 in lymphoma, which is proportional to those found in a previous study; 65 unique out of 787 TADs identified in pro-B cell (Boya et al. 2017).

Collectively, my results show that the majority of TADs among pro-B cell,

lymphoma and ES cell are highly conserved, whereas specific genomic regions are involved in the structural reorganization. In particular, B-cell lymphoma organizes the genome structure to ES-cell like TADs that are different from pro-B cell.

### 3.2.4 TAD boundaries suggest gene regulation function in cancer

Recent studies have revealed that TADs associate with CTCF and the protein complex cohesin (Zuin et al. 2014) by forming relative conserved structures across cell types (Schwarzer et al. 2017) to bring enhancers and specific genes closer (Rao et al. 2014). Also, it has been observed that the disruption of TAD boundaries promotes gene expression leading to a physical malformation in mouse (Lupiáñez et al. 2015), suggesting the importance of CTCF in TAD boundaries. On the other hand, only 15% of TAD boundaries in mammals present CTCFs and 85% reside inside TADs (Ruiz-Velasco and Zaugg 2017). This scattered disposition points out that whereas CTCF can afford flexible adjustment to the chromatin conformation, the 3D chromatin organization is more likely to be influenced by a fine orchestration of cell-specific regulatory program. I then asked whether TAD boundaries of normal and cancer cells would exhibit CTCF enrichment, and whether genes located at boundaries would exhibit any variation of expression levels when compared to those located at intra-TAD regions.

Although I observed CTCF ( $p < 0.01$ ) motif enriched at the boundaries in ES cell, I also found PRDM1 ( $p = 0.01$ ), HRE ( $p < 0.01$ ), and Meis Homeobox 1 ( $p = 0.001$ ) enrichments that play important roles in normal development (Xia et al. 2017). The boundaries in pro-B cell were enriched with Nanog and PRDM1 ( $p = 0.01$ ) motifs. In contrast, I could not profile the CTCF motif enrichment neither in lymphoma cell nor pro-B cell. Interestingly,

the coding gene of PRDM1, that is associated with various cancer developments (Kang et al. 2018; Xia et al. 2017; Zhang et al. 2017; Zhu et al. 2017), exhibits the high expression level only in lymphoma (Figure 3.11), even though PRDM1 was located in active compartment and enriched in TAD boundaries in all samples (Table 3.11).

I next categorized genes based on their proximity to TAD boundaries. Remarkably, genes located around TAD boundaries (<40Kb) show significantly higher expression levels ( $p < 0.0001$ ) in all samples (Figure 3.12). This has been also observed in a high-resolution experiment in fruit flies (Ramírez et al. 2018). Overall, although TADs are highly conserved between cell types and often delimited by CTCF motifs, my results show the relationship of TAD boundaries with cancer-related transcription factors rather than with the CTCF motifs.

### 3.2.5 Inter-chromosomal interactions in B-cell lymphoma

Hi-C identifies two loci that are physically close inside the nucleus. These interactions can be classified in intra- and inter-chromosomal regarding to the location of each pair. A hallmark in tumorigenesis is the characteristic of distinct genetics changes that can be caused by different reasons. In B-cell lymphomas for example, a hallmark is the diverse chromosomal translocations, mostly involving transcriptionally deregulation in oncogenes (Küppers 2005). In Hi-C experiments is typically expected a higher intensity of intra-chromosomal interactions rather than inter-chromosomal. Here I asked whether the chromatin interactions maps in mouse cells have same patterns of interaction pairs. By assessing pairs of regions that have more reads between them than expected by chance, significant interactions ( $p\text{-value} < 0.001$ ) were selected and classified in inter- and inter-interaction. As described in the literature, normal cells present a inexpressive amount of significant inter-chromosomal interactions (Sarnataro et al. 2017). However, in B-cell lymphoma, 34% of the significant interactions were between two different chromosomes (Figure 3.13). Surprisingly, around 70 thousand interactions (23% of all significant interactions) were between chromosome 7 and chromosome 17 (Figure 3.14). Previous research observed this abnormality in those two chromosomes and attributed this karyotype abnormality as a characteristic of a poor survival lymphoma. translocation between other chromosomes in distinct types of lymphoma (Cabanillas et al. 1989). Another research has attributed this abnormality to translocations involving the short arm of chromosome 17 and loss of a copy of the TP53 tumor suppressor (Callet-Bauchu et al. 1999). These results suggest that Hi-C chromatin interaction

maps (Supplementary Figure A.3) can provide means to identify chromosomal translocations in cancer cell samples and may contribute to poorly diagnostics in this type of lymphoma.

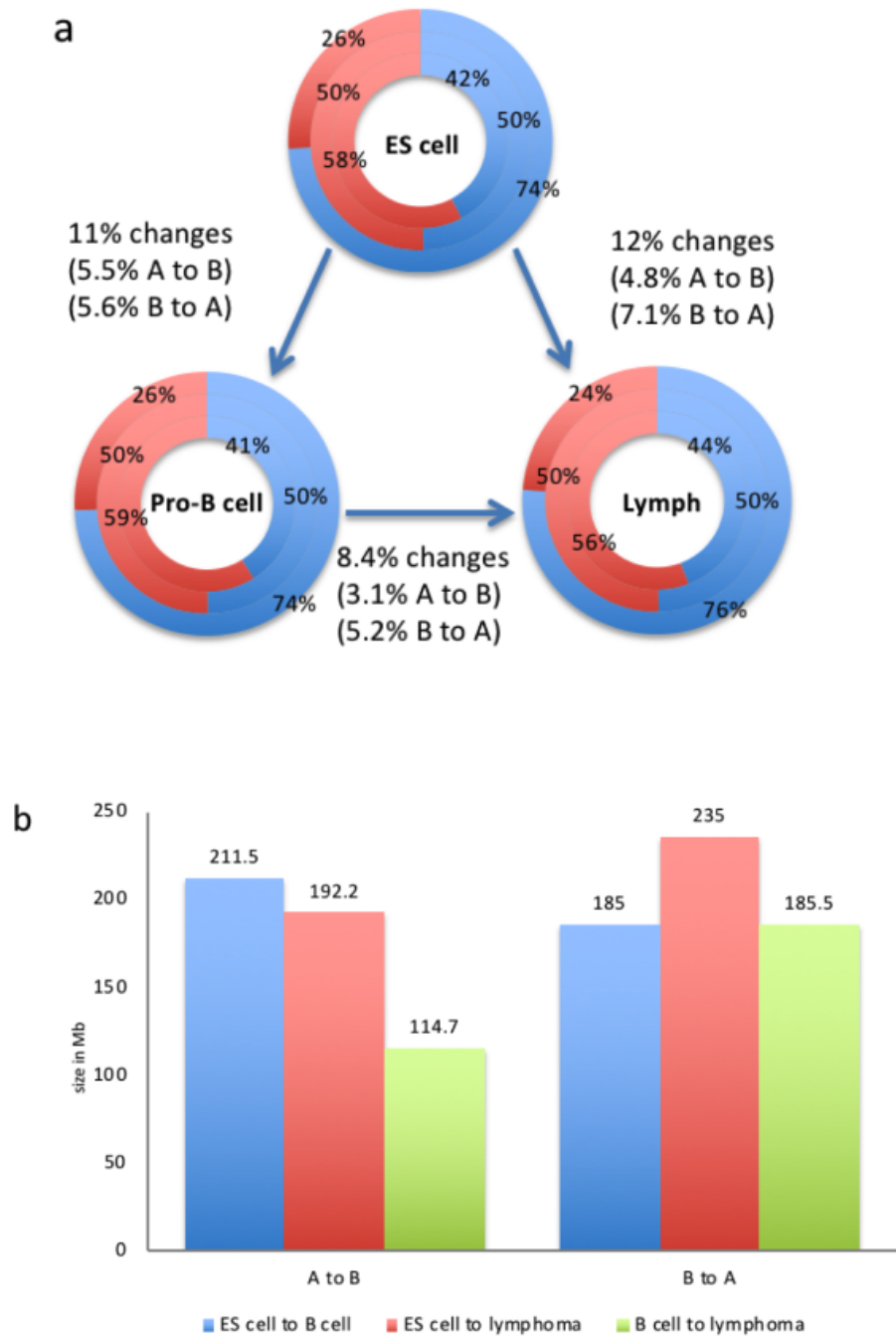


FIGURE 3.2: Chromatin organization at compartment level in ES cell, pro-B cell, and Lymphoma. (a) General information about genome size divided in A (blue) and B compartment (red) per cell type. Outer layer represents genome sequence in A or B compartments; middle layer represents the percentage of A and B compartment units identified in the whole genome; inner layer represents the percentage of genes contained in each compartment. (b) The total size in Mb of switching compartment regions.

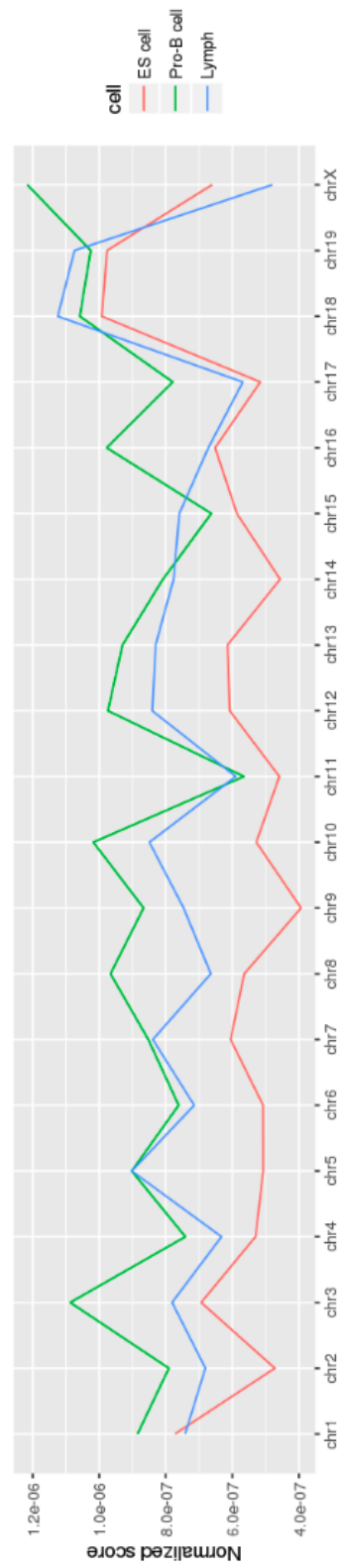


FIGURE 3.3: Compartment normalized score by chromosome size; Red: ES cell, green: B cell, and blue: B-cell lymphoma.



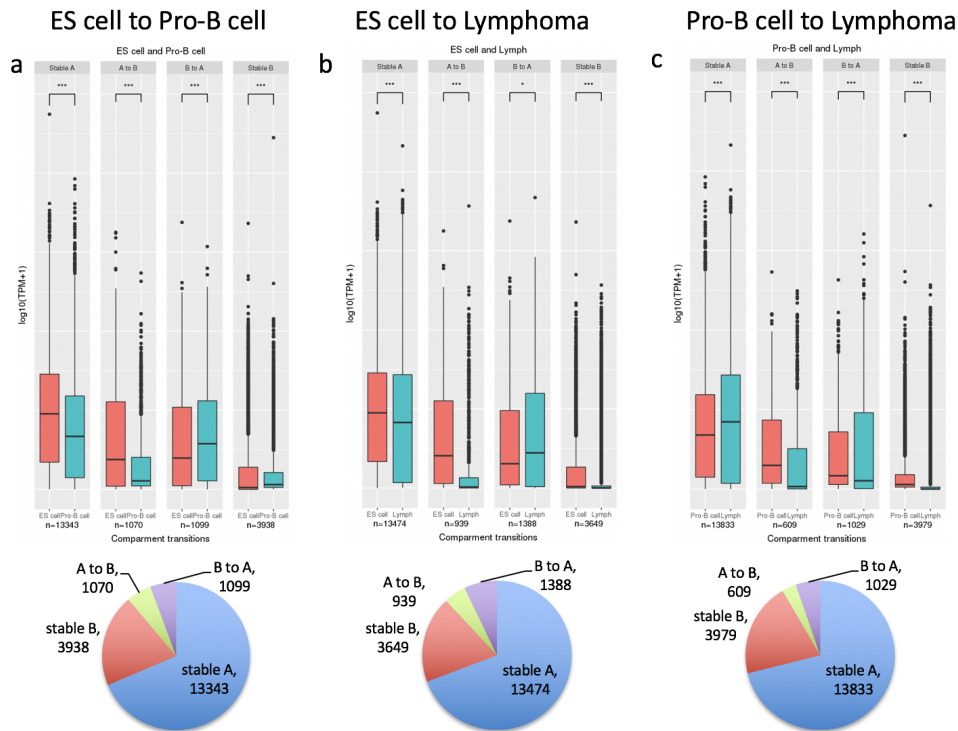


FIGURE 3.4: Profile of gene expression in compartment reorganization comparing ES cell, pro-B cell and lymphoma. The upper panel shows the tendency of gene having higher RNA abundance when located at active A compartment rather than inactive B compartment. Although B compartment is extensively reported as inactive compartment, it still contains some genes with higher levels of RNA molecules. (a) ES cell and pro-B cell comparison. (b) ES cell and lymphoma comparison. (c) pro-B and lymphoma comparison. The bottom panel shows the number of genes in each group.

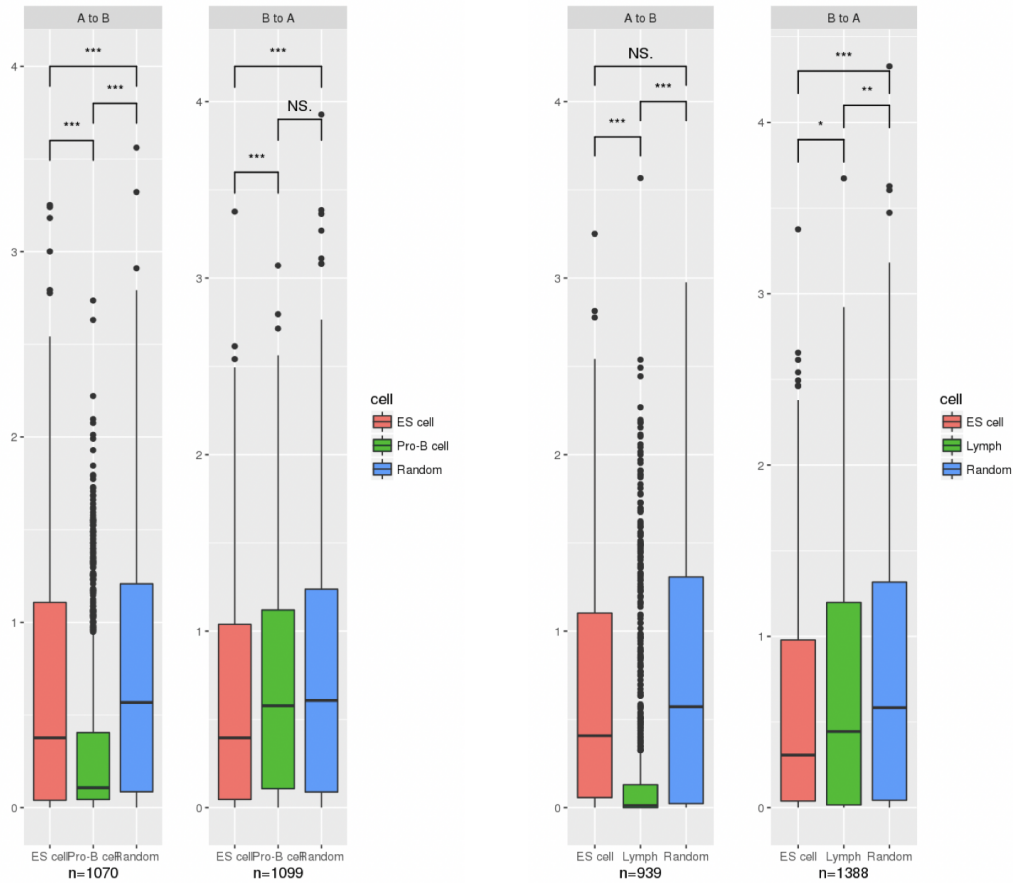


FIGURE 3.5: Profile of gene expression in compartment reorganization in comparison with random genes from stable profile. Left: ES cell, B cell, and random genes from stable regions. Right: ES cell lymphoma, and random genes from stable regions.

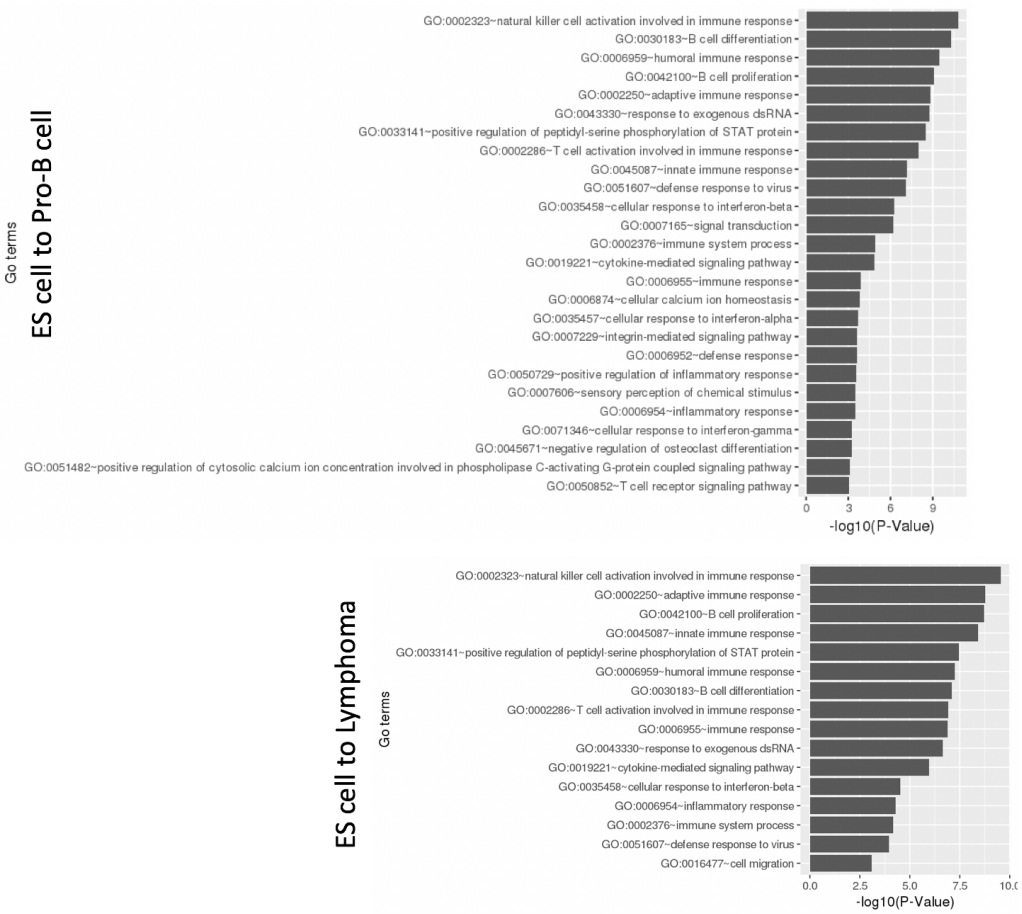


FIGURE 3.6: Enriched GO biological process terms in set of genes from compartment reorganization. Upper: Results from set of all genes in switching compartment region from B (ES cell) to A (pro-B cell). Bottom: Results from set of all genes in switching compartment region from B (ES cell) to A (B-cell lymphoma)

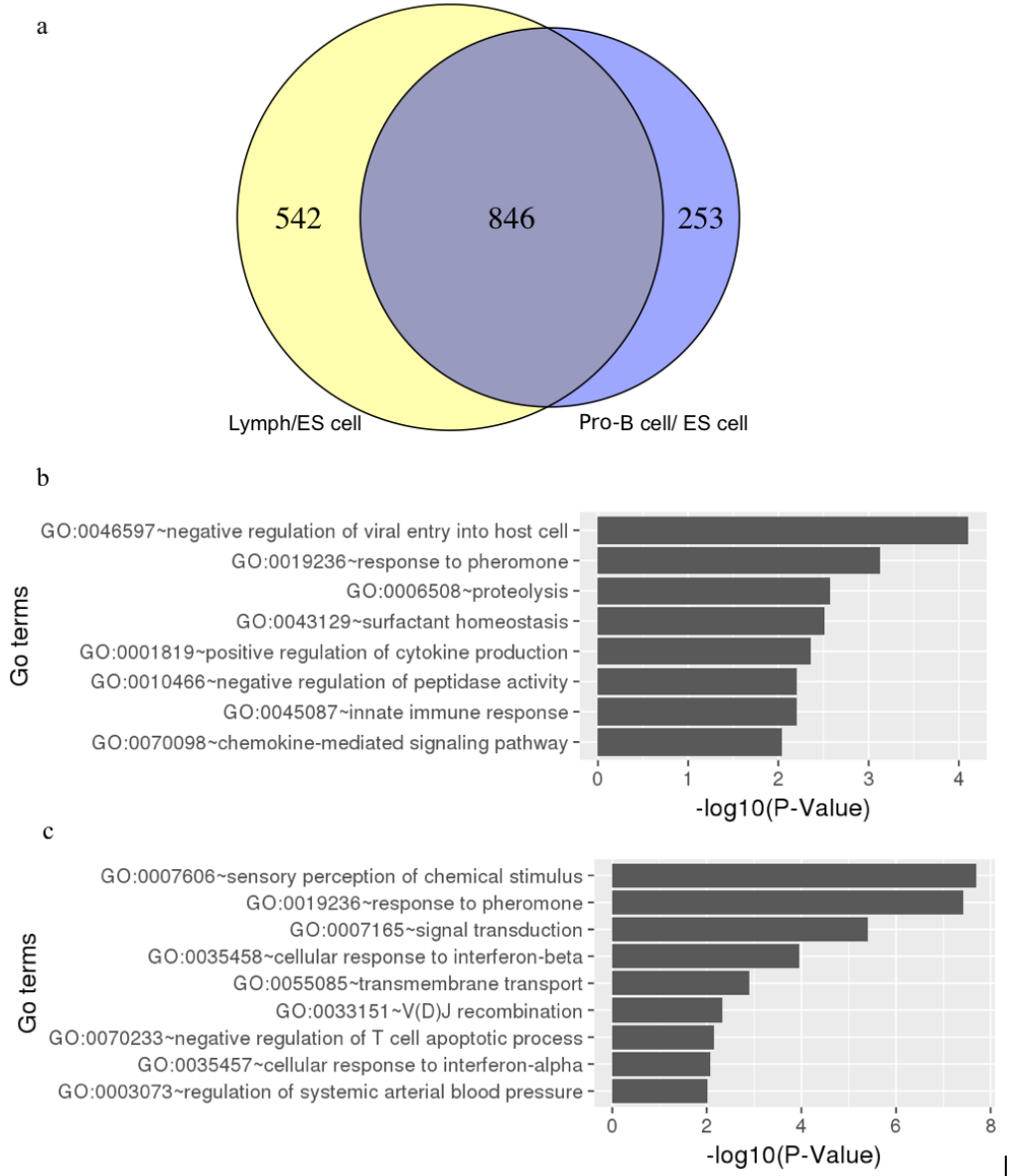


FIGURE 3.7: GO terms enriched in switching regions. (a)Venn diagram of set of all genes in switching compartment region from B to A. Yellow represents the lymphoma/ESC and blue represents pro-B cell/ESC. (b) Results from unique genes in switching compartment region from B (ES cell) to A (pro-B cell). (c) Results from unique set genes in switching compartment region from B (ES cell) to A (B-cell lymphoma).

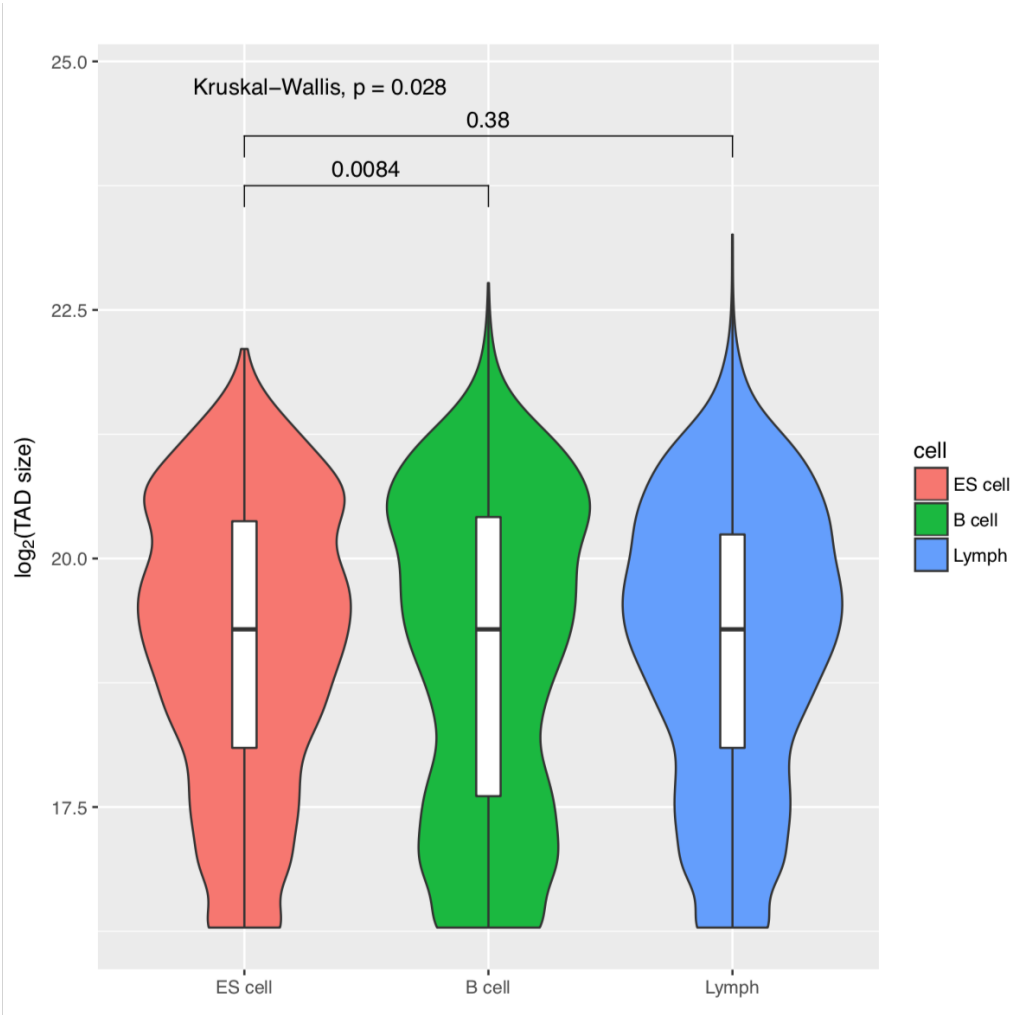


FIGURE 3.8: Violin plot of topologically associating domain size.

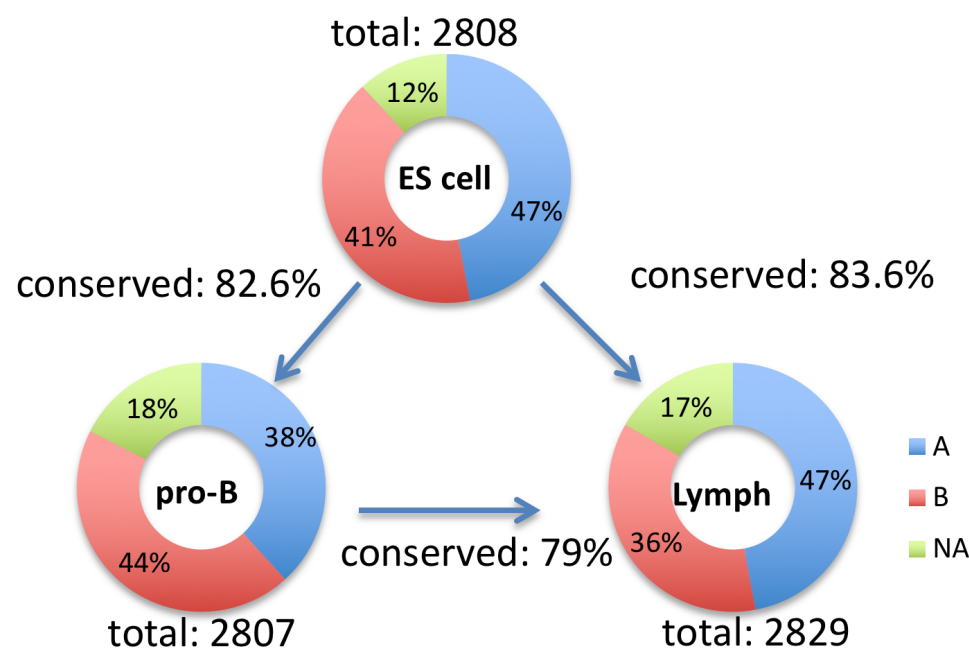


FIGURE 3.9: Chromatin organization at TAD level in ES cell, pro-B cell, and Lymphoma. TADs were classified per compartment and conservation degree is show between two cell types.

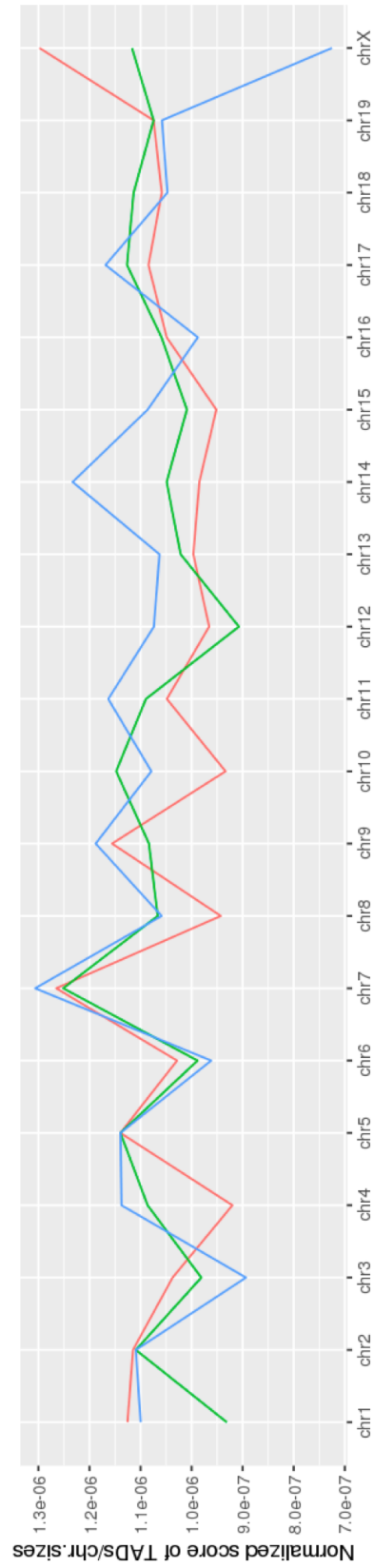


FIGURE 3.10: Normalized TAD number per chromosome size performed in ES cell (red line), B cell (green), and B cell lymphoma (blue).

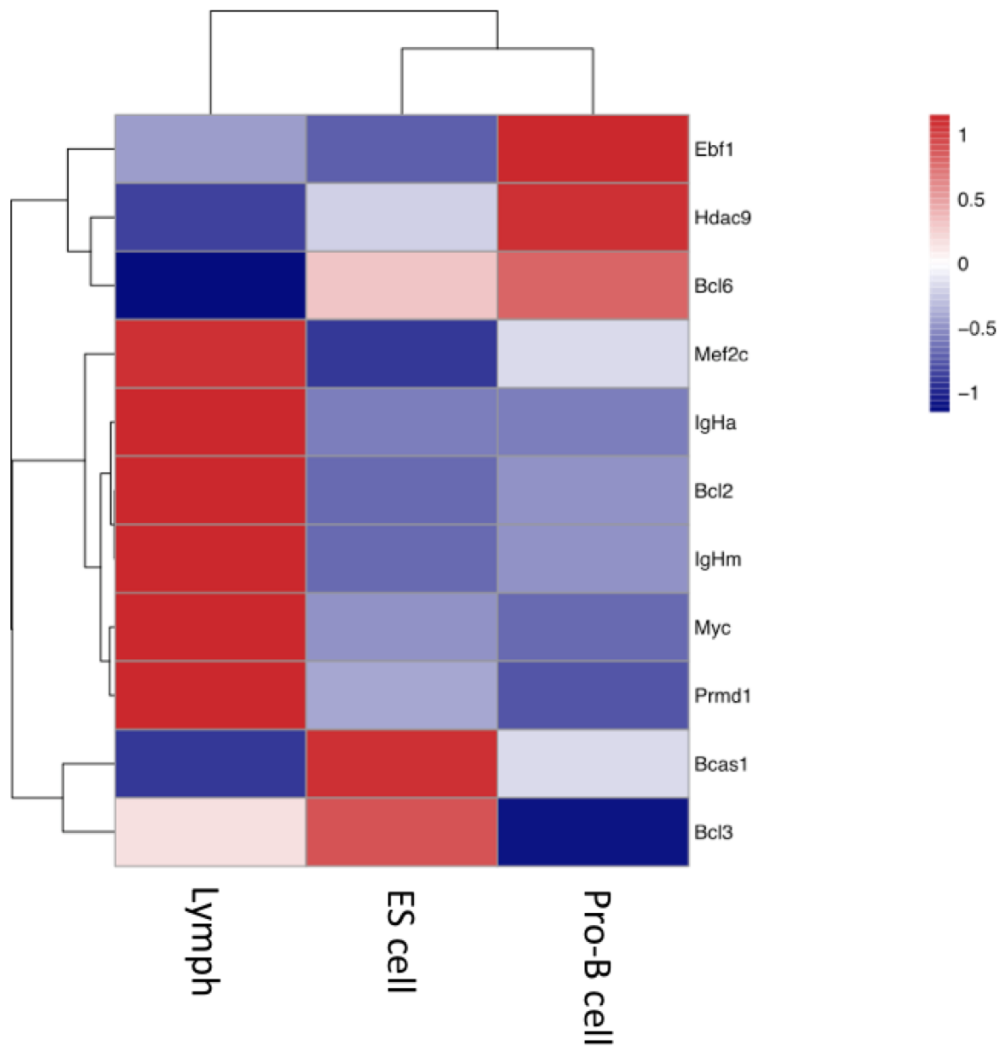


FIGURE 3.11: Heatmap of set of some of genes known to be involved in B-cell fate and B-cell lymphoma. Values are in  $\log_{10} TPM$ .



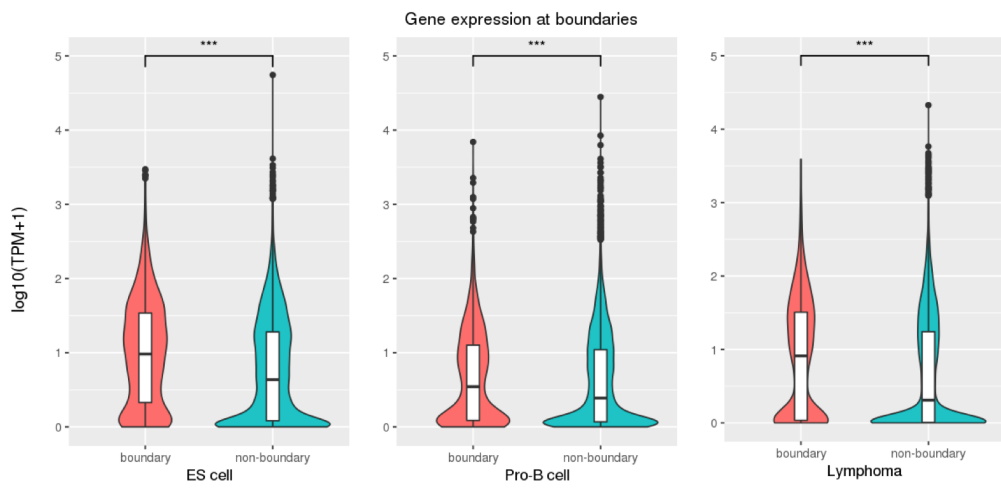


FIGURE 3.12: Comparison of gene expression values based on the chromatin structure disposition of TADs.Red violin plot represents the expression levels of genes located at TAD boundaries and green one represents the expression of genes located in the middle of TADs.

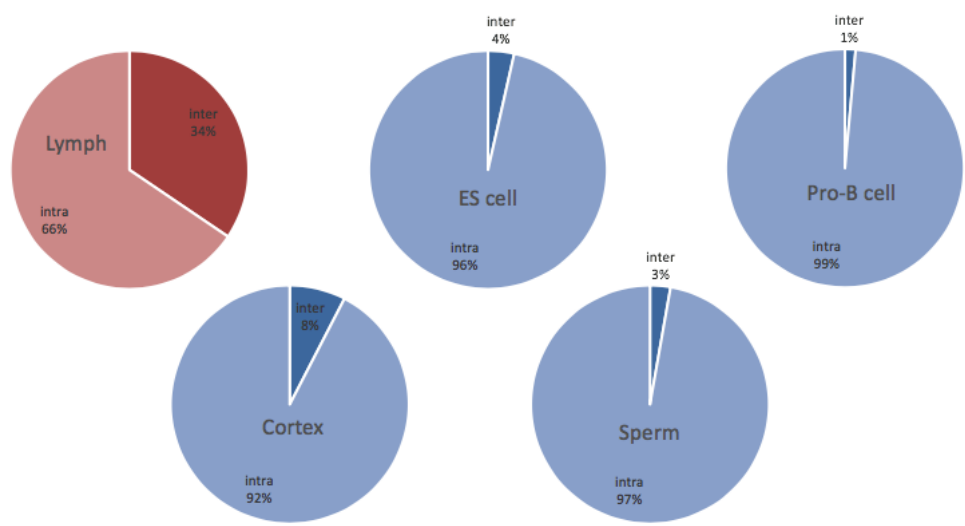


FIGURE 3.13: Significant intra- and inter-chromosomal interaction ratios in mouse cells (P-value < 0.001).

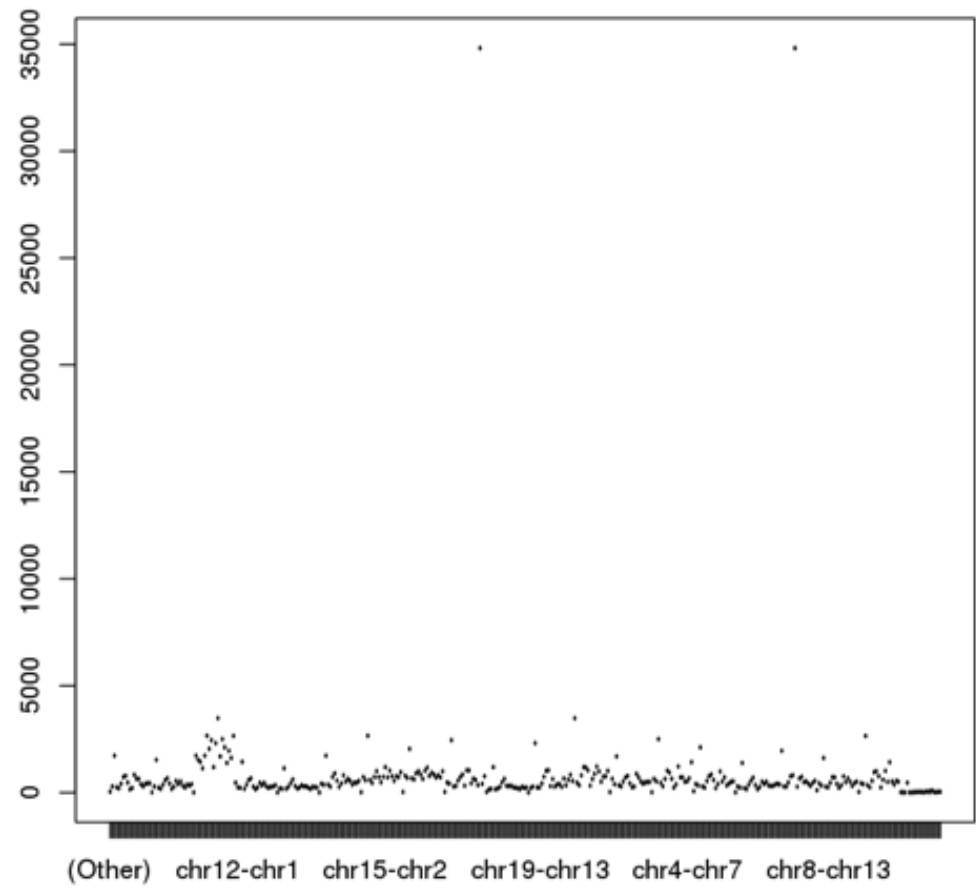


FIGURE 3.14: Distribution of inter-chromosomal interaction pairs in B-cell lymphoma. The two dots with around 35000 interactions are interactions between chromosome 17 and chromosome 7 and vice-versa.

## Chapter 4

# Discussion

Recent studies have revealed that the eukaryotic genome is divided into chromatin compartments and boundary-limited functional units known as TADs. Moreover, this genomic architecture inside the nucleus exhibit conserved patterns across species and cell types. Here, I used public available libraries of Hi-C paired-end sequences from different resources to compare general chromatin structures.

Compartments were identified using the same method as described by the Hi-C developer (Lieberman-Aiden et al. 2009). This compartment identification was previously reported as non-sensitive to higher resolution analysis (Fortin and Hansen 2015), however my results included few small compartments supporting new studies that also found kilo-base-sized compartments (Rowley et al. 2017; Wang et al. 2018). Despite that, my proposed method to quantify chromatin compartments and TADs per chromosome has shown conserved number of compartment and TAD per chromosome. Which is supported by conventional method previously published.

My analysis revealed the genomic regions that switch compartment status in the context of cells; 8% between B-cell and lymphoma, 11% between pro-B cell and ES cell, and 12% between lymphoma and ES cell. This approach classified genes in each compartment state. Despite the large number of genes remaining in the same compartment, cells undergo from ES cell to B cell. Indeed, only 1091 genes were located in compartment that switches from inactivated/silenced status to active transcribed status and the functional annotation analysis significantly defined the set of genes as B-cell functions. B-cell lymphoma transition genes (1376) were also enriched for the same annotation terms. Not all genes identified in the chromatin reorganization status follow this tendency. This suggests that the exact mechanism by which it occurs still remains to be understood. I propose that using data from the same experiment concomitantly would increase the power of gene identification in those transitional sites and reduce noisy genes affected by other undiscovered factors.

I observed that TADs are conserved across cells, which is consistent with the fact that TADs exhibit some reproducibility even using different tools (Forcato et al. 2017). Interestingly, the TAD formation in lymphoma is more similar to that of ES cell than pro-B cell. The similarity between ES cells and induced pluripotent stem cells structures were also detected (Krijger et al. 2016). This suggests that chromatin conformation in cancer cells might induce some specific gene expression control more similar to ES cell than pro-B cell, leading to abnormal gene expression regulations. Furthermore, I presented evidences that the DNA sequences at TAD boundaries are not always related to CTCFs as assumed. In order to confirm these observations, a higher resolution Hi-C or other 3C analyses

are mandatory.

I report in this research that B-cell lymphoma provides statistically inter-chromosomal interactions that was not observed in normal mouse cells. The fact that mostly of those interactions are between chromosome 17 and 7 strongly suggest a cancer-type specific interactions described as a abnormal karyotype associated with poor-survival-ratio lymphoma. This result suggests that Hi-C is an interesting approach to investigate chromosomal abnormalities. Furthermore, in addition to the previous result described with enrichment of PRMD1 motif, this study encourages future efforts to investigate the association of cancer-related motifs with higher-order chromatin structures provided by Hi-C sequencing.

## Chapter 5

# Conclusion

I have integrated publicly available multi-omics data to dissect the 3D structure reorganization loci and observed a subtle coordination with gene expression level changes. Remarkably, the set of genes in dynamic changes are strongly involved in cell-specific biological processes. In particular, the switching compartment regions are coordinated with gene expression increase. Although chromatin reorganization appeared in less impact on the gene regulation, I have discovered an unknown mechanism that impacts the structure and function of chromosomes and cognates with genes in a specific manner. The results presented in this work show that the majority of TADs among pro-B cell, lymphoma and ES cell are highly conserved, whereas specific genomic regions are involved in the compartment change. Overall, my findings suggest the presence of intricate cross-talk between the higher-order chromatin structure and cancer development. I concluded that an unknown mechanism profoundly exists to restrict the structural and functional changes of genomic regions and cognate genes in a specific manner. More efforts using high-resolution data are encouraged for further insights.

## Chapter 6

### Future Directions

This brief research analysis could achieve some knowledge in the emergent area of chromatin dynamics and gene regulation. More studies on this matter are encouraged, specifically on the following contexts:

(i) Role of chromatin organization and its impact on downstream molecular processes; not only in development and differentiation, but also in perturbed pathogenic process; (ii) How to validate findings from multi-omics analysis. Maybe a solution could be perturbation using CRISPR approach; (iii) Since chromatin architecture have shown highly conserved regions, it would be possible to train a model using high-resolution public data to predict chromatin reorganization in other cells at lower resolution.

## Acknowledgements

This dissertation would not be possible to accomplish without many people who supported me and encouraged me. I would like to reflect on the people who have supported and helped me so much throughout this period.

In the first place, I would like to thank Dr. Kenta Nakai for being a great advisor. I have had the privilege to spend three years and half in his laboratory, during which his door has always been open to me. Actually, even before coming to Japan, he had kindly replied my e-mails and touched me with his noble character. His passion for science, kindness, and great attention to details lead to a meaningful Ph.D. training for me.

I would also like to express my sincere gratitude to Dr. Sung-Joon Park, a colleague that I consider my co-advisor. I have had the privilege to work with Park-san for all the period I stayed in the lab but more intensely in the final period of my Ph.D.. His dedication to science, open-mindedness to new ideas, and his ability to keep learning new things greatly impress and influence me.

Special thanks are given to my colleagues and friends in Nakai lab, such as Dr. Ashwini Patil, Dr. Yokomori, Dr. Wei, Raul, Rin, Yang, Zeng,



Ishikawa, Munmee, Fujita, Amir, and those that left before me but contributed to my development/adaptation such as Huiheng Lin, Amartya, Hayashi, Yamada, Ms. Ikeda, Dr. Lee and Dr. Moon. I have special thanks to Ms. Saito, who patiently helped me many times with thousands of documents in Japanese.

I am also indebted to my committee members: Dr. Yoichi Furukawa, Dr. Koichi Matsuda, Dr. Katsuhiko Shirahige and Dr. Martin Frith. Their smart suggestion and insightful comments helped shape my Ph.D. research and this dissertation.

It has been a great privilege to be a student at the Department of Computational Biology and Medical Sciences of the Graduate School of Frontier Science of the University of Tokyo. And the Ministry of Education, Culture, Sports, Science and Technology (MEXT), for awarding me the scholarship that allowed this project to be carried out.

I also would like to thank my dearest friends Tang, Augusto, Marjorie, Rafael, Mabel, Gabriel, Angela, Vincent, Carol, and all members of Tokyo International Group.

At last but not least, I want to express my deepest gratitude to my wonderful family, in special to my parents, Quioji and Iraci, my sisters Tiemy and Akemy, and my beloved girlfriend Thaís Akiko Shirai. Their constant support and love have always sustained me and supported to keep struggling during this Ph.D..

For all that I did not mention here but always will remember the wonderful moments shared together my deepest thank you!

## Chapter 7

# Bibliography

Apostolou, Effie et al. 2013. “Genome-Wide Chromatin Interactions of the Nanog Locus in Pluripotency, Differentiation, and Reprogramming.” *Cell Stem Cell* 12(6): 669–712.

Bannister, Andrew J., and Tony Kouzarides. 2011. “Regulation of Chromatin by Histone Modifications.” *Cell Research* 21(3): 381–95.

Barutcu, A. Rasim et al. 2015. “Chromatin Interaction Analysis Reveals Changes in Small Chromosome and Telomere Clustering between Epithelial and Breast Cancer Cells.” *Genome Biology* 16(1): 1–14.

Berger, Shelley L. 2007. “The Complex Language of Chromatin Regulation during Transcription.” *Nature* 447(7143): 407–12.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. “Genome Analysis Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” 30(15): 2114–20.

Bonev, Boyan et al. 2017. “Multiscale 3D Genome Rewiring during Mouse Neural Development.” *Cell* 171(3): 557–572.e24.

- Boya, Ravi et al. 2017. “Developmentally Regulated Higher-Order Chromatin Interactions Orchestrate B Cell Fate Commitment.” *Nucleic Acids Research*: 1–18.
- Boyle, Shelagh et al. 2011. “Fluorescence in Situ Hybridization with High-Complexity Repeat-Free Oligonucleotide Probes Generated by Massively Parallel Synthesis.” *Chromosome Research* 19(7): 901–9.
- Brown, Christopher R. et al. 2008. “Global Histone Acetylation Induces Functional Genomic Reorganization at Mammalian Nuclear Pore Complexes.” *Genes and Development* 22(5): 627–39.
- Bunting, Karen L. et al. 2016. “Multi-Tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region.” *Immunity* 45(3): 497–512.
- Cabanillas, Fernando et al. 1989. “Refractoriness to Chemotherapy and Poor Survival Related to Abnormalities of Chromosomes 17 and 7 in Lymphoma.” *The American Journal of Medicine* 87(2): 167–72.
- Callet-Bauchu, E. et al. 1999. “Translocations Involving the Short Arm of Chromosome 17 in Chronic B-Lymphoid Disorders: Frequent Occurrence of Dicentric Rearrangements and Possible Association with Adverse Outcome.” *Leukemia* 13(3): 460–68.
- Capelson, Maya et al. 2010. “Chromatin-Bound Nuclear Pore Components Regulate Gene Expression in Higher Eukaryotes.” *Cell* 140(3): 372–83.
- Dekker, Job. 2014. “Two Ways to Fold the Genome during the Cell Cycle: Insights Obtained with Chromosome Conformation Capture.” *Epigenetics and Chromatin* 7(1): 1–12.

- Dekker, Job, and Edith Heard. 2015. “Structural and Functional Diversity of Topologically Associating Domains.” *FEBS Letters* 589(20): 2877–84.
- Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. “Chromosome Conformation.” *Advancement Of Science* 295(5558): 1306–11.
- Denker, Annette, and Wouter De Laat. 2016. “The Second Decade of 3C Technologies: Detailed Insights into Nuclear Organization.” *Genes and Development* 30(12): 1357–82.
- Dixon, J R et al. 2012. “Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions.” *Nature* 485(7398): 376–80.
- Dixon, Jesse R. et al. 2015. “Chromatin Architecture Reorganization during Stem Cell Differentiation.” *Nature* 518(7539): 331–36.
- Dunham, Ian et al. 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome.” *Nature* 489(7414): 57–74.
- Eagen, Kyle P. 2018. “Principles of Chromosome Architecture Revealed by Hi-C.” *Trends in Biochemical Sciences* xx(6): 1–10.
- Finlan, Lee E. et al. 2008. “Recruitment to the Nuclear Periphery Can Alter Expression of Genes in Human Cells.” *PLoS Genetics* 4(3).
- Forcato, Mattia et al. 2017. “Comparison of Computational Methods for Hi-C Data Analysis.” *Nature Methods* 14(7): 679–85.
- Fortin, Jean-Philippe, and Kasper D. Hansen. 2015. “Reconstructing A/B Compartments as Revealed by Hi-C Using Long-Range Correlations in Epigenetic Data.” *Genome Biology* 16(1): 180.
- Fraser, James, Iain Williamson, Wendy A. Bickmore, and Josée Dostie. 2015. “An Overview of Genome Organization and How We Got There: From FISH to Hi-C.” *Microbiology and Molecular Biology Reviews* 79(3): 347–72.

- Ghirlando, Rodolfo, and Gary Felsenfeld. 2016. "CTCF: Making the Right Connections." *Genes and Development* 30(8): 881–91.
- Gibcus, Johan H., and Job Dekker. 2013. "The Hierarchy of the 3D Genome." *Molecular Cell* 49(5): 773–82.
- Gonzalez-Sandoval, A, and S M Gasser. 2016. "On TADs and LADs: Spatial Control Over Gene Expression." *Trends Genet* 32(8): 485–95.
- Gorkin, David U., Danny Leung, and Bing Ren. 2014. "The 3D Genome in Transcriptional Regulation and Pluripotency." *Cell Stem Cell* 14(6): 771–75.
- Gruenbaum, Yosef et al. 2003. "The Nuclear Lamina and Its Functions in the Nucleus." *International Review of Cytology* 226: 1–62.
- Grunstein, Michael. 1997. "EBSCOhost: Histone Acetylation in Chromatin Structure and Transcription." *Nature* 389: 349–52.
- Heinz, S et al. 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities." *Mol Cell* 38(4): 576–89.
- Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. 2009. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." *Nature Protocols* 4(1): 44–57.
- Imakaev, Maxim et al. 2012. "Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization." *Nature Methods* 9(10): 999–1003.
- Javierre, B M et al. 2016. "Lineage-Specific Genome Architecture Links Enhancers and Non-Coding Disease Variants to Target Gene Promoters." *Cell* 167(5): 1369–1384 e19.
- Jin, Fulai et al. 2013. "A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human Cells." *Nature* 503(7475): 290–94.

- Jung, Yoon Hee et al. 2017. Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *CellReports* 18(6): 1366–82.
- Kang, Songhwa et al. 2018. “Adequate Concentration of B Cell Leukemia/Lymphoma 3 (Bcl3) Is Required for Pluripotency and Self-Renewal of Mouse Embryonic Stem Cells via Downregulation of Nanog Transcription.” 51(December 2017): 92–97.
- Kind, Jop et al. 2013. “Single-Cell Dynamics of Genome-Nuclear Lamina Interactions.” *Cell* 153(1): 178–92.
- Klein, Ulf et al. 2001. “Gene Expression Profiling of B Cell Chronic Lymphocytic Leukemia Reveals a Homogeneous Phenotype Related to Memory B Cells.” *J. Exp. Med* 121400(11): 1625–38.
- Koningsbruggen, Silvana van et al. 1992. “Characterization of the Membrane Binding and Fusion Events during Nuclear Envelope Assembly Using Purified Components.” *Journal of Cell Biology* 116(2): 295–306.
- Kosak, Steven T. et al. 2002. “Subnuclear Compartmentalization of Immunoglobulin Loci during Lymphocyte Development.” *Science* 296(5565): 158–62.
- Krijger, Peter Hugo Lodewijk et al. 2016. “Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming.” *Cell Stem Cell* 18(5): 597–610.
- Kumaran, R. Ileng, and David L. Spector. 2008. “A Genetic Locus Targeted to the Nuclear Periphery in Living Cells Maintains Its Transcriptional Competence.” *Journal of Cell Biology* 180(1): 51–65.
- Küppers, Ralf. 2005. “Mechanisms of B-Cell Lymphoma Pathogenesis.” *Nature Reviews Cancer* 5(4): 251–62.

- Lenz, Georg, and Louis M. Staudt. 2010. “Aggressive Lymphomas.” *New England Journal of Medicine* 362(15): 1417–29.
- Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows – Wheeler Transform.” 25(14): 1754–60.
- Lieberman-Aiden, E et al. 2009. “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.” *Science* 326(5950): 289–93.
- Lin, Yin C. et al. 2012. “Global Changes in the Nuclear Positioning of Genes and Intra-and Interdomain Genomic Interactions That Orchestrate B Cell Fate.” *Nature Immunology* 13(12): 1196–1204.
- Lupiáñez, Darío G. et al. 2015. “Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions.” *Cell* 161(5): 1012–25.
- Magistris, Paola De, and Wolfram Antonin. 2018. “The Dynamic Nature of the Golgi-Complex.” *Journal of Cell Biology* 108(2): 277–97.
- Nagano, Takashi et al. 2017. “Cell-Cycle Dynamics of Chromosomal Organization at Single-Cell Resolution.” *Nature* 547(7661): 61–67.
- Narendra, Varun et al. 2015. “CTCF Establishes Discrete Functional Chromatin Domains at the Hox Clusters during Differentiation.” *Science* 347(6225): 1017–21.
- Németh, Attila et al. 2010. “Initial Genomics of the Human Nucleolus.” *PLoS Genetics* 6(3).
- Nguyen, Linh, Peter Papenhausen, and Haipeng Shao. 2017. “The Role of C-MYC in B-Cell Lymphomas: Diagnostic and Molecular Aspects.” : 1–23.

- Nora, E P et al. 2012. “Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre.” *Nature* 485(7398): 381–85.
- Nora, Elphège P. et al. 2017. “Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization.” *Cell* 169(5): 930–944.e22.
- Patro, Rob et al. 2017. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression.” *Nature Publishing Group* 14(4): 417–19.
- Phillips-Cremins, Jennifer E. et al. 2013. “Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment.” *Cell* 153(6): 1281–95.
- Quinlan, Aaron R, and Ira M Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” 26(6): 841–42.
- Ramírez, Fidel et al. 2018. “High-Resolution TADs Reveal DNA Sequences Underlying Genome Organization in Flies.” *Nature Communications* 9(1): 189.
- Rao, S S et al. 2014. “A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.” *Cell* 159(7): 1665–80.
- Rodríguez-Carballo, Eddie et al. 2017. “The HoxD Cluster Is a Dynamic and Resilient TAD Boundary Controlling the Segregation of Antagonistic Regulatory Landscapes.” *Genes and Development* 31(22): 2264–81.
- Rowley, M. Jordan et al. 2017. “Evolutionarily Conserved Principles Predict 3D Chromatin Organization.” *Molecular Cell* 67(5): 837–852.e7.
- Rudan, Matteo Vietri et al. 2015. “Comparative Hi-C Reveals That CTCF Underlies Evolution of Chromosomal Domain Architecture” *CellReports* 10(8): 1297–1309.



- Ruiz-Velasco, Mariana, and Judith B. Zaugg. 2017. "Structure Meets Function: How Chromatin Organisation Conveys Functionality." *Current Opinion in Systems Biology* 1(i): 129–36.
- Sarnataro, Sergio et al. 2017. "Structure of the Human Chromosome Interaction Network." *PLoS ONE* 12(11): 1–15.
- Sauria, Michael E.G., Jennifer E. Phillips-Cremins, Victor G. Corces, and James Taylor. 2015. "HiFive: A Tool Suite for Easy and Efficient HiC and 5C Data Analysis." *Genome Biology* 16(1): 1–10.
- Schermelleh, Lothar, Rainer Heintzmann, and Heinrich Leonhardt. 2010. "A Guide to Super-Resolution Fluorescence Microscopy." *Journal of Cell Biology* 190(2): 165–75.
- Schmitt, Anthony D. et al. 2016. "A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome." *Cell Reports* 17(8): 2042–59.
- Schwarzer, Wibke et al. 2017. "Two Independent Modes of Chromatin Organization Revealed by Cohesin Removal." *Nature* 551(7678): 51–56.
- Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. 2014. "Transcriptional Enhancers: From Properties to Genome-Wide Predictions." *Nature reviews. Genetics* 15(4): 272–86.
- Soneson, Charlotte, Michael I Love, and Mark D Robinson. 2016. "Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences.
- Stevens, Tim J et al. 2017. "3D Structures of Individual Mammalian Genomes Studied by Single-Cell Hi-C." *Nature* 544(7648): 1–21.

- Taberlay, Phillippa C et al. 2016. “Three-Dimensional Disorganisation of the Cancer Genome Occurs Coincident with Long Range Genetic and Epigenetic Alterations.” *Genome Research*: gr.201517.115.
- Thomson, Inga, Susan Gilchrist, Wendy A. Bickmore, and Jonathan R. Chubb. 2004. “The Radial Positioning of Chromatin Is Not Inherited through Mitosis but Is Established de Novo in Early G1.” *Current Biology* 14(2): 166–72.
- Wang, Qi, Qiu Sun, Daniel M. Czajkowsky, and Zhifeng Shao. 2018. “Sub-Kb Hi-C in *D. Melanogaster* Reveals Conserved Characteristics of TADs between Insect and Mammalian Cells.” *Nature Communications* 9(1): 1–8.
- Weinreb, C, and B J Raphael. 2016. “Identification of Hierarchical Chromatin Domains.” *Bioinformatics* 32(11): 1601–9.
- Wu, Pengze et al. 2017. “3D Genome of Multiple Myeloma Reveals Spatial Genome Disorganization Associated with Copy Number Variations.” *Nature Communications* 8(1): 1937.
- Xia, Y. et al. 2017. “Loss of PRDM1/BLIMP-1 Function Contributes to Poor Prognosis of Activated B-Cell-like Diffuse Large B-Cell Lymphoma.” *Leukemia* 31(3): 625–36.
- Yue, F et al. 2014. “A Comparative Encyclopedia of DNA Elements in the Mouse Genome.” *Nature* 515(7527): 355–64.
- Zhang, Zhang et al. 2017. “Hypermethylation of PRDM1/Blimp-1 Promoter in Extranodal NK/T-Cell Lymphoma, Nasal Type: An Evidence of Predominant Role in Its Downregulation.” *Hematological Oncology* 35(4): 645–54.
- Zhu, Zhiyan et al. 2017. “Downregulation of PRDM1 Promotes Cellular Invasion and Lung Cancer Metastasis.” *Tumor Biology* 39(4).

---

Zink, Daniele, Andrew H. Fischer, and Jeffrey A. Nickerson. 2004. “Nuclear Structure in Cancer Cells.” *Nature Reviews Cancer* 4(9): 677–87.

Zuin, J. et al. 2014. “Cohesin and CTCF Differentially Affect Chromatin Architecture and Gene Expression in Human Cells.” *Proceedings of the National Academy of Sciences* 111(3): 996–1001.

## Appendix A

### Supplementary Figures

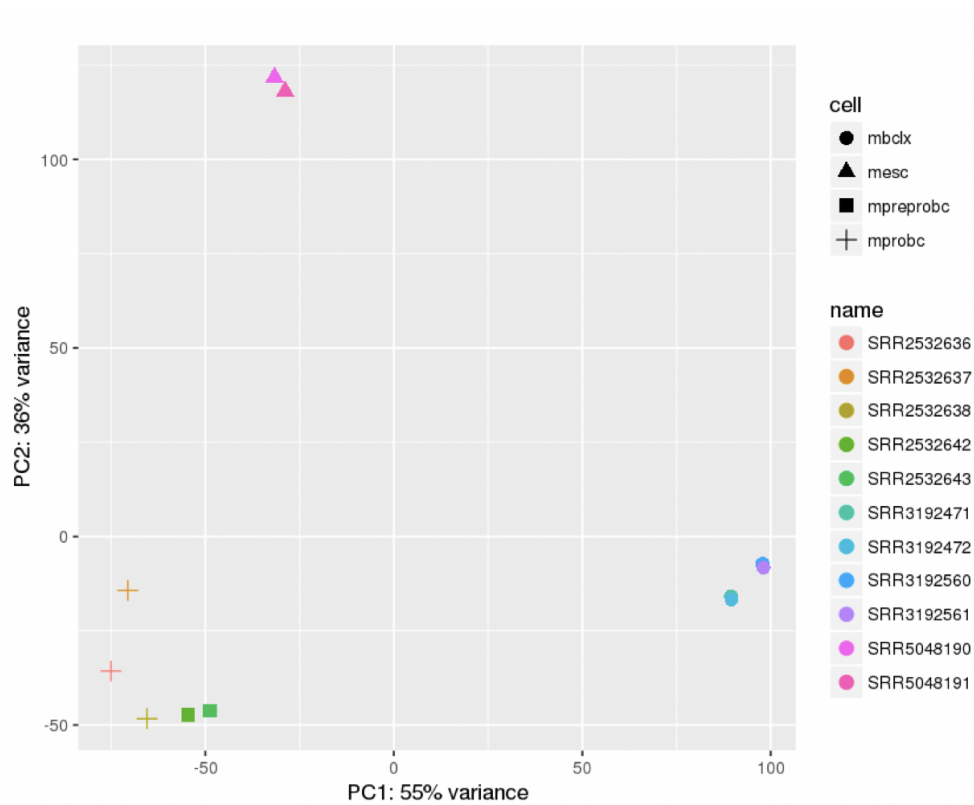


FIGURE A.1: Principal Component Analysis of gene expression values.

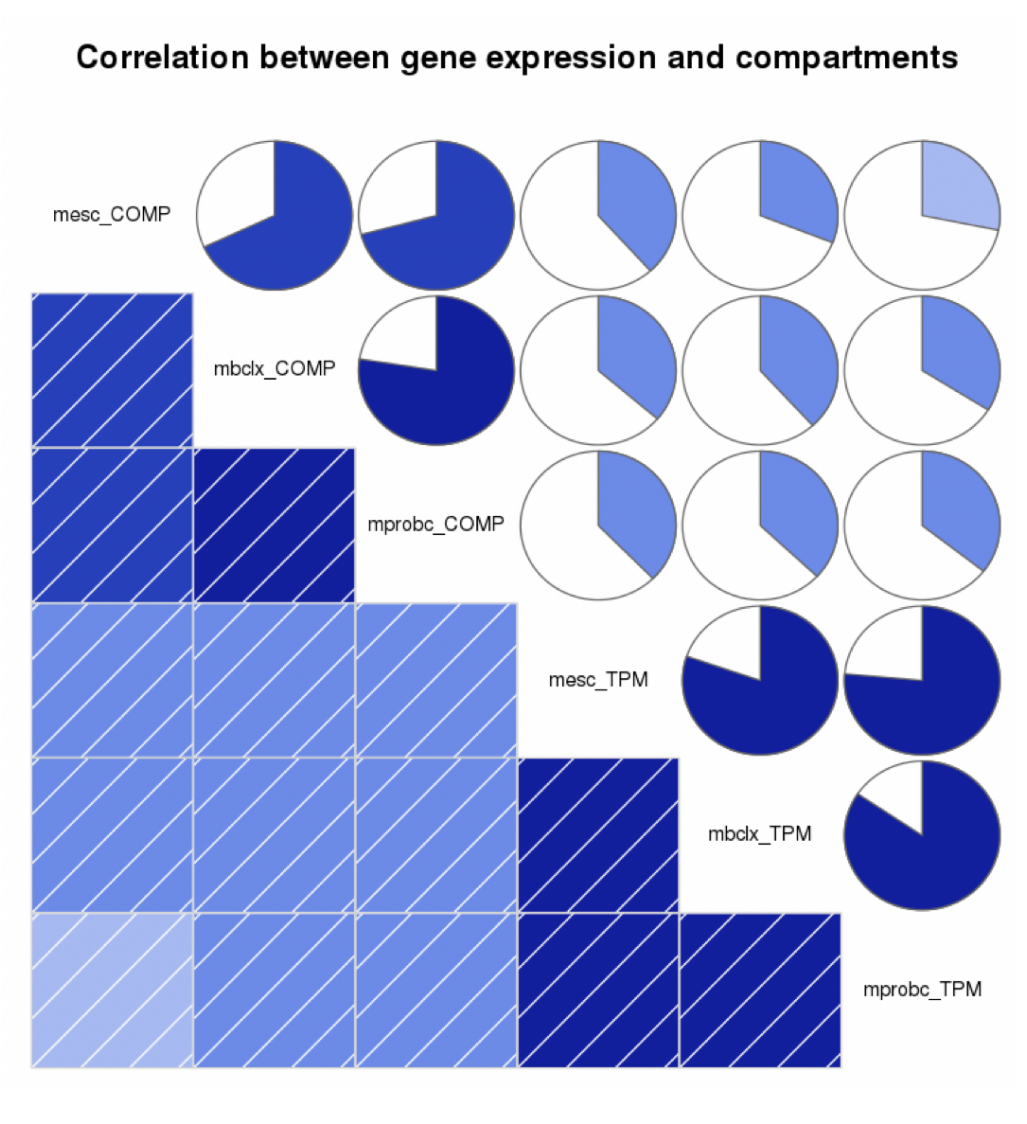


FIGURE A.2: Representation of the correlation between compartments and gene expression.

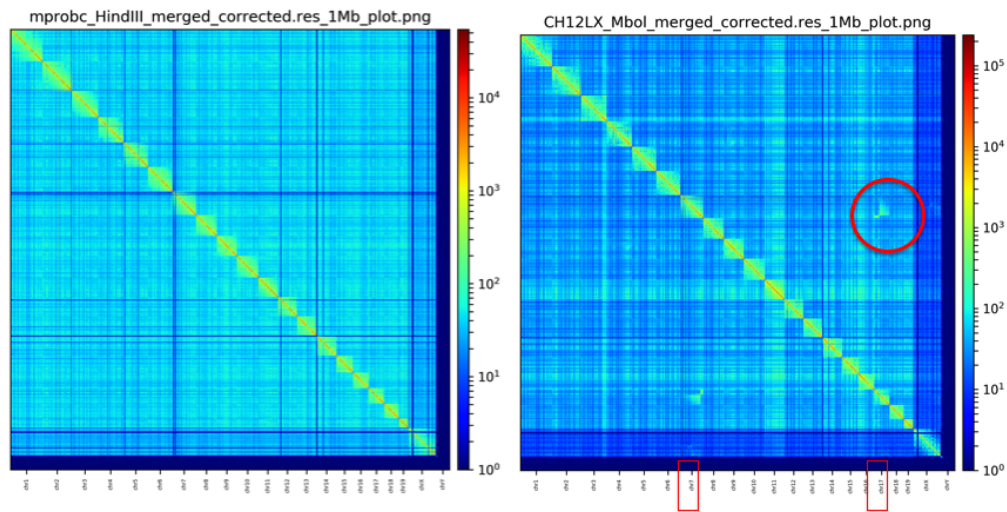


FIGURE A.3: Heatmaps of the chromatin interaction maps of the whole genomes of Pro-B cell and Lymphoma. Red circle indicates the high intensity of interactions between chromosome 7 and chromosome 17 in B-cell lymphoma sample

## Appendix B

# Supplementary Tables

TABLE B.1: RNA-seq detailed data information

id	rep	cell type	sex	biomaterial	description	SRA	GEO
mesc	1	129/Ola ES-E14	Male	Embryonic stem cell	Embryonic stell cell	SRR5048190	GSE90277
mesc	2	129/Ola ES-E14	Male	Embryonic stem cell	Embryonic stell cell	SRR5048191	GSE90277
mprobc	1	RAG1-/-pro-B cells	Male and Female	Pro-B cells	B6.RAG1-/-mice C57Bl/6	SRR2532636	GSM1897405
mprobc	2	RAG1-/-pro-B cells	Male and Female	Pro-B cells	B6.RAG1-/-mice C57Bl/6	SRR2532637	GSM1897406
mprobc	3	RAG1-/-pro-B cells	Male and Female	Pro-B cells	B6.RAG1-/-mice C57Bl/6	SRR2532638	GSM1897407
mbclx	1	CH12-LX	Female	B-cell lymphoma	B10.H-2aH-4bp/Wts	SRR3192471	GSM2072416
mbclx	2	CH12-LX	Female	B-cell lymphoma	B10.H-2aH-4bp/Wts	SRR3192472	GSM2072417

TABLE B.2: Hi-C detailed data information

<i>id</i>	<i>rep</i>	<i>celltype</i>	<i>sex</i>	<i>biomaterial</i>	<i>description</i>	<i>SRA</i>	<i>GEO</i>
<i>mesc</i>	1	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR443883	GSM862720
<i>mesc</i>	1	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR443884	GSM862720
<i>mesc</i>	1	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR443885	GSM862720
<i>mesc</i>	2	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR400251	GSM862721
<i>mesc</i>	2	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR400252	GSM862721
<i>mesc</i>	2	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR400253	GSM862721
<i>mesc</i>	2	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR400254	GSM862721
<i>mesc</i>	2	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR400255	GSM862721
<i>mesc</i>	2	<i>mESClinef1</i>	<i>Male</i>	<i>Embryonicstemcell</i>	129S4/ <i>SoJae</i> ,restrictionenzyme : <i>HindIII</i>	SRR400256	GSM862721
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543432	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543433	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543434	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543435	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543436	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543437	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543438	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543439	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543440	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543441	GSM987818
<i>mprobc</i>	1	<i>Rag1deficientpro - B</i>	<i>NA</i>	<i>Pro - Bcells</i>	C57bl/6,restrictionenzyme : <i>HindIII</i>	SRR543442	GSM987818
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658716	GSM1551633
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658717	GSM1551634
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658718	GSM1551635
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658719	GSM1551636
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658720	GSM1551637
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658721	GSM1551638
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658722	GSM1551639
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658723	GSM1551640
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658724	GSM1551641
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658725	GSM1551642
<i>mbclx</i>	1	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658726	GSM1551643
<i>mbclx</i>	2	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658727	GSM1551644
<i>mbclx</i>	2	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658728	GSM1551645
<i>mbclx</i>	3	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658729	GSM1551646
<i>mbclx</i>	3	<i>CH12 - LX</i>	<i>NA</i>	<i>B - lymphoblasts</i>	Restrictionenzyme : <i>Mbol</i>	SRR1658730	GSM1551647



TABLE B.3: Gene enrichment analysis in Pro-B cell (from B to A compartment). GO terms, biological function and statistical test (FDR).

GO Term	Biological Function	FDR
GO:0002323	natural killer cell activation involved in immune response	2.6439117561949299E-8
GO:0030183	B cell differentiation	9.65546087527968E-8
GO:0006959	humoral immune response	6.1155301933624099E-7
GO:0042100	B cell proliferation	1.4627151601054299E-6
GO:0002250	adaptive immune response	2.5762652122729401E-6
GO:0043330	response to exogenous dsRNA	2.9027335979314698E-6
GO:0033141	positive regulation of peptidyl-serine phosphorylation of STAT protein	5.36493637293133E-6
GO:0002286	T cell activation involved in immune response	1.87680725738381E-5
GO:0045087	innate immune response	1.25743038859038E-4
GO:0051607	defense response to virus	1.52066959802077E-4
GO:0035458	cellular response to interferon-beta	9.7323685754879797E-4
GO:0007165	signal transduction	1.2396494193600499E-3
GO:0002376	immune system process	2.1983337602116099E-2
GO:0019221	cytokine-mediated signaling pathway	2.6825483180570299E-2
GO:0006955	immune response	0.23414565597217299
GO:0006874	cellular calcium ion homeostasis	0.26854410860683797
GO:0035457	cellular response to interferon-alpha	0.350604492056105
GO:0007229	integrin-mediated signaling pathway	0.42504855343122799
GO:0006952	defense response	0.44431661280636098
GO:0050729	positive regulation of inflammatory response	0.52528751261473605
GO:0007606	sensory perception of chemical stimulus	0.549831480156592
GO:0006954	inflammatory response	0.61007409905654497
GO:0071346	cellular response to interferon-gamma	1.0361156726573
GO:0045671	negative regulation of osteoclast differentiation	1.1100687985731299
GO:0051482	positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G-protein coupled signaling pathway	1.35869148440723
GO:0050852	T cell receptor signaling pathway	1.6932073950477899
GO:0006032	chitin catabolic process	1.9287085088917499
GO:0032729	positive regulation of interferon-gamma production	1.95463311904577
GO:0070233	negative regulation of T cell apoptotic process	2.8962518522273402
GO:0045785	positive regulation of cell adhesion	3.3376969489994401
GO:0042832	defense response to protozoan	6.0997865731256002
GO:0014832	urinary bladder smooth muscle contraction	10.310743320221199
GO:0032609	interferon-gamma production	10.310743320221199
GO:0042104	positive regulation of activated T cell proliferation	10.7684571556289
GO:0006919	activation of cysteine-type endopeptidase activity involved in apoptotic process	11.0560328849678
GO:0042110	T cell activation	11.3287138837482
GO:0006935	chemotaxis	11.5800446541518
GO:0001953	negative regulation of cell-matrix adhesion	12.430683029992601
GO:0001816	cytokine production	12.774452214182899
GO:0045672	positive regulation of osteoclast differentiation	12.774452214182899
GO:0044406	adhesion of symbiont to host	14.5584734558403

TABLE B.4: Gene enrichment analysis in B-cell lymphoma (from B to A compartment). GO terms, biological function and statistical test (FDR).

GO Term	Biological Function	FDR
GO:0002323	natural killer cell activation involved in immune response	4.980000000000004E-7
GO:0002250	adaptive immune response	3.019999999999999E-6
GO:0042100	B cell proliferation	3.530000000000001E-6
GO:0045087	innate immune response	7.230000000000002E-6
GO:0033141	positive regulation of peptidyl-serine phosphorylation of STAT protein	6.509999999999997E-5
GO:0006959	humoral immune response	1.05E-4
GO:0030183	B cell differentiation	1.479999999999999E-4
GO:0002286	T cell activation involved in immune response	2.200000000000001E-4
GO:0006955	immune response	2.380000000000001E-4
GO:0043330	response to exogenous dsRNA	3.959999999999998E-4
GO:0019221	cytokine-mediated signaling pathway	2.003284000000001E-3
GO:0035458	cellular response to interferon-beta	5.367185399999998E-2
GO:0006954	inflammatory response	9.254604800000006E-2
GO:0002376	immune system process	0.120304668
GO:0051607	defense response to virus	0.1999990949999999
GO:0016477	cell migration	1.5030658779999999
GO:0006935	chemotaxis	1.8523384060000001
GO:0007155	cell adhesion	1.8660102000000001
GO:0007229	integrin-mediated signaling pathway	3.2470591679999998
GO:0045671	negative regulation of osteoclast differentiation	3.2806466790000002
GO:0070098	chemokine-mediated signaling pathway	3.4338991920000002
GO:0071305	cellular response to vitamin D	4.1818442109999996
GO:0006032	chitin catabolic process	4.1818442109999996
GO:0051482	positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G-protein coupled signaling pathway	4.512252105
GO:0007165	signal transduction	5.8414795679999996
GO:0006874	cellular calcium ion homeostasis	6.3642280729999996
GO:0016525	negative regulation of angiogenesis	8.4096874110000002
GO:0035457	cellular response to interferon-alpha	8.7155417009999994
GO:0033674	positive regulation of kinase activity	8.7155417009999994
GO:0050729	positive regulation of inflammatory response	9.2949332009999992
GO:0010466	negative regulation of peptidase activity	10.57224137
GO:0046597	negative regulation of viral entry into host cell	11.403195459999999
GO:0030195	negative regulation of blood coagulation	11.78609509
GO:0016337	single organismal cell-cell adhesion	12.33142106
GO:0045785	positive regulation of cell adhesion	12.66886966
GO:0006508	proteolysis	14.81437878
GO:0007160	cell-matrix adhesion	14.91569015