PROBING CHANGES IN SPATIAL ORGANIZATION OF GENOME DURING B CELL FATE COMMITMENT

Thesis submitted for the award of the degree of

DOCTOR OF PHILOSOPHY

By

B. Ravi

Regd. No. 09LAPH11

Department of Animal Biology

School of Life Sciences

University of Hyderabad

Hyderabad - 500 046

India

October 2017

University of Hyderabad
(A Central University established in 1974 by an act of parliament)
Hyderabad– 500 046, INDIA

CERTIFICATE

This is to certify that the thesis entitled **"Probing changes in spatial organization of genome during B cell fate commitment"** submitted by **Mr. B. Ravi** bearing registration number **09LAPH11** in partial fulfillment of the requirements for award of Doctor of philosophy in the School of Life Sciences is a bonafide work carried out by him under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other university or Institution for award of any degree or diploma.

The results of this thesis have been:

A. Published in the following publication:

1. Boya, R., Yadavalli, A. D., Nikhat, S., Kurukuti, S., Palakodeti, D., & Pongubala, J. M. (2017). Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment. *Nucleic Acids Research*. doi: 10.1093/nar/gkx722 (ISSN: 1362-4962)

B. Presented in the following conferences:

1. Genome Architecture and Cell Fate Regulation (GACFR) (University of Hyderabad, Hyderabad), 2014 **(International)**

2. All India Cell Biology Conference (IISER, Trivandrum), 2015 **(International)**

Further the student has passed the following courses towards fulfillment of coursework requirement for Ph.D.

Course Code	Name	Credits	Pass/Fail
1. AS 801	Analytical Techniques	4	Pass
2. AS 802	Research Ethics & Management	2	Pass
3. AS 803	Lab Work	4	Pass
4. AS 804	Biostatistics	2	Pass

Supervisor **Head of the Department** **Dean of School**

University of Hyderabad
(A Central University established in 1974 by an act of parliament)
Hyderabad – 500 046, INDIA

DECLARATION

I hereby declare that the results of the study incorporated in the thesis entitled **"Probing changes in spatial organization of genome during B cell fate commitment"** has been carried out by me under the supervision of **Prof. Jagan Pongubala** at Department of Animal Biology, School of Life Sciences. The work presented in this thesis is a bonafide research work and has not been submitted for any degree or diploma in any other University or Institute.

Date:

B. Ravi
(09LAPH11)

Acknowledgements

This thesis would not have been possible without the inspiration and support of a number of wonderful individuals, my thanks and appreciation to all of them for being part of this journey.

Foremost, I would like to thank my supervisor, **Prof. Jagan Pongubala**, for the guidance, support and encouragement he has provided throughout my time as his student. I greatly appreciate the freedom he has given me to find my own path and develop myself as a researcher in the best possible way.

I would like to thank my doctoral committee members, Prof. S. Dayananda and Dr. Sreenivasulu Kurukuti for their suggestions during my work.

I am thankful to Prof. P. Jagan Pongubala, Head, Department of Animal Biology and the former Heads, Prof. B. Senthilkumaran and Prof. Manjula Sritharan for allowing me to use departmental facilities.

I am thankful to Prof. P. Reddanna, Dean, School of Life Sciences and former Deans, Prof. R. P. Sharma, Prof. Aparna Dutta Gupta and Prof. M. Ramanadham for allowing me to use the school facilities.

I thank our collaborator, Dr. Dasaradhi Palakodeti for his support during my stay at inStem.

I would like to thank all my lab members, Sameena Nikhat, Anurupa Devi, Priyanka, Anubhooti and Arpita Prusty for their help and creating a good atmosphere in the laboratory. I gratefully acknowledge the immense support and co-operation of Sameena, throughout this journey particularly during the time of lab establishment. I thankfully acknowledge the contributions of Anurupa Devi, who has immensely helped me in bioinformatics analyses of my experiments.

I am thankful to my friends and all the research scholars in School of Life Sciences for their timely help and co-operation during my stay in the University.

I thank all non-teaching staff of Department of Animal Biology for their help in all the administrative works.

I sincerely thank CSIR for the fellowship during my doctoral studies.

I thank DBT, DST, ICMR, UGC, UPE-II, UGC-SAP and DBT-CREBB for providing financial support to lab and for the infrastructural facility of the School.

It is a pleasure to thank Suresh, Prasad, Madhavi and Keshav, who are like my second family, for always being there for me and also for the wonderful time we have shared. I find myself lucky to have friends like them in my life.

I would like to thank my pals, Papa Rao, Ashok, Vinod, Mastan, Naveen, Kumar Reddy, Mohan Krishna and Sravan for their encouragement and all the humorous conversations.

I am deeply indebted to my school gurus, **Sri G. Sreenivasa Rao** and **Smt. S. Narmada**, for their moral and financial support since my childhood. I would not have come this far without their constant love, blessings and support.

I would like to acknowledge all the teachers who have taught me since my childhood, I would not have been here without their guidance.

I thank the **Almighty God**, for always being there for me and making my life so beautiful with his blessings.

I would like to express a deep sense of gratitude to **Amma** and **Nanna**, whom I owe my life for their constant love, blessings, moral support and encouragement. I thank my brothers **Gopinath** and **Tilak**, who strengthened my morale by standing by me in all situations. Finally, special thanks to my sweet sister **Ragasudha,** who made my life colorful with her love and affection.

B. Ravi

Dedicated to

My Gurus Sri G. Sreenivasa Rao, Smt. S. Narmada,

My Parents and the Almighty God

Table of contents

Abbreviations

3C	Chromosome Conformation Capture
3D	Three-dimensional
^{0}C	Degrees Celsius
BAC	Bacterial artificial chromosome
bp	Base pair
BSA	Bovine serum albumin
CTCF	CCCTC-binding factor
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol
dCTP	Deoxycytidine triphosphate
dNTPs	Deoxynucleotide triphosphates
dATP	Deoxyadenosine triphosphate
dTTP	Deoxythymidine triphosphate
dGTP	Deoxyguanosine triphosphate
Ebf1	Early B-cell factor 1
EDTA	Ethylenediaminetetraacetic acid
ES cell	Embryonic stem cell
Flt3L	FMS-like tyrosine kinase 3 ligand
GFP	Green fluorescent protein
IL-7	Interleukin 7
kb	Kilo base pair
M	Molar

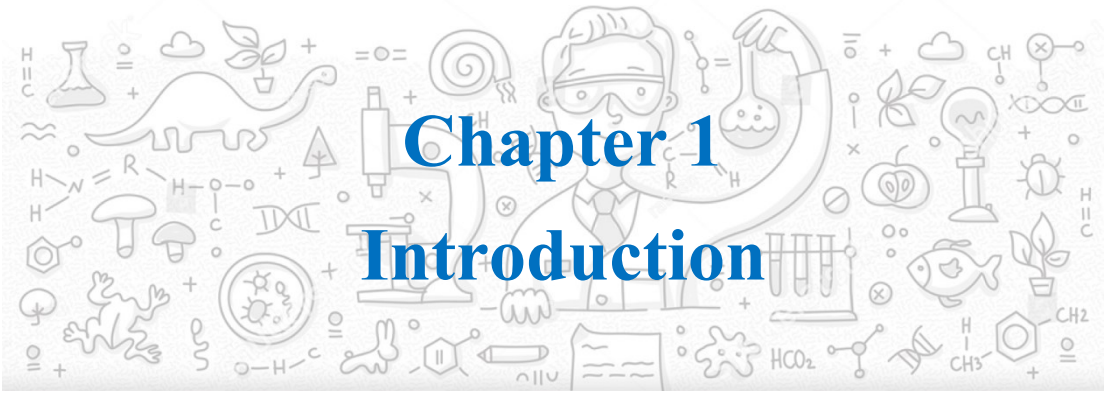
Mb	Megabase
mg	Milligram
$MgCl_2$	Magnesium chloride
min	Minutes
ml	Milliliter
MPPs	Multipotent progenitors
NaCl	Sodium chloride
NP-40	Nonidet P-40
Pax5	Paired box 5
PC1	Principal component 1
PBS	Phosphate-buffered saline
pH	Potential of hydrogen
rpm	Revolutions per minute
RNA PolII	RNA polymerase II
RNase A	Ribonuclease A
SCF	Stem cell factor
SDS	Sodium dodecyl sulfate
TADs	Topologically associating domains
TAE	Tris-acetate-EDTA
Tris-HCl	Tris (hydroxymethyl) aminomethane hydrochloride
µg	Microgram
µM	Micromolar
µl	Microliter

List of figures

List of tables

Abstract

Genome organization in 3D nuclear-space is important for regulation of gene expression. However, the alterations of chromatin architecture that impinge on the B cell-fate choice of multi potent progenitors are still unclear. By integrating *in situ* Hi-C analyses with epigenetic landscapes and genome-wide expression profiles, we tracked the changes in genome architecture as the cells transit from a progenitor to a committed state. We identified the genomic loci that undergo developmental switch between A and B compartments during B-cell fate determination. Furthermore, although, topologically associating domains (TADs) are stable, a significant number of TADs display structural alterations that are associated with changes in *cis*-regulatory interaction landscape. Finally, we demonstrate the potential roles for Ebf1 and its downstream factor, Pax5, in chromatin reorganization and transcription regulation. Collectively, our studies provide a general paradigm of the dynamic relationship between chromatin reorganization and lineage specific gene expression pattern that dictates cell-fate determination.



Chapter 1
Introduction

Growing body of evidence suggests that three-dimensional (3D) organization of genome is closely associated with transcriptional activity and establishment of cell type-specific gene expression program (Bickmore et al., 2013; Gibcus et al., 2013; Gorkin et al., 2014; Misteli, 2007). Although the primary information regarding the organization of chromatin within the nucleus has been known more than a half century ago, the comprehensive 3D organization of chromatin and its association with cell-type specific gene expression pattern still remains unclear. Microscopic studies provide the first insights into the nuclear organization by revealing the presence of subnuclear structures called Cajal bodies (Cajal, 1903). Later, in the interphase nuclei of mosses Heitz, observed differentially stained chromatin and described it as heterochromatin and euchromatin (Heitz, 1928). Further studies demonstrated that during interphase, chromosomes occupy a preferential position, known as chromosome territories (Bolzer et al., 2005; Cremer et al., 2006), where large chromosomes found close to nuclear periphery and small chromosomes found more interiorly. This organization is further complicated by the observation that gene-poor chromatin is localized close to the nuclear periphery, whereas gene-dense chromatin is localized at the center of nuclei (Cremer et al. 2003; Croft et al. 1999). The spatial segregation of chromatin into active and inactive regions raised the possibility that nuclear positioning influences the gene activity. This proposition is supported by a seminal study which shows that immunoglobulin heavy chain (IgH) loci are preferentially localized at the nuclear periphery in multipotent progenitors and pro-T cells where they are transcriptionally silent. On the other hand in pro-B cells, these genomic loci are localized away from the nuclear periphery where undergo large scale compaction and subsequent rearrangement events (Kosak et al., 2002).

Although microscopy studies serve as invaluable tools to understand the nuclear organization of chromatin, they are limited in throughput and resolution. In order to explore the 3D organization of chromatin systematically and understanding its role in transcriptional regulation, requires advent of new techniques that exceed the limits imposed by resolution and throughput. Development of Chromosome Conformation Capture (3C) technique by Dekker revolutionized the field of nuclear organization, by enabling the detection of chromatin interactions with a resolution far beyond that provided by microscopic studies (Dekker et al., 2002). The 3C method involves cross-linking and proximity ligation of chromatin ends to measure the papulation-averaged contact frequencies between pairs of selected genomic loci. The advent of 3C technique

sparked the development of various 3C-based approaches **(Table 1.1)** (Dekker et al., 2002; Dostie et al., 2006; Dryden et al., 2014; Fullwood et al., 2009; Hughes et al., 2014; Kalhor et al., 2011; Lieberman-Aiden et al., 2009; Ma et al., 2015; Nagano et al., 2013; Rao et al., 2014; Simonis et al., 2006). These methods involve the same biochemical steps to capture close proximity chromatin interactions: cross-linking of chromatin, digestion of fixed chromatin with the restriction enzyme, proximity-ligation of digested ends. However, these 3C-based methods differ in the way of detection and quantification of ligation products by a variety of approaches including PCR, DNA microarrays, or high-throughput sequencing. 3C based studies not only reproduce the general features of genome organization that have been observed by microscopy studies but also unravel

S.No	Technique	Number of interrogating loci
1	Chromosome Conformation Capture (3C)	One versus One
2	Chromosome Conformation Capture on Chip (4C)	One versus All
3	Chromosome Conformation Capture Carbon-Copy (5C)	Many versus Many
4	Chromatin Interaction Analysis with Paired-End Tag sequencing (ChIA-PET)	Many versus Many
5	Capture-3C	Many versus All
6	Capture-Hi-C	Many versus All
7	Hi-C	All to All
8	*In situ* Hi-C	All to All
9	Tethered Conformation Capture (TCC)	All to All
10	Single-cell Hi-C	All to All
11	DNase Hi-C	All to All

Table1.1. Summary of chromosome conformation capture based approaches

the new principles of chromatin organization in the 3D nuclear space (Lieberman-Aiden et al., 2009). In line with microscopic observations, 3C based studies have demonstrated that chromatin is organized into A (permissive) and B (repressive) compartments (Lieberman-Aiden et al., 2009). Genome-wide inter-chromosomal interaction matrices reveled that small, gene-dense chromosomes (Chr:16,17,19,20,21 and 22) interact preferentially with each other. Consistently, FISH (Fluorescence in situ hybridization) studies revealed that these chromosomes are often colocalized at the center of the nuclei (Boyle et al., 2001; Tanabe et al., 2002). On the contrary small, gene-poor chromosomes (Chr:18) fail to interact with the other small chromosomes and found to be localized at nuclear periphery (Croft et al., 1999). Compartment A is strongly correlates with the presence of genes, higher transcript levels, DNaseI accessible chromatin and active

methylation marks (H3K36me3), thus it represents permissive chromatin. On the other hand, compartment B showed higher interaction frequency as compared to compartment A and also depleted with DNaseI accessible chromatin, indicating that B represents densely packed heterochromatin (Lieberman-Aiden et al., 2009; Ryba et al., 2010).

Recent studies indicate that at sub-megabase level these compartments are further organized into highly self-interacting regions known as topologically associating domains (TADs) (Dixon et al., 2012). The genomic regions between TADs are termed as 'boundary regions'. These chromatin domains were found to be stable and conserved across various cell types (Dixon et al., 2012). Growing body of evidence suggests that structural organization of TADs play important roles in various biological processes. First, TADs have been implicated as regulatory domains for coordinate gene expression by constraining interactions among genomic loci within the TAD (Flavahan et al., 2016; Nora et al., 2012; Shen et al., 2012). This is further supported by the findings that during cell differentiation and hormonal-induction epigenetic marks as well as gene expression patterns tend to change coordinately within a given TAD (Dixon et al., 2012; Le Dily et al., 2014). Second, TADs serve as functional domains for enhancer action by spatially restricting interactions with its cognate promoter to minimize bystander effect of enhancers (Dekker et al., 2015; Symmons et al., 2014). The spatial restriction of chromatin interactions within TADs is attributed to the boundary regions that are enriched with insulator binding protein, CTCF (Dixon et al., 2012). Deletion of boundary regions results in an increase in inter-TAD interactions as well as enhanced expression of genes that are located next to the deleted boundary region, indicating the critical role of CTCF in the maintenance of these discrete, functional domains (Ibn-Salem et al., 2014; Nora et al., 2012). Strikingly, disruption of CTCF-associated boundary regions across *EPHA4* locus, alters promoter-enhancer interactions, leading to the misexpression of various genes encoding for developmental regulators, resulting in limb deformities (Lupianez et al., 2015). Further, it was demonstrated that loss of CTCF results in dose dependent insulation defects at most of the boundary regions (Nora et al., 2017). TADs also appear to play a critical role in the VDJ recombination during B cell development by constraining the chromatin interactions within the same TAD which facilitate the proper recombination events and suppress potential for deletions as well as specious breaks (Hu et al., 2015). Third, recent studies suggest that TAD boundaries have the potential to serve as a physical barrier to restrict the spread of inactive chromatin into the

active one and vice versa (Austenaa et al., 2015). Although much is known about the function of TADs, the mechanism that underlying the formation these chromatin domains is still unclear. In a recent study, by employing polymer simulations and genome-wide chromatin interaction analyses Fudenberg et al., proposed a model where loop extrusion underlies the formation of TADs (Fudenberg et al., 2016). According to this model, loop extruding factors like cohesins form multiple loops dynamically and are stalled at boundary regions due to interactions with CTCF. The loop extrusion model explains several features of TADs that have been observed experimentally (Fudenberg et al., 2016). Accumulating evidence suggests that although TADs are invariant, the intrinsic chromatin interactions within these TADs show heterogeneity and is closely associated with epigenetic state and gene expression pattern (Dixon et al., 2015; Dixon et al., 2012; Wang et al., 2015).

It is widely accepted that the precise spatiotemporal gene expression depends on interactions between various regulatory elements like promoters, enhancers and locus control regions (LCR) as well as the molecular machinery (cell type-specific transcription factors, chromatin remodeling complexes) that bind to these regulatory elements (Kim et al., 2015; Lee et al., 2013; Levine, 2010; Li et al., 2002; Ong et al., 2011). While the promoter and insulator elements vary little across various cell types, the enhancer repertoire undergoes considerable changes during cell differentiation (Thurman et al., 2012). Thus, enhancer elements dictate the tissue-specific gene expression pattern by communicating with their cognate promoters. Epigenomics have been shown that promoters are typically enriched with H3K4me3 mark whereas H3K4me1 and H3K27Ac marks are enriched at the primed and active enhancers respectively (Bonn et al., 2012; Rada-Iglesias et al., 2011). On the other hand, H3K27me3 mark is typically enriched at silent promoters and enhancers (Cao et al., 2002; Czermin et al., 2002; Simon et al., 2009). Traditionally regulatory interactions have been inferred from genomic proximity, where enhancers are typically assigned to the nearest cognate promoters. However, growing evidence clearly suggests that enhancers can interact with their target promoters that are located tens of kilo-bases away on linear genomic-scale by looping out intervening DNA sequence and regulate their expression (Carter et al., 2002; Pennacchio et al., 2006; Ruf et al., 2011; Spitz et al., 2003). Thus genome-wide mapping of promoter-enhancer interactions is important to understand the gene-regulatory mechanisms that orchestrate the lineage-specific gene expression pattern. The

combination of 3C-based genomic techniques with genome-wide epigenetic signatures, transcriptome analyses and genome-editing approaches have enabled the systematic and comprehensive identification of promoter-enhancer interaction networks (Chepelev et al., 2012; Fullwood et al., 2009; Jin et al., 2013; Kieffer-Kwon et al., 2013; Li et al., 2012; Rao et al., 2014; Sanyal et al., 2012). These studies uncovered the new principles of dynamics of promoter-enhancer interactions across various cell types. RNA PolII ChIA-PET interactome data revealed that promoter-tethered chromatin interactions are classified into three basic groups: intergenic (promoter-promoter), extragenic (promoter-enhancer) and intragenic (promoter-gene body) (Li et al., 2012). These interactions are further aggregated as "single-gene" and "multigene" complexes based on the number of genes involved. Single-gene complexes comprise of single gene promoter interacting with one or multiple enhancers, whereas multigene complexes consist of multiple promoters interacting with single or multiple promoters and/or enhancers. The promoters that bind with PolII but not involved in any chromatin interactions are termed as "basal promoters". The expression levels found to be higher for the genes whose promoters interact with *cis*-regulatory elements as compared to genes of basal promoters. Interestingly, several studies reported that a considerable number of promoters interact with multiple enhancers and the expression level of cognate genes is positively correlated with the number of interacting enhancers (Chepelev et al., 2012; Kieffer-Kwon et al., 2013; Schoenfelder et al., 2015). Comparison of promoter-enhancer interactome between B and ES cell revealed that both cell-type specific as well broadly expressed gene promoters interact with stage-specific enhancers. Thus, it is proposed that modulation of enhancer repertoire enables fine tuning of transcriptional output and also place all the regulatory elements under the control of tissue-specific transcription (Kieffer-Kwon et al., 2013).

The hematopoietic system represents a leading developmental model to decipher the gene-regulatory networks that underlie B-cell lineage specification and commitment. Lymphopoiesis initiates within the bone marrow, where lymphoid-primed multipotent progenitors (LMPPs) differentiate into common lymphoid progenitors (CLPs). Pre-pro-B cells also termed as Fraction A cells are generated from CLPs (Inlay et al., 2009). These pre-pro-B cells further differentiate into pro-B cells or Fraction B cells. Pro-B cells are characterized by the activation of B-lineage specific genes including *Cd79a*, *Cd79b*, *Vpreb1*, *Igll1* and *Cd19*, initiation of ordered immunoglobulin heavy chain gene rearrangements and interleukin-7 (IL-7) dependent

proliferation (Hardy et al., 2007; Murre, 2009). Differentiation of multipotent progenitors into committed pro-B cells, which involves activation of B-lineage specific genes and repression of inappropriate lineage determinants, is orchestrated by coordinated expression of various cell type-specific transcription factors and signaling cascades (Mandel et al., 2010; Singh et al., 2005; Singh et al., 2007). Genetic ablation studies have enabled the identification of a unique set of transcription factors (Ikaros, PU.1, E2A, EBF1 and Pax5) and signaling molecules (Flt3 and IL7) that direct the progression of multipotent progenitors along B-cell pathway (Bain et al., 1994; Georgopoulos et al., 1994; Lin et al., 1995; Peschon et al., 1994; Scott et al., 1994; Urbanek et al., 1994; Yoshida et al., 2006).

In *Ikaros* deficient mice, LMPPs undergo excessive myelopoiesis and are completely lost their B-cell potential, highlighting the role of Ikaros in lymphopoiesis (Yoshida et al., 2006). The differential expression level of PU.1 refines the mixed lineage pattern of LMPPs into a specific cell fate choice. High levels of PU.1 induce differentiation LMPPs into macrophages, whereas low levels PU.1 induce B cell developmental program (DeKoter et al., 2000). Loss of PU.1 leads to a severe block in B lymphopoiesis and also mutant progenitors fail to express IL-7R which is critical for the development of B lymphocytes (DeKoter et al., 2002; Scott et al., 1994). Interestingly, ectopic expression of IL-7Rα in PU.1 mutant cells partially rescued their B-cell developmental potential, suggesting the role of additional critical downstream targets of PU.1 in B-lymphopoiesis (DeKoter et al., 2002). *E2A* gene which encodes for E12 and E47 proteins is critical for proper B-cell development, loss of which leads to defects in early B-linage expression program despite maintaining them under B-lymphoid conditions (Bain et al., 1994). Thus Ikaros, PU.1 and E2A induce lymphoid specific gene expression program in multipotent progenitors. Later, it was found that PU.1 and E47 in conjunction with IL-7 signaling induce the expression of Ebf1, which is a primary determinant of B-lineage specification (Ikawa et al., 2004; Kee et al., 1998; Medina et al., 2004; Reynaud et al., 2008; Roessler et al., 2007). Targeted disruption of Ebf1 in mice results in complete lack of B lymphocytes and mutant progenitors are blocked at the developmental transition from pre-pro-B (fraction A) to pro-B cell stage (fraction B) (Lin et al., 1995). Molecular analysis revealed that these mutant cells fail to express early B-lineage specific genes *Cd79a*, *Cd79b*, *Vpreb1*, *Igll1, RAG1 and RAG2,* also show defects in V_H to DJ_H but not $D\text{-}J_H$ rearrangements(Lin et al., 1995). The imperative role of Ebf1 in B-lineage specification was

further demonstrated by adoptive transfer experiments in which enforced expression of Ebf1 in hematopoietic progenitors skewed their differentiation along the B-cell developmental pathway at the expense of alternate lineages (Zhang et al., 2003). Ebf1 directly regulates the expression of a secondary B-cell fate determinant, Pax5 (Roessler et al., 2007) (Decker et al., 2009).

Unlike E2A and Ebf1 mutant cells, targeted disruption of Pax5 results in developmental arrest at pro-B cell stage (Urbanek et al., 1994). Pax5-/- pro-B cells express most of the early B-lineage genes and undergo productive proximal but not distal V_H-DJ_H gene rearrangements (Hesslein et al., 2003; Nutt et al., 1997). However, transplantation experiments demonstrate that unlike their wild-type counterparts Pax5-/- cells have the potential to differentiate into other hematopoietic cell types (Nutt et al., 1999; Rolink et al., 1999). Molecular analysis revealed that Pax5 antagonizes the alternate lineage choice of B-cells by repressing the various lineage inappropriate genes including *csf1r* and *Notch1* (Delogu et al., 2006; Nutt et al., 1999; Souabni et al., 2002). Thus, Pax5 has been considered as a major factor required for B-cell fate commitment. Interestingly, it was found that ectopic expression of Ebf1 but not Pax5 restores the B-cell developmental potential of *PU.1-/-*, *Il-7ra-/-* and *E2A-/-* progenitor cells (Kikuchi et al., 2005; Medina et al., 2004; Seet et al., 2004).

Later loss-of-function and gain-of-function studies highlight the additional role of Ebf1 in B-cell fate commitment independent of Pax5 (Pongubala et al., 2008). Developmental analysis revealed that *Ebf1-/-* cells blocked at pre-pro-B cell stage and have the potential to differentiate into various other lineages including myeloid, T- and natural killer (NK) cells (Pongubala et al., 2008). Ectopic expression of Ebf1 in these mutant progenitors has been shown to restrict their inappropriate lineage choice and induce B-cell developmental program (Pongubala et al., 2008). Transplantation and reconstitution experiments showed that sustained expression of Ebf1 in *Pax5-/-* fetal liver cells inhibits their myeloid and T-lineage potential *in vivo* (Pongubala et al., 2008). Ectopic expression of Ebf1 but not Pax5 induce differentiation of MPPs along B-cell pathway at the expense of myeloid lineage by downregulating the myeloid determinants including *Cebpa*, *Sfpi1* and *Id2*. Similarly, enforced expression of Ebf1 in *Pax5-/-* pro-B cells impedes their T-cell potential by repressing GATA3 expression (Pongubala et al., 2008). Thus, Ebf1 restrict the alternate lineage choice of multipotent progenitors independent of Pax5. Collectively, these results

clearly demonstrate that differentiation of multipotent progenitors into committed B-cells involves a complex and hierarchical gene regulatory network comprising of various signaling molecules and lineage-specific transcription factors **(Figure 1.1)**.

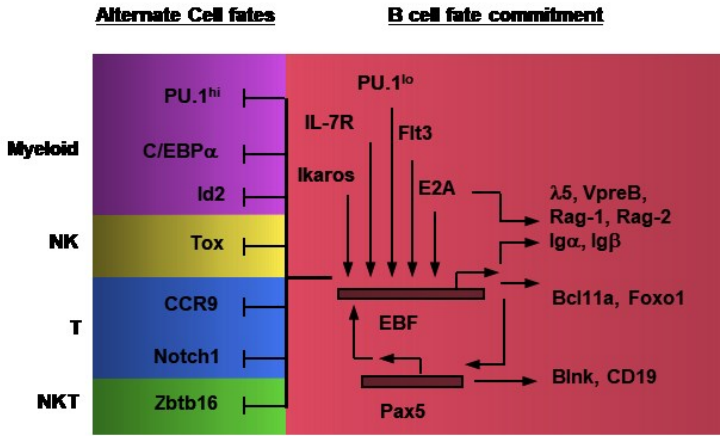
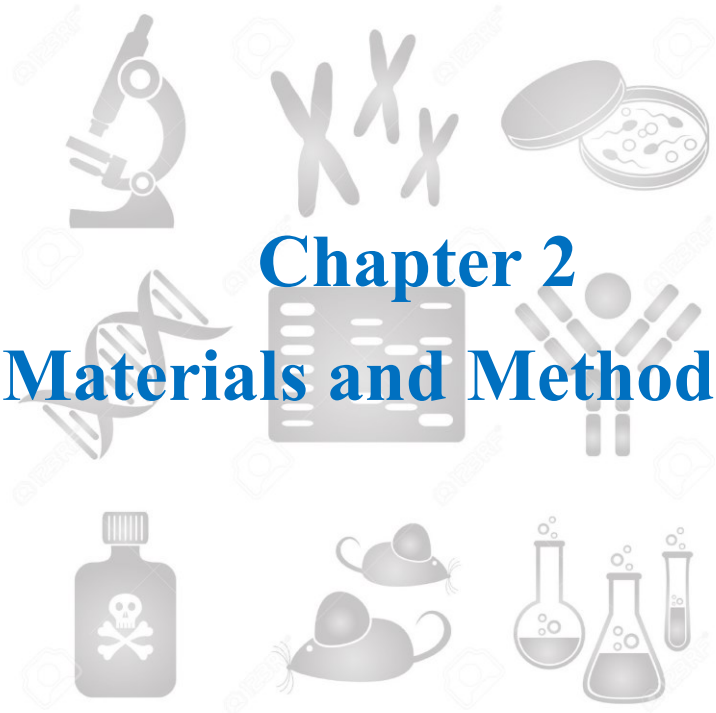


Figure 1.1. Gene regulatory network that orchestrate B-cell fate specification and commitment of multipotent progenitors

Although much is known about the concerted interplay of transcription factors that are important for B cell determination, far less is known about the genome-wide composition of *cis*-regulatory interactions controlling B-lineage-specific gene expression program and relevance of these interactions on hierarchical organization of the chromatin during early B cell developmental transition. To obtain a comprehensive view of the relationship between higher-order chromatin reorganization and induction of developmentally regulated B lineage-specific gene expression program, we have framed the following objectives:

1. Investigating the changes in chromatin compartmentalization during B cell fate commitment
2. Probing the changes in structural organization of topologically associating domains (TADs) during developmental transition of pre-pro-B cells to pro-B cells
3. Deciphering genome-wide *cis*-regulatory interaction maps during differentiation of pre-pro-B cells to pro-B cells
4. Establish the role of Ebf1 in chromatin re-organization during B cell fate commitment

The findings of these studies have been recently published in the journal of Nucleic Acids Research (Boya et al., 2017).



Chapter 2
Materials and Methods

2.1 Cell culture

Pre-pro-B cells (*Ebf1-/-* progenitors) were maintained on stromal layer (OP9 cells) in the presence of Opti-MEM (Gibco) containing 4% (v/v) fetal calf serum, β-mercaptoethanol (50 µM), penicillin (10 U/ml) and streptomycin (10 µg/ml) and supplemented with SCF (10 ng/ml), Flt3L (10 ng/ml) and IL-7 (5 ng/ml). Pro-B cells (*Rag2-/-* cells) were maintained under similar conditions except that the media was supplemented with only IL-7 (5 ng/ml). Both pre-pro-B cells and pro-B cells were used for preparation of RNA for RT-PCR and chromatin for the 3C and Hi-C assays.

2.2 *In situ* Hi-C approach

In situ Hi-C was carried out as described (Lieberman-Aiden et al., 2009), except that chromatin cross-linking, restriction enzyme (HindIII) digestion, biotin fill-in and ligation reactions were performed in intact nuclei as described below **(Figure 2.2.1A)** (Nagano et al., 2013; Rao et al., 2014).

2.2.1 Cell cross-linking and isolation of nuclei

Target cells (Pre-pro-B cells and pro-B cells; $30X10^6$ for each biological replicate) were cross-linked with 2% (v/v) formaldehyde (37%) in a total volume of 40 ml for 5 min at room temperature (RT) with gentle mixing. The crosslinking reaction was stopped by the addition of 5.7 ml of 1 M glycine (0.125 M) and incubated at RT for 5 min. Cells were collected by centrifugation at 1300 rpm at 4^0C for 8 min and washed once with ice-cold 1 X PBS. Cells were lysed in 50 ml ice-cold lysis buffer, (10 mM Tris-HCl (pH8.0), 10 mM NaCl, 0.2% NP-40), supplemented with protease inhibitors (Roche), by incubating on ice for 15 min with intermittent mixing. Cell lysates were centrifuged at 1800 rpm at 4^0C for 5 min, nuclei were resuspended in 1.2 X NEbuffer 2 (60 mM NaCl, 12 mM Tris-HCl (pH7.9), 12 mM $MgCl_2$, 1.2 mM DTT) at a density of $20X10^6$/ml, and aliquoted 500 µl per tube.

2.2.2 Nuclear restriction digestion, fill-in and ligation

To set up intra-nuclear enzymatic reactions, three nuclear aliquots containing $10X10^6$ nuclei (two aliquots for *in situ* Hi-C and one aliquot for 3C control), were taken for each biological replicate of a given target cell type and permeabilized in the presence of 0.3% SDS (7.5 µl of 20% SDS) by incubating at 37^0C for 1hour with constant agitation. SDS was quenched with 50 µl of 20%

Triton X 100 (2%) under constant agitation at 37^0C for 1hour. Subsequently 1500 U (15 µl) of HindIII (New England Biolabs) was added to each nuclear suspension and continued to maintain under constant agitation at 37^0C for overnight. Following restriction digestion, two nuclear aliquots (*in situ* Hi-C, for each biological replicate) were pooled, and nuclei were pelleted by centrifugation at 14000 rpm at RT for 30 sec. Nuclei were washed twice with 500 µl of 1.2 X NEBuffer 2 by brief vortexing followed by centrifugation and digested chromatin ends were marked with biotinylated nucleotide (dCTP) by fill-in reaction. For fill-in, each nuclear pellet was resuspended in 100 µl reaction mix containing 1 X NEBuffer 2, 5 µl (25 U) Klenow polymerase (New England Biolabs), dNTPs: 1.5 µl 10 mM dATP, 1.5 µl 10 mM dTTP, 1.5 µl 10 mM dGTP and 37.5 µl of 0.4 mM biotinylated d-CTP (Invitrogen) and incubated at 37^0C for 45 min with intermittent mixing. Following fill-in reaction, the nuclei were pelleted and washed twice with 1 X NEB T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl2, 1 mM ATP, 10 mM DTT) by brief vortexing and centrifugation at RT. Nuclei were resuspended in 300 µl of 1 X NEB ligase buffer containing 3200 U (8 µl) of T4 DNA ligase (New England Biolabs). In parallel, 3C control nuclei (digested with HindIII) were pelleted by brief centrifugation, washed twice with 500 µl of 1 X T4 DNA ligase buffer, resuspended in 200 µl 1 X T4 DNA ligase buffer containing 1600 U (4 µl) of T4 DNA ligase (New England Biolabs). Both *in situ* Hi-C and 3C control nuclei were incubated at 16^0C for 4 hours followed by 30 min at RT.

2.2.3 Chromatin de-crosslinking and purification of DNA

Following ligation, the nuclear samples were diluted with 300 µl (*in situ* Hi-C) or 400 µl (3C) of NEB elution buffer (10 mM Tris-HCl (pH8.0)) and disruption of nuclei and reversal of chromatin crosslinking was accomplished by incubating at 65^0C for 12-16 hours in the presence of (80 µl for *in situ* Hi-C reaction or 40 µl for 3C control) proteinase K (10 mg/ml, Invitrogen). Following proteinase K digestion, 10 µl or 5 µl of RNase A (0.5 mg/ml, Roche) was added to *in situ* Hi-C and 3C samples respectively and incubated at 37^0C for 1hour. DNA was purified by phenol (600 µl of Tris-Saturated phenol) extraction followed by chloroform (600 µl) extraction, and DNA was precipitated overnight by storing at -80^0C with ethanol (100%) in the presence of sodium acetate (0.3M, pH 5.2). The next day, DNA was pelleted by centrifugation at 14000 rpm, 4^0C for 30 min, washed with 70% ethanol, re-suspended in 100 µl of nuclease-free water (Ambion) and quantitated using a fluorometer (Qubit 2.0; Invitrogen).

2.2.4 Qualitative analysis of DNA libraries

100 ng of *in situ* Hi-C and 3C DNA from each biological replicate was run on a 0.8% agarose gel (1 X TAE) and the integrity of DNA was assessed by molecular weight **(Figure 2.2.1B)** as described previously. The efficiency of nuclear enzymatic reactions, particularly fill-in and ligation, during *in situ* Hi-C were monitored using regeneration of NheI site. During this, 3C DNA sample was included as a negative control, where DNA ends were ligated without end fill-in **(Figure 2.2.1C)**.

2.2.5 Removal of biotinylated-dCTP from un-ligated ends

To avoid capturing of un-ligated DNA fragments carrying biotinylated-dCTPs, the DNA samples were subjected to T4 DNA polymerase treatment. For this, 10 µg of *in situ* Hi-C DNA was mixed with a 100 µl reaction mix containing 2 X NEBuffer 2, 2 µl dATP (10 mM), 2 µl dGTP (10 mM), 2 µl BSA (10%), and 3.3 µl (10 U) of T4 DNA polymerase (New England Biolabs) and incubated for 2 hours at 12^0C. Following T4 DNA polymerase treatment, DNA was extracted with phenol followed by chloroform and then DNA was precipitated with 100% ethanol in the presence of sodium acetate (0.3M, pH 5.2) for overnight at -80^0C. Next day, DNA was pelleted by centrifugation at 14000 rpm, 4^0C for 30 min, washed with 70% ethanol and resuspended in elution buffer (10 mM Tris-HCl, (pH8.0), Qiagen). DNA was quantitated using a fluorometer (Qubit, Invitrogen), diluted with elution buffer and aliquoted at 5 µg/100 µl per tube.

2.2.6 Sonication and end-repair

DNA (5 µg/100 µl) was sheared using ultrasonicator (Covaris S220) with an output setting of 10 duty cycle, 200 burst/cycle and time 85 sec, to generate DNA fragment size ranging from 300-500 bp. The efficiency of sonication and DNA fragment size was assessed using Agilent bio-analyzer. Four aliquots (400 µl) of sonicated DNA samples were pooled and end-repair was performed in 500 µl reaction mix containing 50 µl 10 X T4 DNA ligase buffer, 12.5 µl of dNTPs mix (10 mM), 5 µl (15 U) T4 DNA polymerase, 5 µl (15 U) T4 ploynucleotide kinase and 1 µl (5 U) Klenow DNA polymerase by incubating at 20^0C for 30 min. DNA was purified by phenol extraction followed by chloroform extraction, precipitated using 100% ethanol, DNA pellet was washed, dried and resuspended in DNase- and RNase-free water (Ambion).

2.2.7 Size selection and enrichment of in situ Hi-C library

Sonicated DNA sample of each biological replicate was run on 2% agarose gel (1 X TAE), DNA fragments ranging from 300-500 bps were excised and eluted using gel extraction columns (Qiagen) in DNA LoBind tubes (1.5 ml; Eppendorf) in a total volume of 100 µl elution buffer (10mM Tris-HCl (pH8.0)). For enrichment of biotinylated DNA fragments, which represents the close proximity ligated products, each *in situ* Hi-C DNA sample was mixed with 50 µl of Dynabeads M280 streptavidin beads (M280; Invitrogen) that are pre-washed with 1 X binding/washing buffer (5 mM Tris-HCl (pH7.5), 0.5 mM EDTA, 1 M NaCl) and incubated in the presence of 2X binding/washing buffer (10 mM Tris-HCl (pH7.5), 1 mM EDTA, 2 M NaCl) for 1 hour at RT under constant rotation. Streptavidin beads were washed with 1X binding/washing buffer followed by 400 µl of 1 X NEBuffer 2 to remove non-biotinylated DNA fragments. All downstream steps were carried out using DNA LoBind tubes (1.5 ml; Eppendorf) and washing steps were performed using a DynaMag-spin (Invitrogen).

2.2.8 Adenylation of 3'ends

To avoid self-ligation of DNA fragments and to facilitate complementary 5' overhang ('T') ligation of adaptors, a single 'A' nucleotide was added to 3'ends of the blunt fragments. For this, DNA bound beads were resuspended in 50 µl reaction mix containing 5 µl of 10 X NEBuffer 2, 10 µl dATP (1 mM) and 1 µl (5 U) Klenow exo- and incubated at 37^0C for 20 min. Beads were washed twice with 400 µl of 1 X binding/washing buffer and once with 200 µl of 1 X T4 DNA ligase buffer.

2.2.9 Paired-end adaptor ligation and PCR amplification of in situ Hi-C library

Streptavidin beads carrying DNA fragments were resuspended in a 50 µl reaction mix containing 5 µl of 10 X T4 DNA ligase buffer (New England Biolabs), 2.5 µl multiplex paired-end adapter oligos (True-Seq Paired-end kit, Illumina), and 2 µl T4 DNA ligase (800 U, New England Biolabs) and incubated at RT for overnight at constant rotation. To remove un-ligated adapters, beads were washed twice with 100 µl of 1 X binding/washing buffer. Paired-end PCR amplification was carried out to selectively enrich the adapter ligated DNA fragments with cycling conditions of initial denaturation at 98^0C (30 sec) followed by 98^0C (10 sec), 60^0C (30 sec) 72^0C (30 sec) for 15 cycles and a final extension at 72^0C (5 min) **(Figure 2.2.1D)**. *In situ* Hi-C libraries were collected

using magnetic separator (Invitrogen), and run on agarose gel (2%) for size selection. DNA fragments ranging from 400-550 bp were excised from the gel and purified in a total volume of 20 µl elution buffer (10 mM Tris-HCl, (pH 8.5)) using a MinElute gel extraction kit (Qiagen).

2.2.10 Quality control of in situ Hi-C library

Size and quality of *in situ* Hi-C libraries were monitored using a Bioanalyzer (Agilent). The libraries were quantitated using KAPA Library Quantification Kit (Kapa Biosystems).

2.2.11 High-throughput paired-end sequencing

The *in situ* Hi-C libraries were subjected to 2X 70 bp paired-end high-throughput sequencing using Hi-Seq (Illumina). Sequencing reads were compiled and mapped to the mouse genome, mm10, using Bowtie2.

2.3 Quantitative Chromosome Conformation Capture (3C-qPCR)

3C was performed as described previously (Dekker et al., 2002) except that chromatin cross-linking, restriction digestion (HindIII) and ligation were performed in intact nuclei.

2.3.1 Cell cross-linking and isolation of nuclei

Pre-pro-B ($20X10^6$) and pro-B cells ($20X10^6$) were cross-linked with 2% fix containing 2.596 ml of 37% formaldehyde (Merck) and 48 ml of Opti-MEM (Gibco) and incubated for 5 min at room temperature (RT) with gentle mixing. Cross-linking reaction was quenched with 5.7 ml of 1M glycine (final concentration 0.125M) by incubating at RT for 5 min. Cells were collected by centrifugation at 1300 rpm for 8 min at 4ºC and washed once with 50 ml of ice-cold 1 X PBS (Invitrogen). Cell lysis was carried out by resuspending in 50 ml of ice-cold lysis buffer containing 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP-40 (Roche) and 1 X protease inhibitor cocktail (EDTA-free, Roche) and incubated on ice for 15 min with intermittent mixing. Nuclei were collected by centrifugation of cell lysates at 1800 rpm at 4ºC for 5 min and were re-suspended in 1.2 X NEB buffer 2 (60 mM NaCl, 12 mM Tris-HCl (pH7.9), 12 mM MgCl2 and 2 mM DTT) at a density of $20X10^6$ nuclei/ml. Nuclear suspension were aliquoted, 500 µl into each tube.

2.3.2 Nuclear restriction digestion and ligation

Each nuclear suspension (500 µl), in order to facilitate intra-nuclear restriction enzyme digestion, nuclei was permeabilized in the presence of 7.5 µl of 20% SDS (0.3% final concentration) at 37ºC for 1 hour with constant agitation. SDS was sequestered by adding 50 µl of 20% Triton-X100 and incubated for 1 hour at 37ºC with constant agitation. To determine digestion efficiency, 5 µl aliquot of nuclear suspension was taken and labeled as undigested genomic DNA control (UND). Subsequently 1500 U of HindIII (New England Biolabs) was added to each nuclear suspension and incubated at 37ºC for overnight with constant agitation. To determine digestion efficiency 5 µl aliquot of digested sample was taken and labeled as digested genomic DNA control (D) and was stored at -20ºC. Following restriction enzyme digestion each nuclear suspension was washed twice with 1 X T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl2, 1mM ATP and 10 mM DTT, New England Biolabs) by brief centrifugation and vortexing at room temperature. Ligation was carried out by re-suspending nuclei in 200 µl of 1 X T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl2, 1 mM ATP and 10 mM DTT, New England Biolabs) containing 800 U of T4 DNA ligase (New England Biolabs) and incubated for 4 hours at 16ºC followed by 30 min at room-temperature.

2.3.3 Chromatin de-crosslinking and purification of 3C DNA

300 µl of elution buffer (10 mM Tris-HCl (pH8.0), Qiagen) was added to each ligation sample and de-crosslinking was carried out by incubating with 20 µl of proteinase K (20 mg/ml, Invitrogen) at 65ºC for overnight. Following of proteinase K treatment, 5 µl of RNase (0.5mg/ml, Roche) was added and incubated for 1 hour at 37ºC. DNA was purified by sequential extractions with 500 µl of Tris-saturated phenol (Sigma) and 500 µl of chloroform (Sigma) and was precipitated for 1 hour or more by incubating at -80ºC with 100% ethanol in the presence of sodium acetate pH 5.2 (0.3M, pH 5.2). DNA was pelleted by centrifuging at 14000 rpm at 4ºC for 30 min, washed once with 70% ethanol, air-dried for few min, re-suspended in 100 µl of nuclease-free water (Ambion) and quantitated using fluorometer (Qubit 2.0, Invitrogen).

2.3.4 Quality control of 3C DNA

The integrity of 3C template was confirmed by resolving samples on 0.8% agarose gel where DNA run as a tight band above 10 kb.

2.3.5 Determination of digestion efficiency

In order to determine restriction enzyme digestion efficiency 500 µl of 1 X PK buffer (5 mM EDTA (pH 8.0), 10 mM Tris-HCl (pH 8.0), 0.5% SDS) and 1 µl of proteinase K (20 mg/ml) were added to the saved control DNA aliquots (UND, D) and incubated at 65ºC for overnight. The DNA samples were equilibrated at 37ºC for few minutes, then 2 µl of RNase A (0.5 mg/ml, Roche) was added and incubated at 37ºC for 1 hour. DNA was purified by sequential extractions with 500 µl of Tris-saturated phenol (Sigma) and 500 µl of chloroform (Sigma) and was precipitated for 1 hour or more by incubating at -80ºC with 100% ethanol in the presence of sodium acetate (0.3M, pH 5.2). DNA was pelleted by centrifuging at 14000 rpm at 4ºC for 30 min and pellet was washed once with 70% ethanol, air-dried for few min and re-suspended in 50 µl of nuclease free water (Ambion). Real time PCR quantifications (SYBR green) were performed for both DNA samples (UND, D). To assess digestion efficiencies, a primer set (R) that amplify a region (*Ercc3*) across restriction site of interest (HindIII) was included and to correct for differences in the amount of template added to the PCR, a control primer set (C) was used to amplify regions (ZP3) that do not containing the restriction sites of interest. Digestion efficiency of each sample was calculated according to the following formula,

$$\% \text{ restriction digestion} = 100\text{-}100/2^{\wedge} ((Ct_R\text{-}Ct_C)\ D - (Ct_R\text{-}Ct_C)\ UND)$$

The digestion efficiency of HindIII for pre-pro-B and pro-B cells is 80%, 86% respectively.

2.3.6 Generation of control template

A quantitative comparison of different PCR products requires correction of PCR amplification efficiency of each primer set. Thus, a control template is required in which all possible ligation products are present in equimolar amounts. To generate control template, 20ug of BAC clone containing *Ccl3* locus (RP23-59G9) was digested with HindIII restriction enzyme to a final concentration of 9% by incubating at 37ºC for overnight. DNA was purified with phenol, chloroform extraction followed by ethanol precipitation and resuspended in 161 µl of nuclease free water (Ambion). Ligation was carried out in 200 µl reaction mix containing 161 µl of digested BAC DNA, 20 µl of 1 X T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl2, 1 mM ATP and 10 mM DTT, New England Biolabs) and 7,600 U of T4 DNA ligase (New England Biolabs) by incubating at 16ºC for overnight. Next day, DNA was purified with phenol, chloroform extraction followed by ethanol precipitation and eluted in 100 µl of nuclease-free water (Ambion).

The amount of DNA was quantitated using fluorometer (Qubit 2.0, Invitrogen). HindIII digested genomic DNA was mixed with control DNA template similar to the amount (50 ng) used for 3C template. Then serial dilutions of control template was carried out to determine the proper template concentration that shows optimal amplification of ligation product.

2.3.7 Real-time PCR quantification of cross-linking frequencies

Serial dilutions of 3C template was carried out to determine an optimal amount of DNA template (in our study ~50 ng), that provides linear PCR amplification. Each ligation product was quantitated in triplicate on real-time PCR (ABI step-one plus) using 10 µl reaction mix containing 5 µl of 2 X power SYBR green master mix (1 X final concentration), 1 µl of 10 µM forward primer (1µM final concentration), 1 µl of 10 µM reverse primer and 50 ng of 3C template with cycling conditions of initial denaturation at 94ºC for 10 min, followed by 94ºC (30 sec) and 62ºC (60 sec) for 40 cycles. Simultaneously, in each run standard curves were performed using serial dilution of control template in order to normalize PCR amplification efficiencies between different primers sets. Finally, 3C data was normalized with a "loading control" (*Zp3*) (using a primer set that amplify a regions not containing the restriction sites of interest (HindIII)) and a set of "control interaction frequencies" (*Ercc3* primer set) in order to normalize amount and quality of different 3C samples. Relative cross-linking frequency of each ligation product was calculated using following formula,

$$\text{Value} = 10^{(Ct-b)/a} \text{ (b: intercept, a: slope)}$$

2.4 *In situ* Hi-C data analysis

2.4.1 Iterative Mapping

We have used hiclib (Imakaev et al., 2012) to perform preliminary Hi-C data analysis. The Iterative mapping module of hiclib has a functionality to truncate the raw reads to 25 bp starting at the 5′ end which will be subsequently mapped to the reference genome (mouse: mm10) in a single-end mode using Bowtie2 software. Reads that mapped to multiple regions in the genome were extended by 5 bp and then re-mapped. This process was repeated until either all reads were uniquely mapped to the reference genome or until the reads were extended to their entirety (100 bp). Using this approach we were able to uniquely map more than 85% of the reads to the reference genome. We have discarded the un-alignable and chimeric (aligned at multiple sites) reads.

Further, only paired end reads (around 80% of the total uniquely aligned reads) are considered for subsequent analysis.

2.4.2 Filtering spurious ligation products

The quality of the *in situ* Hi-C library was assessed based on the position and orientation of sequenced read pairs relative to their restriction site (HindIII). Inappropriate ligation products such as self-circularized ligation or unligated "dangling end" products, generated as a result of experimental biases were discarded at fragment level filtering. In total, ~0.1% of self-circles (formed due to less cross linking efficiency) and ~18% of read pairs having dangling-ends (formed due to less ligation efficiency) were discarded from both pre-pro-B and pro-B *in situ* Hi-C libraries. We have obtained a total of 31191614 and 39490989 valid paired-end reads from pre-pro-B and pro-B cells respectively.

2.4.3 Filtering of PCR duplicates and extreme fragments

During *in situ* Hi-C library preparation the redundant molecules, also called as PCR duplicates, are generated as a result of PCR over-amplification. The presence of these duplicates although is negligible, may affect the relative contact probability and thus have to be discarded. Also, we have followed the default parameters to remove fraction of the most and the least-count fragments to account for systemic biases (the presence or absence of nearby restriction site). Further, we have discarded the fragments that are too long (>100 kb) or too short (<100 bp), as characterizing these interacting fragments will be difficult. Also, we have discarded all reads that start within 5 bp near HindIII site.

2.4.4 Quality check for library size

To ensure that the aligned sequence reads represent restriction fragment ends, the distances from mapped *in situ* Hi-C reads to the nearest restriction sites (HindIII) were computed. In a pair, if the sum of distances from mapped R1 read position to its restriction site and distance from mapped R2 read position to its restriction site, $D1+D2 \leq 500$ then they are considered as specific interactions. For all the valid pairs obtained after fragment-level filtering around 90% of the interactions were found within 500 bp **(Figure 2.4.1)**.

2.4.5 Generation of relative contact probability matrices using iterative correction

Valid interaction pairs (non-redundant perfect ligation products) are used to measure the frequency of physical contact between two given regions of each chromosome. In order to gain statistical power, almost all the studies that were previously reported pooled the numbers of reads into bins of larger genomic regions (say 1Mb). Although most of the non-specific interactions were removed at fragment-level filtering, the contact maps generated may still be influenced by several intrinsic properties of the genome and would display different "experimental visibility". Thus we have implemented iterative correction for the binned data to eliminate biases based on the assumption that all loci should have equal visibility. we have removed the poor regions by coverage $\leq$0.05 followed by removing the PCR blowouts and implemented iterative correction to calculate the maximum likelihood solution for biases B_i obtained by iteratively solving a system of equations that convert the observed contact counts into corrected counts. This method ensures specific interactions be precisely pictured which otherwise be concealed by visibility-induced biases (Imakaev et al., 2012).

2.4.6 Chromatin compartmentalization through eigen vector decomposition

Eigen vector decomposition was performed based on the interaction profiles obtained through iterative correction of the chromosomes (corrected at 1 Mb and 100 kb resolutions). Since in most of the cases first principal component (PC1) best explains for as much of the variability in the data and reveals its internal structure we have considered it to partition chromosomes into A and B (permissive and repressive) compartments. In order to check whether eigen vector decomposition truly represents the chromatin state, we have obtained the publicly available pre-processed ChIP-Seq data of active methylation marks (H3K4me1, H3K4me3 and H3K9/14ac) for pre-pro-B and pro-B cell types from Gene Expression Omnibus (GEO) database (Lin et al., 2010). We have used liftover tool from UCSC tools to lift the peak positions of each active methylation patterns from mm8 to mm9 and then to mm10 to make it comparable with our data.

2.4.7 Identification of topologically associated domains

Iteratively corrected relative contact probability matrices at 40 kb resolution, generated by implementing HiResHiC module of hiclib were converted into the format specified by Domain Caller (Dixon et al., 2012), where the first three columns represent the chromosome number

followed by start and end of the bin. Domain Caller is a simple and straightforward approach with greater flexibility to identify biologically relevant domain structures.

2.4.8 Generation of 3D structures of TADs

We have generated 3D structures of TADs in both pre-pro-B and pro-B cells by implementing AutoChrom3D (Peng et al., 2013), which uses a novel sequencing-bias-relaxed parameter to normalize chromatin interactions.

2.4.9 Determination of statistically significant *cis*-regulatory interactions

To discriminate between random polymer loops and specific chromatin loops, we have used Fit-Hi-C (Ay et al., 2014), a tool for assigning statistical confidence estimates to mid-range contacts. We have prepared "FRAGSFILE" containing midpoints (or start indices) of the fragments and "INTERSFILE" containing interactions between fragment pairs from the dict-file obtained through fragment level filtering. The BIASFILE is prepared by using the python code that implements the iterative correction in sparse mode by filtering out loci that are less mappable than the threshold (cut off $\geq$0.5). The significant interactions obtained by implementing Fit-Hi-C, were further integrated with various epigenetic modifications (H3K4me1, H3K4me3, H3K4me2 and H3K9/14ac) to identify potential *cis*-regulatory interactions.

2.4.10 Microarray analysis

Pre-pro-B cells were transduced by spin-infection with retrovirus encoding GFP or Ebf1-GFP or Pax5-GFP and maintained for 2 days in lymphoid culture conditions as previously described (Pongubala et al., 2008). After two days, GFP+ transductants were FACS-sorted, total RNA was isolated with TRIzol reagent (Invitrogen) and further purified on RNeasy columns (Qiagen). RNA quality control analysis was performed as previously described (Pongubala et al., 2008). Biotin-labeled cRNA was generated and hybridized to the Mouse Genome 430 2.0 Array (Affymetrix, Santa Clara, USA) according to the manufacturer's instructions. Data were analyzed as previously described (Pongubala et al., 2008).

2.4.11 Statistical analysis

All statistical analyses except for identifying significant promoter-*Cis*-interactions, were performed using R package. Statistical significance was evaluated by unpaired two sample t-test. For all the tests performed, statistical significance was assessed as *P < 0.05, **P < 0.01, ***P < 0.001.

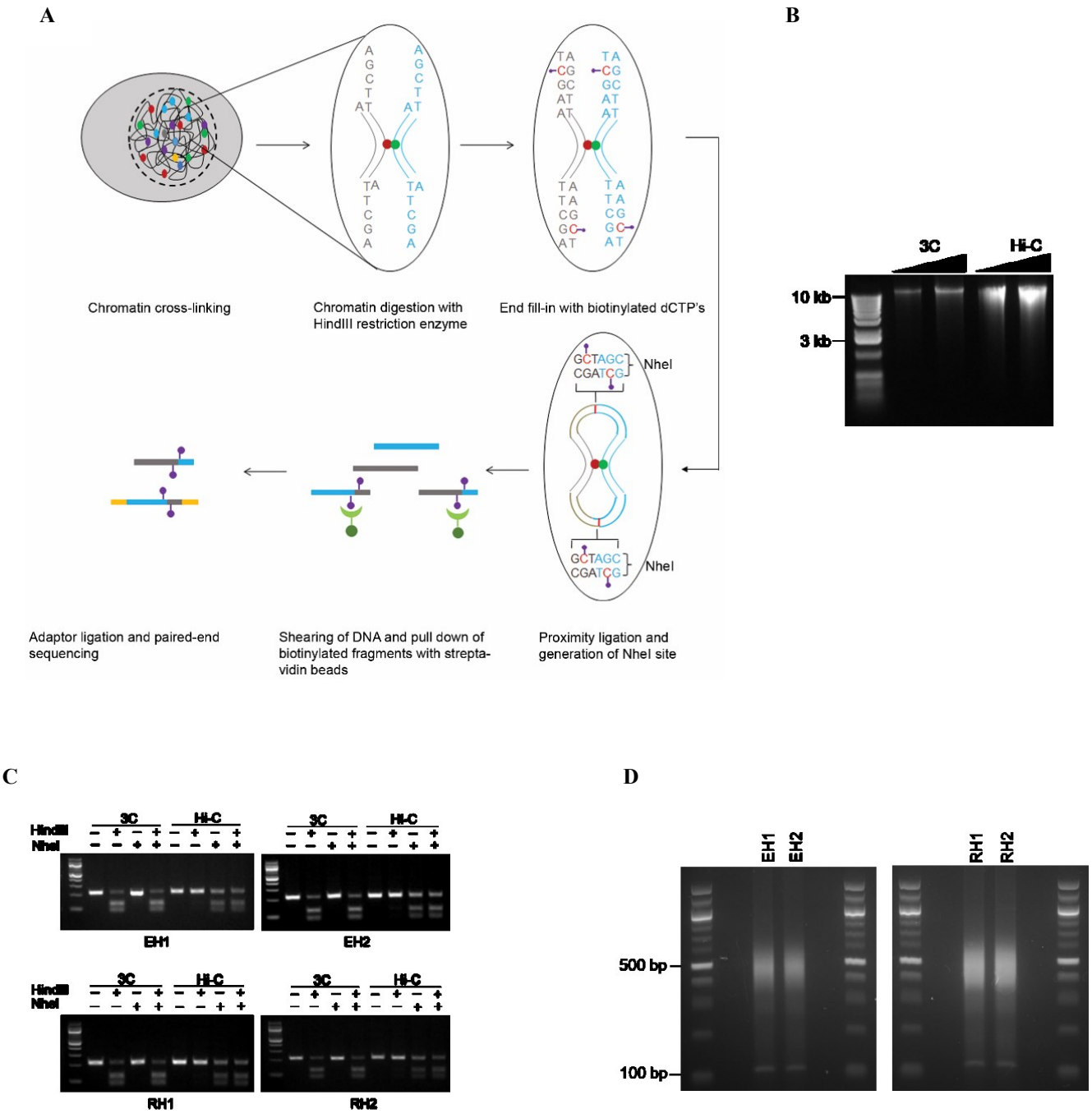


Figure 2.2.1. Quality control of *in situ* Hi-C libraries. (A) Schematic representation of *in situ* Hi-C method. **(B)** Integrity of *in situ* Hi-C libraries was monitored by resolving equal amounts of 3C (Chromosome Conformation Capture) control and *in situ* Hi-C template (100 ng) on a 0.8% agarose gel. Both the templates were visualized as a tight band of a size larger than 10 kb. **(C)** End fill-in efficiency of pre-pro-B (EH1, EH2) and pro-B (RH1, RH2) cells *in situ* Hi-C libraries was monitored by amplifying a ligation junction formed by two HindIII restriction sites (separated by a distance of 5 kb) in *Ercc3* locus. Unlike 3C sample, in *in situ* Hi-C library successful fill-in and ligation of HindIII sites creates NheI site, which is used to assess end fill-in efficiency of Hi-C libraries. **(D)** PCR amplification (15 cycles) of multiplexed *in situ* Hi-C libraries was performed using Illumina PE 1.0 and PE 2.0 primers.

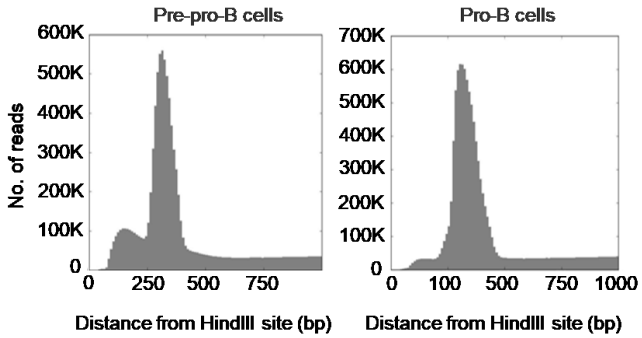
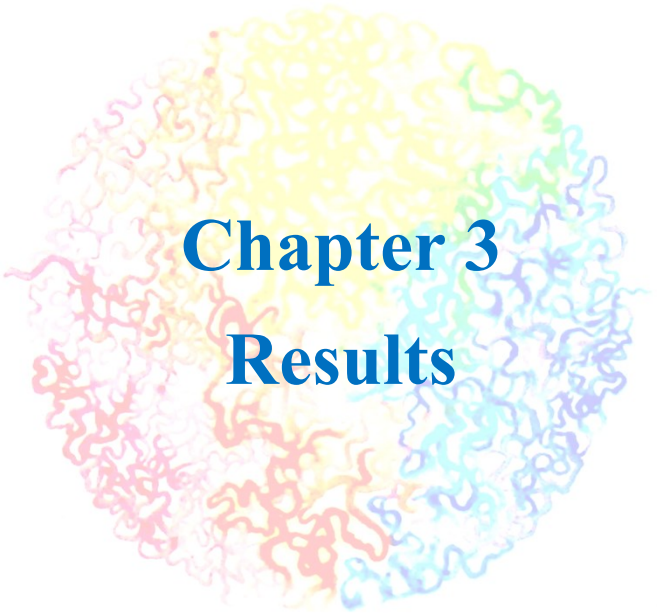


Figure 2.4.1. Quality control of *in situ* Hi-C libraries. Distribution analyses of *in situ* Hi-C paired end reads in the genome with respect to HindIII site, where majority of the paired-end reads are within the distance of 500 bp.



Chapter 3
Results

3.1 Differential chromatin compartmentalization promotes the B lineage gene expression program

To determine programmatic changes in chromatin organization during B cell development, we performed *in situ* Hi-C **(materials and methods)**, a high-throughput molecular approach (Nagano et al., 2013; Rao et al., 2014) that captures genome-wide chromatin interactions, using *Ebf1*-/- progenitors that represent the pre-pro-B cell stage (Pongubala et al., 2008) and *Rag2*-/- cells that represent the pro-B cell stage. The *in situ* Hi-C approach is similar to the previously described dilution Hi-C method (Lieberman-Aiden et al., 2009), except that the reactions: chromatin crosslinking, restriction enzyme digestion (HindIII), fill-in of 5'overhangs and ligation of chromatin ends present in close proximity, were performed in intact nuclei (Nagano et al., 2013). The Hi-C libraries were generated from both pre-pro-B and pro-B cells and then subjected to paired-end sequencing. Following high-throughput sequencing, the uniquely aligned (reference genome mm10) raw-reads were extensively filtered to eliminate various systemic biases originating from experimental procedures and intrinsic properties of the genome (fragment length, GC content and mappability). For this, we employed hiclib that implements filtering at multiple levels to determine the corrected contact counts (Imakaev et al., 2012) **(materials and methods)**. This approach has been known to selectively highlight the specific contacts and to facilitate the generation of corrected relative contact probability matrices, which are critical for determination of changes in chromatin architecture between the two different cell types. Thus, in comparison with similar studies (Lin et al., 2012), our strategy has two major advantages. First, *in situ* Hi-C captures specific DNA-DNA proximity ligations compared to dilution Hi-C (Nagano et al., 2013; Rao et al., 2014). Second, the ICE (Iterative Correction and Eigen vector decomposition implemented by hiclib) approach significantly reduces the frequency of spurious contacts and permits fair comparison of chromatin interactome data between pre-pro-B and pro-B cells.

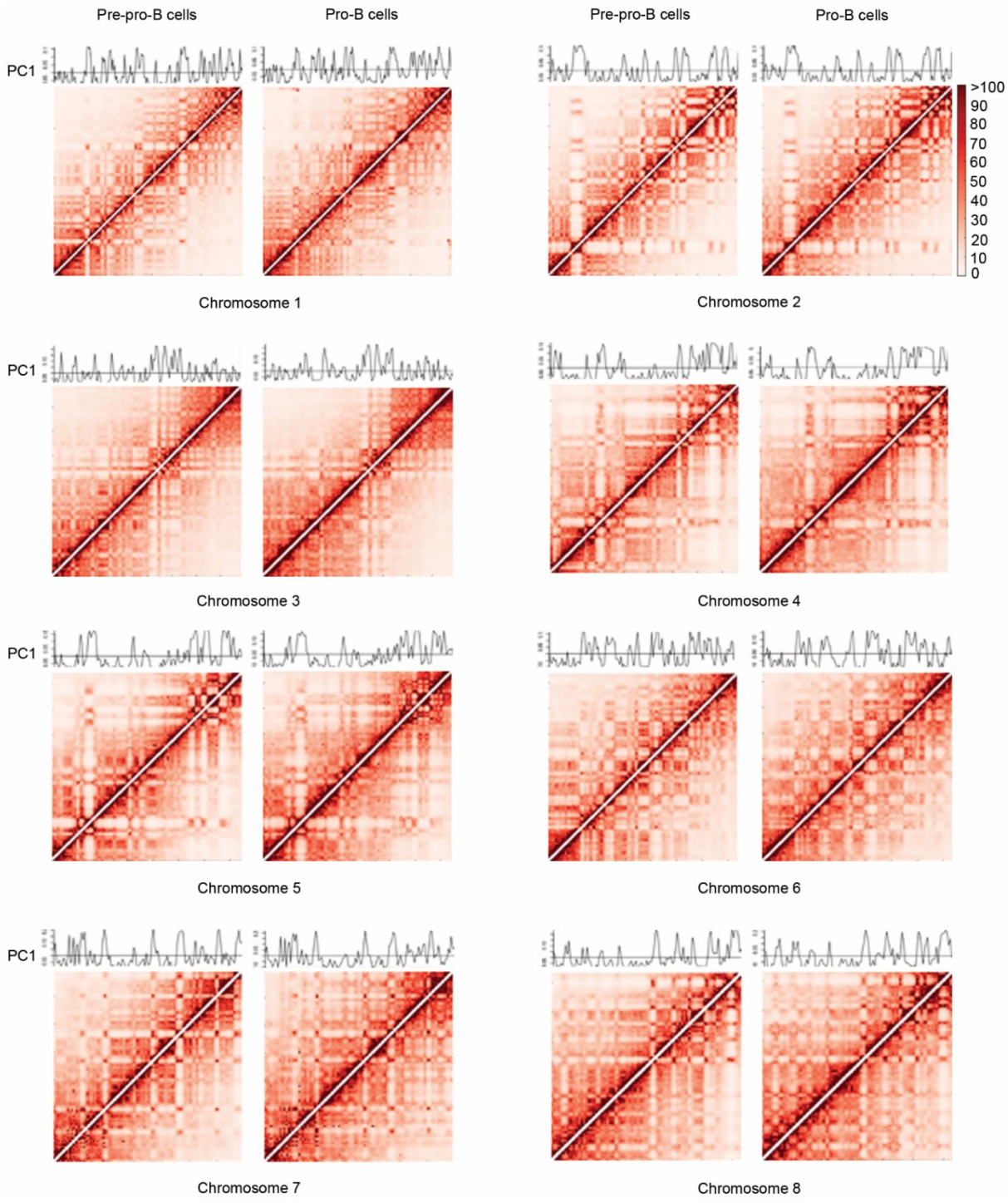
To gain a comprehensive understanding of progressive changes occurring in intra-chromosomal (*cis*) interactions between pre-pro-B and pro-B cells, iteratively corrected contact maps for each chromosome were generated at 1 Mb resolution **(Figure 3.1.1)**. Our analyses captured many of the previously identified long-range chromatin interactions (Kieffer-Kwon et al., 2013; Palstra et al., 2003), indicating that the *in situ* Hi-C approach was performed under optimal conditions and the captured interactions are valid *in vivo* **(Figure 3.1.2A and B)**. In line with

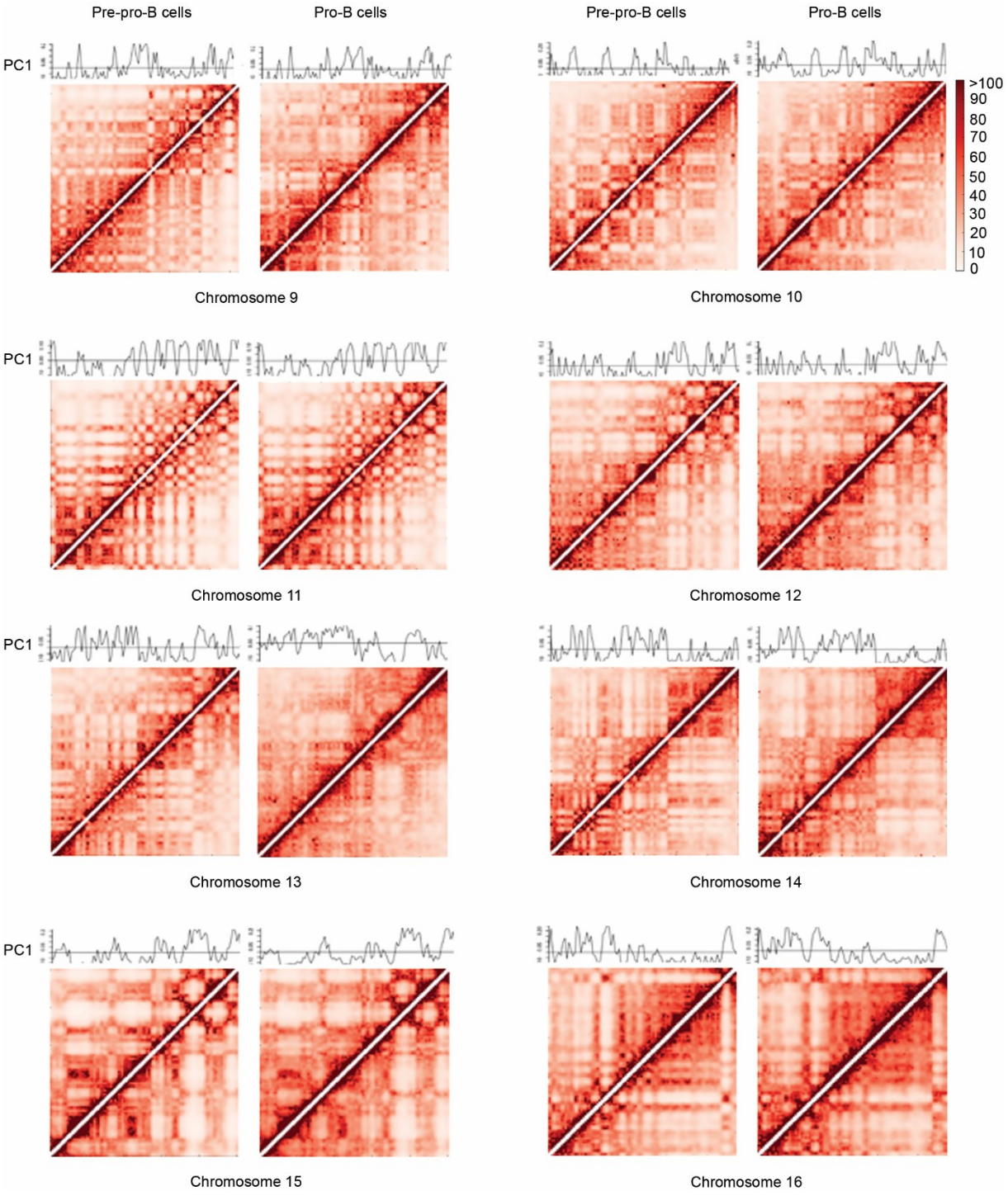
previous studies (Naumova et al., 2013), relative contact probability maps showed an ordered, dense pattern of varying sized blocks spanning across the diagonal **(Figure 3.1.3A; Figure 3.1.1)**. The majority of the interactions (60.0%) were limited to a range of 1-3 Mb and the frequency of such interactions decreased gradually with increasing linear genomic distance. In order to understand the differences in chromatin interaction patterns between the two cell types, we implemented Principal Component Analysis (PCA) at 1 Mb resolution **(materials and methods)**. As expected, these analyses revealed that chromatin is segregated into A or B compartments, which are defined by enriched or minimal interactions respectively **(Figure 3.1.3A; Figure 3.1.1)**. A compartments were found to contain a higher number of genes **(Figure 3.1.3B)** with an increase (4 fold) in CpG islands than the B compartments. Accordingly, the A compartments were substantially enriched for active histone modifications (H3K4me3, H3K4me1 and H3K9/14ac) (Lin et al., 2012) **(Figure 3.1.3C)** and displayed higher transcript levels when compared to the B compartments **(Figure 3.1.3D)**, indicating that chromatin compartmentalization mirrors gene activity in both cell-types.

To investigate the possibility that selective changes in chromatin compartmentalization provide a structural framework for B-lineage gene expression (Kosak et al., 2002; Lieberman-Aiden et al., 2009), we performed PCA analysis at a higher resolution (100 kb). From these analyses, we were able to define the chromatin state of a total number of 22,360 common genes that were captured by *in situ* Hi-C in both the cell types. Of these, 16,045 genes in pre-pro-B cells and 16,643 genes in pro-B cells were found to be present in A compartments, whereas 6,315 genes in pre-pro-B cells and 5,717 genes in pro-B cells were found to be present in B compartments. Further examination of these common genes between pre-pro-B cells and pro-B cells revealed three distinct classes, including a common set of genes that are localized in either A (Group I; 68.44%) or B (Group II; 22.25%) compartments in both cell types. Consistent with previous observations (Lin et al., 2012), although a major fraction (90.69%) of genes remained in the same compartment (Group I or II) in both cell types, a distinct set of genes (Group III; 9.31%) switched between A and B compartments. Of these, 1,339 (5.98%) genes transitioned from B to A compartment, while, 741 (3.31%) genes relocalized from A to B compartment during differentiation of pre-pro-B cells to pro-B cells **(Table 1 and 2)**. These observations demonstrate

that B cell developmental progression from a multipotential progenitor to a specified state encompasses notable changes in chromatin compartmentalization.

In order to test whether the differential chromatin compartmentalization is associated with B-lineage specific gene expression pattern, we compared the abundance of nascent transcript levels as determined by RNA-Seq (GSE52450) of Group III genes in pre-pro-B cells and pro-B cells. We observed that the genes, which switch from the B compartment to the A compartment during differentiation, displayed higher transcript levels in pro-B cells **(Figure 3.1.3E, left panel)**. For instance, *Satb2* **(Figure 3.1.3F)**, *Tead1*, *Pou2af1* and *Tlr4* that are essential for B cell development (Dobreva et al., 2003; Laurenti et al., 2013) are re-localized from the B compartment to the A compartment during pre-pro-B to pro-B cell transition **(Table 1)**. Likewise, genes that relocate from the A compartment to the B compartment displayed lower transcript levels in pro-B cells **(Figure 3.1.3E, right panel)**. Notably, genes that are associated with multipotent progenitors such as *Satb1* **(Figure 3.1.3G)**, *cKit* and *Cd34* as well as key alternate lineage determinants such as *Gata3*, *Zbtb16*, *Klf4*, *Vav3* and *Sox6* are found to be relocated to the B compartment in pro-B cells **(Table 2)**. In comparison with pre-pro-B cells, a significant number of genes within the chromosomes 10, 11 and 16 switch from the B compartment to the A compartment. Similarly, genes that are located in chromosomes 6 and 7, switch from the A compartment to the B compartment in pro-B cells **(Figure 3.1.4)**. Interestingly, our studies reveal that majority of functionally important B-lineage-specific genes (*Ebf1*, *Pax5*, *Foxo1*, *IRF4*, *IRF8*, *Cd79a*, *Cd79b* and *Cd19*) are localized in A compartments in both cell types. However, some of the key alternate lineage genes (*Gata3*, *Zbtb16*, *Klf4*, *Vav3* and *Sox6*) switch to the B compartment in pro-B cells. Thus, these observations indicate that relocalization of alternate lineage genes into B compartments is closely associated with their transcriptional repression. Collectively, our studies demonstrate that switching of selective genomic loci between A and B compartments is closely associated with the B lineage-specific gene expression pattern. However, these studies cannot rule out the possibility that chromatin relocalization and its associated changes may be a result of alteration of transcription.


Pre-pro-B cells
Pro-B cells
Pre-pro-B cells
Pro-B cells
PC1
>100
90
80
70
60
50
40
30
20
10
0
Chromosome 1
Chromosome 2
PC1
Chromosome 3
Chromosome 4
PC1
Chromosome 5
Chromosome 6
PC1
Chromosome 7
Chromosome 8


Pre-pro-B cells
Pro-B cells
PC1
>100
90
80
70
60
50
40
30
20
10
0
Chromosome 9
Chromosome 10
Chromosome 11
Chromosome 12
Chromosome 13
Chromosome 14
Chromosome 15
Chromosome 16

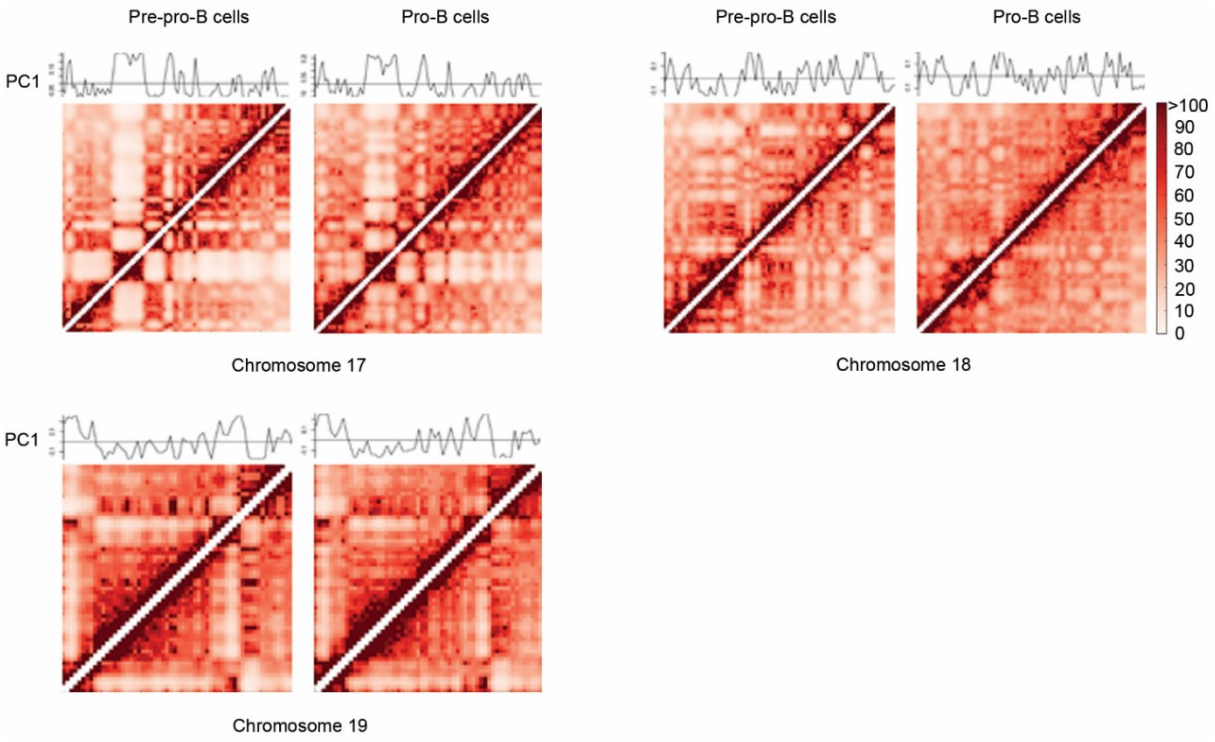


Figure 3.1.1. Contact count matrices of chromosomes (1-19, X). Iteratively corrected intra-chromosomal contact count matrices binned at 1 Mb resolution overlaid with PC1 for both pre-pro-B and pro-B cells. Principal Component Analysis (PCA) is implemented to determine the chromatin state of genomic loci.

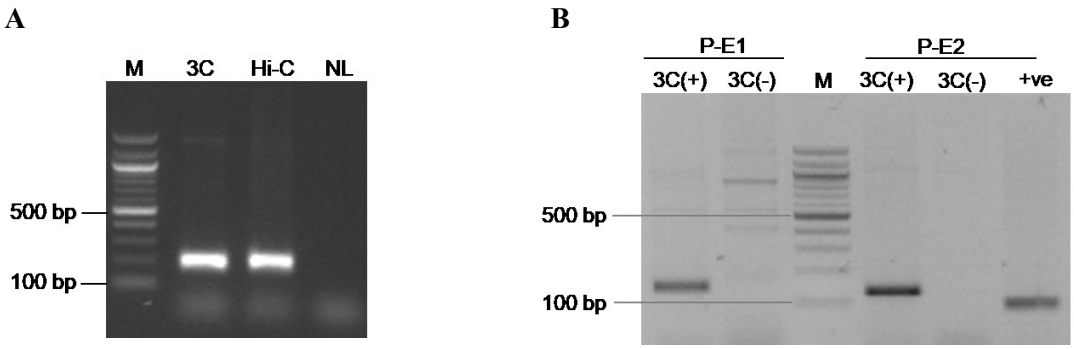


Figure 3.1.2. Validation of *in situ* Hi-C interactions by 3C analysis. (A) Amplification of a ligation junction formed by two HindIII restriction sites (separated by a distance of 5kb) in *Ercc3* locus, which is a ubiquitously expressed gene and has been shown to have identical higher order chromatin architecture in all mouse tissues. When compared to NL (No ligation control), 3C and Hi-C templates have resulted in amplification of specific ligation product. **(B)** In pro-B cells, long-range interactions between *Myc* promoter and its distal putative enhancers (E1, E2) located 700 kb and 800 kb downstream to the promoter were detected by PCR amplification using 3C template (+ve= *Ercc3* control).

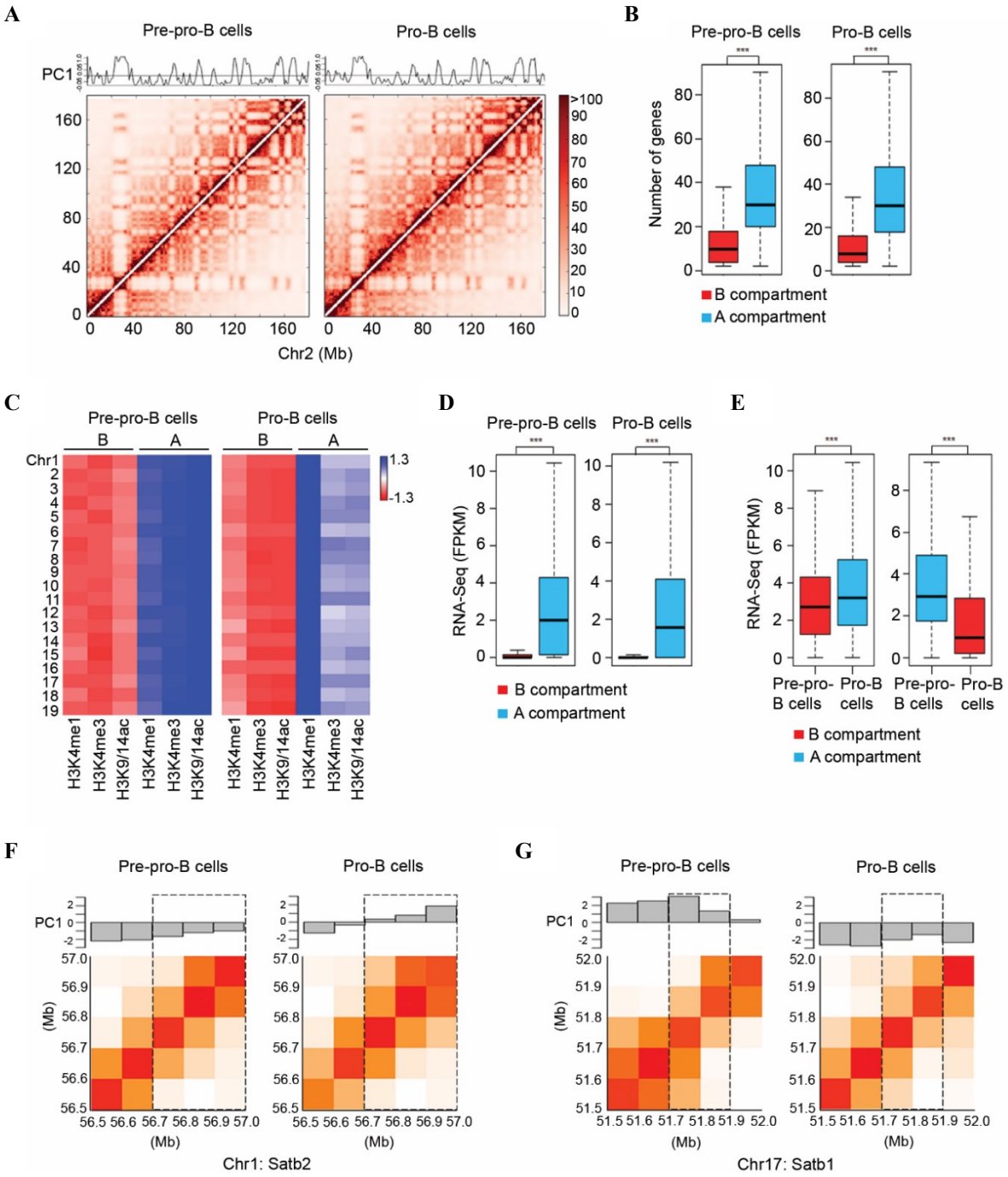


Figure 3.1.3. Chromatin compartmentalization is closely associated with gene activity. (A) Iteratively corrected intra-chromosomal contact count matrix of chromosome 2, representing the frequency of interactions at 1 Mb resolution. The first principal components (PC1) indicate the chromatin state on a linear genomic scale. **(B)** Distribution of genes in A and B compartments for both pre-pro-B and pro-B cell types (***P<0.001). **(C)** A and B compartments that are defined by PC1 were integrated with active methylation marks (H3K4me1, H3K4me3 and H3K9/14ac). Normalized heat-maps were generated by employing Matrix2png. Rows represent individual chromosomes, whereas the columns represent normalized count of respective methylation mark. **(D)** Comparative analysis of transcript levels of genes, based on RNA-Seq, present in A and B compartments (***P<0.001) for both pre-pro-B and pro-B cells. **(E)** Comparative analysis of transcript abundance of genes that relocate from B to A compartment (left panel) and A to B compartment (right panel) during differentiation of pre-pro-B cells into pro-B cells (***P<0.001). **(F, G)** Iteratively corrected contact count matrices derived from genomic regions comprising *Satb2* (chr1) and *Satb1* (chr17) for both pre-pro-B and pro-B cells. The PC1 values indicate the chromatin state of respective genomic loci. Dotted boxes represent genomic regions of *Satb2* and *Satb1*.

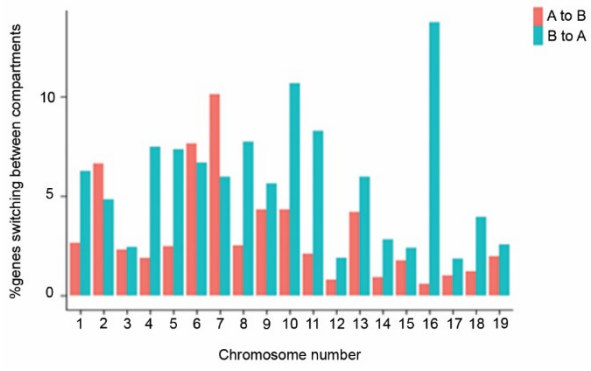


Figure 3.1.4. Percentage of genes that transit between A and B compartments during B cell fate commitment. Histogram representing percentage of genes that switch from A to B and B to A compartments in pro-B cells as compared to pre-pro-B cells. For each chromosome, percentage was calculated based on number of genes that transit from A to B or B to A compartments in relation to total number of genes captured in respective chromosome.

Table 3.1.1. List of differentially expressed genes that are switching from B to A compartments during progression of pre-pro-B cells to pro-B cells.

Chr	Gene	Pre-pro-B cells RNA-Seq	Pro-B cells RNA-Seq
1	ANGPTL1	0	1.06936
1	GM16701	0	2.39649
1	RALGPS2	0.299642	173.374
1	CD55	0.0145898	4.70994
1	9130024F11RIK	0.229238	51.4038
1	SATB2	0.292912	18.2528
1	DST	0.112248	6.17652
1	FAM78B	0.0869397	4.5675
1	BC094916	0.185617	8.45294
1	IFI204	0.258969	9.74153
1	PLXNA2	0.0564396	1.46354
1	4930558J18RIK	0.464516	7.6112
1	PYHIN1	0.902505	12.1828
1	RDH10	5.3855	60.044
1	D1ERTD622E	3.13591	16.4703
1	HSD17B7	6.4267	32.8848
1	3110045C21RIK	0.754319	3.60197
1	PLEKHM3	1.35806	5.07097
1	KIFAP3	9.67066	17.6465
1	1700066M21RIK	6.98649	11.8242
1	POGK	8.43349	13.8726
1	RRP15	12.3287	19.0019
2	ACVR1	0	2.13155
2	PKP4	3.65601	19.2274
2	AURKA	10.583	54.2846
2	DUT	37.2818	188.959
2	FAM98B	6.7997	21.3616
2	TASP1	4.81887	13.7607
2	ARHGAP11A	20.9767	50.0693
2	RIF1	35.9333	84.1126
2	CCDC34	10.5842	22.8999
2	LNP	10.0846	20.9691
2	SLC30A4	25.4202	50.1709
2	NDUFAF5	5.71713	10.5179
2	BLOC1S6	11.6369	18.9454
2	EIF3M	79.6491	128.758
2	ZC3H15	45.7782	73.6512
2	DPH6	11.0244	17.2151
2	ARHGAP15	77.7094	121.093
2	FAM210B	3.19571	4.87643

2	API5	51.4205	77.4806
3	FREM2	0	1.52605
3	WLS	0.398844	69.5851
3	BCAR3	0.062197	7.65024
3	LHFP	0.0887308	8.87904
3	GRIA2	0.236562	4.27381
3	CAMK2D	31.7348	355.765
3	MAP9	0.440882	3.71019
3	VCAM1	0.145905	1.02924
4	DOCK7	0.069132	8.9707
4	NFIA	0.0267694	1.62087
4	PRKAA2	0.0402403	2.09493
4	FGGY	0.0743661	2.30948
4	AKAP2	0.0948782	1.43187
4	TLR4	3.07086	16.6325
4	SMC2	37.4438	130.426
4	USP1	32.3709	70.0155
4	IFT74	7.24771	13.1577
4	NBN	13.0132	23.061
4	TMEM64	16.7031	28.2296
4	ALG6	9.2245	15.5629
5	9330182L06RIK	0.0275777	6.89013
5	PKD2	0.24846	4.47311
5	TGFBR3	0.128204	1.11655
5	FBXL5	36.5019	273.865
5	ZFP11	0.446974	2.41646
5	NCAPG	20.3712	104.363
5	RFC3	16.2787	70.2803
5	SMIM20	9.09848	35.7554
5	ABCG3	3.70231	13.686
5	GBP9	2.31449	7.74641
5	HSD17B11	7.99761	24.3505
5	1600023N17RIK	1.3886	3.90296
5	ZCCHC4	4.7281	12.1723
5	SLC2A9	2.44615	6.11988
5	ATP10D	2.51542	5.91666
5	SPP1	3.16974	6.72129
5	CLOCK	5.79372	10.933
5	TBC1D19	10.9286	19.9124
5	GBP4	4.46603	7.05774
6	ICA1	0	2.8522
6	GLCCI1	0.497737	103.701
6	PON3	0.182468	16.7568
6	MAGI1	0.0273598	1.73813
6	A430035B10RIK	0.164879	6.27741
6	VOPP1	0.115307	2.33803
6	FAM188B	0.171605	2.02898
6	PON2	5.81647	36.4301

6	CXCL12	0.617593	2.92478
6	C1GALT1	28.253	122.589
6	FKBP9	0.397581	1.16786
6	MDFIC	4.96495	12.471
6	ACN9	3.73107	9.24041
6	NDUFA5	73.1752	163.054
6	LANCL2	9.52048	16.4544
7	GP2	0	2.48296
7	UMOD	0.0404613	21.8918
7	TEAD1	0.0158901	3.51603
7	APBB1	0.376543	2.80745
7	LIG1	5.62672	30.8381
7	GAS2	1.89259	7.42244
7	6330408A02RIK	0.561942	1.39736
7	UBA2	56.0173	102.123
7	ZFP667	4.83619	7.68245
8	MFHAS1	0.0624261	1.84316
8	LIG4	1.04098	25.1055
8	CERS4	1.03676	8.08712
8	SHCBP1	13.045	60.0003
8	FRG1	47.6458	87.3816
8	TMEM184C	37.881	62.9321
8	INTS10	18.7447	30.4855
9	MMP10	0	1.28893
9	B430319G15RIK	0	2.64426
9	HMGCLL1	0.0375491	29.9585
9	POU2AF1	1.62648	860.051
9	PLOD2	0.221905	66.9432
9	ELOVL4	0.0154332	2.43243
9	RYK	0.214252	32.0095
9	ACPP	0.0255067	3.3516
9	PLSCR4	0.028257	2.85966
9	TTK	9.72449	62.1997
9	PLSCR1	0.680931	2.2731
9	BCKDHB	5.87157	15.8775
9	ALKBH8	12.2273	25.7896
9	ANAPC13	73.8256	130.131
9	4930579K19RIK	1.81084	3.15536
9	4930526I15RIK	1.06918	1.71792
10	PCBP3	0	2.73672
10	BVES	0.0493622	5.50347
10	E2F7	0.764022	9.35019
10	ULBP1	2.05063	20.9842
10	RTKN2	1.56192	11.2687
10	CDK1	17.6828	62.3272
10	BTG1	28.4686	86.3335
10	SNORA33	82.1751	216.023
10	COL6A2	0.953132	2.03543

10	HACE1	9.04193	19.2505
10	GOPC	12.6797	26.6895
10	SGK1	2.6293	3.97881
11	PTTG1	15.0717	248.217
11	FAM64A	4.21323	24.5212
11	UBLCP1	1.74732	8.3601
11	FBXO48	3.92446	15.0105
11	2810021J22RIK	1.94791	6.836
12	RAPGEF5	0.026154	15.586
12	TRIB2	0.788095	305.568
12	LRRN3	0.0959885	26.9063
12	SGPP1	2.67393	35.394
12	IFI27L2A	10.8232	42.2453
12	ARL4A	10.8625	32.9231
12	FRMD6	0.514077	1.24221
12	TAF1B	13.2909	28.5593
13	RHOBTB3	0.0576545	6.24338
13	FOXC1	0.158433	11.8069
13	GM3604	0.665476	4.41484
13	DIP2C	1.64624	10.7312
13	ATXN1	0.795994	4.98909
13	NAIP7	0.587153	2.23975
13	ZFP759	4.22391	9.65919
13	GM5141	4.14997	8.02769
13	CCNH	33.4202	59.5897
14	BMPR1A	0.289121	36.7922
14	FAM213A	0.146767	9.13802
14	IL3RA	0.586168	1.10756
14	PSMD6	63.5297	108.446
15	LRRK2	0.0461664	7.63185
15	DEPTOR	0.172611	1.0442
16	SLC15A2	7.41606	19.8352
16	FSTL1	0.605659	1.3427
16	BC002163	78.7264	163.204
16	CMSS1	6.56424	13.3587
17	MTCL1	0.00974535	1.36504
17	THBS2	0.201017	2.0761
17	ATL2	15.618	102.186
17	2700099C18RIK	6.84493	36.3958
17	CYP1B1	0.2933	1.52604
17	AGPAT4	0.454192	1.52285
17	GABBR1	1.72764	4.5136
17	CDC42EP3	28.0787	56.8135
18	4930426D05RIK	0	2.06263
18	SPINKL	0	60.2009
18	KLHL14	0.0167426	6.18494
18	SPINK11	0.0936119	6.29432
18	ZFP521	0.085086	5.58952

18	PRELID2	1.51146	87.967
18	CAMK4	0.0310546	1.78986
18	STARD4	12.0258	35.4951
18	YTHDC2	7.47806	15.6333
19	RNLS	0.259824	1.33555
19	TCF7L2	1.18001	3.73683
19	UHRF2	21.1426	42.4151
19	MINPP1	9.83711	18.5564
19	FAM204A	18.9646	34.846
19	2700046G09RIK	0.787547	1.28295
19	FRA10AC1	11.5442	17.3886

Table 3.1.2. List of differentially expressed genes that are switching from A to B compartments during progression of pre-pro-B cells to pro-B cells.

Chr	Gene	Pre-pro-B cells RNA-Seq	Pro-B cells RNA-Seq
1	MARCO	2.04118	0
1	GM7694	2.29597	0
1	GM16897	2.77609	0
1	IL18RAP	87.1834	0.147636
1	IL18R1	200.723	0.54607
1	CD34	533.539	1.52474
1	NRP2	12.7006	0.0669168
1	RGS18	135.289	0.744447
1	NOS1AP	1.60549	0.0154627
1	KHDC1A	1.31913	0.02695
1	AGAP1	2.28352	0.0746374
1	SH3BP4	1.15841	0.0538187
1	TMEM14A	2.90688	0.504574
2	NFATC2	1.19276	0
2	ITGA8	1.46398	0
2	NOSTRIN	2.30518	0
2	MIR669J	64.3396	0
2	CERS6	29.2159	0.0660073
2	SCN3A	10.2744	0.0413349
2	GATA3	2.64488	0.0392445
2	FMNL2	3.92034	0.105547
2	LDLRAD3	7.59802	0.807013
2	IFIH1	6.02355	1.27477
2	SSFA2	11.6528	4.71055
2	ARL5A	47.3703	27.4379
3	GUCY1A3	5.16371	0.0085551
3	CPA3	10.4835	0.0295834
3	SLC25A24	10.8942	0.0473313
3	VAV3	7.02558	0.0436874
3	S100A11	71.1198	4.20063
3	SLC22A15	1.80787	0.525387
3	LMO4	62.1199	20.676
3	S100A10	228.8	97.2808
3	TPD52	23.7815	12.8416
3	TBCK	24.7898	15.6428
4	TMEM51	11.6976	0.0237655
4	AJAP1	2.80085	0.0152035
4	TNFSF8	5.05116	0.0328927
4	CLVS1	1.41831	0.04587
4	KLF4	1.89177	0.357962

5	CPEB2	1.53465	0
5	HS3ST1	7.94804	0
5	CLNK	11.3184	0
5	KIT	107.056	0.377379
5	NIPAL1	3.97433	0.0142303
5	CDS1	6.20893	0.0230933
5	RASGEF1B	4.53638	0.106392
5	FRY	4.51218	0.209581
5	ADAM22	1.8758	0.120937
5	CDK14	1.31599	0.210975
5	SCARB2	49.9996	24.5407
6	KLRC2	1.00817	0
6	KLRA10	1.06938	0
6	KLRA23	1.21279	0
6	KLRI1	1.26621	0
6	PRSS2	1.4639	0
6	ST8SIA1	1.71074	0
6	KLRA13-PS	2.22611	0
6	TES	54.3555	0.0687901
6	KLRK1	4.56162	0.0136367
6	KLRI2	12.979	0.0735156
6	KLRC1	7.1269	0.0582266
6	BCL2L14	1.61588	0.0142404
6	GPRIN3	28.6496	0.84163
6	EPHB6	2.1558	0.0765372
6	ZYX	20.5025	0.926055
6	HPGDS	6.41161	0.468624
6	PLEKHA5	8.5933	3.6728
6	TPK1	11.5343	5.29055
7	MRGPRA1	1.24768	0
7	NCR1	1.89567	0
7	MCTP2	51.9241	0.118911
7	CD163L1	40.4187	0.13421
7	5830411N06RIK	154.707	0.656541
7	SIGLECH	54.2745	0.235825
7	ADAMTS17	1.06831	0.0060798
7	PLEKHA1	58.0944	0.383194
7	TSHZ3	14.1182	0.105217
7	SOX6	2.66652	0.0449762
7	PTPRE	81.3402	1.81641
7	PRCP	18.6812	3.54021
7	TUBGCP5	81.9667	25.7299
7	1600014C10RIK	32.7068	11.89
7	PEPD	14.4153	8.14617
8	KBTBD11	7.69045	0.0134392
8	NRP1	17.0582	0.135697
8	ZFHX3	4.03586	0.0582646
8	RNF150	8.95391	0.183406

8	CDYL2	10.5227	2.36429
9	TMEM158	1.77939	0
9	EOMES	9.95832	0.0248923
9	ZBTB16	15.8803	0.0407465
9	KCNJ1	13.5733	0.0385854
9	DMXL2	2.34928	0.0309904
9	OAF	11.4225	0.411411
9	SLC9A9	34.0744	1.63487
9	PIK3CB	2.51959	0.195365
9	EEPD1	5.47207	0.439633
9	ANXA2	8.00882	0.912337
9	SIAE	33.0719	5.02867
9	MORF4L1	4.37126	0.753983
9	CRTAP	12.9342	3.63079
9	SLC35F2	4.80676	1.37065
9	TBC1D2B	6.50961	2.0249
9	STT3B	92.3649	31.3096
9	GLB1	24.4183	8.83434
9	RAB39	2.67286	1.00051
9	TMPPE	6.36963	3.8339
10	4930444F02RIK	1.09358	0
10	LILRB4	3.17261	0
10	PPM1H	8.75861	0.0136044
10	SLC16A7	6.08008	0.0213333
10	PERP	4.11309	0.0283578
10	HMGA2	44.0214	0.327775
10	ESR1	2.27757	0.0409784
10	GP49A	2.78555	0.0567702
10	CEP85L	16.4446	0.819036
10	SYNE1	16.5741	2.15255
10	EPM2A	2.09075	0.286647
10	NHSL1	2.76631	0.994897
11	TTYH2	8.21671	0.151304
11	SLC39A11	6.771	1.91315
11	SLC36A1	3.92703	1.78595
11	DOCK2	83.581	51.9097
12	9130015A21RIK	1.87208	0
12	NRCAM	2.24995	0
12	AHR	11.671	0.11327
12	ARHGAP5	27.1824	0.696551
12	EGLN3	13.7683	0.45998
12	HPCAL1	7.7666	0.489471
12	DAAM1	4.39215	1.14916
12	ADCK1	9.97327	3.74789
13	GM5086	1.23085	0
13	GZMK	2.38646	0
13	HSPB3	2.45689	0
13	LY86	36.3733	0.0380619

13	AOAH	16.1628	0.0307039
13	F13A1	9.72704	0.0265659
13	GCNT2	21.9358	0.138282
13	ESM1	3.05482	0.0222662
13	ITGA1	9.11274	0.0942137
13	EMB	94.7284	1.21409
13	CDK20	1.85404	0.0484605
13	CTSL	58.156	4.38644
13	AAED1	41.0997	3.50703
13	SNX18	22.129	4.98811
13	PELO	6.43427	1.9051
13	CDYL	13.3558	4.77256
13	ZFP369	33.5559	17.5875
13	ZFP65	11.6495	6.88813
13	ZFP738	14.3576	8.60192
13	A530054K11RIK	13.0466	7.86326
14	DOCK5	3.3284	0.0639458
14	ATP8A2	1.3262	0.0408612
14	FLNB	15.6466	2.17661
15	FAM134B	57.4188	0.112818
15	FYB	8.10207	0.0413415
15	TRIO	10.8631	0.873173
15	OXR1	80.3601	33.2368
15	LRP12	5.78029	2.70821
15	ANGPT1	3.15486	1.72543
15	FAM84B	38.7744	23.8794
15	ZFP622	21.8074	14.0309
16	GM4827	1.17068	0
16	ZDHHC23	2.48371	0
16	DOPEY2	4.52294	0.41575
17	KCNK12	1.16267	0
17	SLC22A3	18.5143	0.0459131
17	SATB1	198.831	1.9932
18	SETBP1	6.61077	0.110675
18	PPIC	5.64475	0.376841
18	MPP7	13.3922	6.3315
19	MPEG1	44.7289	0.204605
19	HECTD2	6.08728	0.0286052
19	GCNT1	30.3873	0.377118
19	DTX4	6.11693	0.431718
19	PIP5K1B	5.26781	0.895916
19	ATRNL1	6.07981	1.85466
19	TJP2	4.21562	1.52531
19	TMEM2	7.10639	2.64035
19	SORBS1	7.35667	3.55583
19	RFK	20.461	12.6326

3.2 Global analysis of topologically associating domains (TADs) during B cell specification

At sub-megabase level, A and B compartments of chromatin are organized into dense and contiguous self-interacting regions termed topologically associating domains, TADs (Dixon et al., 2012). These chromatin domains have been proposed to be stable and conserved across cell types, yet their intrinsic chromatin interactions were found to be varying (Dixon et al., 2015). This raised a possibility that changes in the interaction patterns within TADs may serve as a framework for differential gene activity and contribute to the developmental progression of the cell. In order to capture the changes in chromatin structure within these domains in pre-pro-B and pro-B cells, we employed domain caller software (Dixon et al., 2012) to identify TADs from iteratively corrected relative contact probability matrices generated at 40 kb resolution. Our analyses revealed that the genome of pre-pro-B cells is partitioned into a total of 2,008 TADs, whereas the genome of pro-B cells comprised of 1,810 TADs with a total genomic occupancy of 90.74% and 89.50%, respectively. Strikingly, we found that the median TAD size is higher in pro-B cells (920 kb) as compared to pre-pro-B cells (800 kb). Collectively, these studies provide the first indication that the structural organization of TADs may be subjected to alterations during developmental transition from pre-pro-B to pro-B cell stage.

To gain further insights into changes in structural organization of the chromatin at the sub-megabase level, we cross-compared the TADs between two cell types (pre-pro-B and pro-B cells) based on their linear genomic position. In line with the previous reports (Dixon et al., 2012), a substantial number of TADs, 1,023 (pre-pro-B cells: 50.9%, pro-B cells: 56.5%) were found to be stable, in both cell types **(Figure 3.2.1A)**. The remaining TADs (pre-pro-B: 985, pro-B: 787) exhibited re-organization in terms of their genomic positions and were categorized as ‘dynamic’. It is possible that the stable TADs may maintain persistent chromatin interactions and thus account for uniform gene activity between two cell types. Alternatively, chromatin regions within these stable TADs may be subjected to epigenetic modifications and concomitant changes in intra-molecular interactions, resulting in cell-type specific gene expression pattern. To investigate these possibilities, we have compared the transcript levels of genes present within the stable TADs between pre-pro-B and pro-B cells and observed significant differences in their activities. This differential gene expression pattern may possibly be due to alterations in intrinsic chromatin interaction landscape. To examine this, we have calculated Aggregation Preference (AP), a

parameter that quantitatively measures interaction patterns of TADs (Wang et al., 2015). During this analysis, the local high-frequency chromatin interactions, violating the distance-dependence decay principle, were measured and segregated according to their spatial aggregation by employing DBSCAN. The weighted density of clustered groups, defined as Aggregation Preference (AP), was used to quantitatively measure interaction patterns within each TAD. As expected, we found that TADs with higher AP values were comprised of high density chromatin interaction blocks in both cell types. Accordingly, TADs with higher AP values were enriched with active methylation marks (H3K4me1, H3K4me2 and H3K4me3) and displayed higher nascent transcript levels. On the other hand, TADs with low AP values displayed sparse chromatin interactions and were found to be depleted with active epigenetic marks **(Figure 3.2.1B)**. Furthermore, permissive TADs displayed higher AP values as compared to repressive TADs **(Figure 3.2.1C)**. Thus, AP values define transcriptional status and may serve as an appropriate measure of functional activity of TADs.

Interestingly, of the 1,023 stable TADs, a majority (867, 85%) of them displayed similar AP values (<0.2) between pre-pro-B and pro-B cells, indicating that the cumulative number of chromatin interactions with these TADs are comparable. Thus the differential gene expression pattern observed within stable TADs between the two cell types may possibly be attributed to the combinatorial changes in their promoter and *cis*-regulatory interactions. To test this, we first identified statistically significant (P<0.05) chromatin interactions in both cell types by implementing the spline-fit model (Ay et al., 2014) **(materials and methods)**. Next, these significant interactions were integrated with genome-wide epigenetic marks (H3K4me3, H3K4me1 and H3K4me2) to identify potential promoters and *cis*-regulatory interactions (Lin et al., 2010). Only those promoters located within close proximity ($\pm$ 2.5 kb) of transcription start sites (TSS) **(Figure 3.2.2)** and the *cis*-regulatory elements located $\geq$1 kb away from putative promoters, were considered for further analysis. In total, we have identified 30,150 and 46,263 potential interactions involving promoters and *cis*-regulatory elements in pre-pro-B and pro-B cells, respectively. We have mapped these promoter-*cis*-regulatory interactions to the stable TADs (867, 85%) with similar AP values. From these analyses, we found that 6,678 and 9,468 *cis*-regulatory interactions were associated with stable TADs in pre-pro-B and pro-B cells, respectively. Of these interactions, 1,715 promoters were found to be common in pre-pro-B and

pro-B cells, whereas 1,147 (48.8%) and 952 (57.1%) promoters were found to be unique in pre-pro-B and pro-B cells, respectively. The majority of these promoters (common and unique) were found to interact with developmental stage-specific enhancers **(Figure 3.2.3A)**. For instance, *Polg2*, which is highly expressed in pre-pro-B cells and *Cd79b*, which is induced at pro-B cell stage, are both located in a stable TAD with similar AP values. Interestingly, we observed that *Polg2* interacts with multiple enhancers (3) in pre-pro-B cells, whereas no such promoter and *cis*-regulatory interactions were captured in pro-B cells. Conversely, *Cd79b* promoter interacts with multiple enhancers (4) in pro-B cells, while only one such interaction was observed in pre-pro-B cells **(Figure 3.2.3B)**. These results suggest that although a substantial number of TADs are stable with respect to the genomic position, their intrinsic chromatin interactions involving *cis*-regulatory elements are dynamic. These intrinsic changes in chromatin interactions may have a limited effect on the structural maintenance of the TAD, but may be critical for sustaining the cell type-specific gene expression pattern.

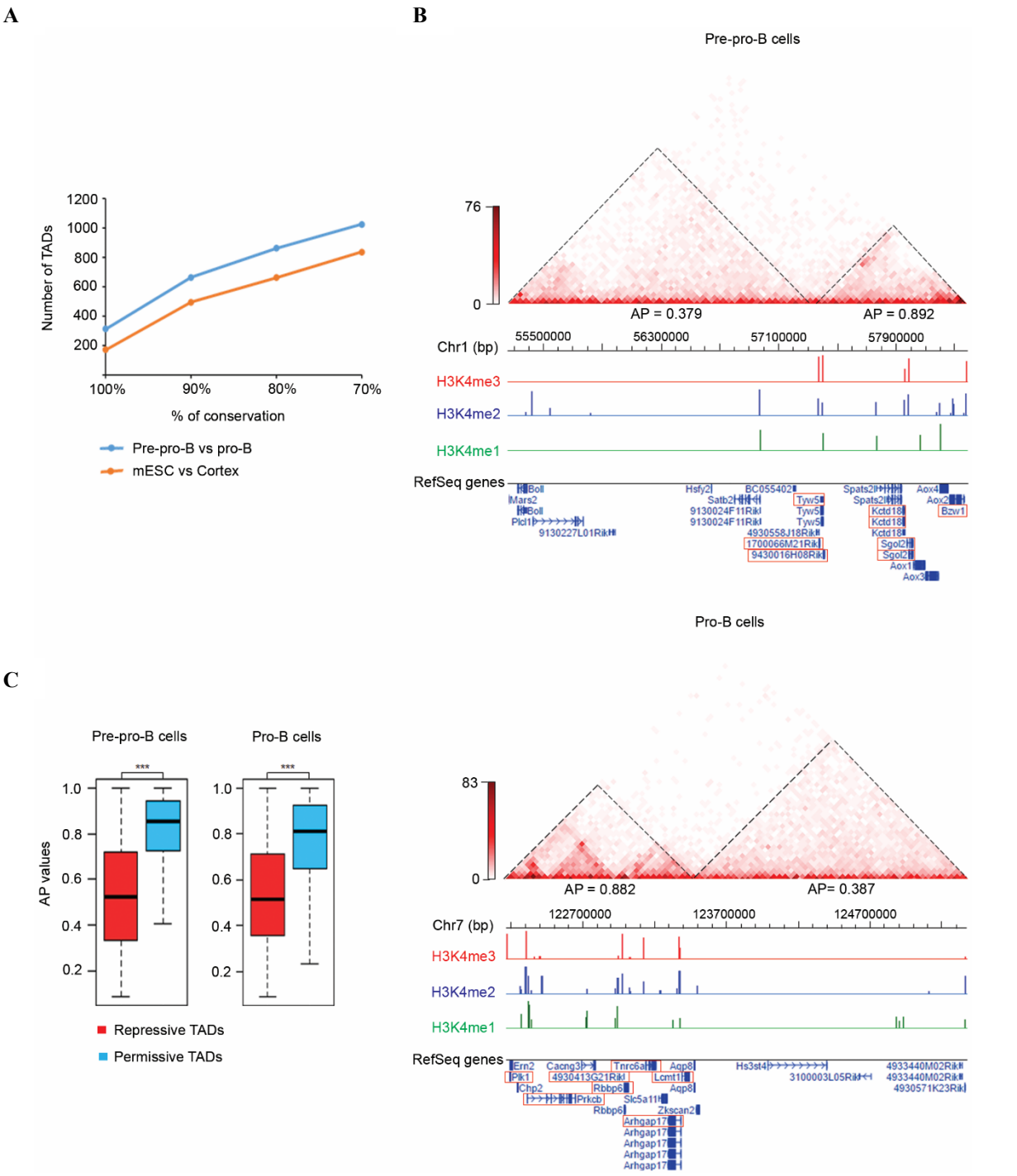


Figure 3.2.1. Aggregation preference is closely associated with chromatin state. (A) Line graph representing the number of TADs conserved between pre-pro-B and pro-B (blue) as well as mESC and cortex cells (orange). The X-axis indicates the percentage of genomic overlap between TADs in two different cell types, whereas Y-axis indicates the number of stable TADs. **(B)** Structural and functional analysis of TADs with variable AP values in pre-pro-B (upper panel, Chr1:55.24-57.320 Mb and 57.36-58.40 Mb) and pro-B cells (lower panel, Chr7:122.16-123.48 Mb and 123.48-125.40 Mb). TADs were mapped with active methylation marks (H3K4me3, H3K4me1 and H3K4me2) as determined by ChIP-Seq. Genes that are transcriptionally active are highlighted by red boxes. TADs are demarcated by dotted lines. **(C)** Box plots showing the relation between aggregation preference (AP) and chromatin state of respective TADs in both pre pro-B and pro-B cells (***P<0.001).

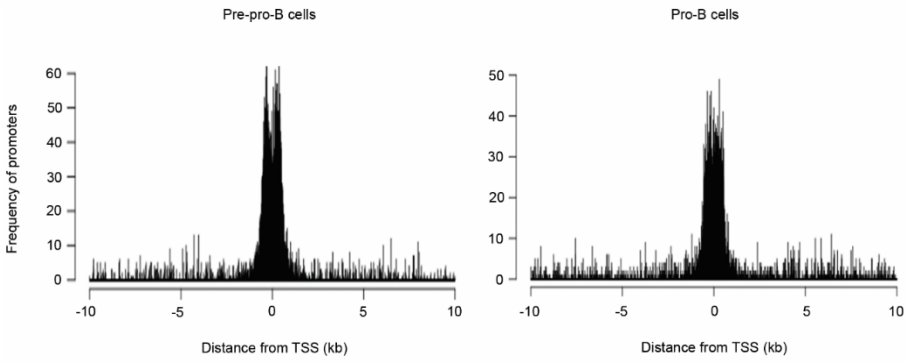


Figure 3.2.2. Genomic distance between transcription start site (TSS) and *in situ* Hi-C captured promoter elements. Distribution plot representing genomic distance between *in situ* Hi-C captured promoter elements and their respective transcription start site (TSS). Majority of the captured promoter elements (~90%) are located within + 1.0 kb of their TSS. Nevertheless, promoters that are located with maximum distance of + 2.5 kb from transcription start site were considered for further analyses. Promoters are identified by integrating *in situ* Hi-C interactome with ChIP-Seq data of H3K4me3. Transcription start sites are retrieved using UCSC genome browser.

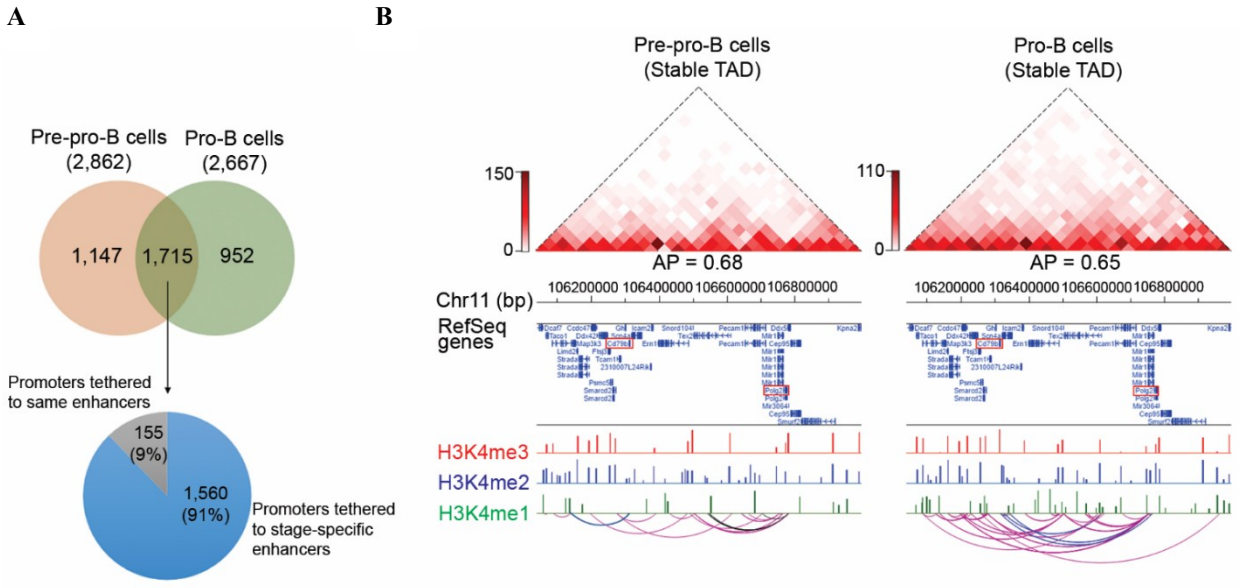


Figure 3.2.3. Figure 2. TADs are dynamic and undergo structural reorganization during early B cell development. **(A)** Venn diagram indicating the number of promoters interacting with cis-regulatory elements that are present in the stable TADs with similar AP values in both pre-pro-B and pro-B cells. The pie chart represents the number of common promoters tethered to same (grey) or cell type-specific (blue) enhancers. **(B)** Comparative analysis of promoter-*cis*-regulatory interactions between stable TADs with similar AP values spanning a genomic region (106.04–107 Mb) of chromosome 11. TADs are mapped with active epigenetic marks, H3K4me3 (enriched at promoter regions), H3K4me1 and H3K4me2 (enriched at enhancers) as determined by ChIP-Seq in both pre-pro-B and pro-B cells. TADs were defined by domain calling approach and are highlighted by dotted lines. The genomic positions of promoter-*cis*-regulatory interactions within the TADs are represented by arcs. *Polg2* (black) and *Cd79b* (blue) interactions are highlighted.

3.3 Structural reorganization of TADs corresponds to changes in *cis*-regulatory interaction landscape

While a substantial number (1,023) of TADs are constant in both pre-pro-B and pro-B cell types, a considerable number of TADs (pre-pro-B:985, pro-B:787) were found to be altered as indicated by changes in their genomic positions. We classified these altered TADs into two groups: unique and merged. Unique TADs are defined as those present in pre-pro-B cells but not in pro-B cells and *vice versa*. We propose that unique TADs (pre-proB:100, pro-B:65) may have been generated as a result of increased local genomic interactions to facilitate cell type-specific gene expression pattern **(Supplementary Table S3 and S4)**. Consistent with this assumption, key alternate lineage genes like *Ccl3*, *Serpini1* and *Vav3* that are highly expressed in multipotent progenitors, were found to be associated with TADs in pre-pro-B cells. On the contrary, in pro-B cells, these genes are located in the boundary regions **(Figure 3.3.1A)**. Besides the unique TADs, we observed that few larger TADs in pre-pro-B cells (110) partitioned into two or more minor TADs in pro-B cells. Conversely, two or more minor TADs in pre-pro-B cells coalesce into a larger 'merged' TAD in pro-B cells (183). We propose that merged TADs may have been formed as a result of increased inter-TAD interactions in pro-B cells**.** Accordingly, the normalized contact frequency of inter-TAD regions of minor TADs in pre-pro-B cells is significantly lower as compared to the counter regions of merged TADs in pro-B cells **(Figure 3.3.1B)**. These observations are in line with increased median TAD size (920kb) of pro-B cells, as compared to the size of pre-pro-B cells (800kb). Correspondingly, we observed a significant increase in the number (pre-pro-B cells: 30,150, pro-B cells: 46,263) as well as in the median distance (pre-pro-B cells: 298kb, pro-B cells: 330kb) between promoter-*cis*-regulatory interactions in pro-B cells. This raises the possibility that inter-TAD promoter-*cis*-regulatory interactions may contribute for re-organization of TADs. To test this, we have mapped promoter-*cis*-regulatory interactions of merged TADs in pro-B cells and compared to their counter TADs in pre-pro-B cells. We found a significant increase in inter-TAD promoter-*cis*-regulatory interactions (2,600) in pro-B cells as compared to those (1,570) in pre-pro-B cells **(Figure 3.3.2)**. Taken together, these analyses provide insight into the dynamic re-organization of TADs which is closely associated with changes in the *cis*-regulatory interaction landscape during developmental transition from pre-pro-B to pro-B cell stage.

To rigorously demonstrate the dynamic organization of TADs observed between pre-pro-B and pro-B cells, we employed two distinct strategies. First, we used Directionality Index (DI), which quantitatively measures the 'interaction bias' of a given genomic region (Dixon et al., 2012), as a parameter to detect the structural variations of TADs between the two cell types. Comparative DI analyses revealed that stable TADs display significantly higher correlation as compared to the dynamic TADs (merged and unique) **(Figure 3.3.3A)**. These observations suggest that, unlike stable TADs, dynamic TADs display dramatic structural alterations. Second, we built 3D models of merged and stable TADs to determine the changes in position order of chromatin using AutoChrom3D (Peng et al., 2013). Compared to conventional 3D modeling methods (Barbieri et al., 2012; Duan et al., 2010; Hu et al., 2013; Kalhor et al., 2012), AutoChrom3D employs a novel sequencing-bias-relaxed parameter to derive 3D chromatin models. Next, we compared the spatial distance between start and end regions of merged TADs in pro-B cells with their counter regions in pre-pro-B cells. We reason that, in pro-B cells, the ends of a merged TAD should be in close spatial proximity compared to their counter regions in pre-pro-B cells. Consistent with this supposition, we found that the spatial distance was significantly lower in pro-B cells as compared to pre-pro-B cells **(Figure 3.3.3B, left panel; Figure 3.3.4A-D)**. In contrast, no significant difference in the spatial distance was observed for stable TADs **(Figure 3.3.3B, right panel; Figure 3.3.4E).** Collectively, these analyses demonstrate that 3D models reflect the changes in 2D interaction maps. To validate these results, we performed 3C-qPCR (Chromosome Conformation Capture) for a merged TAD (Chr12:69720000-71160000) found in pro-B cells formed as a result of coalescence of three minor TADs in pre-pro-B cells. The spatial distance between start and end of this TAD is found to be lower in pro-B cells (2.41Å) compared to pre-pro-B cells (8.00Å) as shown by AutoChrom3D. Correspondingly, our 3C experiment using primers close to start (+1,615 bp) and end regions (-7,004 bp) of merged TAD revealed higher cross-linking frequency in pro-B cells than pre-pro-B cells **(Figure 3.3.5) (Supplementary Table S5)**. These results support our hypothesis that TADs undergo dynamic structural alterations as a result of changes in chromatin interaction patterns that may be important for transcription regulation.

To investigate if our findings could be extrapolated to other cell types, we compared the structural organization of TADs in mESC and cortex cells using publicly available data (Dixon et al., 2012). Similar to what we have observed with our cell types, the total number of TADs (mESC-

2085, cortex-1519) as well as their median size (mESC-880 kb, cortex-1.3 Mb) differed between mESC and cortex cells. Moreover, the comparative analysis of DI and relative contact probabilities suggests that TADs undergo structural reorganization between mESC and cortex **(Figure 3.3.6A and B)**. In concordance with our data for pro-B cells, we noticed that cortex cells have more number of merged TADs (269) than the mESCs (95). In comparison with the pluripotent cells (pre-pro-B and mESC), the observed increase in number of merged TADs and the associated increase in the average size of TADs in differentiated cells (pro-B and cortex) can be attributed to their compact chromatin organization (Gorkin et al., 2014; Meshorer et al., 2006). These findings are further supported by an increase in long-range interactions in pro-B cells compared to those in pre-pro-B cells. Consistently, a recent study suggests that the increase in TAD size as well as long-range interactions in sperm cells may be due to the dense packaging of its genome (Battulin et al., 2015). Collectively, these findings demonstrate that the differences in the TAD organization between various cell types are dependent on the differences in their long-range interactions and chromatin compaction.

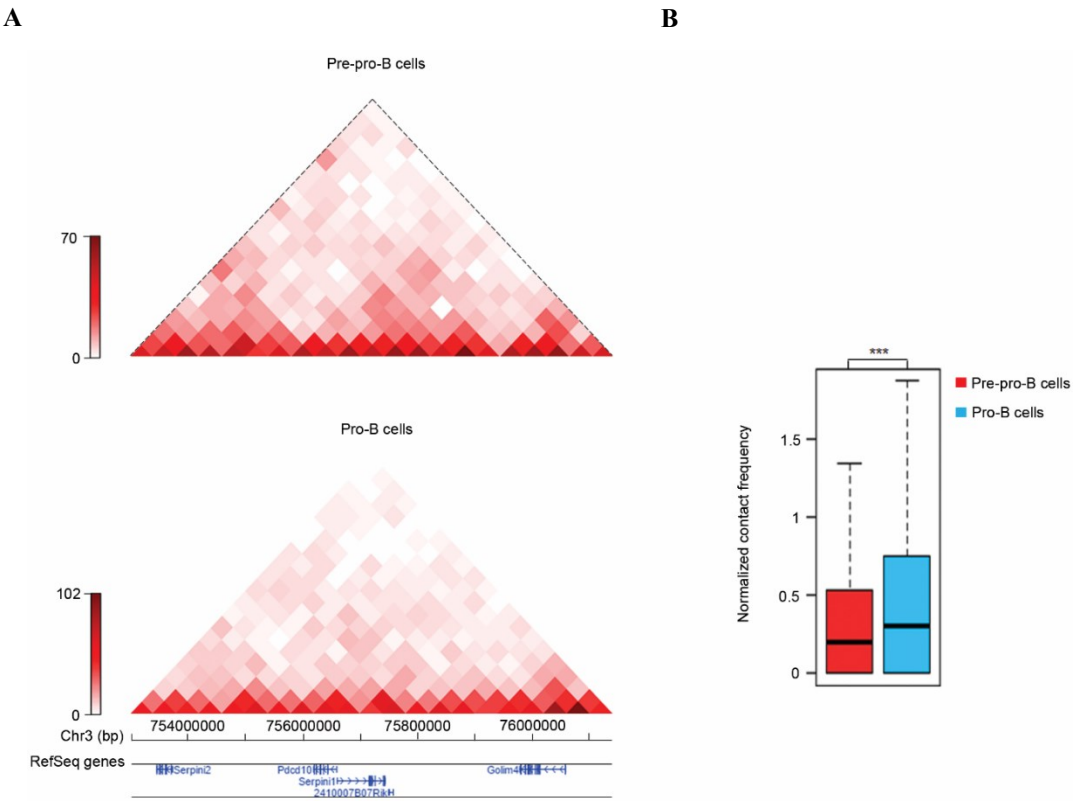


Figure 3.3.1. TADs undergo structural reorganization during early B cell development. (A) 2D interaction map of unique TAD (Chr3:75.2-76 Mb), present in pre-pro-B cells but not in pro-B cells, as defined by domain caller. Unique TAD is highlighted by the dotted line. **(B)** Comparative analysis of normalized contact frequencies between inter-TAD regions of minor TADs in pre-pro-B cells and their counter regions of merged TADs in pro-B cells (***P<0.001).

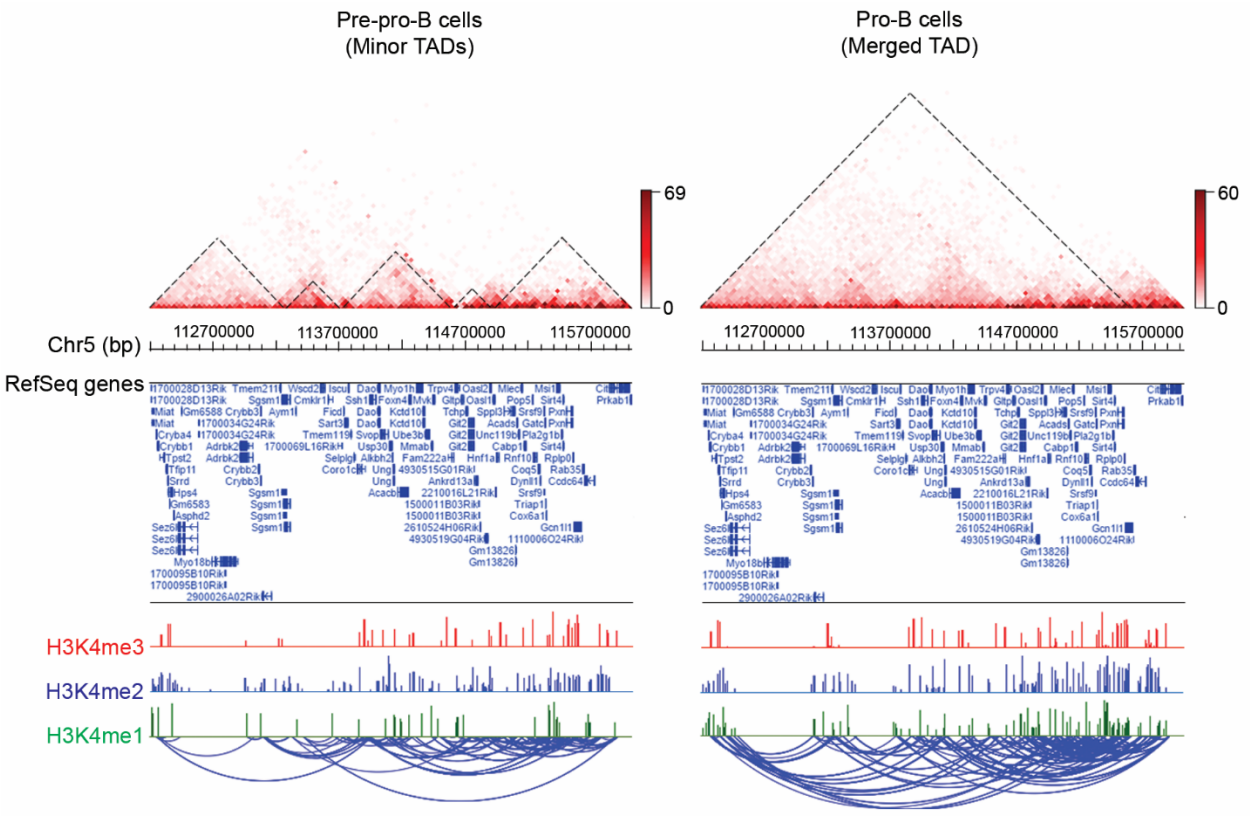


Figure 3.3.2. Structural reorganization of TADs is closely associated with changes in Promoter-*cis*-regulatory interactions. Comparative analysis of promoter-*cis*-regulatory interactions between merged TAD (pro-B cells) and its counter TADs (pre-pro-B cells) spanning the genomic region (112.20-115.64 Mb) of chromosome 5. TADs are demarcated by domain calling approach and highlighted by dotted lines. TADs are mapped with active epigenetic marks, H3K4me3 (enriched at promoter regions), H3K4me1 and H3K4me2 (enriched at enhancers) as determined by ChIP-Seq in both pre-pro-B and pro-B cells. Promoter-*cis*-regulatory interactions are represented by blue arcs.

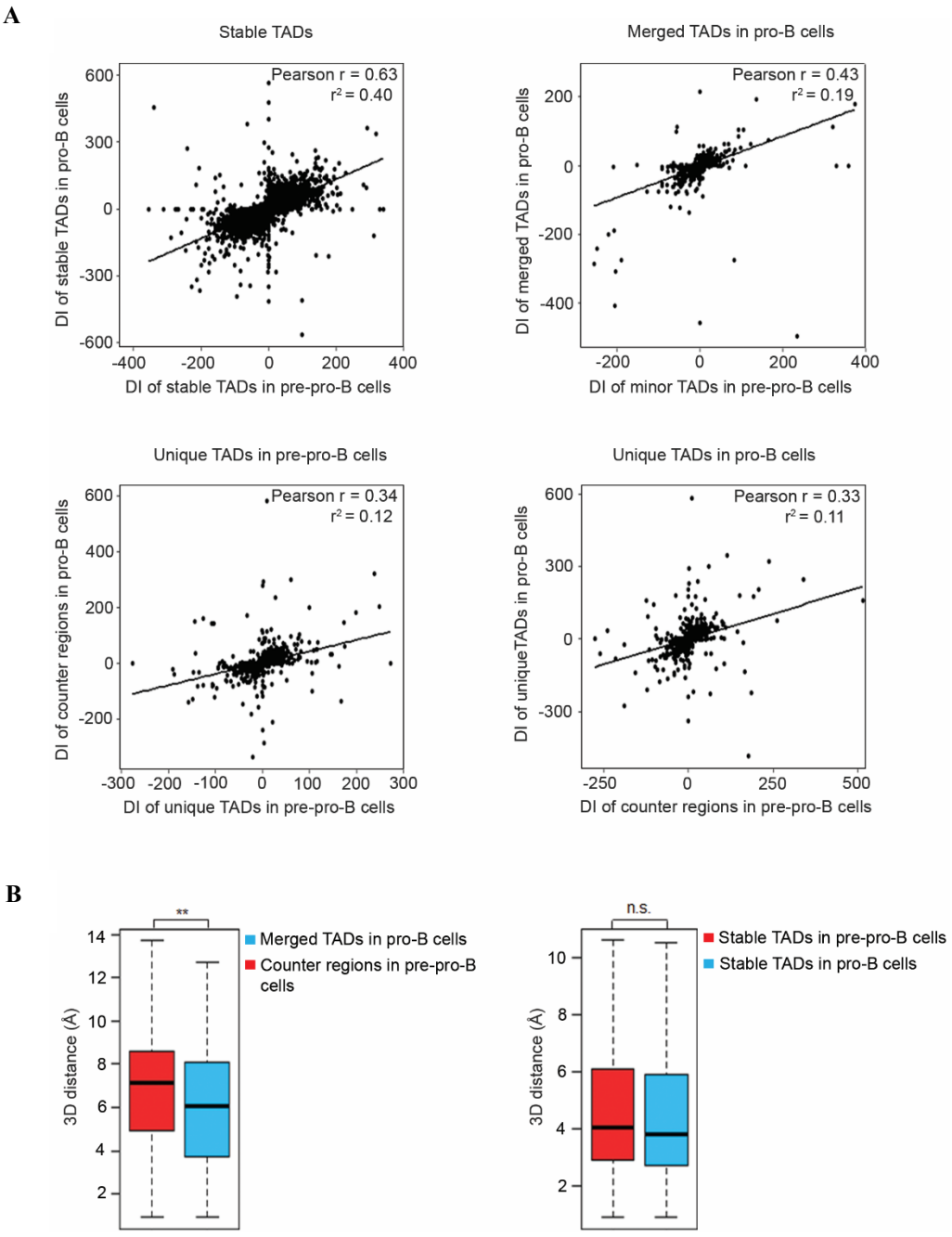
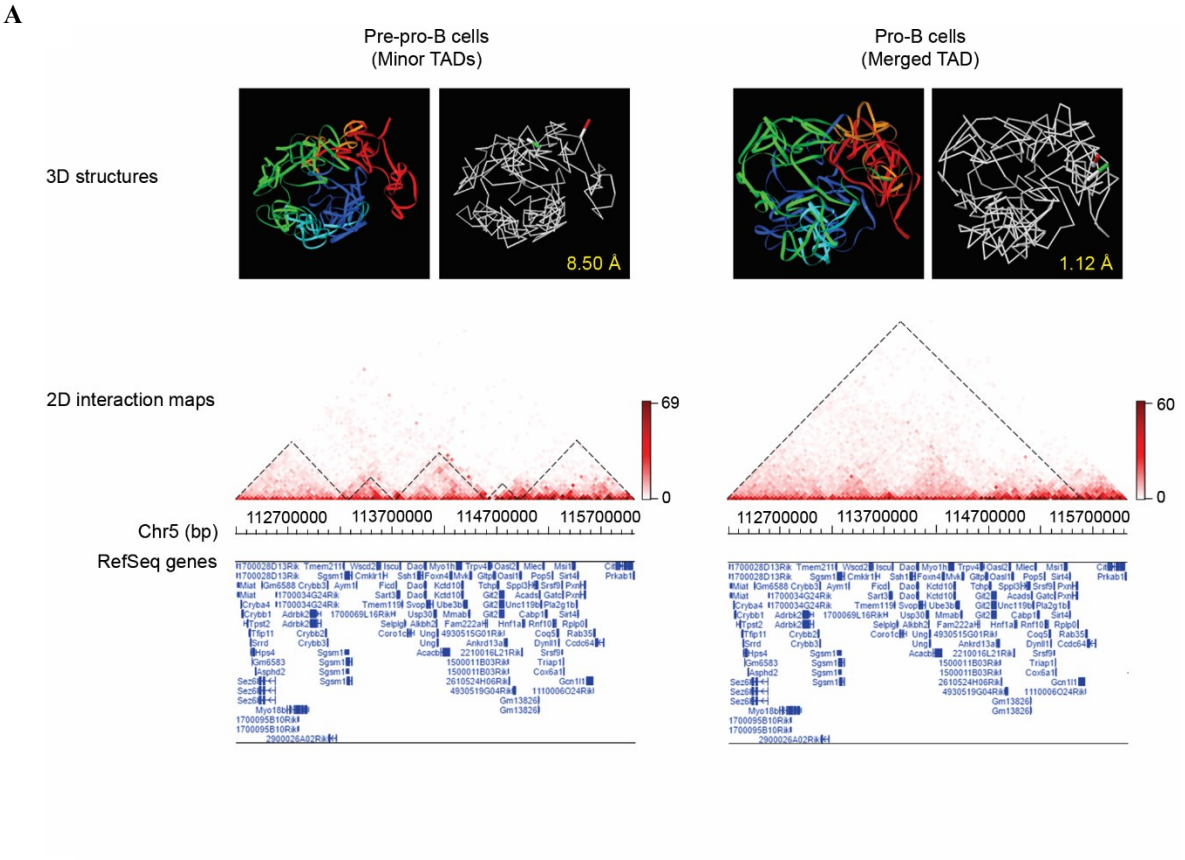
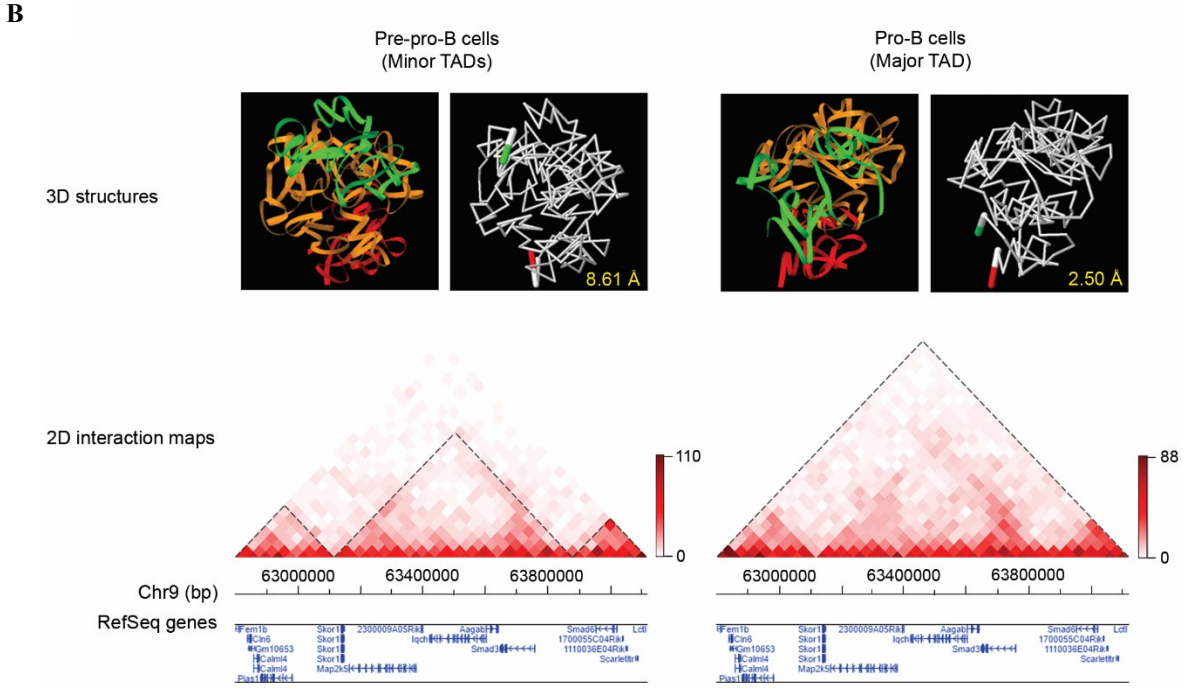
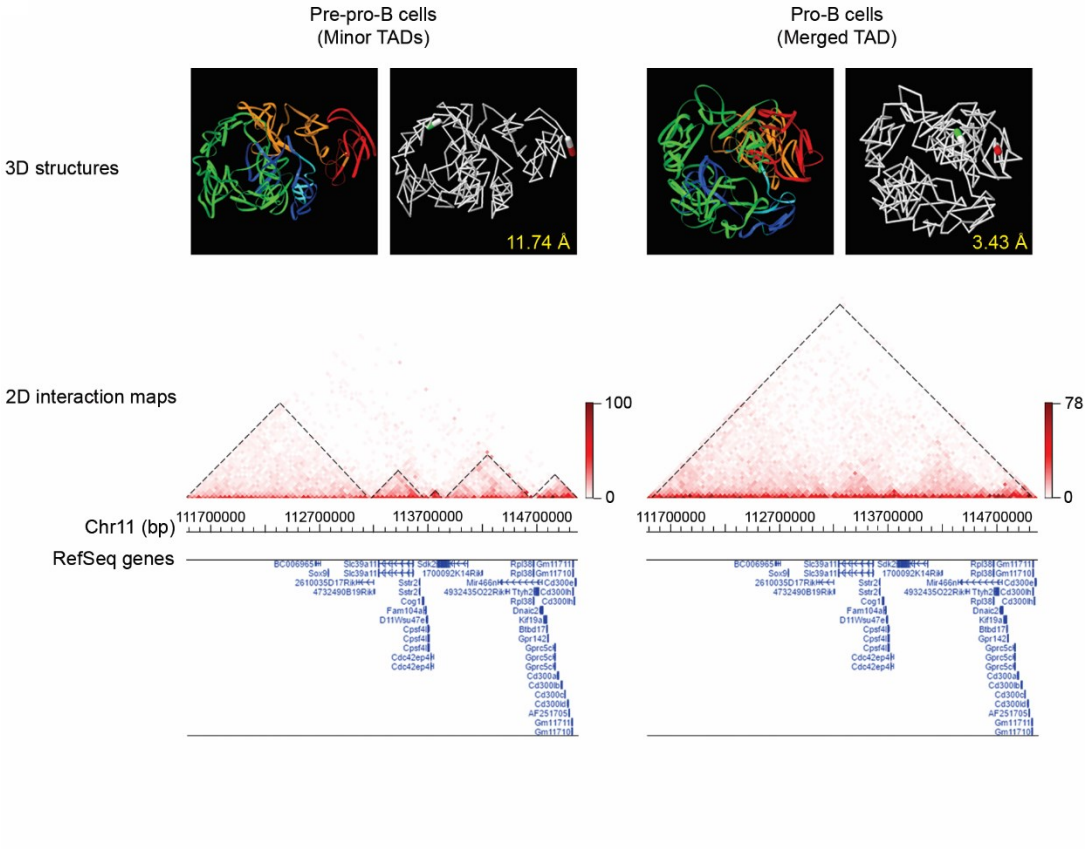


Figure 3.3.3. Comparative analysis of structural organization of TADs between pre-pro-B and pro-B cells. (A) Pearson's Correlation Coefficient for directionality index (DI) calculated for stable as well as dynamic TADs (merged and unique) between pre-pro-B and pro-B cells. **(B)** Genome-wide comparative analysis of 3D spatial distances between start and end regions of merged TADs in pro-B cells and their counter regions in pre-pro-B cells ($^{**}P<0.01$) (left panel). Similar analysis of 3D spatial distances for stable TADs in both pre-pro-B and pro-B cells (n.s. = not significant) (right panel).

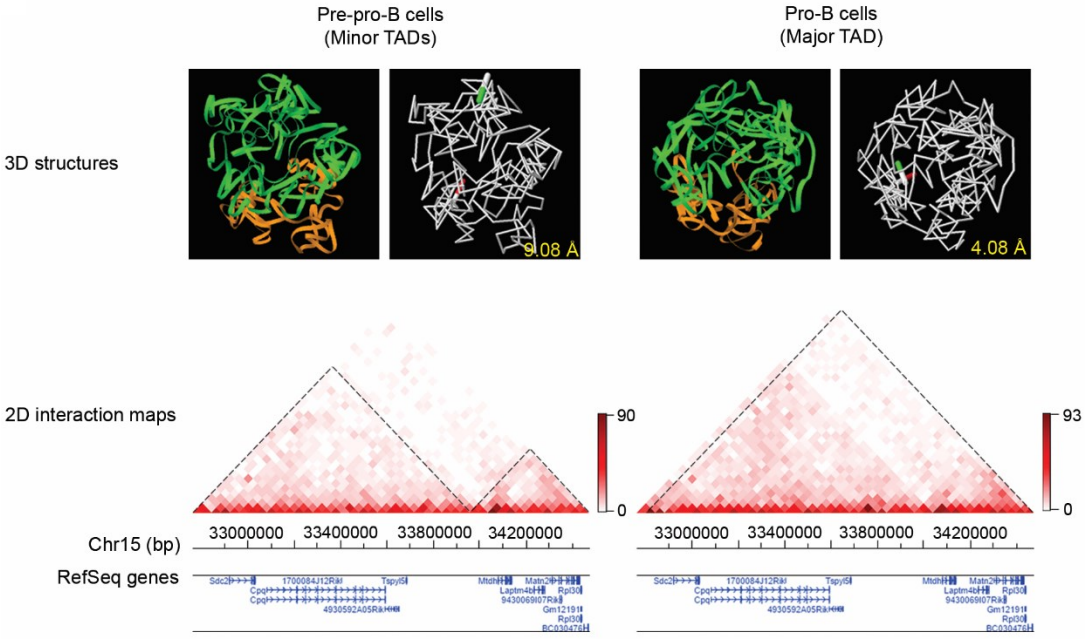

A
Pre-pro-B cells
(Minor TADs)
Pro-B cells
(Merged TAD)
3D structures
8.50 Å
1.12 Å
2D interaction maps
69
0
60
0
Chr5 (bp)
112700000
113700000
114700000
115700000
RefSeq genes


B
Pre-pro-B cells
(Minor TADs)
Pro-B cells
(Major TAD)
3D structures
8.61 Å
2.50 Å
2D interaction maps
110
0
88
0
Chr9 (bp)
63000000
63400000
63800000
RefSeq genes

C


Pre-pro-B cells
(Minor TADs)
Pro-B cells
(Merged TAD)
3D structures
11.74 Å
3.43 Å
2D interaction maps
100
0
78
0
Chr11 (bp)
111700000
112700000
113700000
114700000
RefSeq genes

D


Pre-pro-B cells
(Minor TADs)
Pro-B cells
(Major TAD)
3D structures
9.08 Å
4.08 Å
2D interaction maps
90
0
93
0
Chr15 (bp)
33000000
33400000
33800000
34200000
RefSeq genes

E

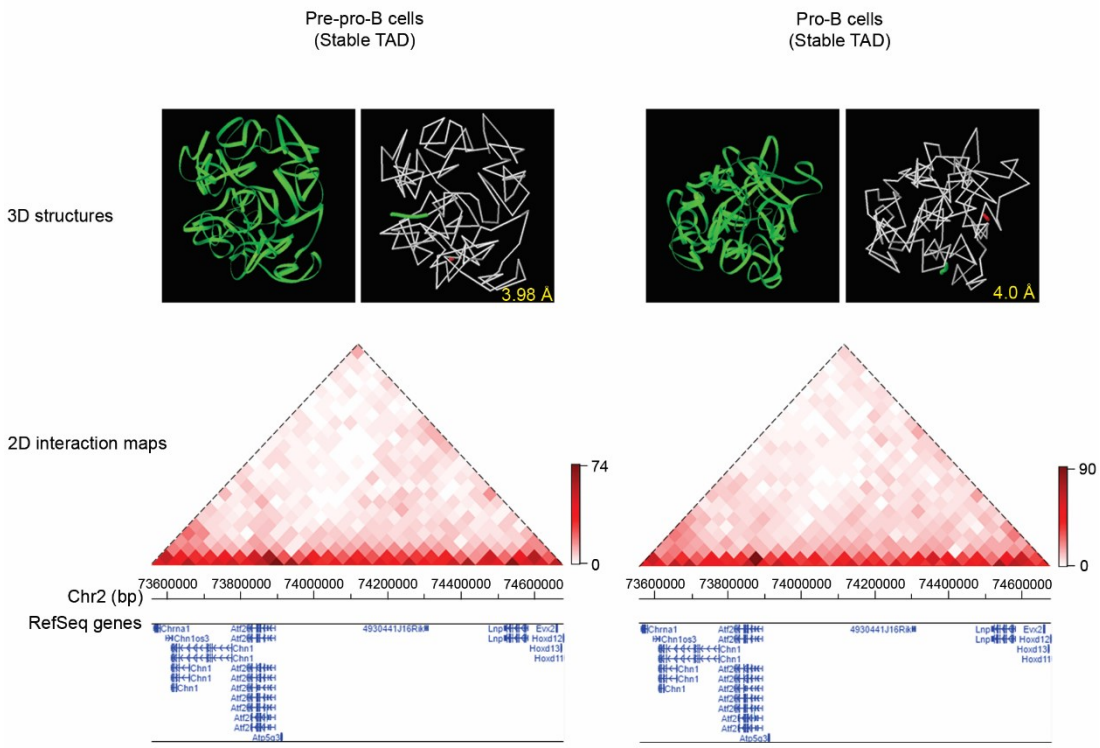


Figure 3.3.4. 3D models reflect the changes in 2D interaction maps. Comparative analysis of 2D interaction maps and 3D models of merged TADs between pre-pro-B and pro-B cells (Chr5: 112.20-115.64 Mb (Fig. A), Chr9:62.8-64.12 Mb (Fig. B), Chr11:111.52-115.04 Mb (Fig. C), and Chr15:32.8-34.48 Mb (Fig. D)). 3D models are generated by AutoChrom3D and each minor TAD in pre-pro-B cell is marked with distinct color and same color code is given for corresponding genomic regions of merged TAD in pro-B cells. The start and end regions of merged TAD in pro-B cells and its counter regions in pre-pro-B cells are highlighted by green and red, respectively in the back bone 3D structure and the spatial distance between these regions is shown in yellow color. TADs are demarcated by domain calling approach and highlighted by dotted lines. 3D models are generated at 8 kb (Chr5 and 15) or 16 kb (Chr 5 and 11) resolution. **(E)** Comparative analysis of 2D interaction maps and 3D models of a stable TAD (Chr2:73.56-74.68 Mb) between pre-pro-B and pro-B. 3D models are generated by AutoChrom3D. The start and end regions of TADs are highlighted by green and red respectively in the back bone 3D structure and the spatial distance between these regions is shown in yellow color. TADs are demarcated by domain calling approach and highlighted by dotted lines. 3D models are generated at 8 kb resolution.

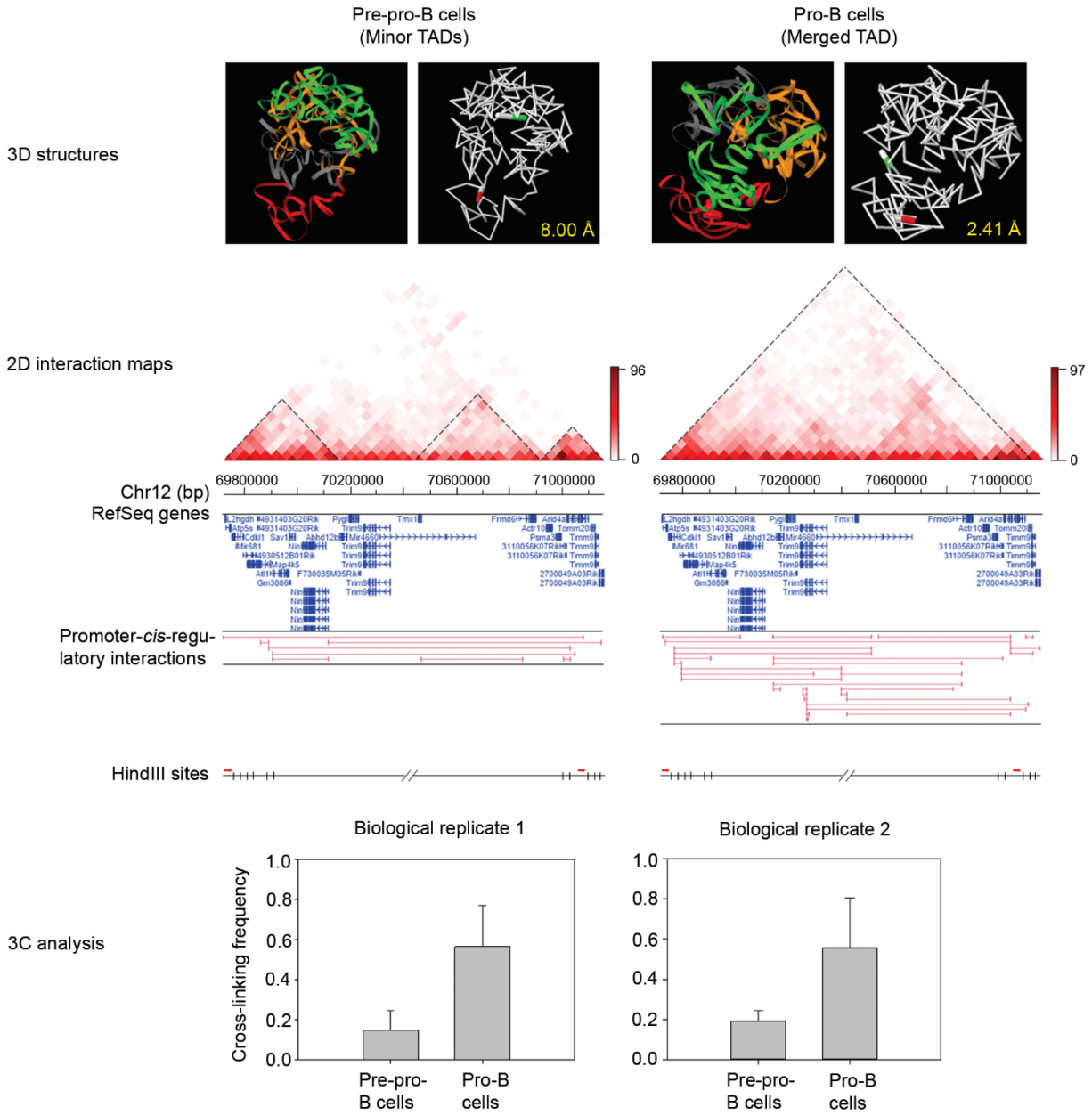


Figure 3.3.5. Validation of TADs reorganization by 3C analysis. Comparative analysis of 3D spatial distances and promoter-*cis*-regulatory interactions between merged TAD (pro-B cells) and its counter TADs (pre-pro-B cells) spanning the genomic region (69.72-71.16 Mb) of chromosome 12. 3D models generated by AutoChrom3D are colored distinctly based on minor TADs in pre-pro-B cells and the same color code is given for corresponding genomic regions of merged TAD in pro-B cells. The start and end regions of merged TAD in pro-B cells and its counter regions in pre-pro-B cells are highlighted by green and red respectively in the back bone 3D structure and the spatial distance between these regions is shown in yellow color. 3D models are generated at 8 kb resolution (upper panel). TADs are demarcated by domain calling approach and highlighted by dotted lines. TADs are mapped with active epigenetic marks (H3K4me3 specific for promoters where as H3K4me1 and H3K4me2 specific for enhancers) as determined by ChIP-Seq in both pre-pro-B and pro-B cells. Promoter-*cis*-regulatory interactions are represented by horizontal lines (middle panel). 3C analysis of interaction frequency between ends of merged TAD (Chr12: 69.72-71.16 Mb) in pro-B cells and its counter regions in pre-pro-B cells. HindIII restriction sites are shown above the 3C plots. The location of primers used for measuring cross-linking frequency is indicated by red arrows (lower panel).

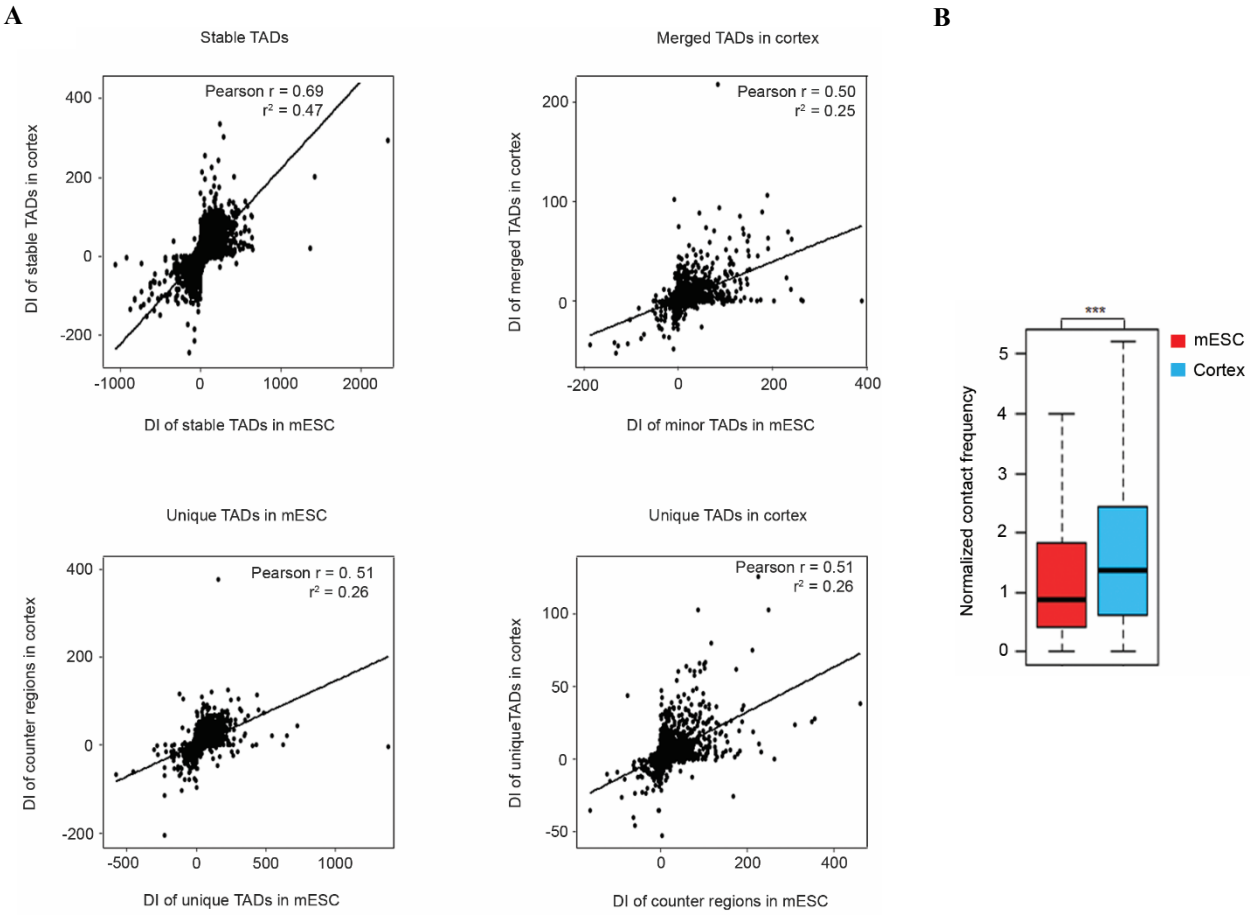


Figure 3.3.6. Comparative analysis of normalized contact frequency and DI between mESC and cortex cells. (A) Box plot representing normalized contact frequency of inter-TAD regions of minor TADs in mESC and their counter regions of merged TADs in cortex (***P<0.001). **(B)** Correlation coefficient analysis of directionality index (DI) for stable as well as dynamic TADs (merged and unique) between mESC and cortex.

3.4 TADs constitute structural frameworks for coordinated gene expression

Next, we sought to determine the relationship between the structural organization of TADs and the differential gene expression pattern in pre-pro-B and pro-B cells. For this, we integrated TAD regions with PC1 values at 100 kb resolution to assess their chromatin state. Interestingly, we found that majority of the TADs are either transcriptionally permissive or repressive. However, a small percentage of TADs are comprised of both permissive and repressive chromatin regions and are referred as uncharacterized **(Figure 3.4.1A).** As expected, genes present in permissive TADs, displayed higher nascent transcript levels compared to those in repressive TADs **(Figure 3.4.1B)**. These observations suggest the possibility that TADs serve as structural frameworks for coordinated regulation of genes. To rigorously demonstrate this, we calculated Pearson's correlation coefficient (PCC) for all possible gene pairs, using publicly available genome-wide expression data sets for pre-pro-B and pro-B cell types (Heng et al., 2008). PCC for pre-pro-B cells was calculated by comparing microarray measurements of hematopoietic stem cells (HSCs) and CLPs as they mimic pre-pro-B cells. Likewise, for pro-B cells, PCC was calculated using microarray measurements of CLPs and pro-B cells (pro-B.FrBC.BM). These analyses revealed that in both pre-pro-B and pro-B cells, genes within a given TAD exhibit significantly higher correlation values ($P<0.001$) in relation to the genes that are present in other TADs, indicating that TADs represent co-regulated sub-units of the genome **(Figure 3.4.1C and D)**. We note that such coordinated regulation of genes within the TADs facilitate activation/repression of gene clusters in a cell type-specific manner. For instance, the HOXA gene cluster (Chromosome 6), which is localized in a single stable TAD is transcriptionally active in pre-pro-B cells, whereas the same cluster is found to be transcriptionally inactive in pro-B cells; suggesting that TADs not only serve as fundamental sub-units for coordinate regulation of genes, but they also provide a framework to sustain lineage-specific gene expression pattern.

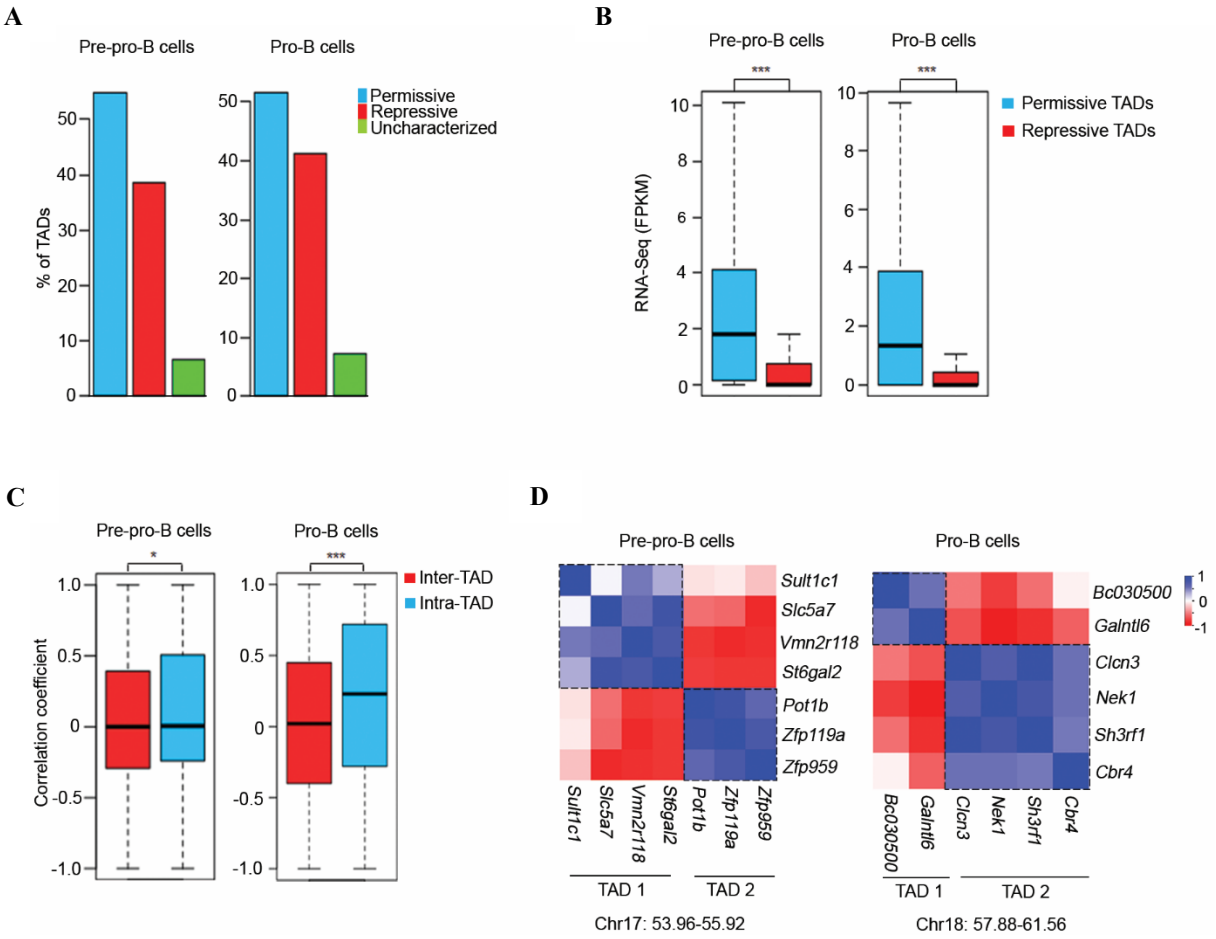
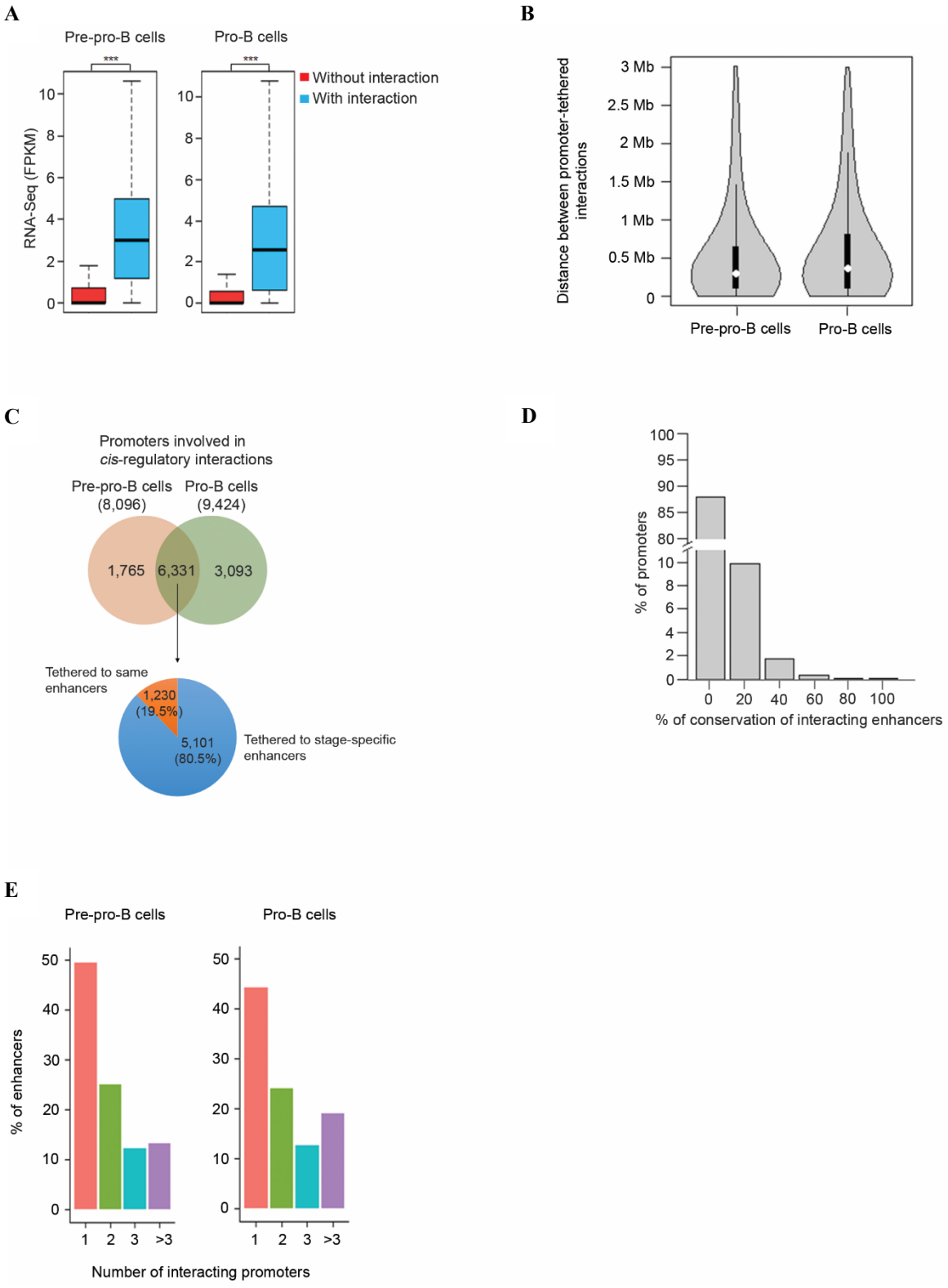


Figure 3.4.1. TADs represent chromatin subunits of coordinate gene expression. (A) Boxplots representing transcript levels of genes present in repressive and permissive TADs for both pre-pro-B and pro-B cells. The chromatin state of the TADs is defined by PC1 values. The transcript levels as measured by RNA-Seq were significantly higher (***P<0.001) in permissive TADs as compared to repressive TADs in pre-pro-B and pro-B cell types. **(B)** Histogram representing chromatin state of TADs as defined by Principal Component Analysis (PCA). **(C)** Comparative analysis of Pearson's correlation coefficient (PCC) for gene-pairs present in the same TAD against gene-pairs present in other TADs (*P<0.05, ***P<0.001) in pre-pro-B and pro-B cells. PCC is calculated by considering microarray measurements of hematopoietic stem cells (HSCs), common Lymphoid Progenitors (CLPs) and pro-B cells (pro-B.FrBC.BM). **(D)** Representation of Pearson's correlation coefficient for gene-pairs present in two different TADs spanning genomic region (53.96-55.92 Mb) of chromosome 17 and (57.88-61.56 Mb) of chromosome 18 for pre-pro-B and pro-B cells respectively. Blue represents positive correlation whereas red represents negative correlation. Each dotted box represents an individual TAD.

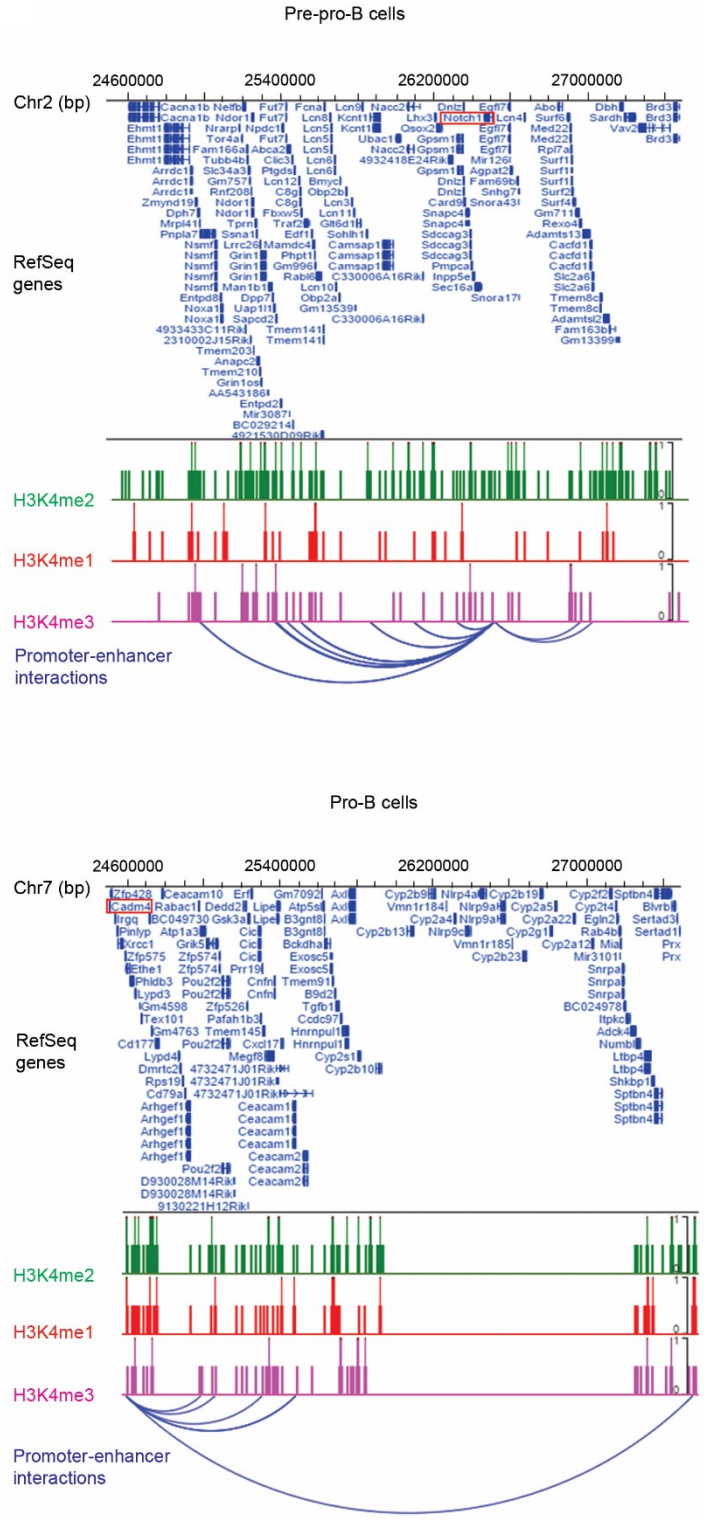
3.5 The *cis*-regulatory interaction landscape undergoes rewiring during B cell fate commitment

Although, it is well established that promoter-*cis*-regulatory interaction landscape determines a lineage-specific gene expression pattern (Carter et al., 2002; Hatzis et al., 2002), much less is known about the genome-wide composition of these interactions during B cell development. From *in situ* Hi-C analyses, we identified a total of 31,190 and 47,711 potential promoter-tethered interactions in pre-pro-B and pro-B cells, respectively. As expected, genes whose promoters are involved in *cis*-regulatory interactions showed significantly higher expression levels than the genes that are not involved in any such interactions **(Figure 3.5.1A)**. The majority of the promoter-*cis*-regulatory interactions (83.5% in pre-pro-B cells and 79% in pro-B cells) are within the range of 1Mb with a median value of 298 kb and 330 kb in pre-pro-B and pro-B cells, respectively **(Figure 3.5.1B)**. We have classified these promoter-tethered interactions into three basic groups: intergenic (promoter-promoter), extragenic (promoter-enhancer) and intragenic (promoter-gene body). We observed 8,410 and 10,556 intergenic interactions in pre-pro-B and pro-B cells, respectively. Similarly, we have captured 1,040 and 1,448 intragenic interactions in pre-pro-B cells and pro-B cells, respectively. Strikingly, we observed a significant increase in promoter-enhancer interactions in pro-B cells compared pre-pro-B cells. We found a total number of 21,740 promoter-enhancer (extragenic) interactions involving 8,096 promoters and 10,637 enhancers in pre-pro-B cells, wherein each promoter on average interacts with 2.68 enhancers. In the case of pro-B cells, 35,707 promoter-enhancer (extragenic) interactions involve about 9,424 promoters and 14,904 enhancers, wherein each promoter on average interacts with 3.79 enhancers. Among 9,424 promoters captured in pro-B cells, 6,331 (67.2%) promoters were also captured in pre-pro-B cells. Interestingly, 5,101 (80.5%) of common promoters were found to interact with cell type-specific enhancers and only 1,230 (19.5%) promoters share common enhancers **(Figure 3.5.1C and D)**. These results reveal that during B-lineage specification, the promoter-enhancer interaction landscape undergoes extensive rewiring. We also note that in pre-pro-B cells, nearly 13.3% of enhancers interact with more than three promoters. Likewise, in pro-B cells, almost 19% of enhancers were found to interact with more than three promoters **(Figure 3.5.1E and F)**. These observations support the assertion that multiple genes interacting with the same enhancer may be co-expressed (Chepelev et al., 2012).

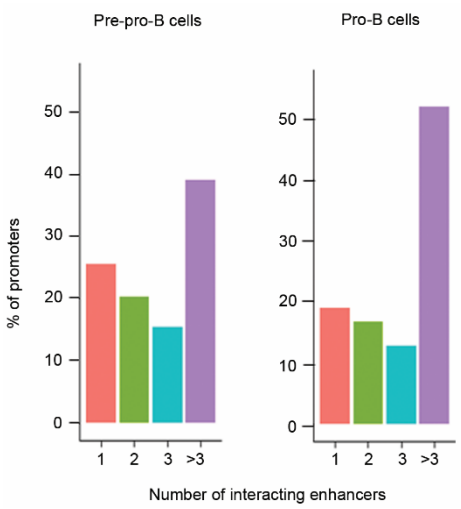
Furthermore, our data revealed that in pre-pro-B cells, nearly 25.4% of promoters interact with only one enhancer, while remaining promoters interact with two or more enhancers. Likewise, in pro-B cells, the majority of the promoters (81.1%) were found to interact with two or more enhancers **(Figure 3.5.1G and H)**. To test whether transcriptional activity of a promoter depends on the number of its *cis*-regulatory interactions, we assessed the transcript levels of the corresponding genes. Interestingly, these studies depicted a positive correlation between gene expression and a number of *cis*-regulatory interactions in both cell types **(Figure 3.5.2A)**. Next, we sought to determine if loss or gain of these interactions induce differential gene expression patterns. For this, we compared the *cis*-regulatory interaction landscape of genes that show + 10 fold differential expression between pre-pro-B and pro-B cells. The analysis showed that the expression pattern is closely associated with an increase in the number of *cis*-regulatory interactions **(Figure 3.5.2B)**. For instance, *Cd24a*, which is highly induced in pro-B cells, interacts with 13 *cis*-regulatory elements, whereas, it is involved in only four such interactions at pre-pro-B cell stage, where its expression is considerably low **(Figure 3.5.2C)**. Correspondingly, the genes: *Flt3* and *Ccl3*, that are important for maintenance of MPPs and differentiation of T-cells, respectively, are transcriptionally active at the pre-pro-B stage. These genes were found to be involved in more number of *cis*-regulatory interactions (*Flt3*:6, *Ccl3*:8) in pre-pro-B cells as compared to pro-B cells (*Flt3*:1, *Ccl3*:0) **(Figure 3.5.2D)**. The examples depicted here demonstrate the prevalence of dynamic promoter-*cis*-regulatory interactions across B cell developmental stages. To rigorously validate these findings, we carried out 3C (Chromosome Conformation Capture) analysis of promoter-enhancer interactions of *Ccl3* locus in pre-pro-B and pro-B cells **(materials and methods).** We observed that the interaction frequency between the *Ccl3* promoter and with its upstream enhancer (located 64 kb away) was higher in pre-pro-B cells compared to that in pro-B cells **(Figure 3.5.3A)**. Correspondingly, the quantitative RT-PCR analysis revealed thirty-fold higher *Ccl3* transcript levels in pre-pro-B cells as compared to the levels in pro-B cells **(Figure 3.5.3B)**. These analyses confirm that reinforcement of lineage-specific gene expression is contingent upon specificity and frequency of interactions between promoters and their *cis*-regulatory elements.


A
Pre-pro-B cells
Pro-B cells
***
***
Without interaction
With interaction
RNA-Seq (FPKM)
0
2
4
6
8
10
B
Distance between promoter-tethered interactions
0
0.5 Mb
1 Mb
1.5 Mb
2 Mb
2.5 Mb
3 Mb
Pre-pro-B cells
Pro-B cells
C
Promoters involved in cis-regulatory interactions
Pre-pro-B cells (8,096)
Pro-B cells (9,424)
1,765
6,331
3,093
Tethered to same enhancers
1,230 (19.5%)
5,101 (80.5%)
Tethered to stage-specific enhancers
D
% of promoters
0
2
4
6
8
10
80
85
90
95
100
0
20
40
60
80
100
% of conservation of interacting enhancers
E
Pre-pro-B cells
Pro-B cells
% of enhancers
0
10
20
30
40
50
1
2
3
>3
Number of interacting promoters

F


Pre-pro-B cells
24600000
25400000
26200000
27000000
Chr2 (bp)
RefSeq genes
H3K4me2
H3K4me1
H3K4me3
Promoter-enhancer interactions
Pro-B cells
Chr7 (bp)
RefSeq genes
H3K4me2
H3K4me1
H3K4me3
Promoter-enhancer interactions

G


Pre-pro-B cells
Pro-B cells
% of promoters
0
10
20
30
40
50
1
2
3
>3
Number of interacting enhancers

H

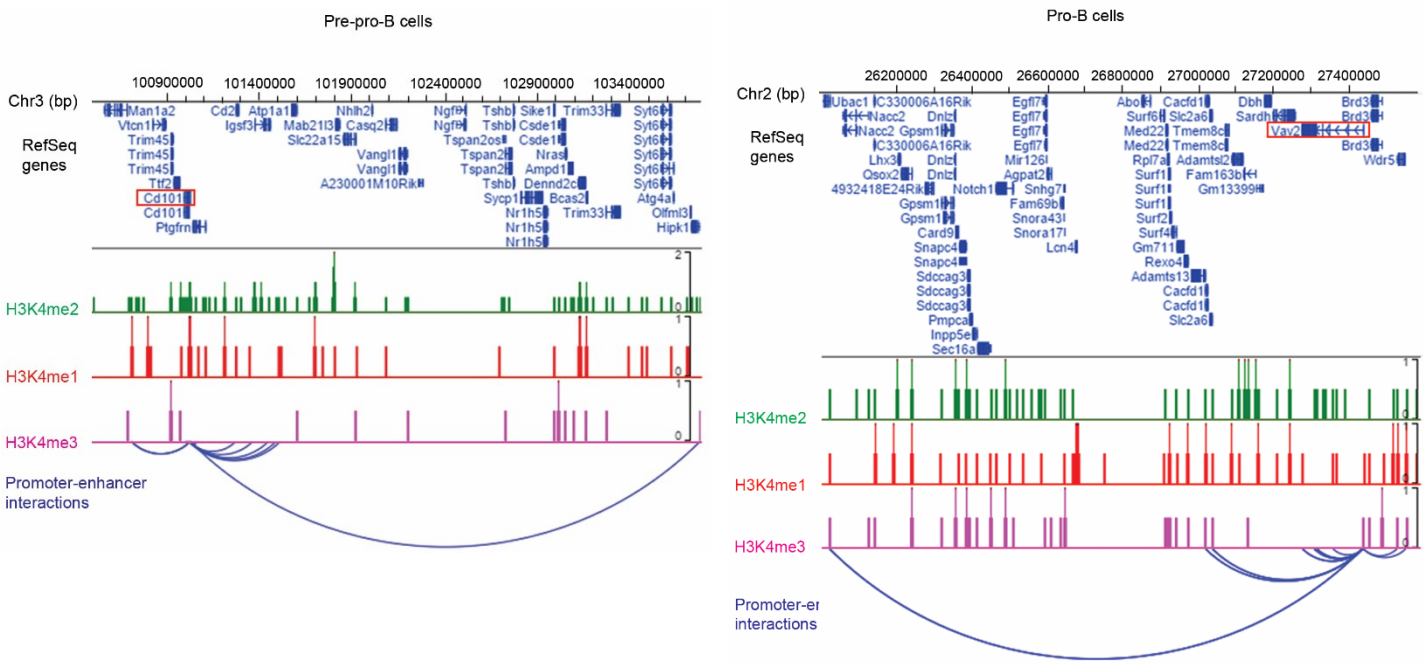


Figure 3.5.1. *Cis*-regulatory interaction landscape undergoes considerable changes during lineage specification. (A) Box plots showing the expression levels (FPKM) of genes whose promoters are associated with (blue) and without (red) *cis*-regulatory interactions (***P<0.001). Transcript levels of genes are measured by RNA-Seq. **(B)** Violin plots showing genome-wide distribution of promoter-*cis*-regulatory interactions (promoter-promoter and promoter-enhancer) in pre-pro-B and pro-B cells. Majority of the promoter-*cis*-regulatory interactions (83.5% in pre-pro-B cells and 79% in pro-B cells) are within the range of 1Mb. Promoter-*cis*-regulatory interactions are identified by integrating *in situ* Hi-C interactome with ChIP-Seq data of various epigenetic marks (H3K4me3, H3K4me1, H3K4me2 and H3K9/14ac). **(C)** Venn diagram representing number of promoters captured in pre-pro-B cells (left), pro-B cells (right) or in both cell types (middle). The pie chart below represents the number of promoters tethered to same (orange) or different enhancers (blue). *In situ* Hi-C interactome is integrated with ChIP-Seq data of various epigenetic marks to identify promoter (H3K4me3) and enhancer (H3K4me1, H3K4me2) elements. **(D)** Conservation of promoter-enhancer interactions between pre-pro-B and pro-B cells. Bar graph representing the percentage of promoters that share 0%, 20%, 40% etc., of their interactions with enhancer elements, in pro-B cells as compared to pre-pro-B cells. **(E)** Histogram representing the percentage of enhancers interacting with the number of promoter elements (1 to >3) that are determined by *in situ* Hi-C in pre-pro-B and pro-B cells. **(F)** Representation of enhancer elements of *Notch1* and *Cadm4* that are interacting with multiple promoters in pre-pro-B (top) and pro-B cells (bottom) respectively. *In situ* Hi-C interactome is integrated with ChIP-Seq data of epigenetic marks to identify promoter (H3K4me3) and enhancer (H3K4me1, H3K4me2) elements. Promoter-enhancer interactions are represented by blue arcs. **(G)** Histogram representing the percentage of promoters interacting with the number of enhancer elements (1 to >3) that are determined by *in situ* Hi-C in pre-pro-B and pro-B cells. **(H)** Representation of promoter elements of *Cd101* and *Vav2* that are interacting with multiple enhancers in pre-pro-B (top) and pro-B cells (bottom) respectively. *In situ* Hi-C interactome is integrated with ChIP-Seq data of epigenetic marks to identify promoter (H3K4me3) and enhancer (H3K4me1, H3K4me2) elements. Promoter-enhancer interactions are represented by blue arcs.

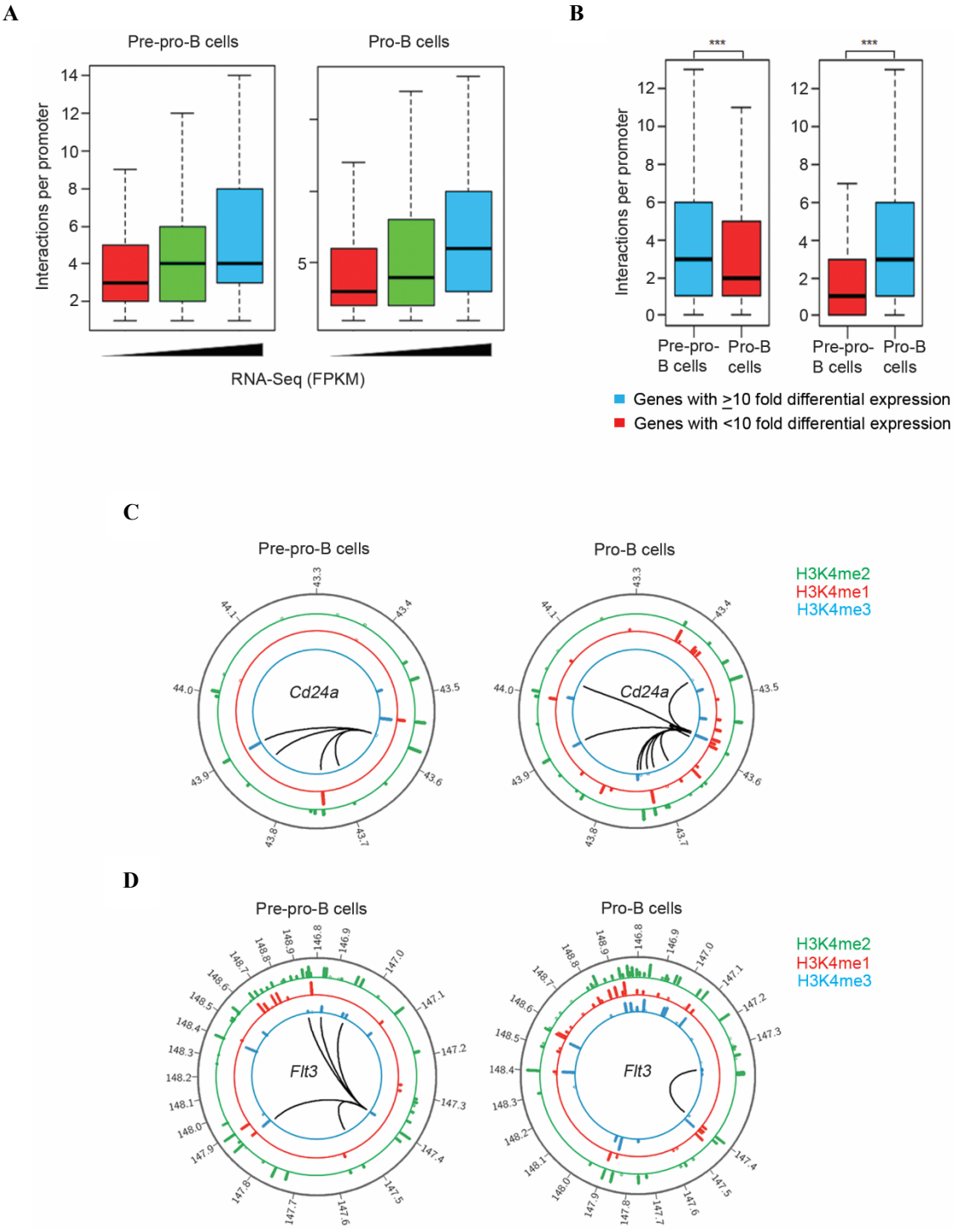


Figure 3.5.2. *Cis*-regulatory interaction landscape determines differential gene expression pattern. (A) Box plots showing the relation between the number of *cis*-regulatory elements that are interacting with promoters and their expression levels, in pre-pro-B and pro-B cells. Transcript levels of genes are measured by RNA-Seq. **(B)** Boxplots representing comparative analysis of promoter-*cis*-regulatory interactions for a set of genes with ≥ 10 fold differential expression in pre-pro-B cells (right panel) and in pro-B cells (left panel) (***P<0.001). **(C, D)** Circos plots showing promoter-*cis*-regulatory interactome of *Cd24a* gene (Chr11:43.3-44.1 Mb) and *Flt3* gene (Chr5:14.68-14.89 Mb) in pre-pro-B (left panel) and pro-B cells (right panel). Black arcs inside circos plot represent promoter-*cis*-regulatory interactions.

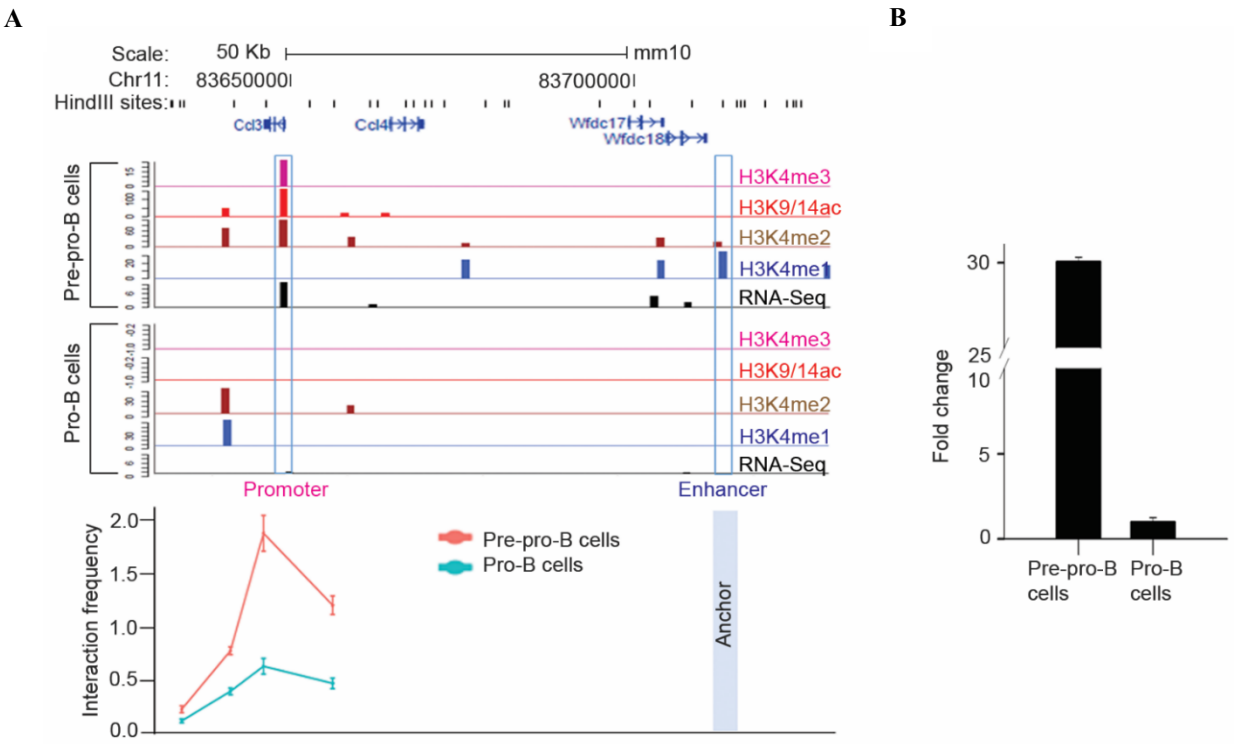


Figure 3.5.3. Validation of promoter-enhancer interactions by 3C analysis. (A) *Ccl3* locus, overlaid with various epigenetic marks, H3K4me3 (enriched at promoter regions), H3K4me1 and H3K4me2 (enriched at enhancers) as determined by ChIP-Seq in both pre-pro-B and pro-B cells (upper panel). 3C-qPCR analysis of interaction frequency between promoter region of *Ccl3* and its distant putative enhancer located 64 kb upstream of promoter in both pre-pro-B and pro-B cells (lower panel). Data are representative of two independent biological experiments (Error bars, S.E.). **(B)** Relative transcript levels of *Ccl3* as measured by quantitative RT-PCR in pre-pro-B cells and pro-B cells. *Hprt* was used as endogenous control and values are normalized against pro-B cells as reference control. Data are representative of two independent biological experiments (Error bars, S.E.).

3.6 Ebf1 coordinates B cell specific *cis*-regulatory interaction landscape

A major goal of this study was to understand the molecular relationship between chromatin architecture and differential transcriptional cascade. It has been shown that Ebf1 is essential for induction of early B lineage gene expression program and targeted inactivation of Ebf1 results in a complete block prior to B cell commitment (Lin et al., 1995; Pongubala et al., 2008). This raises a possible role for Ebf1 in chromatin relocalization and establishment of B lineage-specific *cis*-regulatory interaction landscape. To test this, we scanned for highly specific and significant Ebf1 binding sites in *cis*-regulatory regions of genes that switched to A or B compartments in pro-B cells, using publicly available databases (Jasper, Homer, and Uniprobe). From these analyses, we observed that Ebf1 and/or Pax5 bind to *cis*-regulatory sequences of differentially switched genes (65.3%). Although a subset of genes undergoes differential compartmentalization, this may not solely account for the induction of B lineage expression program. We propose that activation of B lineage-specific genes may be regulated at multiple levels including binding of lineage-specific transcription factors (E2A, Ebf1, Foxo1 and Pax5) to their target promoter-*cis*-regulatory interacting elements. To examine this, we integrated promoter-*cis*-regulatory interactions that are captured in pro-B cells with binding events of these factors. Importantly, we found that in pro-B cells, Ebf1 binds either alone or in combination with Pax5 to 5,390 (57.2%) promoters and 7,629 (51.2%) *cis*-regulatory elements that are involved in long-range interactions as defined by *in situ* Hi-C **(Figure 3.6.1A)**. Consistent with these observations, binding of Ebf1/Pax5 at *cis*-regulatory elements of their target genes is positively correlated with increased expression levels **(Figure 3.6.1B and C)**.

To rigorously demonstrate the induction of B lineage genes in response to Ebf1 and/or Pax5, we carried out genome-wide expression analysis following restoration of Ebf1 or Pax5 in *Ebf1*-/- progenitor cells **(materials and methods)**. As expected, Ebf1 and/or Pax5 induced a spectrum of genes associated with B cell identity, including those that are involved in pre-B and B cell receptor signaling, antigen presentation, DNA recombination, and repair. Conversely, Ebf1 and/or Pax5 repressed a subset of genes that are involved in the development and function of natural killer (NK), dendritic, and T cells **(Figure 3.6.1D and E)**. Integration of *in situ* Hi-C interactome (promoter-*cis*-regulatory interactions) with microarray data sets revealed that the genes that are upregulated (>2 fold) in response to Ebf1 (124, 39.7%) or Pax5 (231, 44%) or both

(37, 72.5%) are involved in long-range interactions **(Figure 3.6.1F)**. We note that activation of these genes could be due to direct binding of Ebf1 and/or Pax5 to their respective promoter or distant regulatory elements that are brought in close proximity by looping-out of intervening DNA sequences. To determine this, we scanned the promoter and their corresponding *cis*-regulatory elements that are upregulated in response to Ebf1 (161) and/or Pax5 (268) for their binding. *De novo* motif analyses revealed that Ebf1 binds to either promoter regions (17.6%) or distant *cis*-regulatory elements (45.3%) or both (36.9%). Similarly, Pax5 binds either at promoter regions (19.2%) or distant *cis*-regulatory elements (54.1%) or both (26%) **(Figure 3.6.1G)**. We note that both Ebf1 and Pax5 co-bind to a number of key B lineage genes such as *Cd19*, *Cd24a*, *Socs3* and *Dtx1*. Taken together, *in situ* Hi-C analyses in combination with genome-wide expression analysis and DNA occupancy studies, we demonstrate that activation of B lineage-specific genes is associated with changes in long-range interactions and many of these genes are potentially regulated by lineage-specific transcription factors, Ebf1 and Pax5.

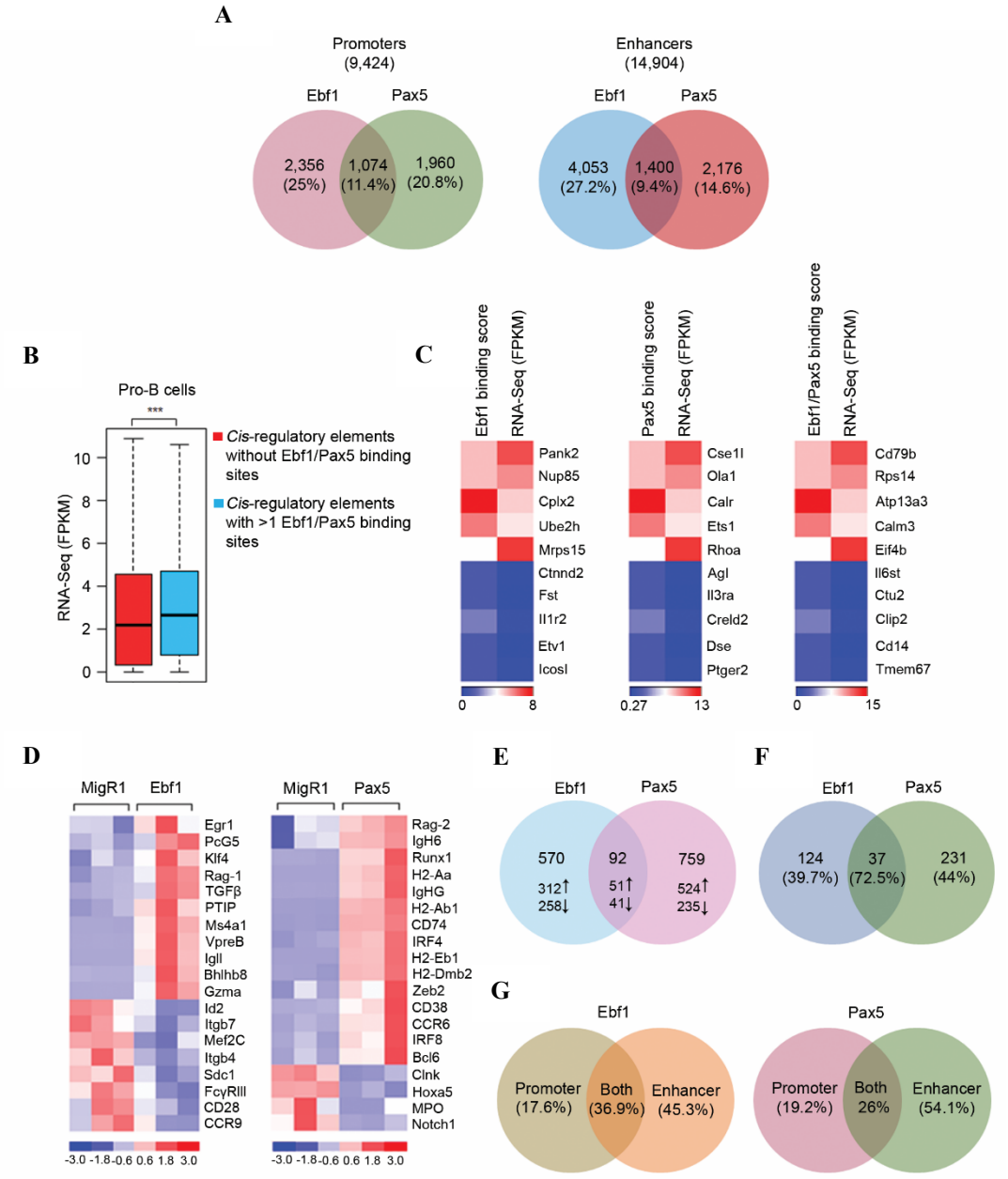
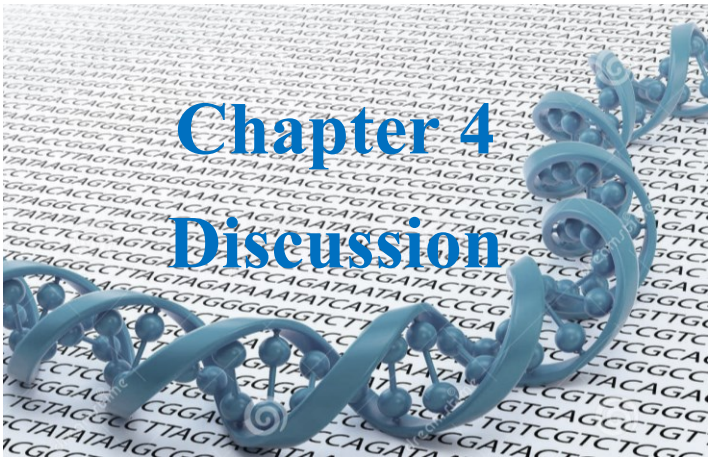


Figure 3.6.1. Ebf1 regulates B lineage specific gene expression pattern in part by binding at *cis*-regulatory interacting elements. (A) Venn diagram representing the motifs of the transcription factors (Ebf1 and Pax5) binding at the promoters and respective *cis*-regulatory elements. **(B)** Boxplot representing genome-wide comparative analysis of transcript levels of genes with or without Ebf1/Pax5 binding sites in the promoter-*cis*-regulatory interacting elements ($^{***}P<0.001$). **(C)** Heat maps showing correlation between transcript levels of genes and Ebf1 and/or Pax5 binding events in the promoter-*cis*-regulatory interacting elements. **(D)** Heat maps showing the genome-wide expression patterns of B-lineage-specific genes (fold change≥2; P value <0.05) obtained by microarray analysis of pre-pro-B cells (*Ebf1*-/- progenitors) transduced with Ebf1 or Pax5. **(E)** Venn diagram indicating the number of genes that are regulated by Ebf1 and/or Pax5. Up headed arrow represents activated, down headed arrow represents repressed genes. **(F)** Venn diagram representing the percentage of upregulated targets of Ebf1 and/or Pax5 that are involved in *cis*-regulatory interactions in pro-B cells. **(G)** Venn diagram representing the percentage of Ebf1 or Pax5 target genes containing Ebf1 and/or Pax5 binding sites within the *cis*-regulatory sequences that are involved in long-range interactions.



Chapter 4
Discussion

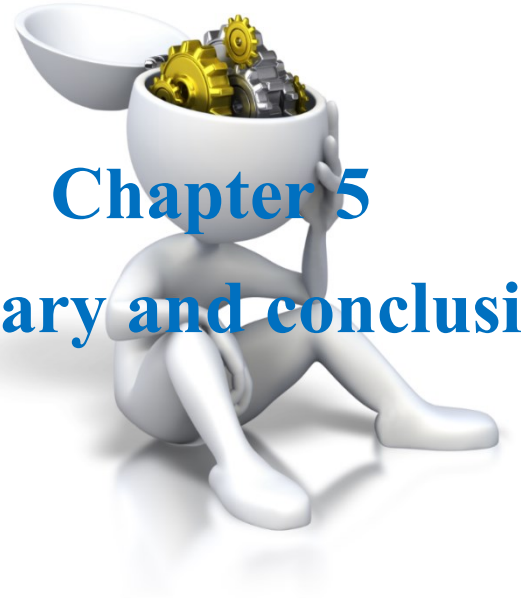
Precise and coordinated control of gene expression is important for the cell fate determination of multipotent progenitors (Bickmore et al., 2013; Gibcus et al., 2013; Gorkin et al., 2014; Misteli, 2007). Recent studies indicate that structural organization of the genome in 3D nuclear space is closely associated with modulation of transcriptional activity and establishment of cell type-specific gene expression program, indicating a potential relationship between nuclear architecture and mechanistic control of transcription. 3C-based studies indicate that genome is organized in a hierarchical manner: folding of chromatin loops, TADs, and large-scale compartments (Dixon et al., 2015; Dixon et al., 2012; Gibcus et al., 2013; Lieberman-Aiden et al., 2009; Lin et al., 2012). However, the comprehensive understanding of how multilayer organization of chromatin regulates cell-type-specific transcriptional activity remains unclear. Specifically, the following questions arise: Does chromatin relocalization precedes lineage commitment? Do chromatin domains undergo structural reorganization? What are the roles of lineage determinants during chromatin reorganization? We have attempted to address these questions by integrating genome-wide chromatin interaction data with epigenetic landscape and transcription profiling of cells that represent two distinct stages (pre-pro-B and pro-B) of B cell development. In line with previous reports, our *in situ* Hi-C analyses revealed that chromatin is non-randomly organized into A and B compartments (Gibcus et al., 2013; Lieberman-Aiden et al., 2009). We identified that a distinct set of genes switch between A and B compartments during the developmental transition from pre-pro-B to pro-B cell stage. For instance, genes that are important for B cell development including *Satb2*, *Tead1*, *Pou2af1* and *Tlr4* switch from the B to A compartment during the pre-pro-B to pro-B transition. Likewise, genes such as *Gata3*, *Klf4*, *Satb1* and *Zbtb16* that are important for disparate lineage differentiation programs localize to B compartment in pro-B cells, where they are silenced. In contrast, the majority of the downstream targets of these master regulators were found to be in A compartments in both cell types. These studies suggest that sequestering master regulators of alternate lineages into B compartments may ensure lineage specification. These observations are further supported by a previous study (Hewitt et al., 2004) wherein Th2-specific regulator, GATA3, was found to be localized in the nuclear periphery in a transcriptionally inactive state in Th1 cells. However, the downstream targets of GATA3 (IL-2, IL-3, and IL-4) were retained in the permissive compartment in both Th1 and Th2 cells. Thus selective relocalization of lineage determinants appears to play an important role during developmental transition from a progenitor to lineage-committed state. However, the mechanistic details of how the genomic loci

switch between A/B compartments is unclear. We propose that binding of transcription factors along with chromatin activation complexes enable the genomic loci to relocate from B to A compartment, whereas binding of polycomb group (PcG) proteins enable the loci to relocate from A to B compartments.

Our findings show that, although, a substantial number of TADs are stable between pre-pro-B and pro-B cells, chromatin interaction patterns particularly promoter-*cis* regulatory interactions within these TADs remodel to facilitate cell type-specific gene expression pattern. Strikingly, in addition to the stable TADs, we found a number of TADs that are dynamic and display structural alterations during B cell developmental transition. Specifically, we uncovered a set of unique TADs that are exclusively present either in pre-pro-B or pro-B cells, which contain genes that are selectively expressed at these stages. Additionally, we found a distinct set of merged TADs in pro-B cells, which were generated by coalesce of contiguous TADs present in pre-pro-B cells. This may account for the presence of fewer TADs with an increase in average TAD size and the corresponding gain of promoter-*cis*-regulatory interactions in pro-B cells as compared to pre-pro-B cells. We propose that the merging of TADs is due to an increase of inter-TAD interactions that are associated with epigenetic modifications and regulation of lineage-specific transcription factors. However, the precise molecular mechanisms that regulate the merging of TADs remains to be elucidated. A recent study (Li et al., 2015), in drosophila cells demonstrates a dramatic increase in inter-TAD promoter-*cis*-regulatory interactions upon heat shock treatment, which was shown to be associated with a corresponding redistribution of architectural proteins from borders to inside of TADs. Moreover, we show that genes that are present in a given TAD exhibit higher correlation of expression compared to genes present in other TADs, supporting the notion that TADs provide a structural framework for coordinated gene regulation (Nora et al., 2012). Thus the studies presented here, provide new insights into the structural organization of TADs and their propensity to undergo alterations during developmental progression.

In this study, we provide a comprehensive map of long-range interactions between promoters and their corresponding *cis*-regulatory elements in pre-pro-B and pro-B cells. We found a significant expansion in promoter-*cis*-regulatory interaction landscape during the developmental transition from pre-pro-B to pro-B stage is evidenced by not only an increase in a total number of

promoter-enhancer interactions, but also by the average increase in the ratio of promoters to enhancers. Additionally, we show that the transcript levels are positively correlated with the number of *cis*-regulatory interactions in both cell types, indicating that enhancer usage dictates transcriptional output. These studies corroborate the observations that modulation of gene activity is regulated by the enhancer landscape (Kieffer-Kwon et al., 2013; Thurman et al., 2012). Furthermore, our studies reveal that a significant number of promoter and enhancer elements that are engaged in long-range interactions contain Ebf1 and/or Pax5 binding sites. Many of these *cis*-regulatory elements are important for expression of developmentally regulated genes during B cell fate commitment. Accordingly, Ebf1 targeted genes displayed high levels of gene expression in pro-B cells. These results are further strengthened by the fact that, a subset of Ebf1 targeted genes, activated upon complementation of *Ebf1*-/- progenitors with Ebf1 or Pax5 were found to be involved in long-range interactions. Our analyses revealed novel molecular functions of Ebf1 and its potential role in the establishment of *cis*-regulatory interactions and activation of B lineage-specific genes. Nevertheless, the molecular mechanisms by which Ebf1 regulates these interactions remains to be understood. Binding of Ebf1 has been shown to recruit chromatin remodeling complexes like SWI-SNF to the *Cd79a* promoter there by increasing local chromatin accessibility for subsequent activation (Hagman et al., 2012). Similar mechanisms may also operate for the establishment of long-range interactions, wherein binding of key lineage-determining transcription factors like Ebf1 to cell type-specific enhancers recruits a distinct combinatorial set of factors, thereby positioning the enhancers in close proximity to their target promoters. This raises the exciting possibility that Ebf1 may mediate lineage-specific long-range interactions crucial for B lineage gene expression program. Collectively, our studies demonstrate that dynamic alterations of chromatin organization associated with changes in *cis*-regulatory interactions that are regulated by lineage determinants impinge on the induction of lineage-specific gene expression. As our understanding of the detailed molecular mechanisms that govern the dynamics of higher-order chromatin organization continues to expand, the relationship between the 3D organization of the genome and lineage-specific gene expression will be better understood.



Chapter 5
Summary and conclusions

Recent studies suggest that 3D organization of the genome is inextricably linked to the cell-type specific gene expression pattern (Bickmore et al., 2013; Gibcus et al., 2013; Gorkin et al., 2014; Misteli, 2007). The advent of chromosome conformation capture (3C) based techniques, have shed new light on three-dimensional organization of chromatin, with increased resolution and throughput (Dekker et al., 2002; Dostie et al., 2006; Dryden et al., 2014; Fullwood et al., 2009; Hughes et al., 2014; Kalhor et al., 2011; Lieberman-Aiden et al., 2009; Ma et al., 2015; Nagano et al., 2013; Rao et al., 2014; Simonis et al., 2006). These studies have demonstrated that within the nucleus chromatin is organized as hierarchical structures, such as compartments, TADs and chromatin loops (Dixon et al., 2015; Dixon et al., 2012; Gibcus et al., 2013; Lieberman-Aiden et al., 2009; Lin et al., 2012). However, the multilayer organization of chromatin and its association with lineage-specific gene expression pattern during B-cell fate commitment poorly understood. By integrating genome-wide chromatin interactome, with epigenetic landscape and transcription profiling, we have attempted to comprehensively analyze the fundamental relationship between nuclear reorganization and transcriptional regulation that orchestrates B cell fate commitment (Boya et al., 2017). To capture the genome-wide chromatin interactome, we have performed *in situ* Hi-C using *Ebf1-/-* progenitors which represents pre-pro-B cells and *Rag2-/-* cells that represent pro-B cells. Consistent with previous reports, principal component analysis analysis (PCA) revealed that chromatin is organized into A (permissive) and B (repressive) compartments (Lieberman-Aiden et al., 2009; Lin et al., 2012). A compartments are enriched with active methylation marks (H3K4me3, H3K4me1 and H3K9/14ac) and displayed higher transcriptional activity as compared to B compartments. Comparative analysis of chromatin state during differentiation of pre-pro-B cells into pro-B cells, revealed that although a major fraction (90.69%) of genes remained in the same compartment in both cell types, a small percentage (9.31%) of genes switch between A and B compartments. We found that the genes that relocate from B compartment to A compartment during differentiation displayed higher transcript levels in pro-B cells as compared to compared to pre-pro-B cells. This group comprised of genes that are essential for B cell development such as *Satb2*, *Tead1*, *Pou2af1* and *Tlr4*. On the other hand genes that switch from A compartment to B compartment displayed lower transcript levels in pro-B cells as compared to pre-pro-B cells. Particularly, genes that are encoding for key alternate lineages determinants such as *Gata3*, *Zbtb16*, *Klf4*, *Vav3* and *Sox6* are found to be relocalized from A compartment to B compartment. However, majority of the downstream targets of these master

regulators were found to be localized in A compartments in both cell types. Thus we propose that sequestering the genes encoding for master regulators of alternate lineages into B compartments may ensure lineage specification during B cell fate commitment. Further, we found that majority of the B lineage-specific genes (*Ebf1*, *Pax5*, *Foxo1*, *IRF4*, *IRF8*, *Cd79a*, *Cd79b* and *Cd19*) are localized in A compartments in both cell types. Collectively, these studies indicate that chromatin relocalization has contributory but not deterministic role in the regulation of gene expression pattern.

At sub-megabase level, A and B compartments are organized as highly self-interacting domains known as topologically associated domains (TADs). These chromatin domains were found to be invariant across various cell types (Dixon et al., 2015; Dixon et al., 2012). Cross-comparison of TADs between pre-pro-B and pro-B cells revealed that although a substantial number of TADs (pre-pro-B cells:50.9%, pro-B cells:56.5%) were stable, a considerable number of TADs undergo re-organization in terms of their genomic position and were categorized as 'dynamic'. Interestingly, comparison of transcript levels of genes within these stable TADs between pre-pro-B and pro-B cells revealed significant differences in their activities. Thus, we propose that the differential gene expression observed within stable TADs can be attributed to the changes in intrinsic chromatin interactions. Integration of promoter-*cis*-regulatory interactions with stable TADs revealed that *cis*-regulatory interaction landscape undergoes rewiring within these TADs to sustain lineage-specific gene expression pattern. Dynamic TADs are categorized as unique and merged TADs. Unique TADs are defined as those that are present only in one cell type but not in other cell types. In pro-B cells, merged TADs are generated as a result of coalescence of two or more minor TADs in pre-pro-B cells. We propose that merged TADs are formed as result of increased inter-TAD interactions in pro-B cells. As expected our analysis revealed that the normalized interaction frequency of inter-TAD regions of minor TADs in pre-pro-B cells is significantly lower as compared to their counter regions of merged TADs in pro-B cells. These results are in line with the increased median TAD size (920 kb) in pro-B cells as compared to pre-pro-B cells (800 kb). Comparative analysis of inter-TAD promoter-*cis*-regulatory interactions between minor TADs of pre-pro-B cells and merged TADs in pro-B cells, revealed a significant increase in the *cis*-regulatory interactome in pro-B cells as compared to pre-pro-B cells. These results indicate that dynamic re-organization of TADs is closely associated with changes in *cis*-

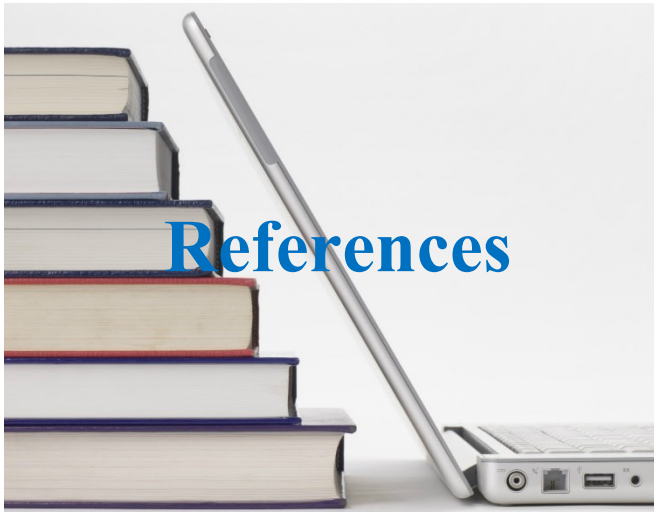
regulatory interaction landscape. To rigorously demonstrate the dynamic organization TADs, first, we have performed directionaly index (DI) analysis and found that stable TADs display higher correlation as compared to altered TADs (merged and unique). Second, by employing AutoChrom3D, we have generated 3D models of merged and stable TADs. Then, we have measured the spatial distance between start and end regions of merged TADs in pro-B cells and compared to their counter regions in pre-pro-B cells. In line with our proposition, our analysis revealed that spatial distance was significantly lower in pro-B cells, compared to pre-pro-B cells. On the other hand, no significant difference in the spatial distance was observed for stable TADs. Taken together, these results demonstrate that during B cell fate specification, TADs undergo dynamic structural re-organization as result of changes in chromatin interaction patterns. Furthermore, we have demonstrated that TADs serves as structural frameworks for coordinate gene expression.

By integrating *in situ* Hi-C interactome with genome-wide epigenetic marks (H3K4me3, H3K4me1 and H3K4me2), we have identified promoter-*cis*-regulatory interactions in both pre-pro-B and pro-B cells. We have observed a significant increase in the promoter-enhancer interaction landscape during differentiation of pre-pro-B cells to pro-B cells. Interestingly, we found that in both pre-pro-B and pro-B cells, majority of the common promoters as well as unique promoters interact with developmental stage-specific enhancers. These results reveal that during B-lineage specification enhancer landscape undergoes extensive rewiring. Further, we show that gene activity is positively correlated with the number of *cis*-regulatory interactions. We also show that gain or loss of these interactions determine the differential gene expression pattern during B-lineage specification. Although much is known about the molecular functions of Ebf1, during B-cell fate commitment, its role in the regulation and establishment of promoter-*cis*-regulatory interactions remains poorly understood. Our studies reveal that a significant number of promoter and enhancer elements that are involved in long-range interactions contain Ebf1 and/or Pax5 binding sites. Binding of Ebf1 and/or Pax5 at *cis*-regulatory elements is positively correlated with the expression levels of their cognate genes. These findings are further reinforced by the fact that, a subset of genes that are activated by Ebf1 and/or Pax5 as measured by microarray analysis, were found to be involved in long-range interactions. These observations demonstrate that activation of Ebf1 may facilitates the establishment of *cis*-regulatory interactions to promote the expression of

B-lineage genes. Collectively, our studies demonstrate that during B cell fate commitment, chromatin undergoes dynamic re-organization to induce of lineage-specific gene expression pattern.

Conclusions:

- Sequestration of lineage inappropriate genes in repressive compartments ensures B cell fate commitment.
- Although a majority of TADs are stable, a significant number of TADs undergo extensive structural reorganization during developmental transition from the pre-pro-B to the pro-B cell stage. These structural changes of TADs are closely associated with alterations in *cis*-regulatory interactions.
- Promoter-*cis*-regulatory interaction landscape undergoes extensive re-wring during B lineage-specification. The specificity and frequency of interactions between promoters and their *cis*-regulatory elements dictate transcriptional output.
- Ebf1 regulates differential gene expression pattern at least in part through establishment or stabilization of long-range interactions between promoter and *cis*-regulatory elements.



References

Austenaa, L. M., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B. A., de Pretis, S., Piccolo, V., Termanini, A., Prosperini, E., Pelizzola, M., de Laat, W., & Natoli, G. (2015). Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Molecular Cell, 60*(3), 460-474. doi: 10.1016/j.molcel.2015.09.018

Ay, F., Bailey, T. L., & Noble, W. S. (2014). Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Research, 24*(6), 999-1011. doi: 10.1101/gr.160374.113

Bain, G., Maandag, E. C., Izon, D. J., Amsen, D., Kruisbeek, A. M., Weintraub, B. C., Krop, I., Schlissel, M. S., Feeney, A. J., van Roon, M., & et al. (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell, 79*(5), 885-892.

Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L. M., Dostie, J., Pombo, A., & Nicodemi, M. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A, 109*(40), 16173-16178. doi: 10.1073/pnas.1204799109

Battulin, N., Fishman, V. S., Mazur, A. M., Pomaznoy, M., Khabarova, A. A., Afonnikov, D. A., Prokhortchouk, E. B., & Serov, O. L. (2015). Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. *Genome Biol, 16*, 77. doi: 10.1186/s13059-015-0642-0

Bickmore, W. A., & van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. *Cell, 152*(6), 1270-1284. doi: 10.1016/j.cell.2013.02.001

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M. R., & Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology, 3*(5), e157. doi: 10.1371/journal.pbio.0030157

Bonn, S., Zinzen, R. P., Girardot, C., Gustafson, E. H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., & Furlong, E. E. (2012). Tissue-specific analysis of

chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics, 44*(2), 148-156. doi: 10.1038/ng.1064

Boya, R., Yadavalli, A. D., Nikhat, S., Kurukuti, S., Palakodeti, D., & Pongubala, J. M. (2017). Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment. *Nucleic Acids Research*. doi: 10.1093/nar/gkx722

Boyle, S., Gilchrist, S., Bridger, J. M., Mahy, N. L., Ellis, J. A., & Bickmore, W. A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human Molecular Genetics, 10*(3), 211-219.

Cajal, S. R. (1903). Un sencillo metodo de coloraco6n selectiva del reticulo protoplasmico y sus efector en los organos nerviosos de vertebrados e invertebrados. *Trab. Lab. Invest. Biol, 2*, 129-221.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., & Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science, 298*(5595), 1039-1043. doi: 10.1126/science.1076997

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., & Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nature Genetics, 32*(4), 623-626. doi: 10.1038/ng1051

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., & Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Research, 22*(3), 490-503. doi: 10.1038/cr.2012.15

Cremer, T., & Cremer, C. (2006). Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem, 50*(4), 223-272.

Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., & Bickmore, W. A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *Journal of Cell Biology, 145*(6), 1119-1131.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., & Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell, 111*(2), 185-196.

Decker, T., Pasca di Magliano, M., McManus, S., Sun, Q., Bonifer, C., Tagoh, H., & Busslinger, M. (2009). Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. *Immunity, 30*(4), 508-520. doi: 10.1016/j.immuni.2009.01.012

Dekker, J., & Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Letters, 589*(20 Pt A), 2877-2884. doi: 10.1016/j.febslet.2015.08.044

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science, 295*(5558), 1306-1311. doi: 10.1126/science.1067799

DeKoter, R. P., Lee, H. J., & Singh, H. (2002). PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors. *Immunity, 16*(2), 297-309.

DeKoter, R. P., & Singh, H. (2000). Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science, 288*(5470), 1439-1441.

Delogu, A., Schebesta, A., Sun, Q., Aschenbrenner, K., Perlot, T., & Busslinger, M. (2006). Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity, 24*(3), 269-281. doi: 10.1016/j.immuni.2006.01.012

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V. V., Ecker, J. R., Thomson, J. A., & Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature, 518*(7539), 331-336. doi: 10.1038/nature14222

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature, 485*(7398), 376-380. doi: 10.1038/nature11082

Dobreva, G., Dambacher, J., & Grosschedl, R. (2003). SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin mu gene expression. *Genes & Development, 17*(24), 3048-3061. doi: 10.1101/gad.1153003

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D., & Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Research, 16*(10), 1299-1309. doi: 10.1101/gr.5571506

Dryden, N. H., Broome, L. R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., Nagano, T., Andrews, S., Wingett, S., Kozarewa, I., Assiotis, I., Fenwick, K., Maguire, S. L., Campbell, J., Natrajan, R., Lambros, M., Perrakis, E., Ashworth, A., Fraser, P., & Fletcher, O. (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Research, 24*(11), 1854-1868. doi: 10.1101/gr.175034.114

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A., & Noble, W. S. (2010). A three-dimensional model of the yeast genome. *Nature, 465*(7296), 363-367. doi: 10.1038/nature08973

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suva, M. L., & Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature, 529*(7584), 110-114. doi: 10.1038/nature16490

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep, 15*(9), 2038-2049. doi: 10.1016/j.celrep.2016.04.085

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., Chew, E. G., Huang, P. Y., Welboren, W. J., Han, Y., Ooi, H. S., Ariyaratne, P. N., Vega, V. B., Luo, Y., Tan, P. Y., Choy, P. Y., Wansa, K. D., Zhao, B., Lim, K. S., Leow, S. C., Yow, J. S., Joseph, R., Li, H., Desai, K. V., Thomsen, J. S., Lee, Y. K., Karuturi, R. K., Herve, T., Bourque, G., Stunnenberg, H. G., Ruan, X., Cacheux-Rataboul, V., Sung, W. K., Liu, E. T., Wei, C. L., Cheung, E., & Ruan, Y. (2009). An

oestrogen-receptor-alpha-bound human chromatin interactome. *Nature, 462*(7269), 58-64. doi: 10.1038/nature08497

Georgopoulos, K., Bigby, M., Wang, J. H., Molnar, A., Wu, P., Winandy, S., & Sharpe, A. (1994). The Ikaros gene is required for the development of all lymphoid lineages. *Cell, 79*(1), 143-156.

Gibcus, J. H., & Dekker, J. (2013). The hierarchy of the 3D genome. *Molecular Cell, 49*(5), 773-782. doi: 10.1016/j.molcel.2013.02.011

Gorkin, D. U., Leung, D., & Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell, 14*(6), 762-775. doi: 10.1016/j.stem.2014.05.017

Hagman, J., Ramirez, J., & Lukin, K. (2012). B lymphocyte lineage specification, commitment and epigenetic control of transcription by early B cell factor 1. *Current Topics in Microbiology and Immunology, 356*, 17-38. doi: 10.1007/82_2011_139

Hardy, R. R., Kincade, P. W., & Dorshkind, K. (2007). The protean nature of cells in the B lymphocyte lineage. *Immunity, 26*(6), 703-714. doi: 10.1016/j.immuni.2007.05.013

Hatzis, P., & Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Molecular Cell, 10*(6), 1467-1477.

Heitz, E. (1928). Das Heterochromatin der Moose. I. Jahrb. wiss. Bot. 69: 762-818. 1931, Die ursache der gesetzmassingen Zahl, Lage, Form und Grosse pflanzlicher Nukleolen. *Planta, 12*, 775-844.

Heng, T. S., & Painter, M. W. (2008). The Immunological Genome Project: networks of gene expression in immune cells. *Nature Immunology, 9*(10), 1091-1094. doi: 10.1038/ni1008-1091

Hesslein, D. G., Pflugh, D. L., Chowdhury, D., Bothwell, A. L., Sen, R., & Schatz, D. G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes & Development, 17*(1), 37-42. doi: 10.1101/gad.1031403

Hewitt, S. L., High, F. A., Reiner, S. L., Fisher, A. G., & Merkenschlager, M. (2004). Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. *European Journal of Immunology, 34*(12), 3604-3613. doi: 10.1002/eji.200425469

Hu, J., Zhang, Y., Zhao, L., Frock, R. L., Du, Z., Meyers, R. M., Meng, F. L., Schatz, D. G., & Alt, F. W. (2015). Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell, 163*(4), 947-959. doi: 10.1016/j.cell.2015.10.016

Hu, M., Deng, K., Qin, Z., Dixon, J., Selvaraj, S., Fang, J., Ren, B., & Liu, J. S. (2013). Bayesian inference of spatial organizations of chromosomes. *PLoS Comput Biol, 9*(1), e1002893. doi: 10.1371/journal.pcbi.1002893

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., & Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics, 46*(2), 205-212. doi: 10.1038/ng.2871

Ibn-Salem, J., Kohler, S., Love, M. I., Chung, H. R., Huang, N., Hurles, M. E., Haendel, M., Washington, N. L., Smedley, D., Mungall, C. J., Lewis, S. E., Ott, C. E., Bauer, S., Schofield, P. N., Mundlos, S., Spielmann, M., & Robinson, P. N. (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol, 15*(9), 423. doi: 10.1186/s13059-014-0423-1

Ikawa, T., Kawamoto, H., Wright, L. Y., & Murre, C. (2004). Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity, 20*(3), 349-360.

Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., Dekker, J., & Mirny, L. A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods, 9*(10), 999-1003. doi: 10.1038/nmeth.2148

Inlay, M. A., Bhattacharya, D., Sahoo, D., Serwold, T., Seita, J., Karsunky, H., Plevritis, S. K., Dill, D. L., & Weissman, I. L. (2009). Ly6d marks the earliest stage of B-cell specification

and identifies the branchpoint between B-cell and T-cell development. *Genes & Development, 23*(20), 2376-2381. doi: 10.1101/gad.1836009

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C. A., Schmitt, A. D., Espinoza, C. A., & Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature, 503*(7475), 290-294. doi: 10.1038/nature12644

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., & Chen, L. (2011). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnology, 30*(1), 90-98. doi: 10.1038/nbt.2057

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., & Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnology, 30*(1), 90-98. doi: 10.1038/nbt.2057

Kee, B. L., & Murre, C. (1998). Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. *Journal of Experimental Medicine, 188*(4), 699-713.

Kieffer-Kwon, K. R., Tang, Z., Mathe, E., Qian, J., Sung, M. H., Li, G., Resch, W., Baek, S., Pruett, N., Grontved, L., Vian, L., Nelson, S., Zare, H., Hakim, O., Reyon, D., Yamane, A., Nakahashi, H., Kovalchuk, A. L., Zou, J., Joung, J. K., Sartorelli, V., Wei, C. L., Ruan, X., Hager, G. L., Ruan, Y., & Casellas, R. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell, 155*(7), 1507-1520. doi: 10.1016/j.cell.2013.11.039

Kikuchi, K., Lai, A. Y., Hsu, C. L., & Kondo, M. (2005). IL-7 receptor signaling is necessary for stage transition in adult B cell development through up-regulation of EBF. *Journal of Experimental Medicine, 201*(8), 1197-1203. doi: 10.1084/jem.20050158

Kim, T. K., & Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. *Cell, 162*(5), 948-959. doi: 10.1016/j.cell.2015.08.008

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., & Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science, 296*(5565), 158-162. doi: 10.1126/science.1068768

Laurenti, E., Doulatov, S., Zandi, S., Plumb, I., Chen, J., April, C., Fan, J. B., & Dick, J. E. (2013). The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nature Immunology, 14*(7), 756-763. doi: 10.1038/ni.2615

Le Dily, F., Bau, D., Pohl, A., Vicent, G. P., Serra, F., Soronellas, D., Castellano, G., Wright, R. H., Ballare, C., Filion, G., Marti-Renom, M. A., & Beato, M. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes & Development, 28*(19), 2151-2162. doi: 10.1101/gad.241422.114

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell, 152*(6), 1237-1251. doi: 10.1016/j.cell.2013.02.014

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Current Biology, 20*(17), R754-763. doi: 10.1016/j.cub.2010.06.070

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., Poh, H. M., Goh, Y., Lim, J., Zhang, J., Sim, H. S., Peh, S. Q., Mulawadi, F. H., Ong, C. T., Orlov, Y. L., Hong, S., Zhang, Z., Landt, S., Raha, D., Euskirchen, G., Wei, C. L., Ge, W., Wang, H., Davis, C., Fisher-Aylor, K. I., Mortazavi, A., Gerstein, M., Gingeras, T., Wold, B., Sun, Y., Fullwood, M. J., Cheung, E., Liu, E., Sung, W. K., Snyder, M., & Ruan, Y. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell, 148*(1-2), 84-98. doi: 10.1016/j.cell.2011.12.014

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C. T., Cubenas-Potts, C., Hu, M., Lei, E. P., Bosco, G., Qin, Z. S., & Corces, V. G. (2015). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Molecular Cell, 58*(2), 216-231. doi: 10.1016/j.molcel.2015.02.023

Li, Q., Peterson, K. R., Fang, X., & Stamatoyannopoulos, G. (2002). Locus control regions. *Blood, 100*(9), 3077-3086. doi: 10.1182/blood-2002-04-1104

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science, 326*(5950), 289-293. doi: 10.1126/science.1181369

Lin, H., & Grosschedl, R. (1995). Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature, 376*(6537), 263-267. doi: 10.1038/376263a0

Lin, Y. C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C. K., & Murre, C. (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. doi: 10.1038/ni.2432

Lin, Y. C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C. A., Dutkowski, J., Ideker, T., Glass, C. K., & Murre, C. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nature Immunology, 11*(7), 635-643. doi: 10.1038/ni.1891

Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., & Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell, 161*(5), 1012-1025. doi: 10.1016/j.cell.2015.04.004

Ma, W., Ay, F., Lee, C., & Gulsoy, G. (2015). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *12*(1), 71-78. doi: 10.1038/nmeth.3205

Mandel, E. M., & Grosschedl, R. (2010). Transcription control of early B cell differentiation. *Current Opinion in Immunology, 22*(2), 161-167. doi: 10.1016/j.coi.2010.01.010

Medina, K. L., Pongubala, J. M., Reddy, K. L., Lancki, D. W., Dekoter, R., Kieslinger, M., Grosschedl, R., & Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. *Developmental Cell, 7*(4), 607-617. doi: 10.1016/j.devcel.2004.08.006

Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T., & Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell, 10*(1), 105-116. doi: 10.1016/j.devcel.2005.10.017

Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. *Cell, 128*(4), 787-800. doi: 10.1016/j.cell.2007.01.028

Murre, C. (2009). Developmental trajectories in early hematopoiesis. *Genes & Development, 23*(20), 2366-2370. doi: 10.1101/gad.1861709

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., & Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature, 502*(7469), 59-64. doi: 10.1038/nature12593

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A., & Dekker, J. (2013). Organization of the mitotic chromosome. *Science, 342*(6161), 948-953. doi: 10.1126/science.1236083

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., & Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell, 169*(5), 930-944.e922. doi: 10.1016/j.cell.2017.05.004

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Bluthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature, 485*(7398), 381-385. doi: 10.1038/nature11049

Nutt, S. L., Heavey, B., Rolink, A. G., & Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature, 401*(6753), 556-562. doi: 10.1038/44076

Nutt, S. L., Urbanek, P., Rolink, A., & Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes & Development, 11*(4), 476-491.

Ong, C. T., & Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews: Genetics, 12*(4), 283-293. doi: 10.1038/nrg2957

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., & de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics, 35*(2), 190-194. doi: 10.1038/ng1244

Peng, C., Fu, L. Y., Dong, P. F., Deng, Z. L., Li, J. X., Wang, X. T., & Zhang, H. Y. (2013). The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. *Nucleic Acids Research, 41*(19), e183. doi: 10.1093/nar/gkt745

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., & Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature, 444*(7118), 499-502. doi: 10.1038/nature05295

Peschon, J. J., Morrissey, P. J., Grabstein, K. H., Ramsdell, F. J., Maraskovsky, E., Gliniak, B. C., Park, L. S., Ziegler, S. F., Williams, D. E., Ware, C. B., Meyer, J. D., & Davison, B. L. (1994). Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *Journal of Experimental Medicine, 180*(5), 1955-1960.

Pongubala, J. M., Northrup, D. L., Lancki, D. W., Medina, K. L., Treiber, T., Bertolino, E., Thomas, M., Grosschedl, R., Allman, D., & Singh, H. (2008). Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nature Immunology, 9*(2), 203-215. doi: 10.1038/ni1555

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature, 470*(7333), 279-283. doi: 10.1038/nature09692

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell, 159*(7), 1665-1680. doi: 10.1016/j.cell.2014.11.021

Reynaud, D., Demarco, I. A., Reddy, K. L., Schjerven, H., Bertolino, E., Chen, Z., Smale, S. T., Winandy, S., & Singh, H. (2008). Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nature Immunology, 9*(8), 927-936. doi: 10.1038/ni.1626

Roessler, S., Gyory, I., Imhof, S., Spivakov, M., Williams, R. R., Busslinger, M., Fisher, A. G., & Grosschedl, R. (2007). Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5. *Molecular and Cellular Biology, 27*(2), 579-594. doi: 10.1128/mcb.01192-06

Rolink, A. G., Nutt, S. L., Melchers, F., & Busslinger, M. (1999). Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. *Nature, 401*(6753), 603-606. doi: 10.1038/44164

Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L., & Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genetics, 43*(4), 379-386. doi: 10.1038/ng.790

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S., & Gilbert, D. M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Research, 20*(6), 761-770. doi: 10.1101/gr.099655.109

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature, 489*(7414), 109-113. doi: 10.1038/nature11279

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B. M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S. W., Dimitrova, E., Dimond, A., Edelman, L. B., Elderkin, S., Tabbada, K., Darbo, E., Andrews, S., Herman, B., Higgs, A., LeProust, E., Osborne, C. S., Mitchell, J. A., Luscombe, N. M., & Fraser, P. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Research, 25*(4), 582-597. doi: 10.1101/gr.185272.114

Scott, E. W., Simon, M. C., Anastasi, J., & Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science, 265*(5178), 1573-1577.

Seet, C. S., Brumbaugh, R. L., & Kee, B. L. (2004). Early B Cell Factor Promotes B Lymphopoiesis with Reduced Interleukin 7 Responsiveness in the Absence of E2A. *Journal of Experimental Medicine, 199*(12), 1689-1700. doi: 10.1084/jem.20032202

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., & Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature, 488*(7409), 116-120. doi: 10.1038/nature11243

Simon, J. A., & Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature Reviews: Molecular Cell Biology, 10*(10), 697-708. doi: 10.1038/nrm2763

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., & de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics, 38*(11), 1348-1354. doi: 10.1038/ng1896

Singh, H., Medina, K. L., & Pongubala, J. M. (2005). Contingent gene regulatory networks and B cell fate specification. *Proc Natl Acad Sci U S A, 102*(14), 4949-4953. doi: 10.1073/pnas.0500480102

Singh, H., Pongubala, J. M., & Medina, K. L. (2007). Gene regulatory networks that orchestrate the development of B lymphocyte precursors. *Advances in Experimental Medicine and Biology, 596*, 57-62. doi: 10.1007/0-387-46530-8_5

Souabni, A., Cobaleda, C., Schebesta, M., & Busslinger, M. (2002). Pax5 promotes B lymphopoiesis and blocks T cell development by repressing Notch1. *Immunity, 17*(6), 781-793.

Spitz, F., Gonzalez, F., & Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell, 113*(3), 405-417.

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., & Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Research, 24*(3), 390-400. doi: 10.1101/gr.163519.113

Tanabe, H., Habermann, F. A., Solovei, I., Cremer, M., & Cremer, T. (2002). Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutation Research, 504*(1-2), 37-45.

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., Frum, T., Giste, E., Johnson, A. K., Johnson, E. M., Kutyavin, T., Lajoie, B., Lee, B. K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Safi, A., Sanchez, M. E., Sanyal, A., Shafer, A., Simon, J. M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M. O., Hansen, R. S., Navas, P. A., Stamatoyannopoulos, G., Iyer, V. R., Lieb, J. D., Sunyaev, S. R., Akey, J. M., Sabo, P. J., Kaul, R., Furey, T. S., Dekker, J., Crawford, G. E., & Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature, 489*(7414), 75-82. doi: 10.1038/nature11232

Urbanek, P., Wang, Z. Q., Fetka, I., Wagner, E. F., & Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell, 79*(5), 901-912.

Wang, X. T., Dong, P. F., Zhang, H. Y., & Peng, C. (2015). Structural heterogeneity and functional diversity of topologically associating domains in mammalian genomes. *Nucleic Acids Research, 43*(15), 7237-7246. doi: 10.1093/nar/gkv684

Yoshida, T., Ng, S. Y., Zuniga-Pflucker, J. C., & Georgopoulos, K. (2006). Early hematopoietic lineage restrictions directed by Ikaros. *Nature Immunology, 7*(4), 382-391. doi: 10.1038/ni1314

Zhang, Z., Cotta, C. V., Stephan, R. P., deGuzman, C. G., & Klug, C. A. (2003). Enforced expression of EBF in hematopoietic stem cells restricts lymphopoiesis to the B cell lineage. *EMBO Journal, 22*(18), 4759-4769. doi: 10.1093/emboj/cdg464

Probing changes in spatial organization of the genome during B cell fate commitment

ORIGINALITY REPORT

55%	13%	53%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ravi Boya, Anurupa Devi Yadavalli, Sameena Nikhat, Sreenivasulu Kurukuti, Dasaradhi Palakodeti, Jagan M. R. Pongubala. "Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment", Nucleic Acids Research, 2017 Publication	43%
2	repub.eur.nl Internet Source	1%
3	media.proquest.com Internet Source	1%
4	Submitted to University of Dehli Student Paper	1%
5	researchspace.auckland.ac.nz Internet Source	<1%
6	Current Topics in Microbiology and Immunology, 2014. Publication	<1%

Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx722

Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment

Ravi Boya[1,†], Anurupa Devi Yadavalli[1,†], Sameena Nikhat[1], Sreenivasulu Kurukuti[1], Dasaradhi Palakodeti[2] and Jagan M. R. Pongubala[1,*]

[1]Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad 500046, India and [2]Institute for Stem Cell Biology and Regenerative Medicine, National Centre for Biological Sciences, Bangalore 560065, India

Received May 25, 2017; Revised July 21, 2017; Editorial Decision August 07, 2017; Accepted August 09, 2017

ABSTRACT

Genome organization in 3D nuclear-space is important for regulation of gene expression. However, the alterations of chromatin architecture that impinge on the B cell-fate choice of multi-potent progenitors are still unclear. By integrating *in situ* Hi-C analyses with epigenetic landscapes and genome-wide expression profiles, we tracked the changes in genome architecture as the cells transit from a progenitor to a committed state. We identified the genomic loci that undergo developmental switch between A and B compartments during B-cell fate determination. Furthermore, although, topologically associating domains (TADs) are stable, a significant number of TADs display structural alterations that are associated with changes in *cis*-regulatory interaction landscape. Finally, we demonstrate the potential roles for Ebf1 and its downstream factor, Pax5, in chromatin reorganization and transcription regulation. Collectively, our studies provide a general paradigm of the dynamic relationship between chromatin reorganization and lineage-specific gene expression pattern that dictates cell-fate determination.

INTRODUCTION

It is increasingly evident that the assembly of higher-order genome structures and their associated sub-nuclear compartments are intimately linked with transcriptional activity (1,2). Recent advances in high-throughput Chromosome Conformation Capture (3C)-derived methods have enabled quantitative measurement of physical interactions of chromatin in 3D nuclear space (2–6). These studies have demonstrated that chromatin is organized into transcriptionally permissive (A) and repressive (B) compartments indicating that chromatin positioning in 3D nuclear space may be associated with gene activity. For instance, B cell specification is associated with relocalization of *Igh* alleles from the nuclear periphery (a repressive compartment) towards center of the nucleus (an active compartment), where they undergo long-range interactions and subsequent rearrangements (7–9). These findings provide a functional link between sub-nuclear localization of the chromatin and gene activity. Recent studies indicate that chromatin compartments are further organized into varying sizes of dense and highly self-interacting regions, known as Topologically Associating Domains (TADs). These chromatin domains have been found to be stable and conserved across various cell types (10). In mammalian cells, insulator binding protein, CTCF, is found to be enriched in TAD boundaries (10). The deletion of boundary regions results in an increase in inter-domain interactions indicating the structural and functional role of insulators in maintenance of discrete, functional chromatin domains (11,12). Further it was demonstrated that loss of CTCF results in dose dependent insulation defects at most of the TAD boundaries (13). However, recent studies suggest that depletion of cohesin-loading factor *Nipbl*, but not CTCF, results in genome-wide disappearance of TADs, reinforcing the critical role of cohesin in the formation of TADs by loop extrusion mechanism (Schwarzer et al., 2016; Kubo et al., 2017, Unpublished). Although TADs are invariant, the intrinsic interactions within these TADs were found to be varying (10,14). Moreover, several studies show that the cell type-specific gene expression is regulated through interactions between promoters and distantly located *cis*-regulatory elements, particularly enhancers, by looping out of intervening DNA sequences (15–18). These long-range interactions were found to be associated with changes in histone modifications and DNA methylation (19–21). Furthermore, the transcriptional output is controlled by a combinatorial binding of transcription factors at *cis*-regulatory elements (22–25). Thus, a number of molecular mechanisms contribute to the precise regulation of gene expression pat-

*To whom correspondence should be addressed. Tel: +91 40 23134580; Fax: +91 40 23010145; Email: jpsl@uohyd.ernet.in
†These authors contributed equally to this work as first authors.