

Regulatory interplay between chromatin organization and lineage determinants during B cell specification

A thesis

Submitted for the degree of

Doctor of Philosophy

By

Anurupa Devi Yadavalli

(Reg. No. 14LAPH11)



Department of Animal Biology

School of Life Sciences

University of Hyderabad

June 2022



UNIVERSITY OF HYDERABAD

SCHOOL OF LIFE SCIENCES

Jagan Pongubala, Ph.D.
Department of Animal Biology

Tel. No: 40-23134580
Email: jpsl@uohyd.ernet.in

CERTIFICATE

This is to certify that the thesis entitled “**Regulatory interplay between chromatin organization and lineage determinants during B cell specification**” submitted by **Ms. Anurupa Devi Yadavalli** bearing registration number **14LAPH11** is her original work carried out by her under my guidance and thus can be submitted for partial fulfilment of therequirements for the award of ‘Doctor of philosophy’ from the University of Hyderabad. I attest that her thesis is free of plagiarism (has not been submitted previously for the award of any degree or diploma).Her thesis work has been published in reputable Journals and presented at conferences.

Selected list of publications:

Boya, R., **Yadavalli AD**, Nikhat, S, Kurukuti, S., Palakodeti, D., and Pongubala, JMR. (2017), Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment, *Nucleic Acids Res.*, 2017 Nov 2;45(19):11070-11087. doi: 10.1093/nar/gkx722 (**Co-first author**).

Helen Beilinson, Rebecca Glynn, **Anurupa Devi Yadavalli**, Jianxiong Xiao, Elizabeth Corbett, HuseyinSaribasak, Rahul Arya, Charline Miot, Anamika Bhattacharyya, Jessica Jones, JaganPongubala, Craig Bassing, and David Schatz. (2021), The RAG1 N-terminal region regulates the efficiency and pathways of synapsis for V(D)J recombination, *Journal of Exp. Med.*, 2021 Oct 4;218(10): e20210250. doi: 10.1084/jem.20210250.

Selected list of conferences:

Ravi Boya, **Anurupa DeviYadavalli**,JaganPongubala. Decoding chromatin dynamics during B cell fate commitment. International conference on Genome Architecture and Cell Fate Regulation (GACFR), University of Hyderabad, Hyderabad, India, 2014.

Anurupa Devi Yadavalli, Ravi Boya, Jagan M.R. Pongubala. Decoding chromatin dynamics during B cell fate commitment. *Bioquest-2015, University of Hyderabad, Hyderabad, India*, 2015.

Ravi Boya,**Anurupa Devi Yadavalli**, SameenaNikhat, Jagan M.R. Pongubala. Programmed changes in higher-order chromatin interactions dictate B-cell fate commitment. 6th Meeting of Asian Forum of Chromosome and Chromatin Biology, Centre for Cellular and Molecular Biology (CCMB), Hyderabad, India, 2017.

Further, the student has successfully completed the course work

<i>Course Code</i>	<i>Name</i>	<i>Credits</i>	<i>Pass/Fail</i>
AS 801	Analytical Techniques	4	Pass
AS 802	Research Ethics, Data Analysis and Biostatistics	3	Pass
AS 803	Lab Work and Seminar	5	Pass


20/06/2022
Supervisor
I. M. R. M. Pongubala
Department of Animal Biology
School of Life Sciences
University of Hyderabad
Gachibowli, Hyderabad-500 046.


Head of the Department
अध्यक्ष / HEAD I/C
जंतु जैविकी विभाग
Department of Animal Biology


Dean
DEAN
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

University of Hyderabad

(A Central University by an Act of Parliament)

Department of Animal Biology

School of Life Sciences

P.O. Central University, Gachibowli, Hyderabad-500046



DECLARATION

I, Anurupa Devi Yadavalli (14LAPH11), hereby declare that this thesis entitled "Regulatory interplay between chromatin organization and lineage determinants during B cell specification" submitted by me under the guidance of Prof. JaganPongubala is an Original and independent research work. I also declare that it has not been submitted previously in part or in full to this University or any other University/Institute for the award of any degree or diploma.

Name: Anurupa Devi Yadavalli

Reg. No: 14LAPH11

Y. Anurupa devi
22-06-2022

Signature of the student

Acknowledgements

I thank my supervisor, Prof. Jagan Pongubala, for his guidance, encouragement, and support, throughout my Ph.D program.

I am thankful to Dr. David G. Schatz, for all his support and encouragement during my stay at Yale as an exchange visiting scholar. I also would like to thank all the members of the Schatz Lab -Helen, Rashu, Liz, Lizhen, and Miriam for the laughs, the science chat, and everything else.

I am thankful to my Doctoral Committee members, Dr. Naresh Babu Sepuri and Dr. Noorudhin Khan, for their guidance and helpful suggestions.

I am thankful to Prof. Srinivasulu Kurukuti, Head, Department of Animal Biology, and the former Heads, Prof. Anita Jagota, Prof. Jagan Pongubala, and Prof. B. SenthilKumaran for allowing me to use departmental facilities.

I am thankful to Prof. S. Dayananda, Dean, School of Life Sciences and the former Deans, Prof. P. Reddanna and Prof. K.V. Ramaiah for allowing me to use the school facilities.

I would like to thank Dr. Dasaradhi Palakodeti from InStem (The Institute of Stem Cell Science and Regenerative Medicine, Bengaluru) for his guidance, support, and valuable advice during the course of my work.

A special thanks to my dear friends Ms. Annapoorni and Ms. Anita who have been my support systems during troubled times.

I would like to thank my former lab members, Dr. Ravi Boya and Dr. Sameena Nikhat, for their helpful suggestions during my early days at the lab. I would like to thank my current lab members, Ms. Arpita Prusty, Mr. Ashok Kumar, Ms. Nidhi Singh, and Ms. Apoorva Soni, for their help and support in the lab, and for all the happy times we have spent together.

A special thanks to my dear friends Ms. Anubhooti and Ms. Priyanka Kriti Narayan, who made every difficult day bearable and every happy day worth celebrating.

I would like to thank all previous and current non-teaching staff of Dept. of Animal Biology: Mr. Ankineedu, Mrs. Jyoti, Mrs. Vijaya Lakshmi, Mr. Nikhil, Mr. Jagan, Mr. Gopi, Mr. Pandu, Mr. Rangaswamy, and Mr. Srinivas, for their help in office and administrative work.

I sincerely thank CSIR for funding my doctoral studies and DBT, DST, UGC – SAP and DBT – CREBB for providing financial support to the lab and for the infrastructural facilities of the school.

I am so gratified to have had the support and encouragement from my parents, Mr. Bhagavath Sastry and Mrs. Rajeswari, all my life. I am obliged to them for their unequivocal love, advice, prayers, and their earnest efforts in teaching and inculcating virtues and good morals. I am so grateful to my sister, Teja and brother-in-law Mr. Ravindranath, who are no less than parents to me in every possible way.

Words fail to express my gratitude towards all the members of the COTR, family especially Ms. Mary (Mema), and Ps. Kearse for supporting and encouraging me through their prayers during my stay at New Haven.

I cherish and treasure the company and affection of my dear friends Ms. Ellen Rachael, Ms. Eva Justin. Finally, Ellen and Eva, the day for which you guys prayed has arrived! I am thankful to Dr. Afsheen Nasir, who has been my confidante and my good friend. Afsheen, your calmness and smile always made me believe that everything would work out, even when it felt like it never would.

Above and beyond all, I am ever grateful to the GOD Almighty, for giving me the health, strength, and courage to endure this colossal journey, and for placing all the above people in my life to help and guide me in one way or another.

LIST OF FIGURES

FIGURE 1.1 SCHEMATIC REPRESENTATION OF HEMATOPOIESIS: THE DEVELOPMENT OF DIFFERENT BLOOD CELLS FROM HEMATOPOIETIC STEM CELLS.	1
FIGURE 1.2 PLURIPOTENT STEM CELLS DIFFERENTIATE INTO VARIOUS CELL TYPES OF IMMUNE SYSTEM IN RESPONSE TO PRIMARY SITES OF ACTION OF THE VARIOUS GROWTH FACTORS.	2
FIGURE 1.3 SCHEMATIC ILLUSTRATION OF TRANSCRIPTION FACTORS REQUIRED AT EACH B CELL DEVELOPMENTAL STAGES AS IDENTIFIED BY GENE ABALATION STUDIES.	4
FIGURE 1.4 B CELL GENE REGULATION NETWORK ASSEMBLED FROM VARIOUS STUDIES.	5
FIGURE 1.5 SUMMARY OF EPIGENETIC CHANGES THAT OCCUR DURING B CELL DEVELOPMENT.	6
FIGURE 1.6 THE ORGANIZATION OF THE EUKARYOTIC GENOME.	9
FIGURE 2.1 THE SCHEMATIC REPRESENTATION OF IN SITU HI-C METHOD.	12
FIGURE 2.2 HI-C PIPELINE FOR IDENTIFICATION OF CHROMATIN INTERACTOME.	13
FIGURE 2.3 FLOW CHART REPRESENTING STEPWISE PROCEDURE INVOLVED IN IDENTIFYING THE TRANSCRIPTIONAL REGULATORS.	16
FIGURE 2.4 FLOW CHART REPRESENTING STEPS INVOLVED IN BULK RNA-SEQ DATA ANALYSIS.	17
FIGURE 2.5 FLOW CHART REPRESENTING STEPS INVOLVED IN CHIP-SEQ DATA ANALYSIS.	18
FIGURE 2.6 SAMPLE-WISE ALIGNMENT PERCENTAGES OF CHIP-SEQ DATA USED IN THIS STUDY.	19
FIGURE 2.7 FLOW CHART OF HTGTS-REP-SEQ PROTOCOL.	21
FIGURE 2.8 FLOW CHART REPRESENTING THE EXPERIMENT DESIGN AND THE WORKFLOW FOR MEASSURING THE VDJ RECOMBINATION EVENTS.	22
FIGURE 3.1 CHROMATIN COMPARTMENTALIZATION IS CLOSELY ASSOCIATED WITH GENE ACTIVITY DURING PRE-PRO-B TO PRO-B CELL DEVELOPMENT.	26
FIGURE 3.2 GENETIC SWITCH BETWEEN A AND B COMPARTMENTS HAS DECREASED DURING PRO-B TO PRE-B CELL DEVELOPMENT	32
FIGURE 3.3 THE 3D MODELING VALIDATES THAT GENES UNDERGO NUCLEAR RE-LOCALIZATION DURING PRE-PRO-B TO PRO-B CELL TRANSITION.	35
FIGURE 3.4 THE 3D MODELING VALIDATES THAT GENES UNDERGO NUCLEAR RE-LOCALIZATION DURING PRO-B TO PRE-B CELL TRANSITION.	36
FIGURE 3.5 TRANSCRIPTIONAL REGULATORS ARE RESPONSIBLE FOR GENETIC SWITCH.	38
FIGURE 3.6 EBF1 AND PAX5 PROMOTE B CELL DEVELOPMENT BY ACTIVATING B-LINEAGE SPECIFIC GENES AND SILENCING ALTERNATE LINEAGE GENES.	40
FIGURE 3.7 EBF1 REGULATES B LINEAGE-SPECIFIC GENE EXPRESSION PATTERN IN PART BY BINDING AT CIS-REGULATORY INTERACTING ELEMENTS.	50
FIGURE 3.8 COVERAGE PLOT SHOWING THE H3K27AC ENRICHMENT IN EBF1-/- AND EBF1-ER. THE ENRICHMENT OF H3K27AC IN EBF1-/-ER CORRELATES IT'S ENRICHMENT IN RAG2-/- PROGENITORS	52
FIGURE 3.9 CIS-REGULATORY INTERACTION LANDSCAPE DETERMINES DIFFERENTIAL GENE EXPRESSION PATTERN.	53
FIGURE 3.10 DH GENE SEGMENT USAGE IN IGH REARRANGEMENTS IN BONE MARROW OBTAINED BY HTGTS-REP-SEQ.	58
FIGURE 3.11 DH GENE SEGMENT USAGE AND ANALYSIS OF PRODUCTIVE AND NONPRODUCTIVE REARRANGEMENTS IN VDJH REARRANGEMENTS IN SPLEEN OBTAINEDBY HTGTS-REP-SEQ.	60
FIGURE 3.12 D _H AND J _H GENE SEGMENT USAGE IN VDJ _H REARRANGEMENTS IN BONE MARROW OBTAINED BY IMMUNOSEQ.	64

LIST OF TABLES

TABLE 2.1 LIST OF HI-C DATA SETS OBTAINED FOR THIS STUDY. EACH DATASET REPRESENTS KNOCKOUT OF A B CELL-SPECIFIC TRANSCRIPTION FACTOR AT DEFINITE STAGE OF THE B CELL DEVELOPMENT.	11
TABLE 2.2 PAIRED-END READ COUNTS AND QUALITY STATISTICS FOR EACH CHIP-SEQ SAMPLE USED IN THIS STUDY.	19
TABLE 3.1 LIST OF DIFFERENTIALLY EXPRESSED GENES THAT ARE SWITCHING FROM B TO A COMPARTMENTS DURING PROGRESSION OF PRE-PRO-B CELLS TO PRO-B CELLS.	27
TABLE 3.2 LIST OF DIFFERENTIALLY EXPRESSED GENES THAT ARE SWITCHING FROM A TO B COMPARTMENTS DURING PROGRESSION OF PRE-PRO-B CELLS TO PRO-B CELLS.	29
TABLE 3.3 LIST OF DIFFERENTIALLY EXPRESSED GENES THAT ARE SWITCHING FROM A TO B COMPARTMENTS DURING PROGRESSION OF PRO-B CELLS TO PRE-B CELLS.	33
TABLE 3.4 LIST OF DIFFERENTIALLY EXPRESSED GENES THAT ARE SWITCHING FROM B TO A COMPARTMENTS DURING PROGRESSION OF PRO-B CELLS TO PRE-B CELLS.	34
TABLE 3.5 LIST OF DIFFERENTIALLY EXPRESSED GENES (LOG2 FC 4; $P < 0.05$) OBTAINED BY ECTOPIC EXPRESSION OF Ebf1 IN Ebf1-/- CELLS	41
TABLE 3.6 LIST OF DIFFERENTIALLY EXPRESSED GENES (LOG2 FC 4; $P < 0.05$) OBTAINED BY ECTOPIC EXPRESSION OF PAX5 IN PAX5-/- CELLS	45
TABLE 3.7 SELECTED LIST OF PROMOTER-ENHANCER INTERACTIONS IN RAG2-/- WITH TRANSCRIPTION FACTOR BINDING SITES	51
TABLE 3.8 SUPER ENHANCER INTERACTION LANDSCAPE OF GENE PROMOTERS THAT ARE CAPTURED IN BOTH Ebf1-/-EBF1-ER AND RAG2-/- PROGENITORS.	54
TABLE 3.9 UTILIZATION OF D _H GENE SEGMENTS AMONG TOTAL VDJ REARRANGEMENTS IN PRIMARY B220+ IgM- BONE MARROW CELLS.	59
TABLE 3.10 UTILIZATION OF D _H GENE SEGMENTS AMONG DJ REARRANGEMENTS IN PRIMARY B220+IgM- BONE MARROW CELL	61
TABLE 3.11 UTILIZATION OF D _H GENE SEGMENTS AMONG TOTAL VDJ REARRANGEMENTS IN PRIMARY B220+ SPLENOCYTES	61
TABLE 3.12 UTILIZATION OF DH GENE SEGMENTS AMONG DJ REARRANGEMENTS IN PRIMARY B220+ SPLENOCYTES	62

LIST OF ABBREVIATIONS

HSC	Hematopoietic Stem Cells
MPP	Multipotent Progenitors
LMPP	Lymphoid primed Multipotent Progenitors
CLP	Common Lymphoid Progenitors
CD19	Cluster of Differentiation 19
EBF1	Early B cell Factor 1
PAX5	Paired Box 5
RAG	Recombination Activating Gene
BCR	B Cell Receptor
FISH	Fluorescence In Situ Hybridization
3C	Chromosome Conformation Capture
4C	Chromosome Conformation Capture-on-Chip
5C	Chromosome Conformation Capture Carbon Copy
2D	Two Dimensional
3D	Three Dimensional
TADs	Topologically Associating Domains
RC	Recombination center
RSS	Recombination Signal Sequence
IGH	Immunoglobulin Heavy chain
PCA	Principal Component Analysis
LORDG	Lorentzian 3D Genome
ICE	Iterative Correction and Eigenvector Decomposition
PC1	Principal Component 1
ChIP	Chromatin Immuno Precipitation
NTR	N-terminal Region
IMGT	International ImMunoGeneTics Information system
BLAST	Basic Local Alignment Search Tool
HTGTS	High-Throughput Genome-wide Translocation Sequencing

Contents

1. INTRODUCTION	1
1.1 The Hematopoietic System	1
1.2 B Lymphocyte development	2
1.3 Transcription factors that dictate B cell fate specification and commitment	3
1.4 Epigenetic regulation of early B cell differentiation	6
1.5 Spatial organization of genome in B cell specification and commitment	7
1.6 V(D)J Recombination and the B lymphocyte	9
1.7 Objectives	10
2. Materials and methods	11
2.1 Data acquisition	11
2.2 Hi-C data analysis	11
2.2.1 Iterative Mapping	12
2.2.2 Filtering spurious ligation products	12
2.2.3 Filtering of PCR duplicates and extreme fragments	13
2.2.4 Quality check for library size	14
2.2.5 Generation of relative contact probability matrices using iterative correction	14
2.2.6 Eigen vector decomposition for chromatin compartmentalization	14
2.3 Three-dimensional chromatin structure generation and visualization	15
2.4 Binding analysis for regulation of transcription using Hi-C	16
2.5 Genome-wide transcriptome data analysis	16
2.5.1 Data acquisition	16
2.5.2 Sample clustering and PCA analysis	17
2.6 ChIP-Seq data analysis	17
2.6.1 Data acquisition	18
2.6.2 Determination of statistically significant <i>cis</i> -regulatory interactions	20
2.7 HTGTS-Rep-seq B cell repertoire	20
3. Results	23
3.1 Chromatin undergoes genetic switch between A and B compartments during B cell developmental transitions	23
3.1.1 Chromatin compartmentalization in E2A ^{-/-} Ebf1 ^{-/-} (pre-pro-B) and Rag2 ^{-/-} (pro-B)	24
3.1.2 Chromatin compartmentalization in Pax5 ^{-/-} (pro-B) and Pax5-ER (pre-B) cells.	31

3.1.3 Validation of genetic switching using LORDG based 3D modeling at various resolutions	34
3.1.4 The molecular basis for genetic switch	36
3.2 Interplay between Ebf1 and Pax5 establishes B-lineage signature	39
3.2.1 Ebf1 governs B lineage specification and commitment in positive feedback manner through Pax5	39
3.2.2 Ebf1 functions as a major B cell architectural factor	49
3.2.3 Ebf1 regulates B lineage-specific gene expression pattern in part by modulating cis-regulatory interactions	52
3.3 Igh repertoire undergoes alterations in R1Δ215 and R1.PG but not in R1.KR mice	56
3.3.1 Reduced secondary D-to-JH recombination contributes to the R1.PG Igh repertoire phenotype	62
4. Discussion	65
5. Summary and Conclusions	69
6. References	72

1. INTRODUCTION

1.1 The Hematopoietic System

Hematopoiesis is the continuous process by which cellular components of the immune system are formed. It is regulated by both stochastic and deterministic mechanisms (Kimmel 2014), resulting in hematopoietic stem cells (HSCs) that possess an immense self-renewal capacity and the ability to differentiate into all mature blood lineages upon loss of self-renewing capacity. They are programmed to allow efficient production of all the cellular components of the blood that serve a variety of functions: from RBCs that serve to carry oxygen, megakaryocytes (which generate platelets) that regulate blood clotting; to the innate as well as adaptive immune cells that protect against infections and cancers (Figure 1.1).

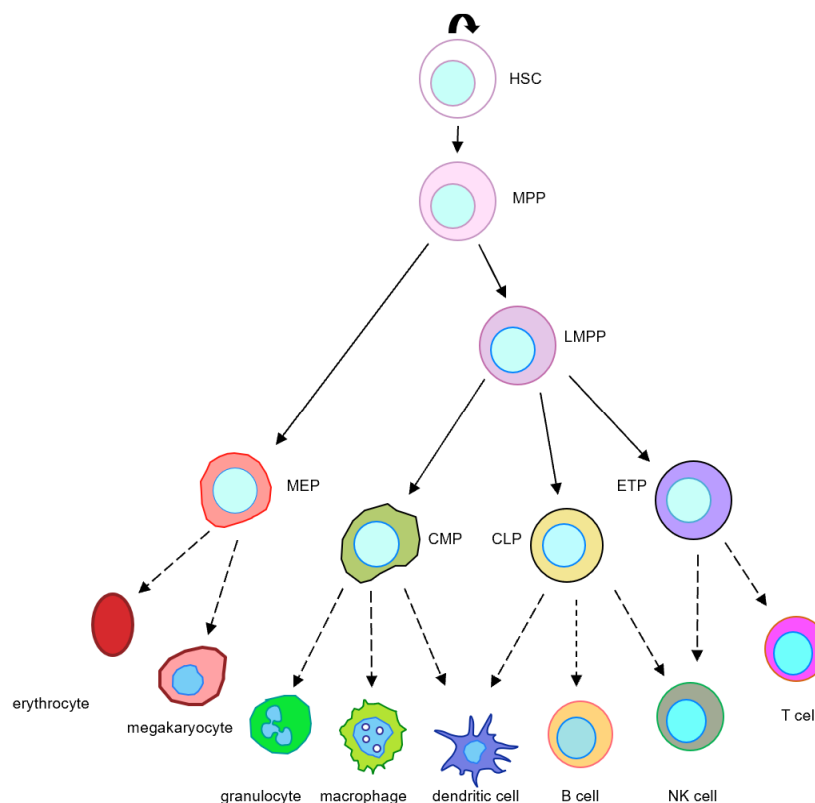


Figure 1.1 Schematic representation of Hematopoiesis: The development of different blood cells from hematopoietic stem cells.

The hematopoietic stem cell niche is extremely complex. It promotes HSC function by facilitating the survival and long-term maintenance of the HSC pool. The mechanisms by which interactions within the niche shape the differentiation journeys and fate of

individual HSC subsets remain unclear. Growth factors have been demonstrated to induce diverse cell fate option of multipotent (MPP) progenitors promote the developmental transitions based on the combination of factors and the stage of differentiation of the cell (Figure 1.2).

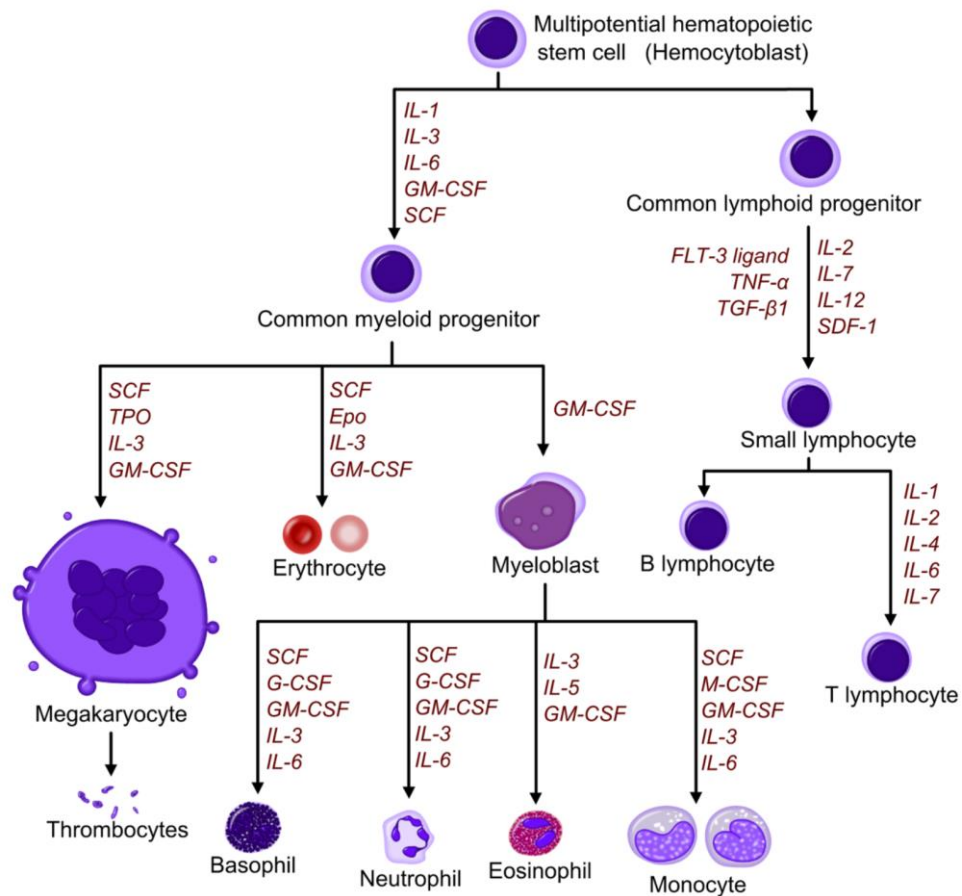


Figure 1.2 Pluripotent stem cells differentiate into various cell types of immune system in response to primary sites of action of the various growth factors.

1.2 B Lymphocyte development

Lymphopoiesis (lymphocyte development) is a recursive process of cellular differentiation beginning with HSCs. Self-renewing HSCs differentiate into multipotent progenitors (MPPs), characterized by loss of self-renewal capacity and acquisition of Flt3 tyrosine kinase expression. Based on Flt3 receptor expression, it was demonstrated that MPPs that are Flt3⁻ differentiate preferentially along the erythroid/megakaryocyte pathway, whereas Flt3⁺ MPPs have significantly reduced megakaryocyte and erythrocyte potential and give rise predominantly to lymphoid (B and T) and myeloid lineages (macrophages and granulocytes) (Adolfsson et al. 2005). According to these findings, MPPs

initially undergo a binary decision to differentiate into a megakaryocyte/erythroid (MEP) progenitor and a lymphoid/myeloid multipotential progenitor (LMPP). LMPP then differentiate into ETP (Early thymic progenitors), CLP (Common lymphoid progenitors) or GMP (Granulocyte macrophage progenitor). The pre-pro-B cells are generated from CLP populations and differentiate into pro-B cells, which are characterized by the up regulation of B lineage cell surface markers such as CD19 and the immunoglobulin (Ig) surrogate light chain genes: VpreB and $\lambda 5$, as well as initiation of Ig heavy chain gene rearrangement. Pre-B cells express the pre-B cell receptor (pre-BCR) and further differentiate into immature B cells upon Ig light chain gene rearrangement and migrate from the bone marrow to the spleen.

1.3 Transcription factors that dictate B cell fate specification and commitment

Gene ablation studies have revealed several transcription factors, including Ikaros, PU.1, E2A, Ebf1, and Pax5 have been implicated in the transcription of lymphoid progenitors into the B lineage precursors (Georgopoulos et al. 1994; Lin and Grosschedl 1995; Peschon et al. 1994; Pongubala et al. 2008; Singh, Medina, and Pongubala 2005). Because the transcription factors E2A and Ebf1 are necessary for the B cell specification, they are considered as the primary B cell fate determinants. They control the early program of B lineage gene expression by induction of lineage specific genes (mb-1, B29, $\lambda 5$, VpreB, and Rag-1,2) and suppressing the lineage-inappropriate genes (Zbtb16, Tox, Id2, Gata3). E2A is a basic helix–loop–helix transcription factor that has two splice variants, E12 and E47 (Bain et al. 1994). E2A is essential for induction and maintenance of expression of Ebf1, Pax5 and the B cell specific program at pro-B cell stage, as well as directing B cell maturation in germinal centers (Kwon et al. 2008). Additionally, E2A promotes the generation of LMPPs by antagonizing the expression of a key erythroid and megakaryocytic lineage determinant (GATA.1) (Ikawa et al. 2004). Early B-cell factor (Ebf1) has been found to restrict alternative lineage 'choice' to promote B cell fate commitment in a Pax5 independent manner (Pongubala et al. 2008). Early reports of Ebf1 have shown that its transcription is controlled by two promoters that are differentially regulated in B cells (Roessler et al. 2007). The distal (α) promoter becomes active in response to interleukin-7

signaling (via STAT5), and binding of E47. The proximal (β) promoter is up regulated by Pax5, Ets1 and PU.1. As a result, Ebf1 activation of the Pax5 gene promotes Ebf1 transcription via a positive feedback loop.

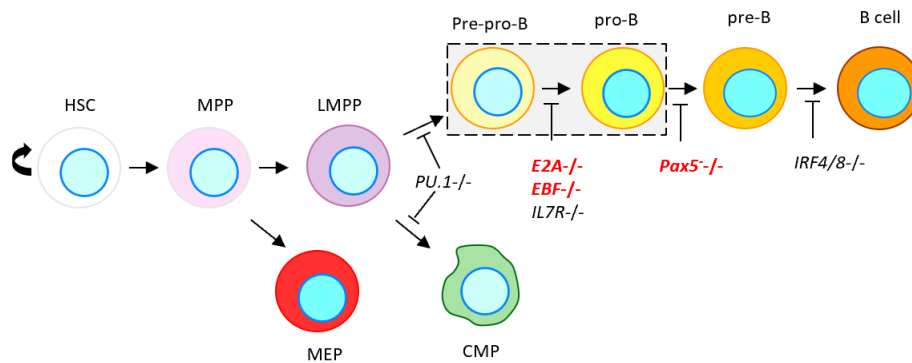


Figure 1.3 Schematic illustration of transcription factors required at each B cell developmental stages as identified by gene ablation studies.

Ebf1 maintains E47 activity by another positive feedback loop through down regulation of Id2 and Id3 (Thal et al. 2009).

Because ectopic expression of Ebf1 is sufficient to overcome developmental block in E2A, IL7, IL7-Ra, Ikaros and Pu.1 mutant progenitors, Ebf1 is considered as a major determinant of the B cell fate (Kikuchi et al. 2005; Medina et al. 2004; Reynaud et al. 2008). The ability of Ebf1 to restore the B cell program is due in part to the fact that Ebf1 target genes encode proteins essential for survival signals that mediate developmental checkpoints. Biochemical studies and microarray analysis revealed functionally important Ebf1 binding sites in promoters of the Cd79a (mb-1; Ig α), Cd79b (B29; Ig β), B lymphoid kinase (Blk), Vpreb1, Igll1 (λ 5) Cd19, Pax5, Cd53, Pou2af and Foxo1 genes (Gisler, Akerblad, and Sigvardsson 1999; Sigvardsson, O'Riordan, and Grosschedl 1997). One of the hallmarks of Ebf1 function is its cooperation with other regulators of B cell development. The well documented regulatory pathway involves functional interactions between Ebf1 and E2A. The two proteins bind together on the promoters of the Cd79a, Vpreb1 and Igll1 genes. Ebf1 and E47 also function cooperatively in the activation of V(D)J recombination. Consistent with these findings, Ebf1^{+/-} and E2a^{+/-} mice display a marked defect in pro-B

cell differentiation at a stage later than that observed in the single homozygous mutant mice (O'Riordan and Grosschedl 1999). Moreover, Ebf1 acts synergistically with E2a, Runx1 and Pax5 to stimulate transcription of the mb-1 and Cd19 genes (Sigvardsson et al. 2002). Previous studies revealed a novel role of Ebf1 in B cell commitment independently of pax5 (Pongubala et al. 2008).

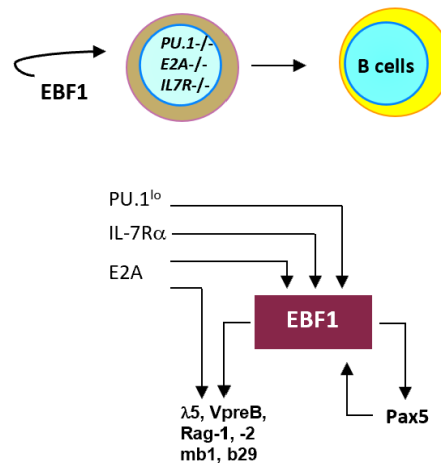


Figure 1.4 B cell gene regulation network assembled from various studies. Prototype of the network is taken from (Medina et al. 2004; Singh, Medina, and Pongubala 2005).

Ebf1 antagonized expression of genes encoding the transcription factors C/EBPα, PU.1 and Id2, all of which are essential for myeloid lineage development. In addition to this, Ebf1 also down regulates many of the genes that are repressed by pax5 and activated by PU.1, independently of pax5. Therefore, Ebf1 is considered as a crucial component of the regulatory network that determines the destiny of B cells vs myeloid cells. Pax5, the major determinant of B cell precursor commitment (Schebesta, Heavey, and Busslinger 2002) is regulated by Ebf1. Pax5 is a transcription factor of the paired box (PAX) family. Its expression is first detected at the early pro-B cell stage and maintained until the mature B cell stage. It is essential for repressing alternate lineage genes. Pax5 inhibits T cell fate by suppressing Notch-1 expression (Nutt et al. 1999). In B-lineage cells, Pax5 actively and continuously represses the expression of various myeloid genes, including M-CSFR in B lineage cells. Pax5 is also essential in reinforcing and expanding the early program of B lineage gene expression by regulating the expression of B cell-specific genes such as Ebf1, CD19, mb-1, Blnk, IglI5, and VpreB1. Collectively, these findings clearly demonstrate that differentiation of multipotent progenitors into committed B-cells is facilitated by a

complex and hierarchical gene regulatory network comprised of various signaling molecules and lineage-specific transcription factors (**Figure 1.4**).

1.4 Epigenetic regulation of early B cell differentiation

An ample amount of evidence suggests that cell-type specific gene expression is controlled by chromatin remodeling factors that would either unwind (Active) or compact (Inactive) the chromatin (Tamaru 2010). Context-specific gene activation necessitates the recruitment of transcriptional machinery to their promoter regions, which is only possible if the chromatin is accessible. Thus, a lot of focus has been shifted towards studying the epigenetic mechanisms that would lead to genetic switching and eventually impact gene expression.

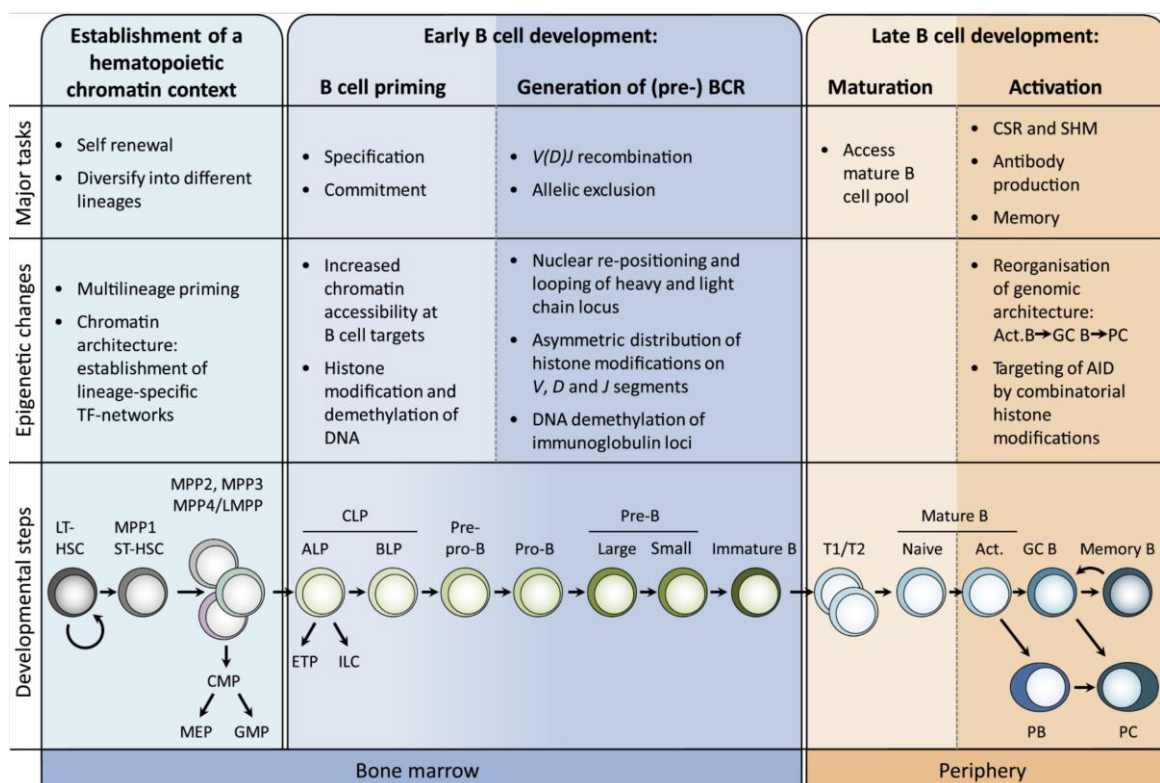


Figure 1.5 Summary of Epigenetic changes that occur during B cell development. Adapted from (Tamaru 2010)

DNA demethylation and chromatin remodeling appear to represent shared mechanisms by which Ebf1 activates transcription of genes essential for B-cell development (Fortin and Hansen 2015). Ebf1, for example, induces DNA demethylation and chromatin remodeling of the Cd79a and other B lineage target genes in conjunction with E2A and Runx1. This

facilitates Pax5 mediated transactivation of these genes (Roessler et al. 2007). The recruitment of the SWI/SNF complex, was required for activation of the mb-1 promoter by Ebf1 and Pax5 and was blocked by the Mi-2/ nucleosome remodeling deacetylase (Mi-2/NuRD), which inhibits chromatin remodeling (Gao et al. 2009). Binding of E2A, Ebf1 and Pax5 at an enhancer region of Cd19 resulted in 'priming' of the locus by chromatin-remodeling factors as early as the MPP stage, followed by the interaction of Pax5 with the promoter and subsequent transcriptional initiation (Heinz et al. 2010). However, it is not known how Ebf1 represses the expression of alternate lineage genes during B cell development.

1.5 Spatial organization of genome in B cell specification and commitment

Understanding the fundamental aspects of gene regulation, chromosome morphogenesis, genome stability and transmission requires the study of the folding principles of chromatin in the nucleus. Microscopic techniques like FISH (Fluorescence *In situ* hybridization) and electron microscopy, can investigate the chromatin structure, but their throughput is currently limited allowing analysis of only a few loci simultaneously (Szydlowski, Go, and Hu 2019; Sivakumar, de Las Heras, and Schirmer 2019). Molecular assays based on chromosome conformation capture (3C) techniques have evolved to assess long-range interactions between specified pairs of loci by using spatially constrained ligation followed by locus-specific polymerase chain reaction (Dekker et al. 2002; Han, Zhang, and Wang 2018). The initial chromosome conformation capture method has sparked the development of various 3C based methods, each with their own set of advantages and limitations. The circularized chromosome conformation capture (4C), a modification of 3C, offers the benefit of requiring only the sequence of target site of interest to be known. Inverse PCR can be used to determine the sequences of all loci that interact with the chosen locus followed by hybridization to microarrays or high-throughput sequencing. However, because only one input sequence can be used per experiment (Han, Zhang, and Wang 2018; Simonis et al. 2006), this method is still limited in terms of high-throughput. Carbon copy chromosome conformation capture (5C) is somewhat high-throughput method that expands on 3C by permitting parallel analysis of the interactions across several selected loci (Dostie et al. 2006; Dostie and Dekker 2007). However, it

involves extensive use of primers and is not suitable for genome-wide analysis of chromosome conformation. Many of these constraints have been solved by Hi-C, which is now widely used to examine the spatial organization of the entire genome in an unbiased manner.

Given that multipotent progenitors require precise and coordinated control of gene expression for cell fate determination (Gibcus and Dekker 2013; Gorkin, Leung, and Ren 2014; Bickmore and van Steensel 2013; Misteli 2007) this study sought to determine how B-lineage specific master regulators tune together to promote the B cell fate choice. It has been demonstrated that Ebf1 and Pax5 are required for induction of early B lineage gene expression program and that targeted inactivation of Ebf1 results in a complete block prior to B cell commitment (Pongubala et al. 2008; Lin and Grosschedl 1995). This suggests that Ebf1 and Pax5 may play a role in chromatin re-localization and establishment of B lineage-specific cis-regulatory interaction landscape. Recent studies indicate that structural organization of the B cell genome in 3D nuclear space is closely associated with transcriptional activity modulation and establishment of cell type-specific gene expression program (Lin et al. 2012), indicating a possible relationship between nuclear architecture and mechanistic control of transcription. 3C-based studies, (Gibcus and Dekker 2013; Lieberman-Aiden et al. 2009; Dixon et al. 2012; Dixon et al. 2015) demonstrate that the genome is composed of hierarchically folded chromatin loops, TADs, and largescale compartments). However, a comprehensive understanding of the transcription factor dependent organization of chromatin remains elusive.

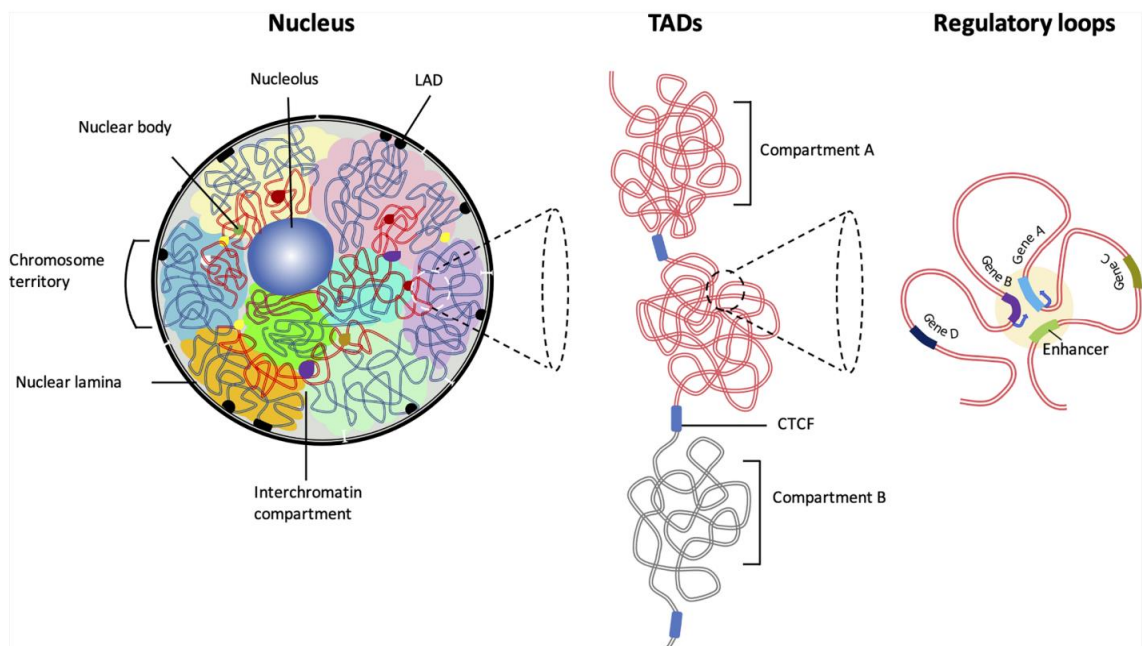


Figure 1.6 The organization of the eukaryotic genome. The hierarchical organization of 3D chromatin suggests that chromatin is organized as transcriptionally permissive and repressive (A and B compartments) at mega-base level. TADs at sub mega-base level, are formed by highly self interacting domains. Within TADs chromatin undergoes multiple folds to form ‘regulatory loops’ which facilitate cell-type specific gene expression; derived from (Pongubala and Murre 2021).

1.6 V(D)J Recombination and the B lymphocyte

Not only does B lymphocyte development necessitates the induction of lineage specific genes and repression of inappropriate genes, but it also needs the assembly and expression of productive antigen receptors (Dudley et al. 2003). Igh recombination occurs before IgL In developing B lymphocytes. RAG has been found to accumulate in recombination centers (RCs) encompassing J and J-proximal D gene segments (Schatz and Ji 2011) and binds to recombination signal sequences (RSS) within the RC. As a result, RSS is brought into the RC for synapsis and cleavage, a process that was previously assumed to involve chromatin looping and random collision. However, a model of RAG chromatin scanning evolved, to explain partner RSS capture, which was predicted on the observation of a strong orientation bias for recombination to cryptic RSSs located more than 5-10 kb away from a reference RSS (Hu et al. 2015). Subsequent analyses of Ig heavy chain gene (Igh) recombination revealed strong evidence that V(D)J recombination can occur via a mechanism involving RAG chromatin scanning driven by cohesin-dependent loop extrusion

(Ba et al. 2020; Dai et al. 2021; Hill et al. 2020; Jain et al. 2018; Zhang et al. 2019). Short-range recombination events between gene segments within the Igh RC were found to be cohesin independent and regulated largely by RSS sequence and not RSS orientation (Ba et al. 2020; Zhang et al. 2019). Long-range recombination events on the other hand, were found to require the cohesin, Rad21 and a convergent orientation of the recombining RSSs (Ba et al. 2020; Dai et al. 2021; Jain et al. 2018; Zhang et al. 2019) and to be regulated by the chromatin looping factor CCCTC-binding factor, CTCF binding elements (Guo et al. 2011; Lin et al. 2015), and the cohesin loader WAPL (Dai et al. 2021; Hill et al. 2020). Altogether, these findings support the existence of two distinct modes of recombination that differ in the mechanism leading to RSS synapsis short-range recombination within or close to an RC, that involves random RSS collision and long-range scanning driven by cohesin-mediated loop extrusion. However, it is not known whether the RAG proteins have any influence over the choice between these two modes of recombination.

1.7 Objectives

1. Determine the molecular mechanisms that regulate chromatin re-localization throughout early B cell stages.
2. Understand the complex interplay between Ebf1 and Pax5 during B cell fate commitment.
3. Delineate the molecular mechanism underlying preferential usage of V(D)J gene segments during IGH recombination.

2. Materials and methods

2.1 Data acquisition

The data for this study were obtained either from Hi-C experiments conducted in-house or from the Geo database that hosts the high-throughput sequencing data. Below table (**Table 2.1**) provides details of the data produced/acquired along with the geo accession numbers.

Table 2.1 List of Hi-C data sets obtained for this study. Each dataset represents knockout of a B cell-specific transcription factor at definite stage of the B cell development.

Cell-type	Stage	Source
E2A-/-	Pre-Pro-B	GSM987816
Ebf1-/-	Pre-Pro-B	Data generated inhouse
Rag2-/-	Pro-B	Data generated inhouse
Pax5-/-	Pro-B	GSM2634267, GSM2634268, GSM2634269
Pax5-ER	Pre-B	GSM2634275, GSM2634276

2.2 Hi-C data analysis

In Situ Hi-C Experiment: Briefly, Ebf1-/- (pre-proB) and Rag2-/- (pro-B cells) cells were harvested, crosslinked with formaldehyde, and nuclei were isolated. Chromatin was subjected to restriction enzyme (HindIII) digestion, overhang filled in the presence of biotinylated nucleotide, and blunt ends were ligated in the intact nuclei. Subsequently, DNA was sheared, size selected, biotinylated fragments were enriched with streptavidin beads and in situ Hi-C libraries were generated using True-Seq paired-end kit (Illumina, San Diego, CA, USA), according to the manufacturer protocol (Boya et al. 2017). The in situ Hi-C libraries were subjected to paired-end high-throughput sequencing using Hi-Seq (C-CAMP, Bangalore) (**Figure 2.1**). The uniquely aligned (reference genome mm10) raw reads were extensively filtered (See below for detailed description) to eliminate various systemic biases originating from experimental manipulations such as restriction enzyme digestion, ligation, DNA fragmentation and PCR amplification. To accomplish this, we employed Iterative Correction and Eigen vector decomposition (ICE) (Imakaev et al. 2012), which implements filtering at multiple levels and determines the corrected contact counts.

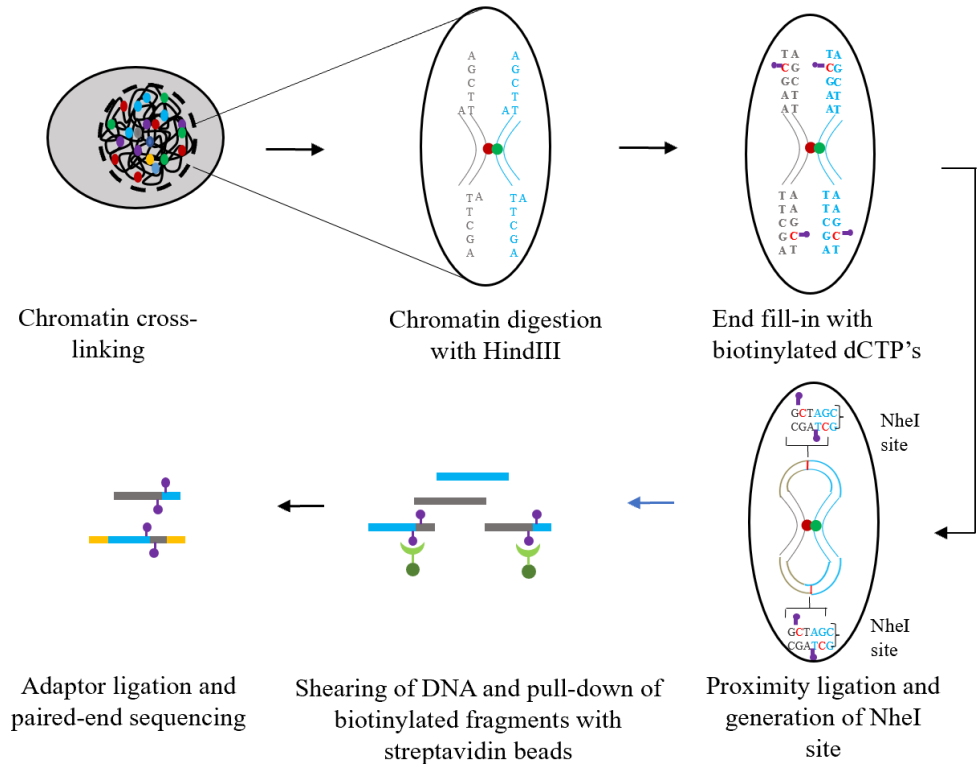


Figure 2.1 The Schematic representation of in situ Hi-C method.

2.2.1 Iterative Mapping

The hiclib package (Imakaev et al. 2012) was used 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 (**Figure2.2A**).

2.2.2 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/MboI). Inappropriate

ligation products such as self-circularized ligations or un-ligated “dangling end” products, generated due to 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. The paired end reads retained after extensive fragment level filtering are considered as valid pairs and are further used for contact map generation and chromatin compartmentalization.

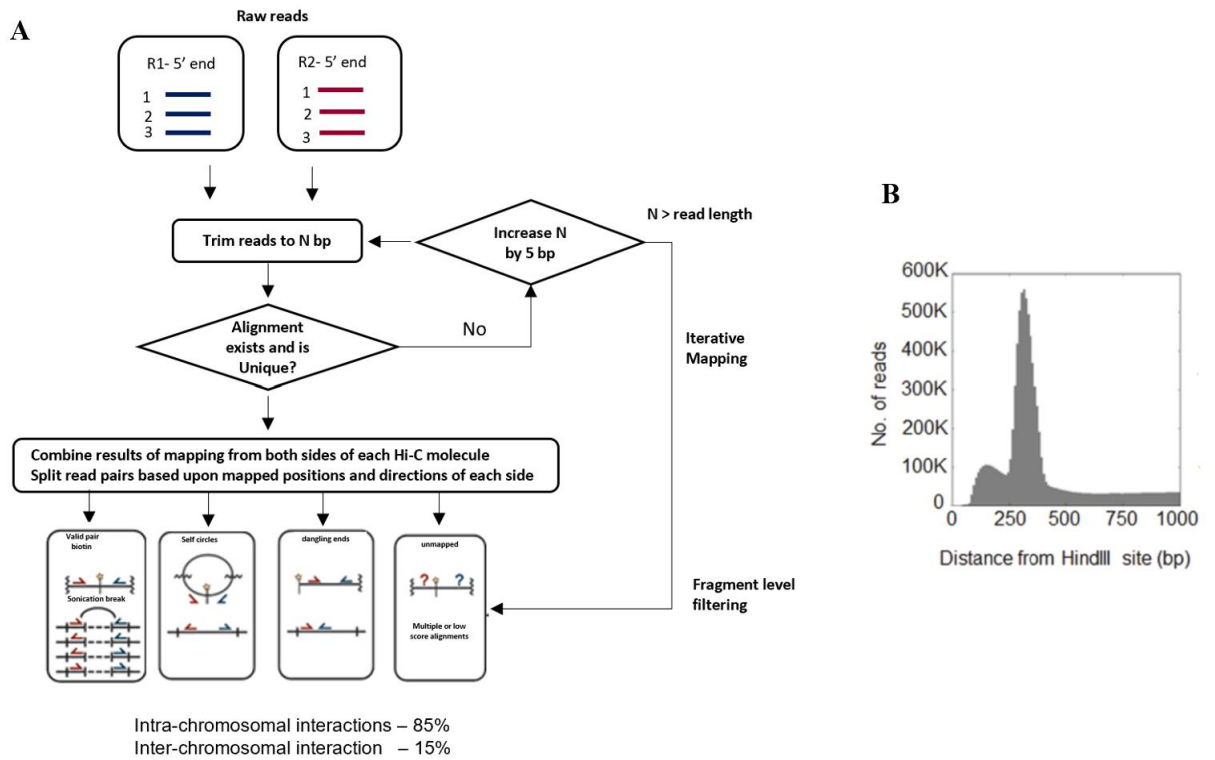


Figure 2.2 Hi-C pipeline for identification of chromatin interactome.

(A) The steps involved in filterig Hi-C data.

(B) 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.

2.2.3 Filtering of PCR duplicates and extreme fragments

During in situ Hi-C library preparation redundant molecules, also called PCR duplicates, are generated because of PCR over-amplification. The presence of these duplicates although is negligible, may affect the relative contact probability and thus must be discarded. Also, we have followed the default parameters to remove a fraction of the most and the least-count fragments to account for systemic biases (the presence or absence of nearby

restriction site). Furthermore, we have discarded the fragments that are too long (>100 kb) or too short (<100 bp), as characterizing these interacting fragments would be difficult. Also, we have discarded all reads that start within 5 bp near HindIII site/Mbol site.

2.2.4 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/Mbol) 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 2.2B**).

2.2.5 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. 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.

2.2.6 Eigen vector decomposition for chromatin compartmentalization

Eigen vector decomposition was performed based on the interaction profiles obtained through iterative correction of the chromosomes (corrected at 100 kb resolution). Since in most of the cases, the first few principal components best explain much of the variability in the data and reveal internal data structure, we calculated the variance for first three principal components. For all the chromosomes across the samples,

we observed that the PC1 (the first principal component), has explained around 60-70% of variance. Thus, we considered PC1 to partition chromosomes into A and B (permissive and repressive) compartments.

2.3 Three-dimensional chromatin structure generation and visualization

The whole genome (20kb resolution) and locus specific 3D chromatin modeling (5kb resolution) 11q was performed at each cell stage using the LORDG method to closely study the effect of genetic switch in 3D. In particular, we focused on the 3D structures of regions that exhibited a developmental switch between transcriptionally permissive and repressive compartments, across the cell stages. These structures were generated at 5kb using the spring 3D modeling (Kadlof, Rozycka, and Plewczynski 2020). Instead of developing a 3D conformation that breaks the least number of restraints or translates frequencies into distances and produces an “average” model the spring method operates on user provided context-dependent chromatin interactions and models them as springs (harmonic bonds) to define constraints between beads that are initially stretched. The ensemble energy is then minimized, the springs shorten and the fiber folds into a structure that satisfies the chosen contacts.

2.4 Binding analysis for regulation of transcription using Hi-C

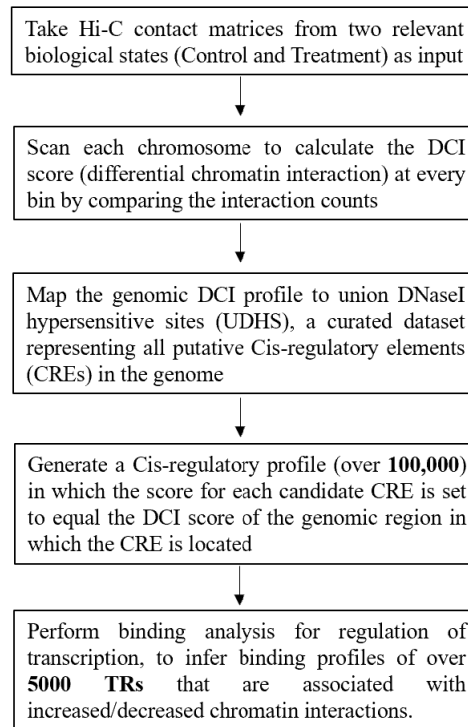


Figure 2.3 Flow chart representing stepwise procedure involved in identifying the transcriptional regulators.

2.5 Genome-wide transcriptome data analysis

2.5.1 Data acquisition

The entire RNA-Seq data used in this study is obtained from (GSE159957) where murine B-cell progenitor cell lines or cultured primary cells of Wt (Pro-B), Ebf1-KO or Pax5-KO genotype were subjected to RNA-seq after transduction vectors expressing Ebf1 or Pax5. The raw reads obtained after sequencing were filtered for quality and aligned to the mouse transcriptome (mm10) using Tophat (V2). Read quantification was performed using htseq-count with UCSC RefSeq gene annotation. The raw count normalization and differentially expressed genes were identified by DESeq2 (V 1.24.0). For DESeq2, we used the DESeq function, which estimates size factors and dispersions and finally fits a model to perform differential expression tests using negative binomial generalized linear models. Genes showing (Log 2) $FC \geq +1.5$ or ≤ -1.5 with $p\text{-val} \leq 0.01$, were differentially up- or down-regulated, respectively.

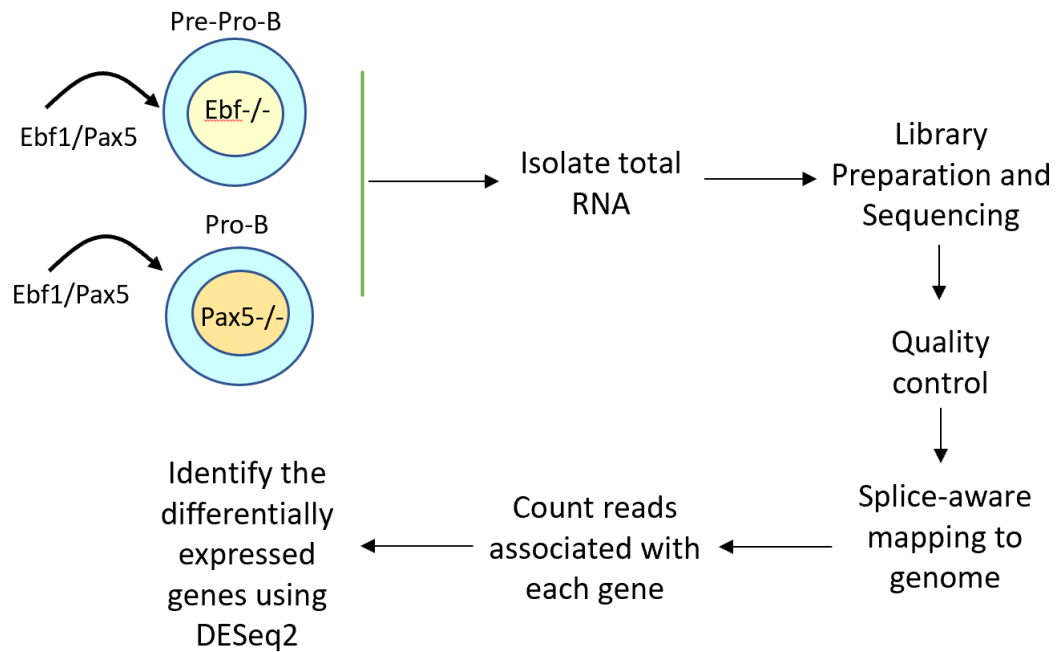


Figure 2.4 Flow chart representing steps involved in bulk RNA-Seq data analysis.

2.5.2 Sample clustering and PCA analysis

Hierarchical clustering and principal component analysis (PCA) of DESeq2 normalized genes were performed using R statistical package with the custom R script.

2.6 ChIP-Seq data analysis

In recent years, ChIP-Seq has emerged as a cost-effective tool that could provide insights into fundamental aspects involved during gene regulation. By combining massive parallel sequencing with ChIP, ChIP-Seq delivers genome-wide profiling of transcription factors and epigenetic modifications in an unbiased manner. Briefly, it involves chromatin crosslinking and sonication followed by immunoprecipitation and purification. The purified DNA is subjected to high-throughput sequencing.

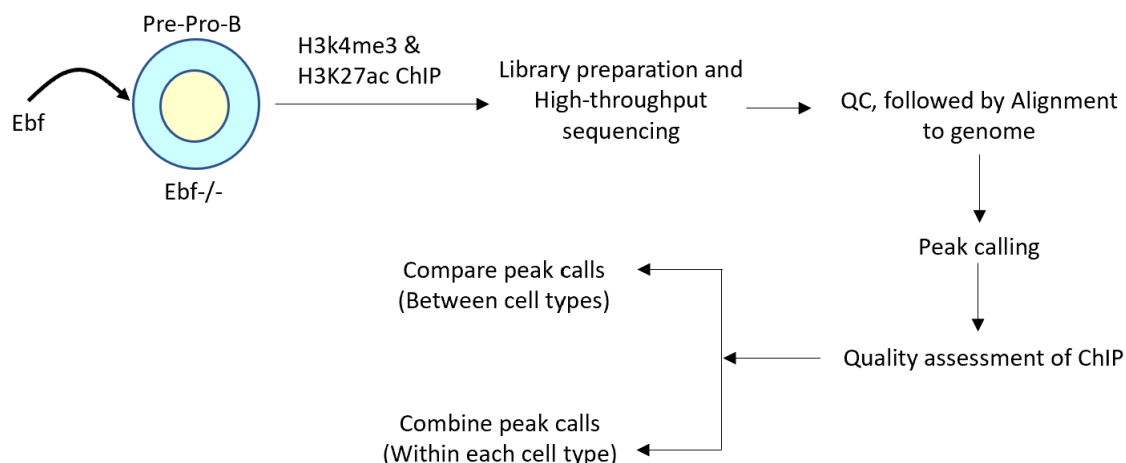


Figure 2.5 Flow chart representing steps involved in ChIP-Seq data analysis.

We wanted to compare the ChIP-Seq datasets of H3K4me3 (serving as active promoter mark) and H3K27ac (representing active enhancers) from Ebf1^{-/-} progenitors with ChIP-Seq data sets obtained using the same antibodies, but after transducing vector expressing Ebf1 in Ebf1^{-/-} (addressed as Early-B cells here after) to identify regions that gain H3K4me3 and H3K27ac peaks upon Ebf1 transduction in Ebf1^{-/-} progenitors. For this, the paired-end raw reads from each sample were checked for quality using standard FastQC tool followed by adaptor trimming using Cutadapt (Martin, 2011). The reads that passed the quality control filters were then aligned to the reference genome (mm10) using bowtie1.2.2 software (Langmead, 2009). We allowed up to 2 mismatches and performed alignment in paired-end mode. Through this, on an average about 90% of the reads were successfully mapped to the reference database.

2.6.1 Data acquisition

The ChIP-Seq data used in this study were generated in house by following standard ChIP-Seq protocol. The H3K4me3 (serving as active promoter mark) and H3K27ac (representing active enhancers) ChIP-Seq datasets from Ebf1^{-/-} progenitors were compared with ChIP-Seq data sets obtained after transducing vector expressing Ebf1 in Ebf1^{-/-} progenitors. These analyses enabled us to identify regions that gained H3K4me3 and H3K27ac peaks upon Ebf1 transduction in Ebf1^{-/-} progenitors.

Table 2.2 Paired-end read counts and quality statistics for each ChIP-Seq sample used in this study.

S.NO	Sample Name	Paired-end reads in Million	% >= Q30	Mean Quality Score
1	Input DNA Rep1 Ebf1null	28.76	95.9	37.81
2	Input DNA Rep2 Ebf1null	36.56	95.9	37.82
3	H3K4me3 Rep1 Ebfnull	35.46	95.68	37.75
4	H3K4me3 Rep2 Ebfnull	31.74	95.69	37.75
5	H3K27ac Rep1 Ebfnull	31.84	95.69	37.73
6	H3K27ac Rep2 Ebfnull	31.98	95.62	37.72
7	Input DNA Rep1 Ebf1-ER	28.84	95.48	37.48
8	Input DNA Rep2 Ebf1-ER	36.86	95.9	37.81
9	H3K4me3 Rep1 Ebf1-ER	31.68	95.63	37.73
10	H3K4me3 Rep2 Ebf1-ER	30.24	95.64	37.73
11	H3K27ac Rep1 Ebf1-ER	32.9	95.62	37.71
12	H3K27ac Rep2 Ebf1-ER	32.88	95.78	37.77

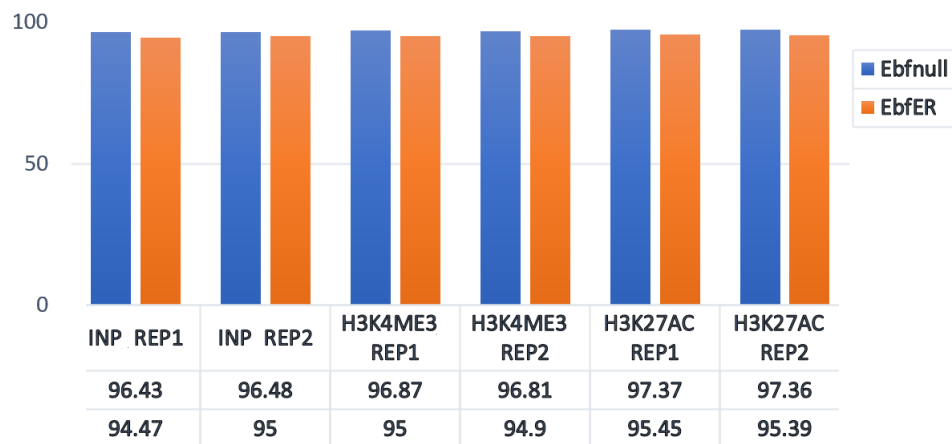


Figure 2.6 Sample-wise alignment percentages of ChIP-Seq data used in this study. On an average all the samples have above 95% alignment rate.

2.6.2 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 "INTERSEFILE" 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 >0.5). The significant interactions obtained by implementing Fit-Hi-C, were further integrated with epigenetic modifications (H3K4me3 and H3K27ac) to identify potential *cis*-regulatory interactions.

2.7 HTGTS–Rep-seq B cell repertoire

To investigate the possibility that the RAG1 NTR mutations alter gene segment usage, High-throughput genome-wide translocation sequencing–adapted repertoire sequencing (HTGTS–Rep-seq) data was analyzed. Briefly, genomic DNA from B-cell populations are sonicated and linearly amplified with a biotinylated primer that anneals downstream of one specific J segment. The biotin-labeled single-stranded DNA products are enriched with streptavidin beads, and 3' ends are ligated in an unbiased manner with a bridge adaptor containing a 6-nucleotide random nucleotide (highlighted in the rectangular box). Products were then prepared for 2 × 300-bp sequencing on an Illumina Mi-Seq.

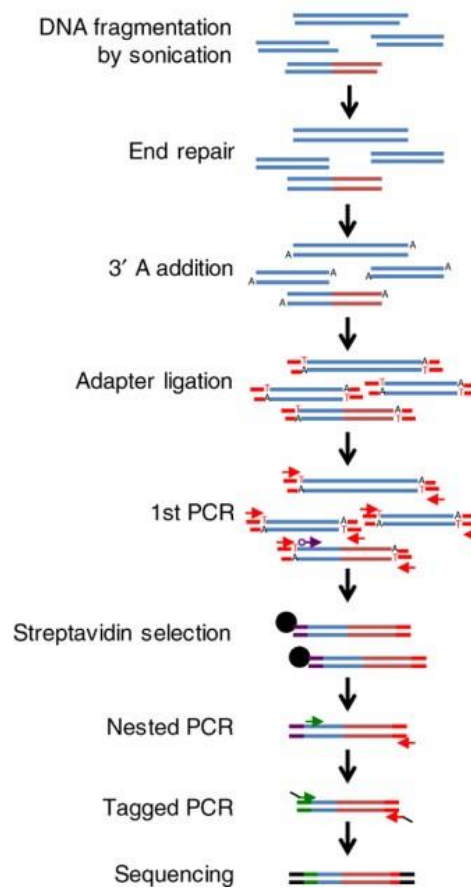


Figure 2.7 Flow chart of HTGTS-Rep-Seq protocol.

To quantify *Igh* repertoires, an HTGTS pipeline was implemented that provides the all-inclusive steps for the analysis of HTGTS-Rep-seq data. In brief, the pipeline involves fastq-multx tool-based (<https://expressionanalysis.github.io/ea-utils/>) demultiplexing followed by the implementation of cutadapt (<https://cutadapt.readthedocs.io/en/stable/>) to trim the adaptors in the metadata file for each sample. The high-quality (Phred score >20) adaptor-trimmed reads were then joined using fastq-join tool if, and only if, the reads from both sides had at least a 10-bp overlap with mismatch rate $\leq 8\%$. Both joined and unjoined reads were mapped against the mouse (mm10) IGH V(D)J gene database (IMGT) using default parameters of the IgBLAST. The aligned reads (now assigned with V, D, and J genes) were further filtered to have high IgBLAST score >150 with total alignment length >100, comprising overall mismatch ratio <0.1. The usage of V and D genes was computed based on the processed IgBLAST results as provided by the pipeline.

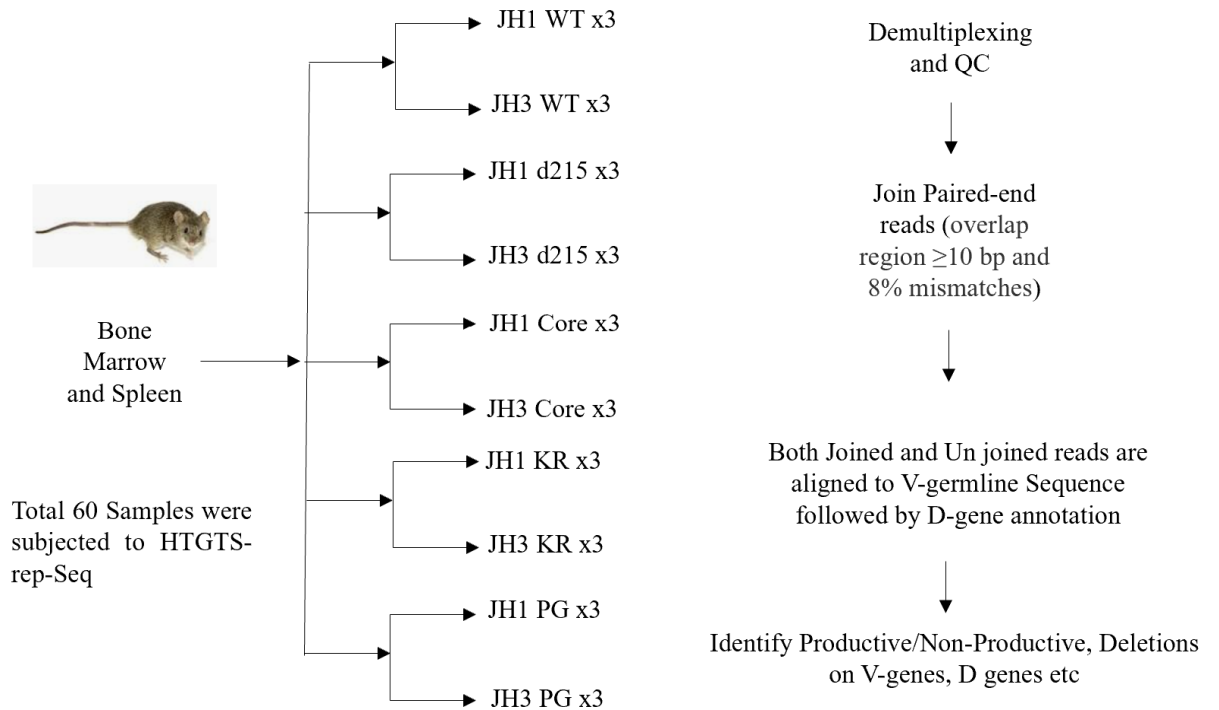


Figure 2.8 Flow chart representing the experiment design and the workflow for measuring the VDJ recombination events.

To quantitate DJ_H rearrangements, we implemented the transloc pipeline (https://github.com/robinmeyers/transloc_pipeline) that involves D_H -to- J_H junction detection based on the Optimal Query Coverage algorithm. The identified junctions were then intersected with expanded RSS locations (± 40 bp) to annotate the region with their respective D and J gene segments. After gene segments were called, all quantitative analysis was conducted in R.

3. Results

3.1 Chromatin undergoes genetic switch between A and B compartments during B cell developmental transitions

To determine programmatic changes in chromatin organization during B cell development, we performed in situ Hi-C, a high-throughput molecular approach (Nagano et al. 2013) 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 in situ Hi-C experiments were performed in two biological replicates for both pre-proB and pro-B cells, and libraries were subjected to deep-sequencing and generated unique read pairs. We also obtained Hi-C data from the Geo database for E2A^{-/-} (representing pre-pro-B cell stage), Pax5^{-/-} (representing pro-B cells) and Pax5-ER (representing pre-B cell stage) from Geo database (accession numbers provided in the materials and methods section). 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). 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 the 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). 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 samples, iteratively corrected

contact maps for each chromosome were generated at 1 Mb resolution (**Figure 3.1A**). Our analyses captured many of the previously identified long-range chromatin interactions (Kieffer-Kwon et al. 2013). 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 3.1A**). Most 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. To understand the differences in chromatin interaction patterns between the B cell developmental stages, we implemented Principal Component Analysis (PCA) at 1 Mb resolution. As expected, these analyses revealed that chromatin is segregated into A (Permissive) or B (Repressive) compartments, which are defined by enriched or minimal interactions, respectively. Permissive (A) compartments were found to contain a higher number of genes 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; ChIP-Seq data (Data not shown) obtained from (Lin et al. 2012)) and displayed higher transcript levels when compared to the B compartments (**Figure 3.1B**), indicating that chromatin compartmentalization mirrors gene activity. To investigate the possibility that selective changes in chromatin compartmentalization provide a structural framework for B-lineage gene expression (Lin et al. 2012), 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 22,461 common genes that were captured by Hi-C.

3.1.1 Chromatin compartmentalization in E2A^{-/-} Ebf1^{-/-} (pre-pro-B) and Rag2^{-/-} (pro-B)

Because the phenotypes of E2a^{-/-} and Ebf1^{-/-} are identical, and because the intra-chromosomal contact maps and PCA values of both samples showed little difference, we considered Ebf1^{-/-} data generated in our lab to represent pre-pro-B cell stage. We used PCA analysis to define the chromatin state and cross-compared them to find the selective changes in chromatin between pre-proB and pro-B. Of the 22,360 common genes, 16,045 genes were found to be present in A compartments 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. 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 (**Figure 3.1D**). Of these, 1,339 (5.98%) genes transitioned from B to A compartment, while 741 (3.31%) genes relocated from A to B compartment during differentiation of pre-pro-B cells to pro-B cells (**Table 3.1 and 3.2B**). These observations demonstrate that B cell developmental progression from a multipotential progenitor to a specified state encompasses notable changes in chromatin compartmentalization. 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 3.1C**). For instance, *Satb2*, *Tead1*, *Pou2af1* and *Tlr4* that are essential for B cell development (Dobrev, Dambacher, and Grosschedl 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. Likewise, genes that relocate from the A compartment to the B compartment displayed lower transcript levels in pro-B cells. Notably, genes that are associated with multipotent progenitors such as *Satb1*, *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. 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 in chromosomes 6 and 7, switch from the A compartment to the B compartment in pro-B cells (**Figure 3.1E**). 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 compartment in both cell types. However, some of the key alternate lineage genes (*Gata3*, *Zbtb16*, *Klf4*, *Vav3* and *Sox6*) switch to the B compartment in proB cells. Thus, these observations indicate that re-localization 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 re-localization and its associated changes may be a result of alteration of transcription.

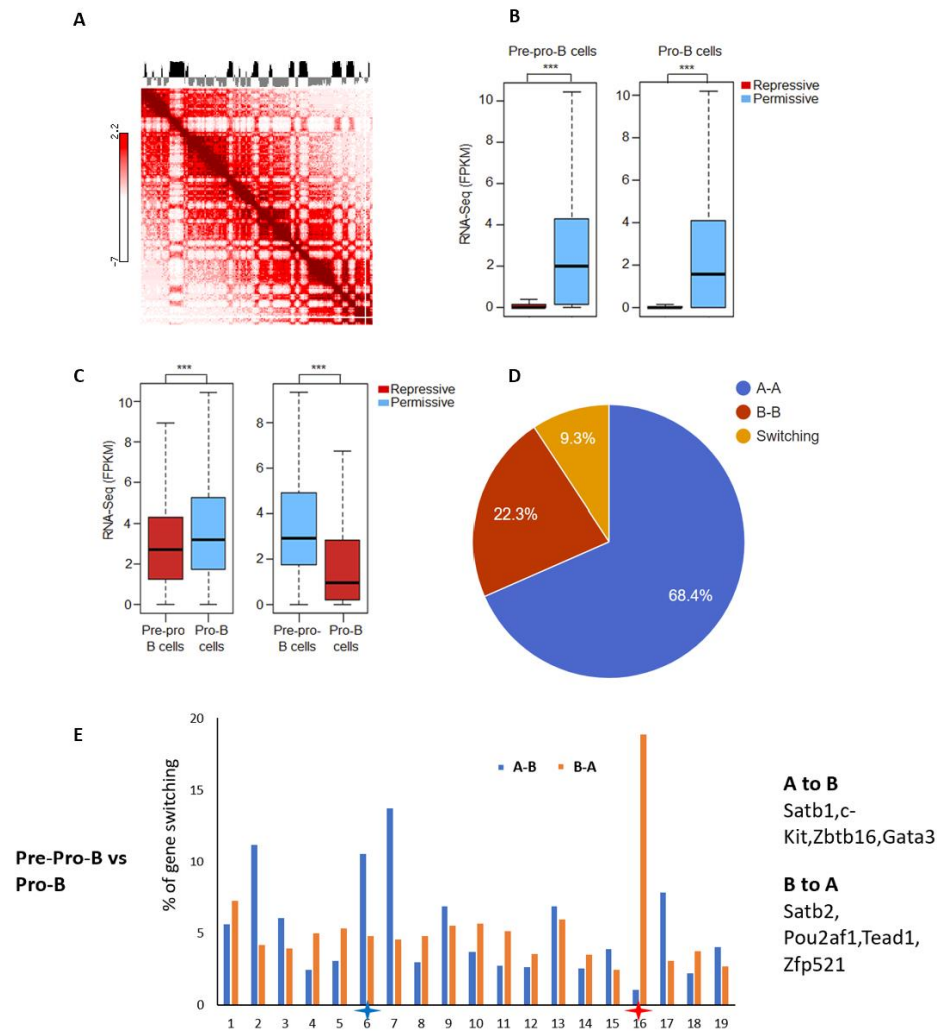


Figure 3.1 Chromatin compartmentalization is closely associated with gene activity during pre-pro-B to pro-B cell development.

- (A) Iteratively corrected intrachromosomal 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) 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.
- (C) 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 (** $P < 0.001$).
- (D) Pie graph representing the chromatin state of genes as the cells transit from pre-pro-B to pro-B.
- (E) 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 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	Chr	Gene	Pre-pro-B cells RNA-Seq	Pro-B cells RNA-Seq
1	ANGPTL1	0	1.06936	7	APBB1	0.376543	2.80745
1	GM16701	0	2.39649	7	LIG1	5.62672	30.8381
1	RALGPS2	0.299642	173.374	7	GAS2	1.89259	7.42244
1	CD55	0.0145898	4.70994	7	6330408A02RIK	0.561942	1.39736
1	9130024F11RIK	0.229238	51.4038	7	UBA2	56.0173	102.123
1	SATB2	0.292912	18.2528	7	ZFP667	4.83619	7.68245
1	DST	0.112248	6.17652	8	MFHAS1	0.0624261	1.84316
1	FAM78B	0.0869397	4.5675	8	LIG4	1.04098	25.1055
1	BC094916	0.185617	8.45294	8	CERS4	1.03676	8.08712
1	IFI204	0.258969	9.74153	8	SHCBP1	13.045	60.0003
1	PLXNA2	0.0564396	1.46354	8	FRG1	47.6458	87.3816
1	4930558J18RIK	0.464516	7.6112	8	TMEM184C	37.881	62.9321
1	PYHIN1	0.902505	12.1828	8	INTS10	18.7447	30.4855
1	RDH10	5.3855	60.044	9	MMP10	0	1.28893
1	D1ERTD622E	3.13591	16.4703	9	B430319G15RIK	0	2.64426
1	HSD17B7	6.4267	32.8848	9	HMGCLL1	0.0375491	29.9585
1	3110045C21RIK	0.754319	3.60197	9	POU2AF1	1.62648	860.051
1	PLEKHM3	1.35806	5.07097	9	PLOD2	0.221905	66.9432
1	KIFAP3	9.67066	17.6465	9	ELOVL4	0.0154332	2.43243
1	1700066M21RIK	6.98649	11.8242	9	RYK	0.214252	32.0095
1	POGK	8.43349	13.8726	9	ACPP	0.0255067	3.3516
1	RRP15	12.3287	19.0019	9	PLSCR4	0.028257	2.85966
2	ACVR1	0	2.13155	9	TTK	9.72449	62.1997
2	PKP4	3.65601	19.2274	9	PLSCR1	0.680931	2.2731
2	AURKA	10.583	54.2846	9	BCKDHB	5.87157	15.8775
2	DUT	37.2818	188.959	9	ALKBH8	12.2273	25.7896
2	FAM98B	6.7997	21.3616	9	ANAPC13	73.8256	130.131
2	TASP1	4.81887	13.7607	9	4930579K19RIK	1.81084	3.15536
2	ARHGAP11A	20.9767	50.0693	9	4930526I15RIK	1.06918	1.71792
2	RIF1	35.9333	84.1126	10	PCBP3	0	2.73672
2	CCDC34	10.5842	22.8999	10	BVES	0.0493622	5.50347
2	LNP	10.0846	20.9691	10	E2F7	0.764022	9.35019
2	SLC30A4	25.4202	50.1709	10	ULBP1	2.05063	20.9842
2	NDUFAF5	5.71713	10.5179	10	RTKN2	1.56192	11.2687
2	BLOC1S6	11.6369	18.9454	10	CDK1	17.6828	62.3272
2	EIF3M	79.6491	128.758	10	BTG1	28.4686	86.3335
2	ZC3H15	45.7782	73.6512	10	SNORA33	82.1751	216.023
2	DPH6	11.0244	17.2151	10	COL6A2	0.953132	2.03543
2	ARHGAP15	77.7094	121.093	10	HACE1	9.04193	19.2505
2	FAM210B	3.19571	4.87643	10	GOPC	12.6797	26.6895
2	API5	51.4205	77.4806	10	SGK1	2.6293	3.97881

3	FREM2	0	1.52605	11	PTTG1	15.0717	248.217
3	WLS	0.398844	69.5851	11	FAM64A	4.21323	24.5212
3	BCAR3	0.062197	7.65024	11	UBLCP1	1.74732	8.3601
3	LHFP	0.0887308	8.87904	11	FBXO48	3.92446	15.0105
3	GRIA2	0.236562	4.27381	11	2810021J22RIK	1.94791	6.836
3	CAMK2D	31.7348	355.765	12	RAPGEF5	0.026154	15.586
3	MAP9	0.440882	3.71019	12	TRIB2	0.788095	305.568
3	VCAM1	0.145905	1.02924	12	LRRN3	0.0959885	26.9063
4	DOCK7	0.069132	8.9707	12	SGPP1	2.67393	35.394
4	NFIA	0.0267694	1.62087	12	IFI27L2A	10.8232	42.2453
4	PRKAA2	0.0402403	2.09493	12	ARL4A	10.8625	32.9231
4	FGGY	0.0743661	2.30948	12	FRMD6	0.514077	1.24221
4	AKAP2	0.0948782	1.43187	12	TAF1B	13.2909	28.5593
4	TLR4	3.07086	16.6325	13	RHOBTB3	0.0576545	6.24338
4	SMC2	37.4438	130.426	13	FOXC1	0.158433	11.8069
4	USP1	32.3709	70.0155	13	GM3604	0.665476	4.41484
4	IFT74	7.24771	13.1577	13	DIP2C	1.64624	10.7312
4	NBN	13.0132	23.061	13	ATXN1	0.795994	4.98909
4	TMEM64	16.7031	28.2296	13	NAIP7	0.587153	2.23975
4	ALG6	9.2245	15.5629	13	ZFP759	4.22391	9.65919
5	9330182L06RIK	0.0275777	6.89013	13	GM5141	4.14997	8.02769
5	PKD2	0.24846	4.47311	13	CCNH	33.4202	59.5897
5	TGFBR3	0.128204	1.11655	14	BMPRI1A	0.289121	36.7922
5	FBXL5	36.5019	273.865	14	FAM213A	0.146767	9.13802
5	ZFP11	0.446974	2.41646	14	IL3RA	0.586168	1.10756
5	NCAPG	20.3712	104.363	14	PSMD6	63.5297	108.446
5	RFC3	16.2787	70.2803	15	LRRK2	0.0461664	7.63185
5	SMIM20	9.09848	35.7554	15	DEPTOR	0.172611	1.0442
5	ABCG3	3.70231	13.686	16	SLC15A2	7.41606	19.8352
5	GBP9	2.31449	7.74641	16	FSTL1	0.605659	1.3427
5	HSD17B11	7.99761	24.3505	16	BC002163	78.7264	163.204
5	1600023N17RIK	1.3886	3.90296	16	CMSS1	6.56424	13.3587
5	ZCCHC4	4.7281	12.1723	17	MTCL1	0.0097454	1.36504
5	SLC2A9	2.44615	6.11988	17	THBS2	0.201017	2.0761
5	ATP10D	2.51542	5.91666	17	ATL2	15.618	102.186
5	SPP1	3.16974	6.72129	17	2700099C18RIK	6.84493	36.3958
5	CLOCK	5.79372	10.933	17	CYP1B1	0.2933	1.52604
5	TBC1D19	10.9286	19.9124	17	AGPAT4	0.454192	1.52285
5	GBP4	4.46603	7.05774	17	GABBR1	1.72764	4.5136
6	ICA1	0	2.8522	17	CDC42EP3	28.0787	56.8135
6	GLCCI1	0.497737	103.701	18	4930426D05RIK	0	2.06263
6	PON3	0.182468	16.7568	18	SPINKL	0	60.2009
6	MAGI1	0.0273598	1.73813	18	KLHL14	0.0167426	6.18494
6	A430035B10RIK	0.164879	6.27741	18	SPINK11	0.0936119	6.29432
6	VOPP1	0.115307	2.33803	18	ZFP521	0.085086	5.58952
6	FAM188B	0.171605	2.02898	18	PRELID2	1.51146	87.967

6	PON2	5.81647	36.4301	18	CAMK4	0.0310546	1.78986
6	CXCL12	0.617593	2.92478	18	STARD4	12.0258	35.4951
6	C1GALT1	28.253	122.589	18	YTHDC2	7.47806	15.6333
6	FKBP9	0.397581	1.16786	19	RNLS	0.259824	1.33555
6	MDFIC	4.96495	12.471	19	TCF7L2	1.18001	3.73683
6	ACN9	3.73107	9.24041	19	UHRF2	21.1426	42.4151
6	NDUFA5	73.1752	163.054	19	MINPP1	9.83711	18.5564
6	LANCL2	9.52048	16.4544	19	FAM204A	18.9646	34.846
7	GP2	0	2.48296	19	2700046G09RIK	0.787547	1.28295
7	UMOD	0.0404613	21.8918	19	FRA10AC1	11.5442	17.3886
7	TEAD1	0.0158901	3.51603				

Table 3.2 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	Chr	Gene	Pre-pro-B cells RNA-Seq	Pro-B cells RNA-Seq
1	MARCO	2.04118	0	9	ZBTB16	15.8803	0.0407465
1	GM7694	2.29597	0	9	KCNJ1	13.5733	0.0385854
1	GM16897	2.77609	0	9	DMXL2	2.34928	0.0309904
1	IL18RAP	87.1834	0.147636	9	OAF	11.4225	0.411411
1	IL18R1	200.723	0.54607	9	SLC9A9	34.0744	1.63487
1	CD34	533.539	1.52474	9	PIK3CB	2.51959	0.195365
1	NRP2	12.7006	0.0669168	9	EEPD1	5.47207	0.439633
1	RGS18	135.289	0.744447	9	ANXA2	8.00882	0.912337
1	NOS1AP	1.60549	0.0154627	9	SIAE	33.0719	5.02867
1	KHDC1A	1.31913	0.02695	9	MORF4L1	4.37126	0.753983
1	AGAP1	2.28352	0.0746374	9	CRTAP	12.9342	3.63079
1	SH3BP4	1.15841	0.0538187	9	SLC35F2	4.80676	1.37065
1	TMEM14A	2.90688	0.504574	9	TBC1D2B	6.50961	2.0249
2	NFATC2	1.19276	0	9	STT3B	92.3649	31.3096
2	ITGA8	1.46398	0	9	GLB1	24.4183	8.83434
2	NOSTRIN	2.30518	0	9	RAB39	2.67286	1.00051
2	MIR669J	64.3396	0	9	TMPPE	6.36963	3.8339
2	CERS6	29.2159	0.0660073	10	4930444F02RIK	1.09358	0
2	SCN3A	10.2744	0.0413349	10	LILRB4	3.17261	0
2	GATA3	2.64488	0.0392445	10	PPM1H	8.75861	0.0136044
2	FMNL2	3.92034	0.105547	10	SLC16A7	6.08008	0.0213333
2	LDLRAD3	7.59802	0.807013	10	PERP	4.11309	0.0283578
2	IFIH1	6.02355	1.27477	10	HMGA2	44.0214	0.327775
2	SSFA2	11.6528	4.71055	10	ESR1	2.27757	0.0409784
2	ARL5A	47.3703	27.4379	10	GP49A	2.78555	0.0567702
3	GUCY1A3	5.16371	0.0085551	10	CEP85L	16.4446	0.819036

3	CPA3	10.4835	0.0295834	10	SYNE1	16.5741	2.15255
3	SLC25A24	10.8942	0.0473313	10	EPM2A	2.09075	0.286647
3	VAV3	7.02558	0.0436874	10	NHSL1	2.76631	0.994897
3	S100A11	71.1198	4.20063	11	TTYH2	8.21671	0.151304
3	SLC22A15	1.80787	0.525387	11	SLC39A11	6.771	1.91315
3	LMO4	62.1199	20.676	11	SLC36A1	3.92703	1.78595
3	S100A10	228.8	97.2808	11	DOCK2	83.581	51.9097
3	TPD52	23.7815	12.8416	12	9130015A21RIK	1.87208	0
3	TBCK	24.7898	15.6428	12	NRCAM	2.24995	0
4	TMEM51	11.6976	0.0237655	12	AHR	11.671	0.11327
4	AJAP1	2.80085	0.0152035	12	ARHGAP5	27.1824	0.696551
4	TNFSF8	5.05116	0.0328927	12	EGLN3	13.7683	0.45998
4	CLVS1	1.41831	0.04587	12	HPCAL1	7.7666	0.489471
4	KLF4	1.89177	0.357962	12	DAAM1	4.39215	1.14916
5	CPEB2	1.53465	0	12	ADCK1	9.97327	3.74789
5	HS3ST1	7.94804	0	13	GM5086	1.23085	0
5	CLNK	11.3184	0	13	GZMK	2.38646	0
5	KIT	107.056	0.377379	13	HSPB3	2.45689	0
5	NIPAL1	3.97433	0.0142303	13	LY86	36.3733	0.0380619
5	CDS1	6.20893	0.0230933	13	AOAH	16.1628	0.0307039
5	RASGEF1B	4.53638	0.106392	13	F13A1	9.72704	0.0265659
5	FRY	4.51218	0.209581	13	GCNT2	21.9358	0.138282
5	ADAM22	1.8758	0.120937	13	ESM1	3.05482	0.0222662
5	CDK14	1.31599	0.210975	13	ITGA1	9.11274	0.0942137
5	SCARB2	49.9996	24.5407	13	EMB	94.7284	1.21409
6	KLRC2	1.00817	0	13	CDK20	1.85404	0.0484605
6	KLRA10	1.06938	0	13	CTSL	58.156	4.38644
6	KLRA23	1.21279	0	13	AAED1	41.0997	3.50703
6	KLRI1	1.26621	0	13	SNX18	22.129	4.98811
6	PRSS2	1.4639	0	13	PELO	6.43427	1.9051
6	ST8SIA1	1.71074	0	13	CDYL	13.3558	4.77256
6	KLRA13-PS	2.22611	0	13	ZFP369	33.5559	17.5875
6	TES	54.3555	0.0687901	13	ZFP65	11.6495	6.88813
6	KLRK1	4.56162	0.0136367	13	ZFP738	14.3576	8.60192
6	KLRI2	12.979	0.0735156	13	A530054K11RIK	13.0466	7.86326
6	KLRC1	7.1269	0.0582266	14	DOCK5	3.3284	0.0639458
6	BCL2L14	1.61588	0.0142404	14	ATP8A2	1.3262	0.0408612
6	GPRIN3	28.6496	0.84163	14	FLNB	15.6466	2.17661
6	EPHB6	2.1558	0.0765372	15	FAM134B	57.4188	0.112818
6	ZYX	20.5025	0.926055	15	FYB	8.10207	0.0413415
6	HPGDS	6.41161	0.468624	15	TRIO	10.8631	0.873173
6	PLEKHA5	8.5933	3.6728	15	OXR1	80.3601	33.2368
6	TPK1	11.5343	5.29055	15	LRP12	5.78029	2.70821
7	MRGPR1	1.24768	0	15	ANGPT1	3.15486	1.72543

7	NCR1	1.89567	0	15	FAM84B	38.7744	23.8794
7	MCTP2	51.9241	0.118911	15	ZFP622	21.8074	14.0309
7	CD163L1	40.4187	0.13421	16	GM4827	1.17068	0
7	5830411N06RIK	154.707	0.656541	16	ZDHHC23	2.48371	0
7	SIGLECH	54.2745	0.235825	16	DOPEY2	4.52294	0.41575
7	ADAMTS17	1.06831	0.0060798	17	KCNK12	1.16267	0
7	PLEKHA1	58.0944	0.383194	17	SLC22A3	18.5143	0.0459131
7	TSHZ3	14.1182	0.105217	17	SATB1	198.831	1.9932
7	SOX6	2.66652	0.0449762	18	SETBP1	6.61077	0.110675
7	PTPRE	81.3402	1.81641	18	PPIC	5.64475	0.376841
7	PRCP	18.6812	3.54021	18	MPP7	13.3922	6.3315
7	TUBGCP5	81.9667	25.7299	19	MPEG1	44.7289	0.204605
7	1600014C10RIK	32.7068	11.89	19	HECTD2	6.08728	0.0286052
7	PEPD	14.4153	8.14617	19	GCNT1	30.3873	0.377118
8	KBTBD11	7.69045	0.0134392	19	DTX4	6.11693	0.431718
8	NRP1	17.0582	0.135697	19	PIP5K1B	5.26781	0.895916
8	ZFHX3	4.03586	0.0582646	19	ATRNL1	6.07981	1.85466
8	RNF150	8.95391	0.183406	19	TJP2	4.21562	1.52531
8	CDYL2	10.5227	2.36429	19	TMEM2	7.10639	2.64035
9	TMEM158	1.77939	0	19	SORBS1	7.35667	3.55583
9	EOMES	9.95832	0.0248923	19	RFK	20.461	12.6326

3.1.2 Chromatin compartmentalization in Pax5^{-/-} (pro-B) and Pax5-ER (pre-B) cells.

The publicly available Pax5^{-/-} and Pax5-ER Hi-C datasets were obtained to study the Pax5 dependent changes in chromatin organization as the cells transit from pro-B to pre-B. The Johanson et al (Johanson et al. 2018) generated these Hi-C datasets using Mbol (A 4bp cutter), which has frequent restriction sites as compared to HindIII (a 6bp cutter). Thus, the Hi-C data for Pax5^{-/-} and Pax5-ER has a greater number of paired-end reads and the greater genome-wide coverage. Of the 22775, common genes, 15711 genes are found to be in permissive compartments whereas 5368 genes are present in repressive compartments. Further examination of these common genes between pro-B cells and pre-B cells revealed three distinct classes, including a common set of genes that are localized in either A (Group I; 69%) or B (Group II; 23.6% compartments in both cell types (**Figure 3.2**).

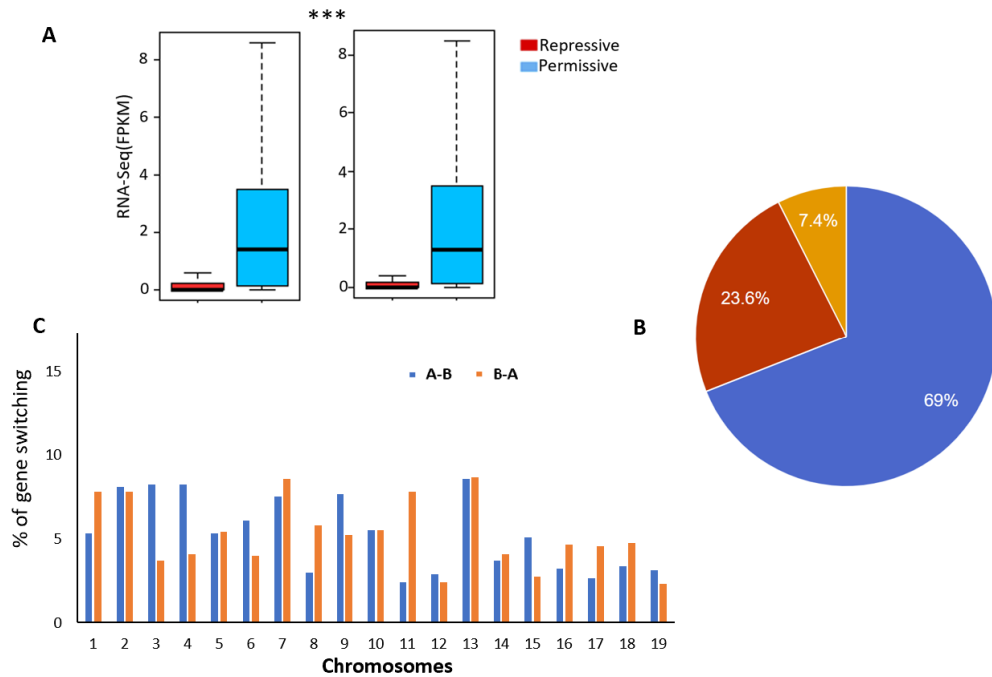


Figure 3.2 Genetic switch between A and B compartments has decreased during pro-B to pre-B cell development

(A) Comparative analysis of transcript levels of genes, based on RNA-Seq, present in A and B compartments (***)P<0.001) for both pro-B and pre-B cells.

(B) Pie graph representing the chromatin state of genes as the cells transit from pre-pro-B to pro-B.

(C) 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.

Although a major fraction (92.6%) of genes remained in the same compartment (Group I or II) in both cell types, a distinct set of genes (Group III; 7.4%) switched between A and B compartments. Of these, 768 (3.372%) genes transitioned from B to A compartment, while 928 (4.074%) genes re-localized from A to B compartment during differentiation of pro-B cells to pre-B cells (**Table 3.3 and 3.4**). These observations demonstrate that chromatin switch at pro-B to pre-B stage is slightly minimal when compared to pre-pro-B vs pro-B stage.

Table 3.3 List of differentially expressed genes that are switching from A to B compartments during progression of pro-B cells to pre-B cells.

Chr	Gene	Pax5null RNA-Seq	Pax5null Pax5-ER RNA-Seq	Chr	Gene	Pax5null RNA-Seq	Pax5null Pax5-ER RNA-Seq
1	FAM135A	19.203566	9.84608519	9	CRTAM	5.7391734	2.21324456
1	FMN2	3.8571742	0.74746208	9	MYZAP	5.1882877	3.33621786
1	GM7694	11.997185	5.1450643	9	TLN2	21.945099	8.95481585
1	HLX	8.9543032	5.28091812	9	UBASH3B	24.099219	11.850507
1	NOS1AP	11.560813	4.26861076	9	ZBTB16	7.7552041	4.94896528
1	NRP2	14.824076	5.51203987	10	FAM26E	2.4455607	0.58628665
2	FAM171B	7.8022236	2.53045927	10	FAM26F	9.0219615	5.28515069
2	NMI	34.080469	20.2440852	10	HEBP2	8.3256374	1.43610313
2	RIN2	14.829766	8.87998875	10	TSPYL4	22.805712	11.870426
3	CASQ2	1.5752008	0.82913453	12	9130015A21RIK	3.3289308	0.82913453
3	GM5538	1.8706976	0.74746208	12	AHR	25.085565	10.2336106
3	GUCY1A3	21.072507	9.78705844	12	AKAP6	8.7544302	3.08858769
3	MED12L	22.04903	13.123653	12	EGLN3	46.36985	19.4603714
3	NBEA	19.67341	9.81133271	12	ETV1	34.840729	12.1512061
3	PDGFC	1.7077055	0.58628665	12	LAMB1	4.490665	2.87998096
3	SLC22A15	13.06599	7.02992531	13	GAPT	14.188843	1.85400117
4	ABCA1	13.54085	6.90952901	13	GM5086	4.8514779	2.51336848
4	CDKN2B	3.6450024	1.83090069	13	GPR141	14.405709	6.76501308
4	KLF4	12.335569	7.69226568	13	ITGA1	5.913934	1.34345196
4	TOX	37.600104	16.7076645	13	MAP1B	9.2062929	5.63557199
5	CLDN12	11.404481	5.1072804	13	MBOAT1	10.34023	5.68463993
5	SLC4A4	6.6190744	4.21277358	14	1700112E06RIK	5.3254911	2.71075953
5	SOWAHB	9.2347473	0.58628665	14	IL3RA	7.5628499	4.90121626
6	CLEC5A	12.266945	2.15428003	14	KLF5	15.051006	8.14548479
6	COL28A1	1.5274182	0.74746208	14	LMO7	9.6381047	5.54706743
6	IL12RB2	39.714582	14.4456078	14	NEK10	12.714029	5.37370736
6	IL23R	20.886407	8.76066473	15	BASP1	9.1896978	4.45094748
6	MET	7.955294	3.27098373	15	GDNF	4.8769428	2.43530852
6	MGAM	9.2272983	2.57237471	15	LAPTM4B	3.3214513	1.05707101
6	PTPRO	2.7535923	0.74746208	15	NOV	55.890941	36.3152409
6	SOX5	6.7908682	4.52257405	15	SEPP1	53.9589	35.0053946
7	1700003G18RIK	4.6058207	1.67137601	15	SNTB1	4.6930867	2.63834108
7	DKK3	2.0205844	1.05707101	16	CD200R2	1.986017	0.58628665
7	FAM169B	17.901743	5.91797482	16	CD200R4	5.1288248	1.83090069
7	PLEKHF1	9.5007133	3.88926336	17	CYP1B1	1.8706976	1.05707101
7	PRCP	42.840126	28.4478205	17	DLL1	13.624183	7.48862624
7	SOX6	2.0205844	0.82913453	17	FNDC1	2.9343272	1.86573359
7	SYNM	7.6369637	3.88392688	17	SLC8A1	15.414119	6.21538995
8	RNF150	19.506947	12.2499017	18	LAMA3	3.3407758	1.05707101

9	AQP9	15.809827	4.09908138	19	4931403E22RIK	3.5821132	1.53738828
---	------	-----------	------------	----	---------------	-----------	------------

Table 3.4 List of differentially expressed genes that are switching from B to A compartments during progression of pro-B cells to pre-B cells.

Chr	Gene	Pax5null RNA-Seq	Pax5null Pax5-ER RNA-Seq	Chr	Gene	Pax5null RNA-Seq	Pax5null Pax5-ER RNA-Seq
1	BC094916	4.885036	10.97716	9	EXPH5	0.763709	1.746719
1	BOLL	1.322783	2.223082	9	GM2087	1.080048	1.746719
1	CD55	2.49159	16.22099	11	KRT27	1.527418	2.695013
1	GM16897	1.454478	3.554288	11	MIR6989	2.883362	4.740737
1	HECW2	2.076058	4.333559	11	NLRP1C-PS	0.763709	2.589285
1	PCMTD1	48.87944	75.85608	12	RAPGEF5	5.219587	25.16812
1	PLCL1	4.193603	8.142739	13	4930525G20RIK	3.828876	5.759868
1	PYDC4	9.67245	15.05287	13	KLHL3	2.194143	3.771464
2	CCDC73	7.321778	11.26873	13	MIR713	0.763709	1.260911
2	E130215H24RIK	1.527418	4.588684	13	POU5F2	5.413948	8.466437
2	ITIH5	13.19254	22.00256	13	RHOBTB3	1.707705	5.01794
2	NUTM1	1.504579	3.708294	14	KCNMA1	0.604774	3.046435
4	AK4	1.504579	4.992321	14	SH2D4B	5.340396	9.618759
4	SAMD11	4.762097	7.852872	16	TMPRSS2	2.354945	4.503833
5	HTR5A	0.974168	2.280268	18	PCDHGA1	2.160096	3.811324
6	FAR2	9.056489	14.21492	18	PCDHGA11	1.870698	3.032432
6	PON2	29.82375	47.43006	18	PCDHGA12	4.183007	6.653272
6	PON3	4.75482	10.42457	18	PCDHGA5	3.940849	8.307316
6	PPP1R9A	2.808652	4.431538	18	PCDHGA7	2.871328	6.372722
7	ZFP108	1.377681	4.014484	18	PCDHGB1	1.833306	6.897277
7	ZFP112	1.966027	3.644029	18	PCDHGB6	3.148853	7.695032
8	4933408N05RIK	4.086171	6.466299	18	VAULTRC5	1.047499	4.04117
8	CDH5	1.322783	2.811924	19	E330013P04RIK	0.763709	2.504887
8	EFNB2	4.908241	7.805975	19	PIP5K1B	6.930614	12.25155
8	FREM3	1.527418	2.59749				

3.1.3 Validation of genetic switching using LORDG based 3D modeling at various resolutions

To re-validate the switching genes, we have performed 3D modeling of the entire genome at 20kb resolution. We implemented LorDG, a method capable of modeling the entire genome without compromising on its robustness against the noise. LorDG is optimized by the highly scalable adaptive step-size gradient descent method and uses nonlinear Lorentzian function as the objective function. It is useful for constructing the

structure of the entire genome that contains inconsistent inter-chromosomal interactions and spurious intra-chromosomal contacts. Its resilience against noisy contacts and scalability makes it a suitable method for the current study.

Our 3D modeling indicated that Satb1, which is important for sustaining multipotency is localized away from the nuclear periphery while it is found to be close to periphery at Pro-B stage conversely, Satb2 is observed to be present close to the nuclear periphery in Pre-Pro-B and re-localized away from the periphery in the Pro-B cell stage (**Figure 3.3A, B, C**) Similarly, in Pax5^{-/-} cells, Zbtb16 which is important for NK cell development is found to be away from the nuclear periphery whereas Ikaros (Ikzf1) a known important factor for B cell development is close to the periphery in the absence of Pax5, a B cell master regulator. On the other hand, the ectopic expression of Pax5 in Pax5^{-/-} cells shows that Ikaros has moved away from the periphery whereas Zbtb16 has moved close to the periphery (**Figure 3.4A, B, C**).

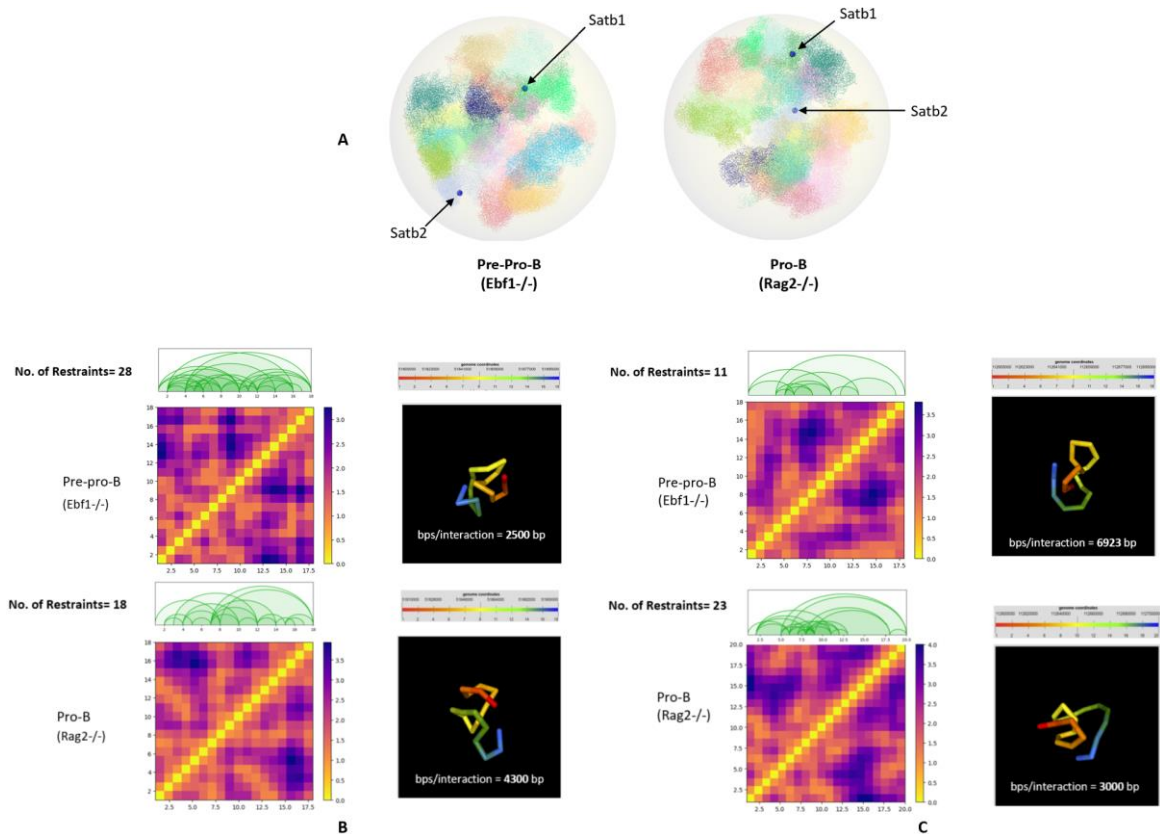


Figure 3.3 The 3D modeling validates that genes undergo nuclear re-localization during pre-pro-B to pro-B cell transition. A) Satb1 and Satb2 undergo nuclear re-localization during Pre-pro-B to pro-B transition. (B,C) Heatmap and 3D structures of respective genes generated at 5kb resolution.

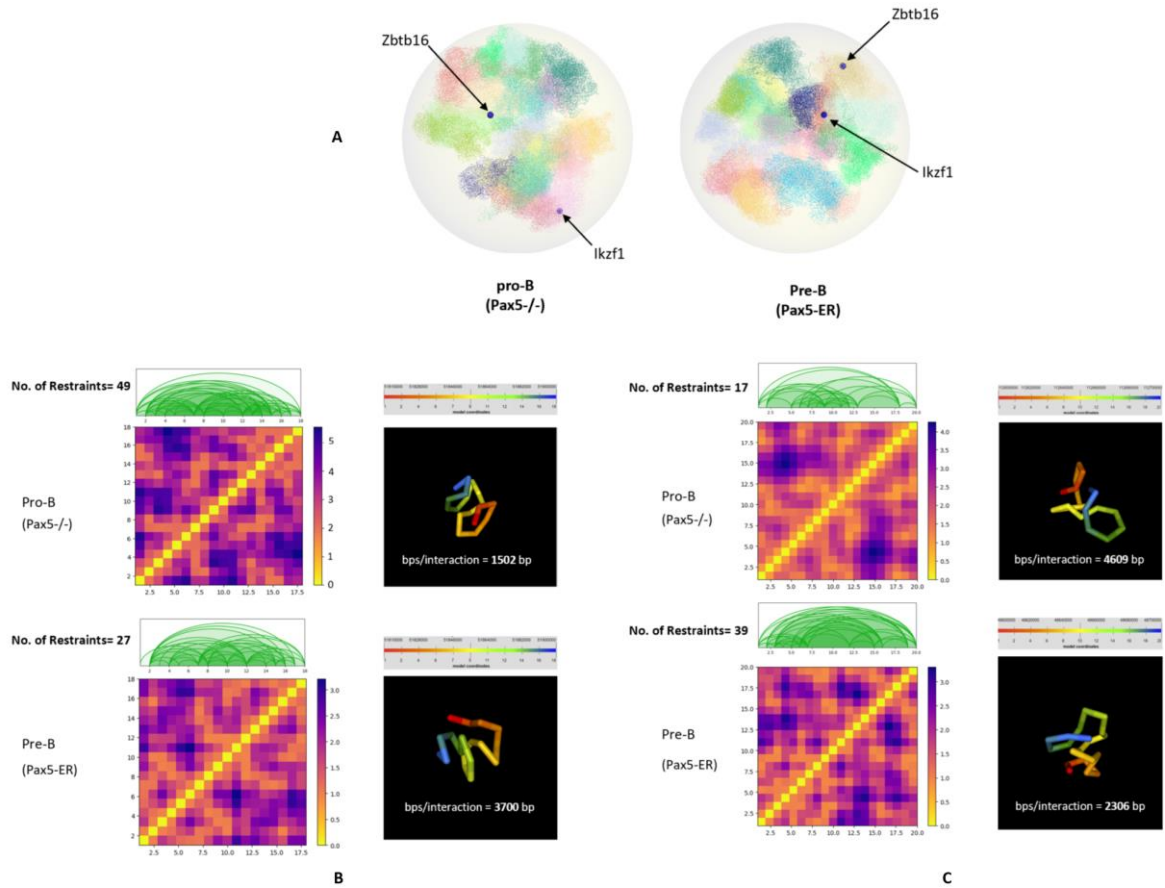


Figure 3.4 The 3D modeling validates that genes undergo nuclear re-localization during pro-B to pre-B cell transition. A) Zbtb16 and Ikzf1 undergo nuclear re-localization during pro-B to pre-B transition. (B,C) Heatmap and 3D structures of respective genes generated at 5kb resolution.

3.1.4 The molecular basis for genetic switch

With Hi-C like techniques, knowledge about 3D genome organization is being accumulated at an enormous rate, yet what determines cell-type-specific gene regulation is still a grey area which needs to be explored. The chromatin-remodeling factors and the transcriptional factors aid in facilitating the structural rearrangement of the chromatin. It is possible that delineating the combinatorial and cooperative interactions between these molecular factors would provide a consensus model for gene regulation. Here, in this section, we have attempted to understand the molecular basis for genetic switch. Since chromatin architecture is maintained by both transcription factors and chromatin regulators (Kim et al. 2016; Stadhouders et al. 2018; van Steensel and Furlong 2019), we reasoned that understanding the combinatorial functioning of the molecular factors would be vital for unravelling the spatial organization of chromatin. For this, we used a

unique approach that combines the differential chromatin interaction profile information from Hi-C with binding profiles obtained from several hundreds of ChIP-Seq data sets to identify the functionally relevant TRs associated with chromatin dynamics. These analyses have led to identification of the molecular basis for chromatin switch. Binding site enrichment analysis revealed that among the regions that undergo genetic switch from B to A compartments, while cells transit from pre-pro-B to pro-B cell stage, are associated with a distinct set of regulatory molecules including Ebf1, Ezh2 and chromatin modifiers like CTCF and Rad21 (**Figure 3.5A and Figure 3.5B**). Conversely, among the regions that showed genetic switch from A to B compartments (representing decreased number of interactions in pro-B), as the cells transited from pre-pro-B to pro-B cell stage, contain TFs such as Sox2 and FosB and the chromatin modifiers like RAD51 (**Figure 3.5C and Figure 3.5D**). These results indicate a direct relationship between the increase or decrease of chromatin interactions with respect to the binding of transcriptional regulators in a cell-type specific manner. The same approach was used to identify transcriptional regulators from the Hi-C data of pro-B and pre-B cells. These analyses revealed that the regions with increased interactions associated a significant number of known lineage-specific molecular regulators like Ebf1, Pax5 and Rag1, whereas the regions with decreased interactions are enriched with Nanog, Gata6, and Sox2 as the cells transited from pro-B to pre-B.

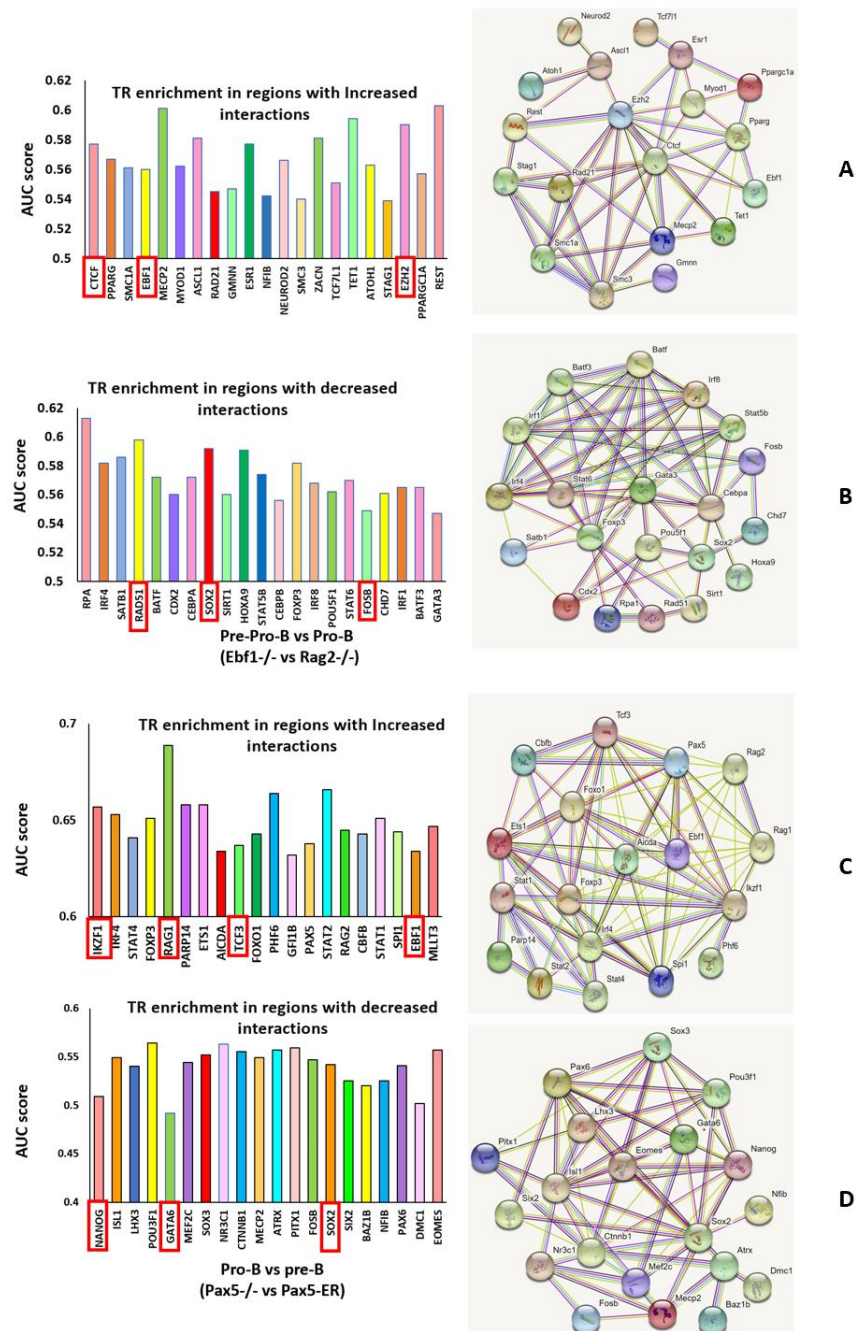


Figure 3.5 Transcriptional regulators are responsible for genetic switch.

A) The histogram representing enrichment of top 20 transcriptional regulators at regions that have increased Hi-C interactions in pro-B when compared to pre-pro-B.

B) The histogram representing enrichment of top 20 transcriptional regulators at regions that have decreased Hi-C interactions in pro-B when compared to pre-pro-B.

C) and D) indicate the TR enrichment in regions with increased and decreased interactions during Pro-B to Pre-B transition; On the right panel of each histogram, the TR networks obtained from string database indicate that all the top 20 TRs interact with each other.

3.2 Interplay between Ebf1 and Pax5 establishes B-lineage signature

One of the major goals of this study was to understand the molecular relationship between TF-dependent chromatin architecture and the differential transcriptional cascade. The formation of B-cell specific progenitor cells during hematopoiesis is under the tight control of gene-regulatory networks controlled by a distinct set of indispensable transcription factors (Master regulators). It has been shown that Ebf1 and Pax5 are essential for induction of early B lineage gene expression program and their targeted inactivation results in a complete block prior to B cell commitment (Lin and Grosschedl 1995; Pongubala et al. 2008; Schebesta, Heavey, and Busslinger 2002). Both Ebf1 and Pax5 establish B cell program by activating a spectrum of genes vital for B cell development and by restricting alternate lineage genes.

3.2.1 Ebf1 governs B lineage specification and commitment in positive feedback manner through Pax5

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^{-/-} or Pax5^{-/-} progenitor cells. As expected, Ebf1 and/or Pax5 induced a spectrum of genes associated with B cell identity (e.g., Cd79a, Cd79b, Foxo1, Vpreb1) including those that are involved in pre-B and B cell receptor signaling, antigen presentation, VDJ recombination (Rag1 and Rag2), and DNA repair. Conversely, Ebf1 and/or Pax5 repressed a subset of genes that are involved in the development and function of natural killer cells (e.g., Zbtb16), dendritic and T cells (e.g., Gata3, Bcl11b) (**Table 3.5 and Table 3.6; Figure 3.5**).

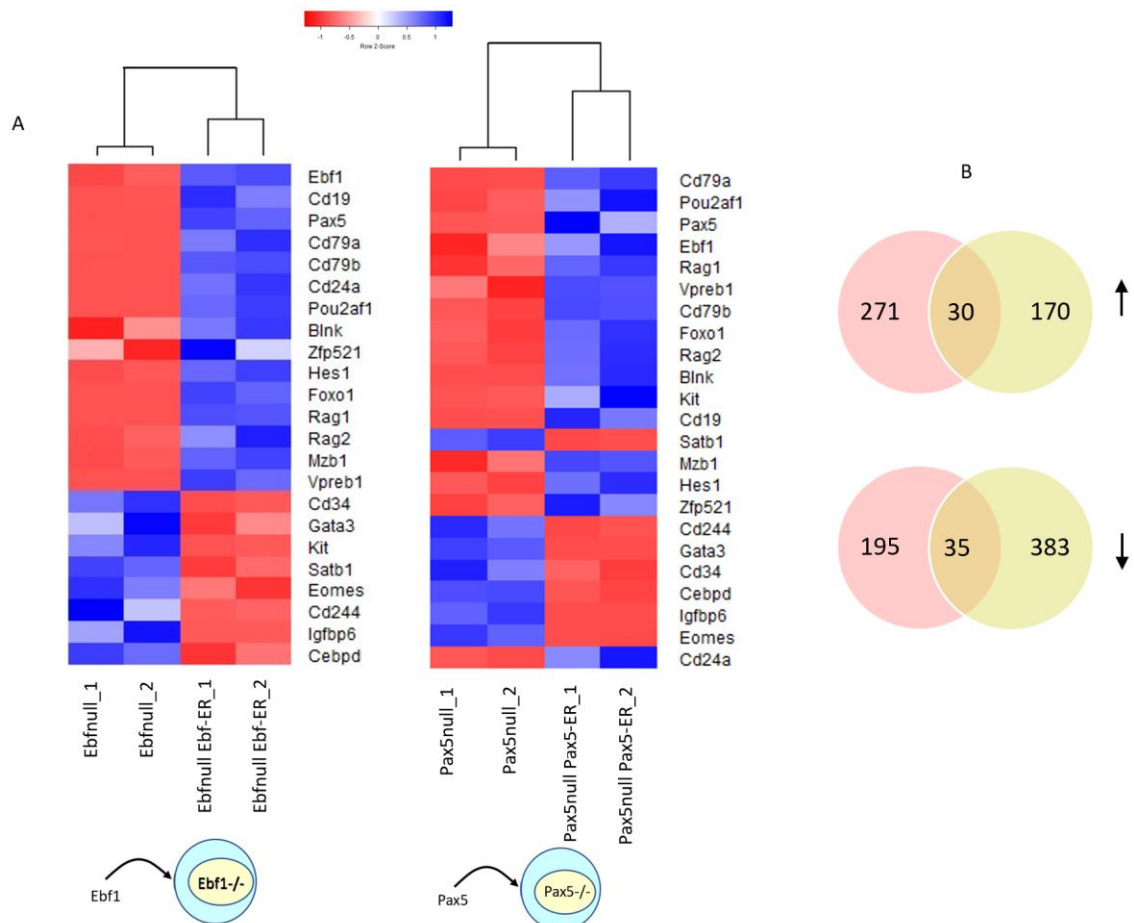


Figure 3.6 Ebf1 and Pax5 promote B cell development by activating B-lineage specific genes and silencing alternate lineage genes.

A) Heatmap showing the genome-wide expression patterns of B lineage-specific genes (fold change ≥ 2 ; P-value < 0.05) obtained by transcriptome analysis of pre-pro-B cells (Ebf1^{-/-} progenitors) transduced with Ebf1.

B) Heatmap showing the genome-wide expression patterns of B lineage-specific genes (fold change ≥ 2 ; P-value < 0.05) obtained by transcriptome analysis of pro-B cells (Pax5^{-/-} progenitors) transduced with Pax5.

C) Venn diagram representing the number of commonly and uniquely upregulated targets of Ebf1 and/or Pax5

Table 3.5 List of differentially expressed genes (log2 FC 4; P<0.05) obtained by ectopic expression of Ebf1 in Ebf1-/- cells

Genes	baseMean	log2 Fold Change	lfcSE	stat	Pvalue	Padj
Mreg	249.3	5.4	0.5	10.6	0.00	0.00
Slc11a1	30.0	4.1	1.2	3.3	0.00	0.01
Neu4	17.9	7.6	2.0	3.8	0.00	0.00
Slc45a3	55.4	6.2	1.2	5.2	0.00	0.00
Lgr6	811.7	11.6	1.4	8.1	0.00	0.00
Tnnt2	2476.5	9.7	0.5	20.1	0.00	0.00
Pkp1	349.6	11.9	1.5	8.0	0.00	0.00
Pcp4l1	10.4	6.8	2.3	3.0	0.00	0.03
Atf3	16.5	6.0	2.1	2.8	0.00	0.05
Akap12	5702.8	5.7	0.2	25.7	0.00	0.00
Enpp3	46.1	8.9	1.7	5.2	0.00	0.00
Cd24a	25604.1	5.0	0.2	24.0	0.00	0.00
1700027J07Rik	143.4	10.6	1.5	6.9	0.00	0.00
Ank3	133.7	10.5	1.5	6.8	0.00	0.00
Chchd10	3002.0	4.8	0.3	15.4	0.00	0.00
Gm867	47.5	5.8	1.3	4.6	0.00	0.00
Vpreb3	55.7	9.2	1.6	5.6	0.00	0.00
Smtnl2	25.3	4.2	1.3	3.1	0.00	0.02
Tusc5	129.8	10.4	1.5	6.8	0.00	0.00
Arl5c	531.7	4.8	0.3	13.8	0.00	0.00
Myl4	2040.6	4.3	0.3	15.9	0.00	0.00
Gh	18.8	7.6	1.9	3.9	0.00	0.00
Cd79b	1803.4	5.2	0.3	19.9	0.00	0.00
Sdk2	43.3	7.4	1.6	4.5	0.00	0.00
Kif19a	996.9	5.8	0.4	14.5	0.00	0.00
Btbd17	32.2	6.0	1.4	4.3	0.00	0.00
Gprc5c	295.6	5.6	0.5	11.5	0.00	0.00
Evpl	139.1	5.7	0.7	8.5	0.00	0.00
Gcgr	26.6	5.7	1.5	3.7	0.00	0.00
Edaradd	14.9	4.8	1.7	2.8	0.00	0.05
Rbm24	199.6	9.6	1.5	6.5	0.00	0.00
Cplx2	119.8	10.3	1.5	6.7	0.00	0.00
Trpc7	59.3	9.3	1.6	5.7	0.00	0.00
Tppp	183.4	4.1	0.6	6.7	0.00	0.00
Adgrv1	76.5	6.6	1.1	6.2	0.00	0.00
Cdc20b	83.3	6.3	1.0	6.4	0.00	0.00
Gzma	788.6	5.6	1.1	5.3	0.00	0.00
Thrb	2458.3	5.8	0.3	22.9	0.00	0.00
Rnase13	18.6	7.6	2.0	3.9	0.00	0.00
Mmp14	470.7	5.7	0.4	14.3	0.00	0.00
Gfra2	2055.0	8.6	0.4	21.1	0.00	0.00

Ankrd33b	60.7	5.5	1.0	5.8	0.00	0.00
Tmprss6	118.4	6.1	0.8	7.8	0.00	0.00
1500009C09Rik	10.4	6.8	2.3	2.9	0.00	0.04
6030408B16Rik	95.4	7.0	1.0	6.8	0.00	0.00
Ppl	1571.8	5.4	0.3	16.8	0.00	0.00
Emp2	23.3	6.5	1.8	3.6	0.00	0.01
Igl1	11815.8	11.3	0.4	28.9	0.00	0.00
Vpreb1	2518.5	8.7	0.4	23.1	0.00	0.00
Vpreb2	1495.6	10.9	0.9	12.6	0.00	0.00
Fetub	126.2	10.4	1.6	6.7	0.00	0.00
A230028O05Rik	12.2	7.0	2.2	3.2	0.00	0.02
Hes1	453.5	4.1	0.4	10.8	0.00	0.00
Mfi2	49.7	6.0	1.2	5.2	0.00	0.00
Adcy5	339.2	5.2	0.4	11.8	0.00	0.00
Clic6	49.8	5.6	1.1	5.0	0.00	0.00
Ggnbp1	145.3	4.1	0.6	6.6	0.00	0.00
Ip6k3	51.2	9.1	1.7	5.4	0.00	0.00
Rsph9	568.8	4.7	0.3	13.6	0.00	0.00
Galnt14	24.9	4.2	1.3	3.4	0.00	0.01
Fhod3	94.8	5.3	0.8	7.1	0.00	0.00
Siglec15	11.5	6.9	2.2	3.2	0.00	0.02
Gal	464.0	9.2	0.9	10.3	0.00	0.00
Nrxn2	731.4	4.4	0.3	14.5	0.00	0.00
Ms4a1	229.3	11.2	1.5	7.5	0.00	0.00
Fam189a2	11.2	6.9	2.3	3.0	0.00	0.04
Gucy2g	22.9	7.9	1.9	4.2	0.00	0.00
Ttll11	65.1	4.7	0.9	5.2	0.00	0.00
Ccdc141	98.6	7.6	1.2	6.4	0.00	0.00
Gm13889	67.2	9.5	1.6	5.8	0.00	0.00
Rag1	7153.0	4.6	0.2	22.0	0.00	0.00
Tcf15	152.8	10.7	1.5	6.9	0.00	0.00
Myl9	42.1	7.4	1.7	4.5	0.00	0.00
Car13	103.3	6.3	0.9	7.2	0.00	0.00
Car1	108.2	7.1	1.1	6.7	0.00	0.00
Car3	388.7	8.3	0.7	11.1	0.00	0.00
Kirrel	74.0	6.6	1.1	6.1	0.00	0.00
S100a3	15.3	7.3	2.2	3.4	0.00	0.01
Hmgcs2	157.4	4.8	0.6	8.1	0.00	0.00
Fam46c	49.0	4.8	1.1	4.6	0.00	0.00
Cfi	14.0	5.8	2.0	2.9	0.00	0.05
Sit1	124.9	4.6	0.8	6.0	0.00	0.00
5730488B01Rik	162.4	4.7	0.6	7.7	0.00	0.00
Gabbr2	18.1	6.1	2.0	3.1	0.00	0.02
Svep1	22.9	6.4	1.9	3.5	0.00	0.01
Col27a1	37.1	8.6	1.8	4.9	0.00	0.00
Slc5a9	552.2	6.7	2.3	2.9	0.00	0.04

Heyl	387.5	7.0	0.6	12.7	0.00	0.00
Slc45a1	27.6	8.2	1.8	4.5	0.00	0.00
Steap4	162.6	4.8	0.6	7.4	0.00	0.00
Htra3	2099.4	4.0	0.3	14.3	0.00	0.00
Bst1	438.1	4.7	0.4	12.0	0.00	0.00
Msi1	20.3	5.4	1.8	3.0	0.00	0.03
Upk3bl	11.3	6.9	2.2	3.1	0.00	0.03
Col26a1	975.8	13.3	1.5	9.1	0.00	0.00
Ephb4	61.1	5.8	1.0	5.7	0.00	0.00
Elfn1	45.5	7.5	1.6	4.6	0.00	0.00
Bhlha15	978.0	5.1	0.3	17.1	0.00	0.00
Kpna7	36.3	8.6	1.7	5.0	0.00	0.00
Aqp1	249.1	4.2	0.4	9.6	0.00	0.00
Clec4f	19.6	6.2	1.9	3.3	0.00	0.01
D6Ertd527e	35.0	7.1	1.7	4.2	0.00	0.00
Plxna1	3900.9	5.0	0.3	18.9	0.00	0.00
Tmem72	26.6	6.7	1.8	3.7	0.00	0.00
Rasgef1a	15.8	7.4	2.0	3.6	0.00	0.00
1700072O05Rik	39.7	5.2	1.2	4.5	0.00	0.00
Cecr2	21524.9	4.2	0.2	21.4	0.00	0.00
Clec4a3	10.3	6.8	2.3	3.0	0.00	0.04
Kcna5	19.6	7.7	1.9	4.0	0.00	0.00
Ssc5d	1340.5	9.0	0.5	17.2	0.00	0.00
Sbk2	20.6	6.3	1.9	3.4	0.00	0.01
Ckm	104.8	8.7	1.5	5.7	0.00	0.00
Dmrtc2	92.2	7.5	1.2	6.3	0.00	0.00
Cd79a	879.9	6.2	0.4	17.8	0.00	0.00
Chst8	23.6	5.6	1.6	3.4	0.00	0.01
Nav2	5641.5	5.2	0.2	24.1	0.00	0.00
Pde2a	620.8	5.7	0.4	14.7	0.00	0.00
Apbb1	198.2	4.3	0.5	8.5	0.00	0.00
Cd19	77.0	6.6	1.1	6.3	0.00	0.00
Eps8l2	1668.3	7.5	0.4	20.8	0.00	0.00
B230206H07Rik	83.2	8.3	1.5	5.4	0.00	0.00
Ifitm10	90.5	5.8	0.9	6.5	0.00	0.00
H19	79823.3	4.3	0.2	19.4	0.00	0.00
Ano1	584.6	5.0	0.3	14.5	0.00	0.00
Csgalnact1	48.6	4.8	1.1	4.6	0.00	0.00
Clgn	615.6	10.2	1.1	9.6	0.00	0.00
Mt2	12.7	7.1	2.2	3.2	0.00	0.02
Cpne2	1921.3	4.3	0.3	17.3	0.00	0.00
Bean1	61.0	9.3	1.6	5.8	0.00	0.00
Mmp13	56.0	9.2	1.6	5.6	0.00	0.00
Mpzl2	52.3	9.1	1.7	5.4	0.00	0.00
Scn2b	89.7	7.4	1.2	6.2	0.00	0.00
Scn4b	70.3	9.5	1.6	6.0	0.00	0.00

Tmprss4	14.8	5.8	2.0	2.9	0.00	0.04
Apoa4	515.1	12.4	1.5	8.4	0.00	0.00
Plet1	169.2	6.6	0.7	9.1	0.00	0.00
Pou2af1	712.7	8.6	0.6	14.0	0.00	0.00
Gm684	18.7	4.6	1.6	3.0	0.00	0.03
Stra6	90.7	9.9	1.6	6.2	0.00	0.00
I730028E13Rik	17.2	5.0	1.7	3.0	0.00	0.03
Adamts7	372.3	9.5	1.1	8.8	0.00	0.00
Xirp1	10.6	6.8	2.3	3.0	0.00	0.03
Zcchc5	15.8	7.4	2.3	3.3	0.00	0.02
Chst10	12.2	-7.1	2.2	-3.3	0.00	0.01
Sec16b	15.1	-7.4	2.0	-3.6	0.00	0.00
Xcl1	97.0	-4.1	0.7	-5.8	0.00	0.00
Pbx1	10.3	-6.9	2.3	-3.0	0.00	0.03
Palld1	30.8	-4.5	1.2	-3.8	0.00	0.00
Il9r	20.2	-6.3	1.8	-3.4	0.00	0.01
Havcr1	20.0	-7.8	1.9	-4.1	0.00	0.00
Alox15	15.7	-5.2	1.8	-2.9	0.00	0.04
Pipox	9.9	-6.8	2.3	-3.0	0.00	0.04
Wfdc17	296.5	-4.1	0.5	-8.7	0.00	0.00
Wfdc18	10.5	-6.9	2.3	-3.0	0.00	0.03
Krt24	18.3	-7.7	2.0	-3.8	0.00	0.00
Fam171a2	16.6	-4.5	1.5	-2.9	0.00	0.04
Fn3k	19.5	-4.5	1.5	-3.0	0.00	0.03
Calm4	14.1	-7.3	2.3	-3.2	0.00	0.02
Pxdc1	15.4	-7.4	2.1	-3.6	0.00	0.01
Lincenc1	20.7	-7.9	1.9	-4.1	0.00	0.00
Gm5086	46.1	-4.2	0.9	-4.5	0.00	0.00
Cdhr1	10.7	-6.9	2.2	-3.1	0.00	0.03
Egr3	20.9	-4.3	1.5	-2.9	0.00	0.04
Sntb1	40.8	-4.1	1.0	-4.1	0.00	0.00
Col2a1	20.3	-4.2	1.4	-2.9	0.00	0.04
Ccdc184	20.2	-4.8	1.5	-3.3	0.00	0.01
Krt8	10.4	-6.9	2.3	-3.0	0.00	0.03
Upk1b	93.5	-5.4	0.8	-6.5	0.00	0.00
Cd200	14.6	-7.4	2.1	-3.6	0.00	0.01
Grm4	15.4	-5.1	1.8	-2.9	0.00	0.04
Prmt1	11.8	-7.1	2.2	-3.2	0.00	0.02
C2	9.9	-6.8	2.3	-3.0	0.00	0.04
Slc6a7	17.2	-6.1	1.9	-3.2	0.00	0.02
BC021614	16.5	-4.9	1.7	-2.8	0.00	0.05
Ctsf	21.4	-5.3	1.6	-3.4	0.00	0.01
Cd6	9.9	-6.8	2.3	-3.0	0.00	0.04
Apba1	26.9	-5.2	1.3	-3.9	0.00	0.00
Adrb1	29.3	-4.5	1.3	-3.5	0.00	0.01
Sohlh1	11.8	-7.1	2.2	-3.2	0.00	0.02

Olfm1	188.1	-4.2	0.6	-7.4	0.00	0.00
Cerkl	25.3	-4.1	1.2	-3.3	0.00	0.01
Gm10804	18.4	-7.7	2.0	-3.9	0.00	0.00
Ddit4l	66.5	-5.7	0.9	-6.1	0.00	0.00
Enho	135.6	-4.7	0.6	-7.6	0.00	0.00
Slc24a2	12.4	-7.1	2.2	-3.3	0.00	0.01
Gm6455	65.0	-4.2	0.8	-5.2	0.00	0.00
Gm5861	62.8	-4.1	0.8	-5.2	0.00	0.00
Gm6460	63.1	-4.2	0.8	-4.9	0.00	0.00
Mapre3	23.2	-5.7	1.6	-3.7	0.00	0.00
Lrrc66	9.5	-6.8	2.3	-2.9	0.00	0.04
C130026L21Rik	52.2	-6.8	1.3	-5.1	0.00	0.00
Lrtm2	18.0	-5.4	1.7	-3.2	0.00	0.02
Klri1	10.7	-6.9	2.3	-3.0	0.00	0.03
Apoc1	31.0	-4.6	1.2	-3.7	0.00	0.00
4732471J01Rik	36.4	-4.6	1.1	-4.2	0.00	0.00
Kcnj11	11.3	-7.0	2.2	-3.2	0.00	0.02
Vegfc	18.1	-5.4	1.7	-3.2	0.00	0.02
Asb5	18.2	-7.7	2.0	-3.9	0.00	0.00
Hsf4	15.3	-7.4	2.0	-3.6	0.00	0.00
Cyp11a1	39.4	-5.1	1.2	-4.1	0.00	0.00
Igdcc4	21.4	-6.5	1.8	-3.6	0.00	0.01
Smim10l2a	15.3	-7.4	2.0	-3.6	0.00	0.00
Ar	10.1	-6.8	2.3	-2.9	0.00	0.04

Table 3.6 List of differentially expressed genes (log2 FC 4; P<0.05) obtained by ectopic expression of Pax5 in Pax5-/- cells

Genes	baseMean	log2 Fold Change	lfcSE	stat	Pvalue	Padj
Cd55b	30.7	4.6	1.3	3.6	0.00	0.01
Cd55	150.5	5.4	0.7	8.0	0.00	0.00
Pcp4l1	125.8	6.1	0.8	7.4	0.00	0.00
Raet1e	47.0	4.6	1.1	4.2	0.00	0.00
1700027J07Rik	189.6	7.2	0.8	8.7	0.00	0.00
Adam19	41.4	7.3	1.7	4.2	0.00	0.00
Lrrc48	38.7	4.4	1.3	3.3	0.00	0.01
Ccdc42	395.6	12.1	1.5	8.1	0.00	0.00
Myh10	10096.3	4.3	0.3	16.3	0.00	0.00
Scn4a	214.7	5.9	0.7	9.0	0.00	0.00
Itgb4	126.3	4.2	0.7	6.2	0.00	0.00
Nptx1	20.9	7.8	2.1	3.8	0.00	0.00
Gm266	120.9	10.3	1.6	6.6	0.00	0.00
Tdrd9	12.1	7.0	2.4	3.0	0.00	0.03
Rd3l	39.7	8.7	1.8	4.9	0.00	0.00
Rapgef5	369.1	4.6	0.4	10.3	0.00	0.00

Edaradd	772.5	6.8	0.5	13.9	0.00	0.00
Gmpr	321.0	4.4	0.5	9.7	0.00	0.00
Wnk2	96.1	6.1	1.0	6.0	0.00	0.00
Diras2	213.9	4.1	0.6	7.4	0.00	0.00
Adcy2	20.9	7.8	2.0	3.8	0.00	0.00
Crhbp	287.0	6.6	0.6	10.3	0.00	0.00
Dnase1l3	22.8	4.7	1.6	2.9	0.00	0.04
Rnase12	34.5	8.5	1.9	4.6	0.00	0.00
Phyhip	42.5	7.4	1.7	4.3	0.00	0.00
Sema5a	15.3	7.4	2.2	3.3	0.00	0.01
Cacna1i	45.4	7.5	1.7	4.4	0.00	0.00
Sox8	25.4	5.5	1.6	3.3	0.00	0.01
Pde9a	21.6	7.9	2.0	3.9	0.00	0.00
Sapcd1	41.1	4.3	1.2	3.7	0.00	0.00
Mtcl1	25.0	8.1	2.1	3.9	0.00	0.00
9430020K01Rik	114.6	5.9	0.9	6.8	0.00	0.00
Klhl14	785.7	13.0	1.5	8.8	0.00	0.00
4930426D05Rik	75.3	9.7	1.6	5.9	0.00	0.00
Flrt1	275.6	4.8	0.6	8.4	0.00	0.00
Ms4a1	104.0	6.6	1.0	6.6	0.00	0.00
Blnk	2911.9	7.5	0.3	21.4	0.00	0.00
Cnnm1	16.3	7.5	2.2	3.4	0.00	0.01
Gchfr	16.7	5.0	1.8	2.8	0.01	0.05
Bfsp1	16.5	7.5	2.3	3.3	0.00	0.01
Kcnb1	13.2	7.1	2.3	3.0	0.00	0.02
Smad9	32.6	8.5	1.9	4.6	0.00	0.00
Csf1	2704.5	4.8	0.3	16.2	0.00	0.00
Bcar3	201.9	4.3	0.6	7.3	0.00	0.00
Enpep	997.2	12.0	1.4	8.5	0.00	0.00
Gm5712	21.1	7.8	2.0	3.9	0.00	0.00
Pax5	12643.8	5.4	0.3	18.1	0.00	0.00
Heyl	80.2	5.3	1.0	5.6	0.00	0.00
Smpdl3b	63.0	4.6	1.0	4.4	0.00	0.00
Alpl	548.9	12.5	1.5	8.4	0.00	0.00
Fblim1	86.8	9.9	1.7	5.8	0.00	0.00
Psapl1	40.5	4.4	1.3	3.3	0.00	0.01
Uchl1	788.0	4.4	0.3	12.9	0.00	0.00
Crybb1	22.3	7.9	2.0	4.0	0.00	0.00
Tpst1	276.0	4.7	0.5	9.1	0.00	0.00
Fam131b	49.7	9.1	1.8	5.1	0.00	0.00
Mmrn1	17.0	7.5	2.2	3.5	0.00	0.01
Hdac11	15.3	7.4	2.2	3.3	0.00	0.01
Pianp	102.4	5.0	0.9	5.8	0.00	0.00
Shisa7	19.9	6.2	2.0	3.2	0.00	0.02
Ehd2	53.1	4.6	1.2	3.9	0.00	0.00
Ceacam9	448.1	8.2	0.7	11.7	0.00	0.00

Pinlyp	39.5	8.7	1.8	4.9	0.00	0.00
Ffar1	27.9	6.7	1.8	3.6	0.00	0.00
Cd22	3163.4	4.6	0.3	16.9	0.00	0.00
Aldh1a3	14.4	7.3	2.3	3.1	0.00	0.02
Tmc7	55.6	9.2	1.7	5.3	0.00	0.00
Cd19	5376.8	7.4	0.3	23.6	0.00	0.00
5830411N06Rik	57.5	4.0	1.2	3.5	0.00	0.01
Muc5ac	25.6	8.1	2.0	4.0	0.00	0.00
Cdkn1c	19.7	7.7	2.3	3.4	0.00	0.01
Hmgb1-rs17	117.5	4.2	0.7	5.8	0.00	0.00
Pkd1l2	16.6	7.5	2.2	3.5	0.00	0.01
Col5a3	15.0	7.3	2.2	3.3	0.00	0.01
B430319G15Rik	21.2	5.2	1.8	2.9	0.00	0.03
Col6a6	33.3	5.0	1.3	3.8	0.00	0.00
Trank1	17.5	7.6	2.5	3.0	0.00	0.03
Scn11a	24.4	8.0	2.0	4.1	0.00	0.00
Ackr2	783.5	7.4	2.4	3.0	0.00	0.03
Cyp8b1	54.5	9.2	1.7	5.4	0.00	0.00
Rps6ka6	21.0	5.2	1.8	2.9	0.00	0.04
Mum1l1	93.2	10.0	1.6	6.2	0.00	0.00
Col19a1	48.9	-5.1	1.3	-3.9	0.00	0.00
Gpr35	62.8	-5.5	1.0	-5.4	0.00	0.00
C1ql2	37.9	-5.5	1.4	-3.9	0.00	0.00
Nckap5	41.6	-8.8	1.8	-4.9	0.00	0.00
Tnni1	51.0	-4.4	1.1	-4.1	0.00	0.00
Pkp1	45.3	-9.0	1.7	-5.1	0.00	0.00
Gpr25	281.5	-4.4	0.5	-8.2	0.00	0.00
Sec16b	20.1	-4.8	1.6	-3.0	0.00	0.03
Cd247	298.5	-5.0	0.5	-10.0	0.00	0.00
Fcer1g	2419.8	-4.4	0.4	-11.5	0.00	0.00
Hebp2	39.2	-4.9	1.3	-3.9	0.00	0.00
Rnf217	16.6	-6.0	2.1	-2.9	0.00	0.04
Gstt1	31.0	-6.1	1.6	-3.9	0.00	0.00
Mybpc1	62.9	-5.2	1.1	-4.7	0.00	0.00
Nr1h4	21.4	-7.9	2.0	-3.8	0.00	0.00
Slc17a8	976.3	-5.3	0.4	-13.8	0.00	0.00
Tbc1d30	18.7	-5.0	1.8	-2.8	0.00	0.05
Apof	51.6	-6.8	1.4	-4.8	0.00	0.00
Wnt9a	40.8	-7.4	1.8	-4.2	0.00	0.00
Adora2b	118.6	-5.8	0.8	-7.2	0.00	0.00
1110028F11Rik	26.1	-4.3	1.5	-2.9	0.00	0.04
Chad	254.9	-4.1	0.5	-7.8	0.00	0.00
Ccr10	89.7	-5.6	0.9	-6.2	0.00	0.00
Adam11	32.0	-5.0	1.3	-3.8	0.00	0.00
Mrc2	125.9	-7.4	1.1	-6.7	0.00	0.00
Nrcam	72.3	-6.6	1.2	-5.7	0.00	0.00

Flrt2	22.4	-8.0	2.1	-3.9	0.00	0.00
Pom121l2	28.9	-4.9	1.4	-3.6	0.00	0.01
Hk3	38.0	-4.1	1.2	-3.4	0.00	0.01
Neurog1	57.0	-4.4	0.9	-4.7	0.00	0.00
Rasgrf2	81.6	-4.1	0.9	-4.6	0.00	0.00
Gap1	113.4	-5.7	1.0	-5.9	0.00	0.00
Itga1	20.1	-4.3	1.5	-2.8	0.00	0.05
Ogdhl	27.7	-4.7	1.4	-3.3	0.00	0.01
1700011H14Rik	1258.2	-4.6	0.3	-13.5	0.00	0.00
Slc35f4	18.9	-6.3	2.0	-3.1	0.00	0.02
Mcpt1	14.7	-7.4	2.6	-2.8	0.01	0.05
Timm8a2	27.3	-6.7	2.1	-3.2	0.00	0.02
Tmprss6	31.5	-7.0	1.8	-3.8	0.00	0.00
Cdc42ep1	10.8	-6.9	2.5	-2.8	0.00	0.05
Fam109b	24.4	-8.1	2.2	-3.8	0.00	0.00
5430421N21Rik	49.5	-4.5	1.4	-3.2	0.00	0.02
Sp7	42.6	-4.3	1.3	-3.3	0.00	0.01
Dgkg	36.3	-4.5	1.2	-3.7	0.00	0.00
Cpn2	33.8	-8.6	1.9	-4.5	0.00	0.00
Lrrc15	1246.3	-4.3	0.3	-14.2	0.00	0.00
Adcy5	523.4	-4.4	0.4	-11.2	0.00	0.00
Ccdc80	89.9	-5.2	0.9	-5.6	0.00	0.00
Morc1	43.1	-4.9	1.3	-3.9	0.00	0.00
Retnlb	22.1	-7.9	2.0	-3.9	0.00	0.00
Ggnbp1	62.4	-4.1	0.9	-4.6	0.00	0.00
Ptcra	253.2	-5.1	0.6	-8.3	0.00	0.00
Lrg1	12.3	-7.1	2.4	-3.0	0.00	0.03
Sall3	13.0	-7.2	2.4	-2.9	0.00	0.03
BC021614	12.8	-7.2	2.3	-3.1	0.00	0.02
Ms4a6c	40.9	-4.6	1.3	-3.4	0.00	0.01
Ms4a6b	46.3	-4.7	1.2	-3.8	0.00	0.00
Vldlr	32.4	-7.0	1.9	-3.8	0.00	0.00
Gm10768	113.8	-23.0	4.8	-4.8	0.00	0.00
Ptgds	78.0	-4.5	0.9	-4.8	0.00	0.00
Tcf15	51.1	-6.3	1.3	-4.7	0.00	0.00
Tox2	61.0	-4.9	1.2	-4.2	0.00	0.00
Chrna4	3005.9	-5.2	0.3	-17.8	0.00	0.00
Trim55	12.2	-7.1	2.4	-2.9	0.00	0.03
Qrfpr	19.2	-6.3	2.0	-3.2	0.00	0.02
Igsf3	39.8	-4.4	1.2	-3.8	0.00	0.00
Fam19a3	50.1	-6.5	1.5	-4.5	0.00	0.00
Snx7	48.2	-4.3	1.3	-3.2	0.00	0.02
Slc44a5	61.6	-4.5	0.9	-4.9	0.00	0.00
Ptger3	22.0	-5.3	1.7	-3.1	0.00	0.02
Tnfsf15	13.4	-7.2	2.3	-3.1	0.00	0.02
Slc5a9	2784.7	-5.2	0.3	-18.3	0.00	0.00

Ermap	14.5	-7.3	2.6	-2.8	0.01	0.05
Bmp8a	257.4	-5.6	0.5	-10.2	0.00	0.00
Grhl3	55.2	-5.0	1.2	-4.2	0.00	0.00
Myom3	37.7	-8.7	1.8	-4.8	0.00	0.00
Hspg2	856.8	-4.2	0.3	-13.1	0.00	0.00
Kcnh2	21.8	-6.5	2.0	-3.2	0.00	0.02
2900005J15Rik	23.4	-8.0	2.0	-4.1	0.00	0.00
Gm9899	11.9	-7.0	2.4	-2.9	0.00	0.03
Zfyve28	29.3	-4.7	1.4	-3.4	0.00	0.01
Cxcl9	50.3	-4.7	1.2	-4.0	0.00	0.00
Sowahb	47.3	-7.6	1.7	-4.5	0.00	0.00
Tmem119	521.7	-5.4	0.4	-13.3	0.00	0.00
Msi1	50.9	-5.2	1.1	-4.8	0.00	0.00
Cyp3a13	109.3	-8.8	1.5	-5.8	0.00	0.00
Prkar1b	12.7	-7.1	2.3	-3.0	0.00	0.03
Grifin	21.5	-5.4	1.7	-3.2	0.00	0.01
Clec5a	84.1	-5.0	1.0	-5.1	0.00	0.00
D6Ertd527e	17.4	-7.6	2.2	-3.5	0.00	0.01
Klrb1	37.7	-4.6	1.5	-3.0	0.00	0.03
Klre1	16.8	-6.0	2.0	-2.9	0.00	0.03
Lair1	24.6	-4.4	1.4	-3.1	0.00	0.02
Psg17	21.5	-7.9	2.0	-3.9	0.00	0.00
Klc3	39.5	-4.2	1.3	-3.3	0.00	0.01
Lrfrn3	73.9	-7.3	1.3	-5.6	0.00	0.00
Chst8	58.0	-4.3	1.0	-4.5	0.00	0.00
Siglec	59.2	-4.5	1.0	-4.4	0.00	0.00
Siglece	43.4	-4.8	1.2	-4.1	0.00	0.00
Itgax	86.9	-4.9	1.0	-4.7	0.00	0.00
Ifitm10	74.2	-5.5	1.0	-5.6	0.00	0.00
Col4a2	30.0	-7.0	1.8	-3.9	0.00	0.00
Slc18a1	50.7	-4.5	1.0	-4.5	0.00	0.00
Tox3	14.0	-7.3	2.4	-3.0	0.00	0.03
Cdh1	1850.5	-4.9	0.4	-13.3	0.00	0.00
Tmed6	958.1	-5.0	0.4	-12.8	0.00	0.00
Cdh13	54.6	-4.9	1.1	-4.5	0.00	0.00
Zfp872	22.1	-5.5	1.7	-3.3	0.00	0.01
Ccr9	8826.0	-4.9	0.3	-17.4	0.00	0.00
C77370	26.9	-4.8	1.5	-3.2	0.00	0.02
Tlr13	42.5	-5.2	1.3	-4.1	0.00	0.00

3.2.2 Ebf1 functions as a major B cell architectural factor

Since targeted inactivation of Ebf1 results in a complete block prior to B cell commitment and Ebf1 is essential for induction B lineage gene expression programs (Lin and Grosschedl

1995; Pongubala et al. 2008). It raised the possibility that chromatin re-localization and associated changes in cis-regulatory interaction landscape may be regulated by Ebf1. To identify the molecular basis for switching (**See section 3.1.4 under Results**), we scanned for highly specific and significant Ebf1 binding sites in cis-regulatory regions of genes that switched to permissive or repressive compartments in pro-B cells, using publicly available databases (Jasper, Homer and Uniprobe). From these analyses, we observed that Ebf1 and/or Pax5 binds (65.3%) to cis-regulatory sequences of differentially switched genes (**Figure 3.6**). Although a subset of genes undergoes differential compartmentalization, this may not account for induction of all B lineage genes. We propose that B lineage gene expression may be regulated by binding of lineage specific transcription factors (E2A, Ebf1, Foxo1, and Pax5) to their target promoter-cis-regulatory interacting elements. To rigorously test this, we integrated promoter-cisregulatory interactions that are captured in pro-B cells (**see section 2.6.2 for more details; Table 3.7**) with binding events of these factors. Importantly, we found that Ebf1 binds either alone or in combination with Pax5 to 57.2% of the promoters and 51.2% of cis-regulatory elements that are involved in long-range interactions as defined by in situ Hi-C. Below table (**Table 3.7**) represents the selected list of functionally relevant Promoter-Enhancer interactions in pro-B cells (Rag2^{-/-}) annotated with binding sites present in either promoter or enhancer regions of each interacting elements.

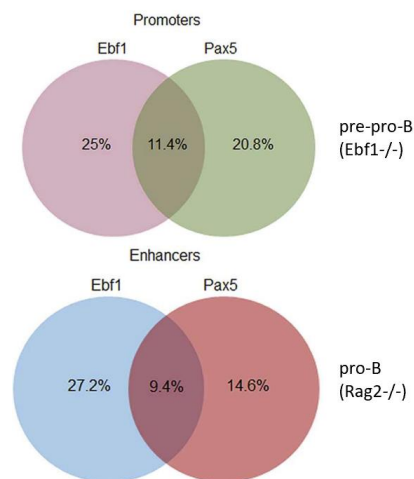


Figure 3.7 Ebf1 regulates B lineage-specific gene expression pattern in part by binding at cis-regulatory interacting elements. Venn diagrams represent the TFs (Ebf1 and Pax5) binding at the promoters and respective cis-regulatory elements.

Table 3.7 Selected list of Promoter-Enhancer interactions in rag2-/- with transcription factor binding sites

Chr	Promoter start	Promoter end	Gene	Chr	Enhancer start	Enhancer end	Ebf1-/- RNA-Seq	Rag2-/- RNA-Seq	Binding sites
10	43579305	43580305	Cd24a	10	43671219	43671619	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43724369	43731244	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43733644	43734894	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43744344	43744369	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43776494	43777219	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43781569	43789644	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43790919	43800319	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43885719	43886269	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43887544	43887994	23.814	3381.22	Ebf1, Pax5
10	43579305	43580305	Cd24a	10	43964344	43965819	23.814	3381.22	Ebf1, Pax5
7	126413832	126414832	Cd19	7	126409386	126411636	0.312599	212.326	Pax5
11	106313538	106314538	Cd79b	11	106287136	106290211	1.81475	180.508	Ebf1, Foxo1, Pax5
11	106313538	106314538	Cd79b	11	106374036	106375361	1.81475	180.508	Ebf1, Foxo1, Pax5
11	106313538	106314538	Cd79b	11	106382236	106382861	1.81475	180.508	Ebf1, Foxo1, Pax5
11	106313538	106314538	Cd79b	11	106383886	106384561	1.81475	180.508	Ebf1, Foxo1, Pax5
11	106313538	106314538	Cd79b	11	106384836	106385161	1.81475	180.508	Ebf1, Foxo1, Pax5
13	30749319	30750319	Irf4	13	30734756	30736856	9.89646	272.093	Pax5
13	30749319	30750319	Irf4	13	30727956	30731456	9.89646	272.093	Pax5
13	30749319	30750319	Irf4	13	30567906	30570356	9.89646	272.093	Pax5
19	40992248	40993248	Blnk	19	41014260	41019510	1.82809	147.128	Ebf1, Pax5
19	40992248	40993248	Blnk	19	40685935	40688235	1.82809	147.128	Ebf1, Pax5
11	98545351	98546351	Ikzf3	11	98470211	98470461	0.471258	71.6364	foxo1
11	98545351	98546351	Ikzf3	11	98443761	98453511	0.471258	71.6364	foxo1
16	23987215	23988215	Bcl6	16	24135414	24135764	7.12781	16.7176	Pax5
16	23987215	23988215	Bcl6	16	24135989	24139239	7.12781	16.7176	Pax5
16	23987215	23988215	Bcl6	16	24262264	24265289	7.12781	16.7176	Pax5
9	32696685	32697685	Ets1	9	32690440	32691015	25.9113	95.4548	Ebf1, Foxo1, Pax5
2	27425875	27426875	Vav2	2	27371655	27372630	6.91188	17.2326	Ebf1, Pax5
2	27425875	27426875	Vav2	2	27343380	27344455	6.91188	17.2326	Ebf1, Pax5
2	27425875	27426875	Vav2	2	27311130	27313955	6.91188	17.2326	Ebf1, Pax5
8	120736424	120737424	Irf8	8	120705800	120706050	66.9074	26.9505	Pax5
8	120736424	120737424	Irf8	8	120706475	120706725	66.9074	26.9505	Pax5
8	120736424	120737424	Irf8	8	120684925	120687450	66.9074	26.9505	Pax5

3.2.3 Ebf1 regulates B lineage-specific gene expression pattern in part by modulating cis-regulatory interactions

It is increasingly becoming evident that the dynamic state of the chromatin is a result of interplay between the regulatory elements, epigenetic modifiers, and physical gene networks. Ebf1 is the primary B cell fate determinant, which promotes B cell fate choice of MPPs by blocking the alternate cell lineages. However, the precise role of Ebf1 in establishment of genome-wide chromatin interactions during B cell development remains unclear. We carried out H3K4me3 and H3K427ac ChIP-Seq in Ebf1^{-/-} and Ebf1^{-/-}-ER. By overlaying the significant loop interactions obtained from Rag2^{-/-} Hi-C data (**See methods section**) with Ebf1-ER H3K27ac peak information, we identified the novel Ebf1 dependent active enhancer elements and their interacting promoter partners. Surprisingly, these analyses revealed that the majority of the Ebf1-dependent cis-regulome is retained in pro-B cells indicating that Ebf1 is responsible for most of the functionally relevant loops in pro-B cells. These studies not only unraveled the key Ebf1 dependent regulatory elements but also aided in assembling physical gene regulatory networks. The newly identified components within the regulatory networks will further be functionally examined for their contributions to B cell development. Altogether, Identification of previously unknown Ebf1 dependent regulatory elements that are important for inducing B cell gene expression program will set stage for future experiments of characterization and understanding of their function.

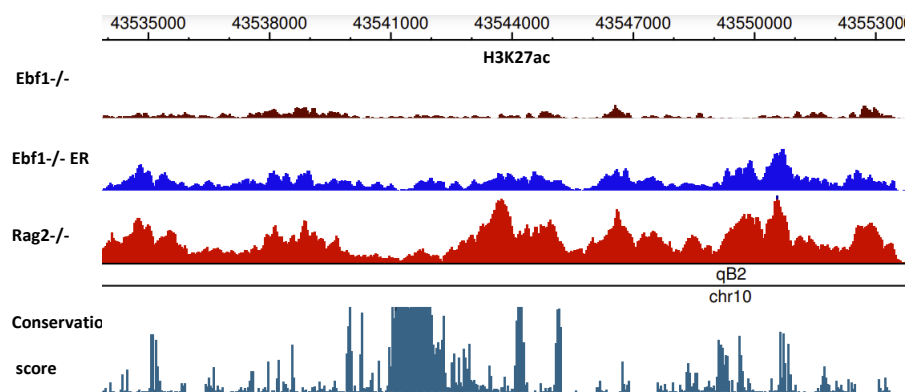


Figure 3.8 Coverage plot showing the H3K27ac enrichment in Ebf1^{-/-} and Ebf1^{-/-}-ER. The enrichment of H3K27ac in Ebf1^{-/-}-ER correlates its enrichment in Rag2^{-/-}. The region near Cd24a gene (Chr10: 43535000-43553000) was selected for representation purpose.

Next, we sought to determine if loss or gain of cis-regulatory 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 (Ebf1^{-/-}) and (Rag2^{-/-}) pro-B cells in addition to interactions identified by overlaying Ebf1-ER H3k27ac. This analysis showed that the expression pattern is closely associated with an increase in the number of cis-regulatory elements. For instance, Cd24a, which is highly induced in pro-B cells, interacts with 8 super-enhancers, of which 7 of them are mediated by Ebf1, whereas it is involved in only 2 such interactions at pre-pro-B cell stage, where its expression is considerably low. 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.

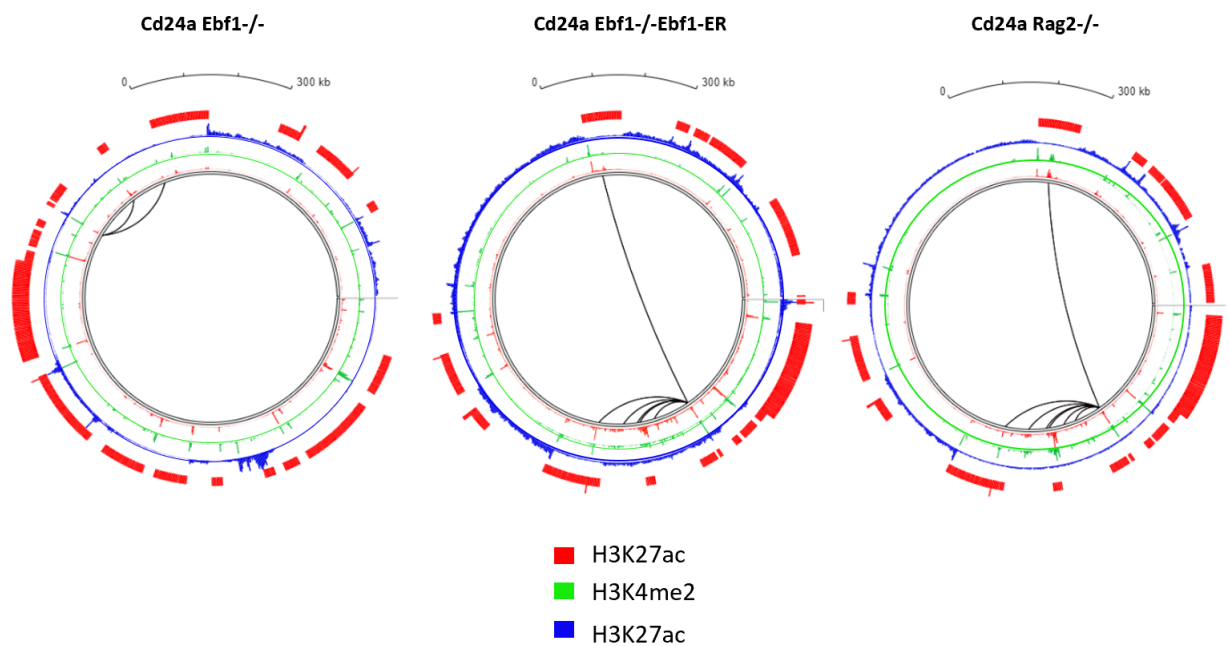


Figure 3.9 Cis-regulatory interaction landscape determines differential gene expression pattern. Circos plots showing promoter-cis-regulatory interactome of Cd24a (Chr11:43.3–44.1 Mb) in Ebf1^{-/-} Ebf1-ER and Rag2^{-/-} cells.

Table 3.8 Super enhancer interaction landscape of gene promoters that are captured in both Ebf1-/-Ebf1-ER and rag2-/- progenitors. (Only top 100 interactions are provided here)

Chr	Promoter start	Promoter end	Promoter gene	Chr	Enhancer start	Enhancer end	Enhancer gene
10	43579305	43580305	CD24A	10	43587460	43589535	CD24A
10	43579305	43580305	CD24A	10	43707260	43707535	1700027J07RIK
10	43579305	43580305	CD24A	10	43755760	43762635	1700027J07RIK
10	43579305	43580305	CD24A	10	43765035	43766285	1700027J07RIK
10	43579305	43580305	CD24A	10	43822310	43831710	QRSL1
10	43579305	43580305	CD24A	10	43898235	43899060	QRSL1
10	43579305	43580305	CD24A	10	45106410	45111385	PREP
12	113151794	113152794	CRIP1	12	1.11E+08	1.11E+08	WDR20
12	113151794	113152794	CRIP1	12	1.13E+08	1.13E+08	GPR132
12	113151794	113152794	CRIP1	12	1.13E+08	1.13E+08	GPR132
12	113151794	113152794	CRIP1	12	1.13E+08	1.13E+08	GPR132
12	113151794	113152794	CRIP1	12	1.13E+08	1.13E+08	4930427A07RIK
2	180256878	180257878	RPS21	2	1.80E+08	1.80E+08	SLCO4A1
9	64173204	64174204	RPL4	9	65072919	65076594	DPP8
9	64173204	64174204	RPL4	9	66280244	66282619	DAPK2
19	5800493	5801493	MALAT1	19	5786950	5788125	MALAT1
19	5800493	5801493	MALAT1	19	5817500	5818975	MALAT1
14	50807684	50808684	RPPH1	14	50969349	50971824	PNP2
14	50807684	50808684	RPPH1	14	51059599	51060974	RNASE12
9	61913856	61914856	RPLP1	9	60991044	60992394	GM5122
9	61913856	61914856	RPLP1	9	61009344	61010294	GM5122
9	61913856	61914856	RPLP1	9	61022044	61022944	GM5122
9	61913856	61914856	RPLP1	9	61053419	61055419	GM5122
9	61913856	61914856	RPLP1	9	61385344	61386519	TLE3
9	61913856	61914856	RPLP1	9	61959969	61961619	PAQR5
6	115807826	115808826	RPL32	6	1.14E+08	1.14E+08	EMC3
6	115807826	115808826	RPL32	6	1.15E+08	1.15E+08	ATG7
6	115807826	115808826	RPL32	6	1.15E+08	1.15E+08	ATG7
6	115807826	115808826	RPL32	6	1.16E+08	1.16E+08	MBD4
6	115807826	115808826	RPL32	6	1.16E+08	1.16E+08	IFT122
3	96219604	96220604	HIST2H2AC	3	95971020	95974370	GM9054
5	100570329	100571329	PLAC8	5	1.01E+08	1.01E+08	PLAC8
5	100570329	100571329	PLAC8	5	1.01E+08	1.01E+08	COQ2
5	100570329	100571329	PLAC8	5	1.01E+08	1.01E+08	COQ2
5	100570329	100571329	PLAC8	5	1.01E+08	1.01E+08	AGPAT9
7	45123473	45124473	RPS11	7	44754683	44755383	VRK3
7	45123473	45124473	RPS11	7	45635133	45636133	RASIP1
7	100493711	100494711	UCP2	7	1.00E+08	1.00E+08	C2CD3
7	100493711	100494711	UCP2	7	1.00E+08	1.00E+08	C2CD3

7	100493711	100494711	UCP2	7	1.00E+08	1.00E+08	C2CD3
7	100493711	100494711	UCP2	7	1.00E+08	1.00E+08	UCP3
7	100493711	100494711	UCP2	7	1.01E+08	1.01E+08	DNAJB13
7	100493711	100494711	UCP2	7	1.01E+08	1.01E+08	P2RY6
7	100493711	100494711	UCP2	7	1.01E+08	1.01E+08	FCHSD2
7	100493711	100494711	UCP2	7	1.01E+08	1.01E+08	ARAP1
7	100493711	100494711	UCP2	7	1.01E+08	1.01E+08	PDE2A
7	81344304	81345304	RPS17	7	80918937	80919512	SEC11A
7	81344304	81345304	RPS17	7	81194087	81200962	PDE8A
11	48800006	48801006	GNB2L1	11	50161177	50164427	TBC1D9B
7	99482678	99483678	SNORD15A	7	99316017	99318692	MAP6
7	99482678	99483678	SNORD15A	7	99544567	99547317	ARRB1
10	128625560	128626560	RPS26	10	1.29E+08	1.29E+08	ERBB3
10	128625560	128626560	RPS26	10	1.29E+08	1.29E+08	ERBB3
7	99482678	99483678	RPS3	7	99316017	99318692	MAP6
7	99482678	99483678	RPS3	7	99544567	99547317	ARRB1
10	80292392	80293392	RPS15	10	77506934	77507709	FAM207A
10	80292392	80293392	RPS15	10	80443271	80445246	TCF3
11	99023171	99024171	TOP2A	11	99104060	99107710	TNS4
3	58524826	58525826	SERP1	3	58501771	58502846	SERP1
17	47593843	47594843	CCND3	17	47526634	47533134	CCND3
17	47593843	47594843	CCND3	17	47552734	47559984	CCND3
17	47593843	47594843	CCND3	17	47566584	47566859	CCND3
17	47593843	47594843	CCND3	17	48106883	48109233	1700067P10RIK
12	111966174	111967174	2010107E04RIK	12	1.12E+08	1.12E+08	PPP1R13B
12	111966174	111967174	2010107E04RIK	12	1.12E+08	1.12E+08	PPP1R13B
12	111966174	111967174	2010107E04RIK	12	1.12E+08	1.12E+08	RD3L
12	111966174	111967174	2010107E04RIK	12	1.12E+08	1.12E+08	TDRD9
12	111966174	111967174	2010107E04RIK	12	1.13E+08	1.13E+08	GPR132
12	111966174	111967174	2010107E04RIK	12	1.13E+08	1.13E+08	GPR132
12	111966174	111967174	2010107E04RIK	12	1.13E+08	1.13E+08	GPR132
11	78983779	78984779	LGALS9	11	79000280	79009430	KSR1
18	35598523	35599523	PAIP2	18	34890076	34896026	ETF1
18	35598523	35599523	PAIP2	18	35745051	35748176	TMEM173
6	145121840	145122840	LRMP	6	1.45E+08	1.45E+08	BCAT1
6	145121840	145122840	LRMP	6	1.45E+08	1.45E+08	BCAT1
6	145121840	145122840	LRMP	6	1.45E+08	1.45E+08	LRMP
6	145121840	145122840	LRMP	6	1.45E+08	1.45E+08	LRMP
6	145121840	145122840	LRMP	6	1.45E+08	1.45E+08	LRMP
6	145121840	145122840	LRMP	6	1.45E+08	1.45E+08	LRMP
19	5490581	5491581	CFL1	19	5786950	5788125	MALAT1
19	5490581	5491581	CFL1	19	6155625	6157425	BATF2
2	62411135	62412135	DPP4	2	62459236	62462011	GCG
17	45572348	45573348	HSP90AB1	17	45580844	45586494	SLC29A1

10	75935732	75936732	CHCHD10	10	75304134	75304984	SPECC1L
10	75935732	75936732	CHCHD10	10	75446784	75449084	4933407G14RIK
10	75935732	75936732	CHCHD10	10	75489859	75492309	GUCD1
10	75935732	75936732	CHCHD10	10	75564984	75566534	LRRC75B
1	55026388	55027388	SF3B1	1	55208005	55212455	RFTN2
5	32459278	32460278	PPP1CB	5	32359344	32361369	PLB1
10	21159676	21160676	MYB	10	21079859	21080934	AHI1
10	21159676	21160676	MYB	10	21081884	21086484	AHI1
10	21159676	21160676	MYB	10	21091909	21094259	AHI1
10	21159676	21160676	MYB	10	21231284	21238659	HBS1L
15	9528277	9529277	IL7R	15	9488525	9494875	IL7R
2	164769834	164770834	UBE2C	2	1.64E+08	1.64E+08	SLPI
2	164769834	164770834	UBE2C	2	1.64E+08	1.64E+08	SLPI
10	45066883	45067883	PREP	10	44447410	44448635	PRDM1
17	80206222	80207222	SRSF7	17	79979574	79984199	HNRNPLL
9	7764106	7765106	TMEM123	9	7829454	7830879	BIRC2

3.3 Igh repertoire undergoes alterations in R1Δ215 and R1.PG but not in R1.KR mice

The murine Igh locus spans 2.7 Mb, with more than 100 VH gene segments located ~100 kb upstream of 13 DH and 4 JH gene segments. Igh recombination is initiated in proB cells by recruitment of RAG to the Igh RC, which encompasses the JH segments and the most downstream DH gene segment, DQ52 (Ji et al. 2010; Teng et al. 2015). DQ52-to-JH recombination occurs through a topologically unconstrained collision-based mechanism (Zhang et al. 2019), whereas recombination of other DH gene segments to JH involves a distinct mechanism in which a JH-bound RAG complex scans chromatin, driven by cohesin-mediated loop extrusion (Zhang et al. 2019; Ba et al. 2020). JH-bound RAG chromatin scanning is arrested by the CBE containing element IGCR1, which lies immediately upstream of DFL16.1, explaining why DFL16.1 is the most frequently used DH gene segment (Ba et al., 2020; Jain et al., 2018; Zhang et al., 2019). Here, in this study, we performed the HTGTS analysis of three new mouse strains in which the RAG1 NTR has been mutated. The mutations eliminate RAG1 ubiquitin ligase activity (P326G, or PG), eliminate the major autoubiquitination site (K233R, or KR), or remove the first 215 amino acids of the NTR (D215). To investigate the possibility that the RAG1 NTR mutations alter gene segment usage we characterized the Igh repertoire in B220+IgM⁺ bone marrow B-lineage

cells (predominantly pre-B cells) and B220+ spleen cells (predominantly naive, mature B cells) using JH1 and JH3 as baits (Lin et al., 2016). This approach reveals the repertoire of VH and DH gene segments that become joined to these JH gene segments. We did not observe major alterations in VH gene segment usage in the RAG1 NTR mutant mice generated for this study. Thus, we focus here on the DH gene segment repertoire, with usage frequency presented for DQ52, DFL16.1, and the intervening DH gene segments considered as a group (DSP; Figure 3. 10(A)). The full dataset of VH and DH gene segment usage is provided in Table 3.9 Table 3.10, Table 3.11, and Table 3.12. R1Δ215 mice exhibited a striking decrease of up to ~10-fold in the frequency of usage of DQ52 in rearrangements involving JH1 and JH3 and in both the bone marrow and spleen relative to WT mice (data for bone marrow and spleen are shown in Figure. 3.10 (B-D)). This was true in total VDJ rearrangements, nonproductive (VDJ-) rearrangements, and DJ rearrangements. Data for productive (VDJ+) rearrangements closely resemble those for total VDJ rearrangements and are not shown. In contrast, usage of DFL16.1 was not substantially or consistently altered in R1Δ215 relative to WT mice (Figure 3.10 B-D).

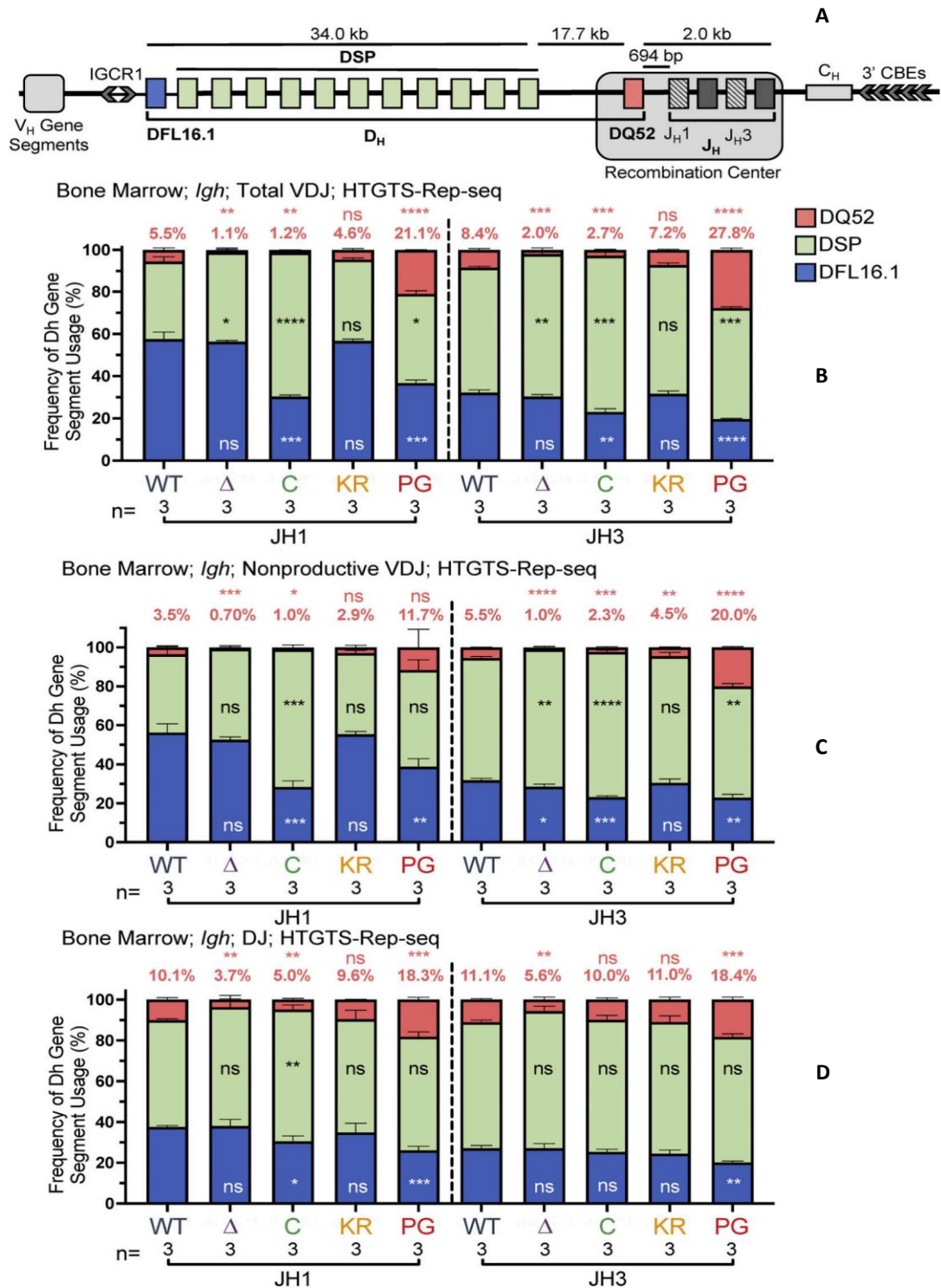


Figure 3.10 DH gene segment usage in *Igh* rearrangements in bone marrow assessed by HTGTS-Rep-seq.

(A) Schematic of murine *Igh* with the RC containing the JH-proximal DQ52 gene segment and four JH segments highlighted. Figure not to scale. Chevrons indicate the orientation of CBEs. CBEs in *VH* portion of locus not depicted.

(B–D) *Igh* HTGTS-Rep-seq data from mice 5–7 wk of age were analyzed for frequency of DFL16.1 (blue), DQ52 (salmon), and intervening 11 DH gene segments (green) in total (B), nonproductive VDJH

(C), or DJ recombination events (D) with JH1 or JH3 in bone marrow. Percentage of DQ52 usage is indicated above each bar. n, number of mice analyzed per genotype. Data are presented as mean with error bars indicating SEM. Statistical significance relative to WT determined by two-tailed unpaired t tests (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$). DSP, intervening DH gene segments considered as a group.

DFL16.1 was the most frequently used DH gene segment in all genotypes analyzed, with usage higher in JH1 than JH3 rearrangements, as reported previously (Lin et al., 2016). We conclude that deletion of the first 215 aa of RAG1 strongly reduces DQ52 usage and that this phenotype arises at the step of DH-to-JH rearrangement and is not due to selection for productive rearrangements.

Table 3.9 Utilization of D_H gene segments among total VDJ rearrangements in primary B220+ IgM- bone marrow cells (JH1 and JH3 primers were used as baits).

D genes	JH1				JH3			
	WT	$\Delta 215$	K233R	P326G	WT	$\Delta 215$	K233R	P326G
Ighd1-1/DFL16.1	13894	7534	4297	2122	6751	5208	3270	1030
Ighd6-1	0	0	0	0	0	0	0	0
Ighd2-3	1907	1774	781	548	2308	2655	1270	502
Ighd6-2	35	17	28	17	28	42	22	15
Ighd2-4	2216	1565	740	533	3660	3854	1859	780
Ighd5-2	46	22	18	22	40	26	11	7
Ighd2-5	1281	1107	484	540	1665	2213	951	381
Ighd5-3	0	0	0	0	0	0	0	0
Ighd6-3	0	0	0	0	0	0	0	0
Ighd5-7	158	65	42	71	165	110	65	62
Ighd5-4	0	0	0	0	0	0	0	0
Ighd6-4	0	0	0	0	0	0	0	0
Ighd5-8	0	0	0	0	0	0	0	0
Ighd2-7	649	322	254	326	1118	789	514	279
Ighd5-5	36	26	12	17	49	49	29	16
Ighd2-8	1355	664	515	449	1678	1195	880	319
Ighd5-6	0	0	0	0	0	0	0	0
Ighd3-2	264	146	77	126	1100	419	446	353
Ighd4-1/DQ52	1013	98	414	1270	1528	243	697	1383
Total	22854	13340	7662	6041	20090	16803	10014	5127

R1.PG mice also exhibited perturbations in Igh gene segment repertoire. DQ52 usage was increased and DFL16.1 usage was decreased relative to WT in total VDJ, VDJ-, and DJ rearrangements in bone marrow (Figure. 3.11 (B-D)). A similar trend was observed in spleen rearrangements, although the magnitudes of the changes were smaller and were

not statistically significant. In contrast, DH segment usage in R1.KR mice closely resembled that of WT in DJH and VDJH joints in bone marrow and spleen (Figure. 3.11 (B-D); and Figure. 3.11 (E-G)). We conclude that loss of RAG1 ubiquitin ligase activity leads to increased representation of DQ52 and decreased representation of DFL16.1 in the IgH repertoire, most strikingly in pre-B cells, and that this phenotype is not due to loss of RAG1 autoubiquitination at K233.

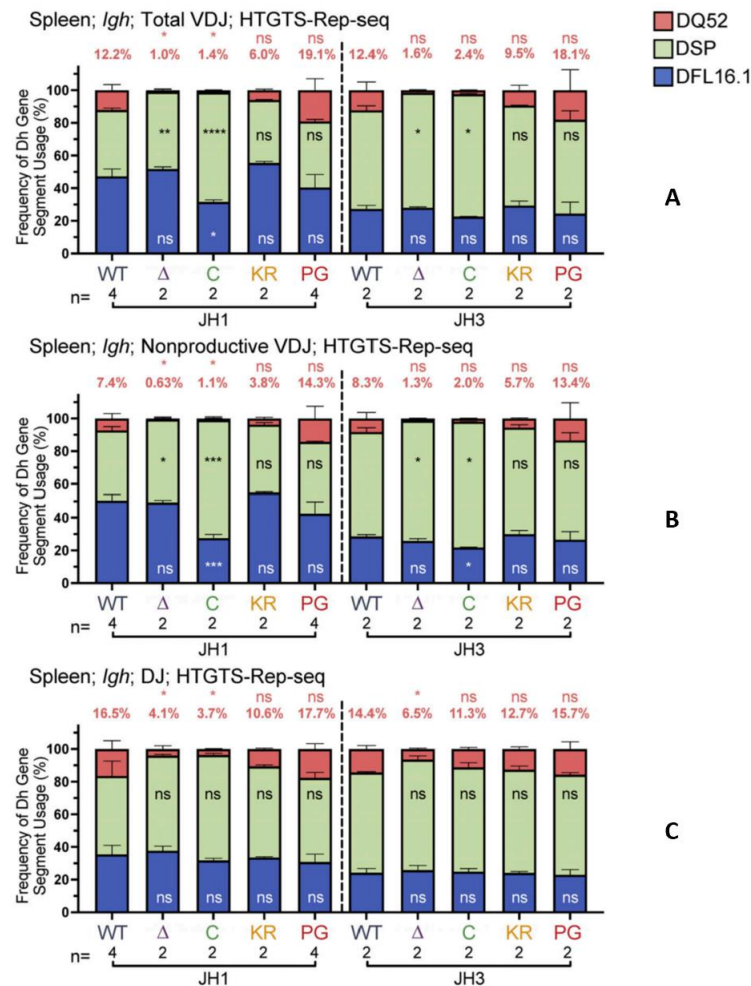


Figure 3.11 DH gene segment usage and analysis of productive and nonproductive rearrangements in VDJH rearrangements in spleen assessed by HTGTS-Rep-seq. (A-C) IgH HTGTS-Rep-seq data from mice 5–7 wk of age were analyzed for frequency of DFL16.1 (blue), DQ52 (salmon), and intervening 11 DH gene segments (green) in total (E) or nonproductive (F) VDJH or DJ recombination events (G) with JH1 or JH3 in spleen. Percentage of DQ52 usage is indicated above each bar. n, number of mice analyzed per genotype.

Table 3.10 Utilization of D_H gene segments among DJ rearrangements in primary B220+IgM- bone marrow cells (JH1 and JH3 primers were used as baits).

	JH1				JH3			
D genes	WT	Δ215	K233R	P326G	WT	Δ215	K233R	P326G
Ighd1-1/DFL16.1	587	519	361	344	605	612	455	325
Ighd6-1	10	12	0	6	19	38	11	18
Ighd2-3	201	177	117	135	253	290	200	151
Ighd6-2	0	0	1	0	6	12	2	3
Ighd2-4	241	227	144	184	352	423	301	266
Ighd5-2	0	0	0	0	0	0	0	0
Ighd2-5	39	52	15	30	75	102	42	32
Ighd5-3	0	0	0	0	0	0	0	0
Ighd6-3	0	0	0	0	0	0	0	0
Ighd5-7	0	0	0	0	1	4	0	2
Ighd5-4	7	0	0	0	5	2	0	0
Ighd6-4	0	0	0	0	0	0	0	0
Ighd5-8	0	0	0	0	1	8	6	4
Ighd2-7	97	73	42	102	161	134	112	118
Ighd5-5	0	1	0	4	2	1	2	0
Ighd2-8	192	136	123	165	250	229	206	167
Ighd5-6	5	0	0	4	8	4	5	6
Ighd3-2	34	9	13	28	254	99	167	188
Ighd4-1/DQ52	142	42	86	241	258	106	213	316
Total	1555	1248	902	1243	2250	2064	1722	1596

Table 3.11 Utilization of D_H gene segments among total VDJ rearrangements in primary B220+ Splenocytes (JH1 and JH3 primers were used as baits).

	JH1				JH3			
D genes	WT	Δ215	K233R	P326G	WT	Δ215	K233R	P326G
Ighd1-1/DFL16.1	56063	35289	52331	36902	6230	8028	6022	8809
Ighd6-1	0	0	0	0	0	0	0	0
Ighd2-3	9258	8872	9225	7262	2454	4570	2553	3829
Ighd6-2	235	57	238	76	38	57	37	52
Ighd2-4	11250	9152	9821	6824	4541	6838	4338	5690
Ighd5-2	144	51	168	60	23	41	21	49
Ighd2-5	6832	7742	6709	5127	2007	4337	1998	3005
Ighd5-3	0	0	0	0	0	0	0	0
Ighd6-3	0	0	0	0	0	0	0	0
Ighd5-7	713	351	569	331	160	202	190	162
Ighd5-4	0	0	0	0	0	0	0	0
Ighd6-4	0	0	0	0	0	0	0	0
Ighd5-8	0	0	3	3	1	0	0	1
Ighd2-7	3244	1822	2888	2183	1128	1489	1161	1355

Ighd5-5	218	166	174	80	21	83	57	86
Ighd2-8	7174	4613	6178	4955	1695	2145	1785	2529
Ighd5-6	0	0	0	0	0	0	0	0
Ighd3-2	1428	856	1223	666	1397	914	1313	1634
Ighd4-1/DQ52	7287	488	6112	6180	1912	415	2560	2775
Total	103846	69459	95639	70649	21607	29119	22035	29976

Table 3.12 Utilization of DH gene segments among DJ rearrangements in primary B220+ Splenocytes

	JH1				JH3			
D genes	WT	Δ 215	K233R	P326G	WT	Δ 215	K233R	P326G
Ighd1-1/DFL16.1	757	577	628	429	614	646	573	522
Ighd6-1	37	20	30	12	62	78	56	56
Ighd2-3	246	189	211	136	243	283	238	202
Ighd6-2	1	0	1	0	13	19	15	14
Ighd2-4	227	202	223	128	282	370	282	255
Ighd5-2	3	6		1	7	14	9	6
Ighd2-5	195	156	225	100	153	259	207	199
Ighd5-3	0	4	2	4	0	8	7	3
Ighd6-3	0	0	0	0	0	0	0	0
Ighd5-7	2	3	2	1	4	11	8	10
Ighd5-4	1	0	1	0	6	2	5	1
Ighd6-4	0	0	0	0	0	0	0	0
Ighd5-8	0	2	1	0	6	7	3	2
Ighd2-7	122	71	87	54	118	107	115	105
Ighd5-5	3	0	3	0	1	3	1	3
Ighd2-8	246	171	225	127	260	231	245	229
Ighd5-6	8	2	3	3	9	6	10	4
Ighd3-2	67	13	48	15	271	135	226	199
Ighd4-1/DQ52	216	38	194	149	303	141	318	261
Total	2131	1454	1884	1159	2352	2320	2318	2071

3.3.1 Reduced secondary D-to-JH recombination contributes to the R1.PG Igh repertoire phenotype

An attractive explanation for increased DQ52 and decreased DFL16.1 usage in R1.PG mice, is decreased recombinase activity leading to a decrease in secondary D-to-JH recombination, which should preserve DQ52-JH alleles at the expense of DFL16.1. If this explanation is correct, then R1.PG mice should exhibit increased usage of 5' and decreased usage of 3' JH gene segments relative to WT. To test this prediction, we assessed Igh

repertoires using ImmunoSEQ (Adaptive Biotechnologies), an approach that provides the repertoire of VH, DH, and JH segments in VDJH rearrangements. The analysis was performed on bone marrow B220+IgM⁻ cells from WT, R1Δ215, R1.PG, and R1core mice (two mice of each genotype). The ImmunoSEQ results (Figure 3.32) recapitulate the HTGTS-Rep-seq findings regarding DH segment usage in R1Δ215 and R1.PG mice: in R1Δ215, DQ52 usage was reduced ~10 fold compared with WT, with little change observed for DFL16.1, while in R1.PG, DQ52 usage was increased and DFL16.1 usage was decreased compared with WT. These differences were seen in total and nonproductive VDJH rearrangements (Figure 3.11A) and were recapitulated at each of the four JH segments (Figure 3.11B). JH repertoire analysis revealed increased JH1 and JH2 and decreased JH3 and JH4 usage in R1.PG mice relative to WT in total VDJ and VDJ⁻ rearrangements, as predicted by the model that R1.PG mice perform secondary D-to-JH recombination inefficiently (Figure 3.11C). The JH repertoire of R1Δ215 mice exhibited only minor differences from that of WT, with small increases in JH3 and JH4, decreased JH2, and slightly decreased JH1 usage, consistent with a normal or slightly elevated efficiency of secondary D-to-JH recombination (Figure 3.11B). Together, our data argue strongly that the altered DH segment repertoire in R1.PG mice is due in part or entirely to reduced secondary D-to-JH recombination and that altered secondary D-to-JH recombination makes at most a minor contribution to the R1Δ215 phenotype.

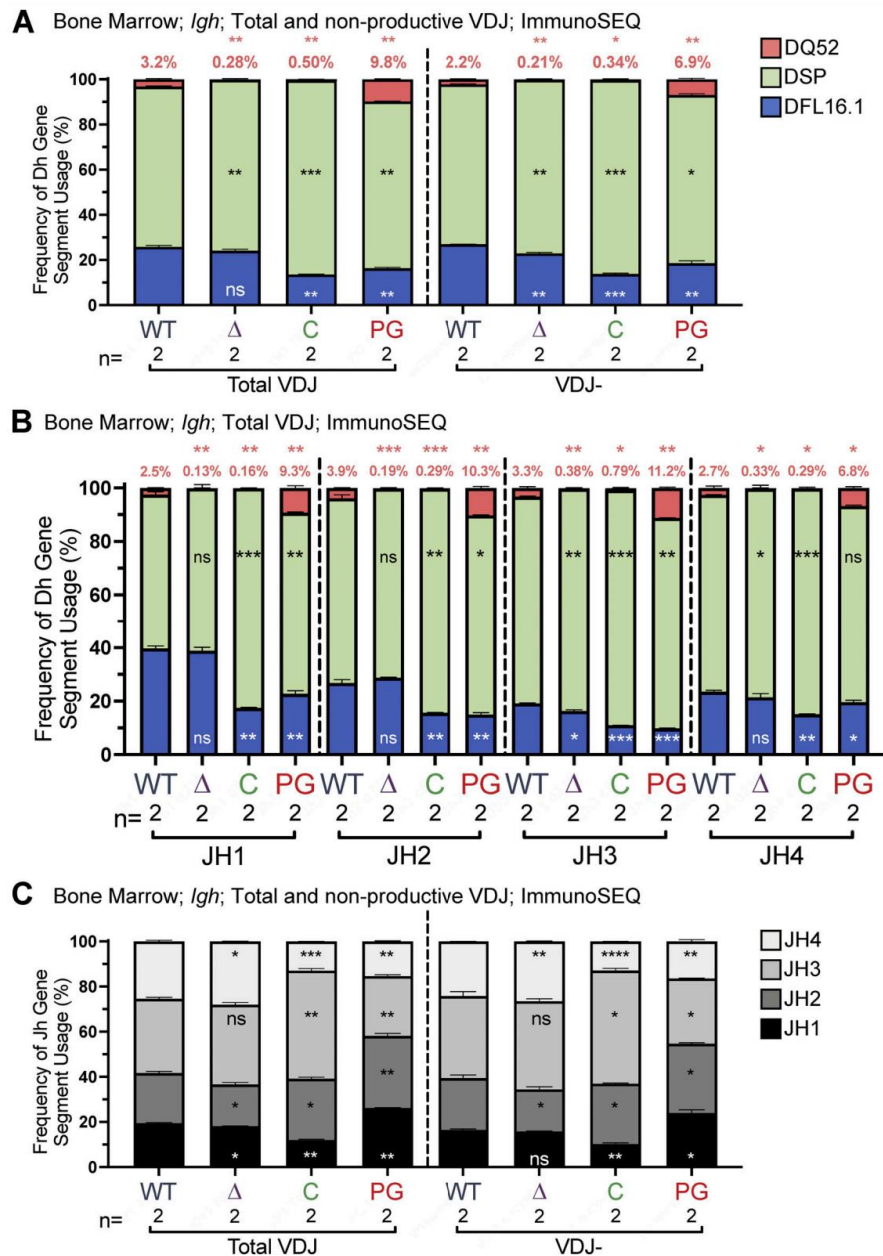


Figure 3.12 D_H and J_H gene segment usage in VDJ_H rearrangements in bone marrow assessed by ImmunoSEQ. (A-C) ImmunoSEQ data from mice 5–7 wk of age were analyzed for frequency of DFL16.1 (blue), DQ52 (salmon), and intervening 11 DH gene segments (green) in total or nonproductive VDJ_H rearrangements (A) or VDJ_H rearrangements to individual JH gene segments (B) and for JH gene segments in total and nonproductive VDJ_H recombination events (C).

Percentage of DQ52 usage is indicated above each bar in A and B. n, number of mice analyzed per genotype. Data are presented as mean with error bars indicating SEM. Statistical significance relative to WT determined by two-tailed unpaired t tests (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$).

4. Discussion

In this study, by integrating genome-wide chromatin interactions, 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. In addition, we attempted to determine the role of RAG1 during long-range vs short-range IgH recombination events using various Rag1-transgenic lines and High-throughput genome-wide translocation sequencing-adapted repertoire sequencing (HTGTS-Rep). Previous reports revealed that chromatin is non-randomly organized into transcriptionally permissive and repressive compartments (Lieberman-Aiden et al. 2009; Gibcus and Dekker 2013). The permissive compartments were found to be transcriptionally active as compared to repressive compartments. The Detection of chromatin compartments through Hi-C has raised several fundamental questions regarding the induction of lineage specific genes emerge. What is the nature of genes prior to activation? How is activation of these genes regulated? What are the roles of transcription factors in the induction of these genes? These questions were addressed by analyzing the chromatin interaction landscape obtained from Hi-C during the early stages of B -cell development. First, we attempted to determine the role of key transcriptional regulators, (Transcription factors and chromatin modifiers) during genetic switch between A and B compartments during B cell developmental transitions. Furthermore, efforts were undertaken to comprehend the regulatory interplay between Ebf1 and the chromatin in the establishment of the lineage specific cis-regulome. For this, we used HiC data generated from our own lab (pre-pro-B cells – Ebf1-/- progenitors, and pro-B cells – Rag2-/-) or publicly available (pro-B – Pax5-/-, and pre-B – Pax5-ER) datasets to see how lineage determinants trigger the genetic switch. The chromatin interactome data at each cell stage were integrated with their respective epigenetic and transcriptome profiles to ascertain changes in chromatin organization and their functional state during the developmental transition from pre-pro-B to pro-B stage as well as from pro-B to Pre-B cell stage. In addition, at each cell stage, LORDG method was used to perform whole-genome (20kb resolution) and locus specific 3D chromatin modeling (5kb resolution). This has helped to rigorously map the genetic switch on a genome-wide scale as well on a specific locus. We observed B lineage genes like Satb2, Tead1, Pou2af1 and Tlr4 switch from the repressive compartment to the permissive compartment during the (Ebf1-/-) pre-

pro-B to (Rag2^{-/-}) pro-B transition. Likewise, genes such as Gata3, Klf4, Satb1 and Zbtb16 that are important for disparate lineage differentiation programs localize to repressive compartment in pro-B cells. In addition, we note that at Pax5^{-/-} pro-B cell stage we observed that alternate lineage genes like Sox6, Zbtb16, Etv1, Gzmk were found to be in permissive compartments. However, upon ectopic expression of Pax5 they have moved to repressive compartments. While the genome-wide 3D modeling at 20kb resolution, confirmed the genetic switch of these genes, locus-specific modeling as revealed increase in the per-base interaction frequencies in the permissive compartment as compared to the repressive compartment. Moreover, we found that majority of the genes that encode master regulators of B lineage choice, are localized in permissive compartments throughout early B cell developmental stages. These findings are in line with earlier studies that Ebf1^{-/-} progenitors express lineage inappropriate genes and differentiate into various hematopoietic lineages in vitro and in vivo, despite maintaining them under B lymphoid conditions (Pongubala et al. 2008). Thus, chromatin localization appears to be central to B cell development. Further, to understand the molecular basis for genetic switch, we mapped the differential chromatin interaction profile (DCI) calculated at each bin (40kb) by comparing the Hi-C contact maps between two developmental stages with union DNaseI hypersensitive sites and performed the TR (transcriptional regulator) enrichment analysis. These analyses revealed the top TRs enriched at regions with increased/decreased interactions (Wang, Zhang, and Zang 2021) during B cell developmental transitions providing the molecular evidence for such genetic switch.

Chromatin looping is a general mechanism for establishing long-range interactions between cis-regulatory elements, such as promoters and enhancers (Kleinjan and van Heyningen 2005; Bulger and Groudine 2011). However, several questions remain to be answered, including: are these cis-interactions cell-type specific and functionally important for establishment of B-lineage gene expression programs? What factors regulate these interactions during B-lineage differentiation? In this study, we have comprehensively analyzed the Ebf1 mediated long-range interactions between in pre-pro-B and pro-B cells and found a significant increase in promoter-promoter and promoter-enhancer interactions during the transition from pre-pro-B to pro-B stage. Notably, we found that the enhancer landscape of a given promoter undergoes extensive rewiring

during B-lineage commitment. Additionally, we showed that transcript levels of a given promoter are proportionate to the number of its cis-regulatory interacting elements. These findings are consistent with previous studies that the enhancer repertoire undergoes considerable changes during development in a spatiotemporal manner (Chepelev et al. 2012; Li et al. 2012; Thurman et al. 2012). More importantly, our studies highlight that promoter-cis-regulatory interactions are cell-type specific and loss- or gain- of these interactions is closely associated with repression of alternate lineage genes and induction of B-lineage specific genes, respectively. For instance, the promoter of B-lineage gene, Cd24a, is involved in interactions with 8 cis-regulatory elements (as determined by H3k27ac ChIP-Seq) in pro-B cells (Rag2^{-/-}) of which 7 were captured by Ebf1-ER H3k27ac ChIP-Seq compared to only two such elements in pre-pro-B cells. This explains the underlying reason for Cd24a higher expression at the pro-B stage and the role of Ebf1 in inducing its expression. Furthermore, our studies propose that combinatorial binding of cell type specific transcription factors to the cis-regulatory elements constitute a general paradigm for induction of lineage specific gene expression programs. Next, we focused to determine the role of RAG1 during long-range vs short-range IgH recombination events using various Rag1-transgenic lines and High-throughput genome-wide translocation sequencing-adapted repertoire sequencing (HTGTS-Rep). Our results suggested that deletion of the first 215 aa of RAG1 results in strongly decreased usage of DQ52 in IgH rearrangements. The reduced usage of DQ52 is not due to increased secondary D-to-JH recombination. In particular, R1Δ215 mice exhibit levels of DFL16.1 usage that are, if anything, lower than in WT and levels of JH1 usage that are only slightly reduced relative to WT (Figure. 3.11 and Figure. 3.32), neither of which is consistent with a large increase in secondary D-to-JH recombination. The possibility that reduced DQ52 usage in R1Δ215 mice is due to a small increase in secondary D-to-JH recombination that selectively targets DQ52-JH alleles is ruled out by the finding that DQ52 usage is strongly reduced in rearrangements involving all four JH segments, including JH4 (Fig. 5 B); D-JH4 alleles cannot be targets of secondary recombination. We conclude that decreased usage of DQ52 in R1Δ215 mice is due primarily to mechanism(s) distinct from altered secondary D-to-JH recombination. The high levels of expression of the R1Δ215 protein might have been predicted to lead to a hyper-recombination phenotype, but we see little evidence for this. While we cannot rule out the possibility that deletion of the first 215 aa perturbs folding

of other portions of the RAG1 protein, aa 215 corresponds closely to an NTR domain boundary (Arbuckle et al., 2011), and robust RAG1 protein accumulation and normal lymphocyte development in R1 Δ 215 mice argue against major structural perturbations.

5. Summary and Conclusions

The hallmark of B cell development is characterized by induction of early B lineage genes ($\lambda 5$, Cd19, VpreB, Ig α and Ig β) and ordered rearrangement of immunoglobulin (heavy and light chain) gene segments. Sequential rearrangement of Ig genes involves nuclear re-localization, long-range compaction and contraction. In addition, multiple lines of evidence indicate that the long-range intra- and inter-chromosomal genomic interactions between regulatory elements, such as promoters and enhancers as well as insulators, is often tightly coupled with modulation of specific transcription programs. Thus, dynamic changes of higher-order DNA loops, which help bring widely separated gene segments into proximity, play an important role in tissue specific gene expression pattern. Genetic studies demonstrate that Ebf1 and Pax5 are primary B cell determinants and they play an important role during B cell fate specification and commitment. However, the precise role of Ebf1 and Pax5 in establishment of genome wide chromatin interactions during B cell development remain unclear. The current study addresses these fundamental questions and aims to integrate the regulatory role of Ebf1 in establishing long-range chromatin interactions and investigates the influence of those interactions on B-lineage gene expression program. First, by combining Hi-C with other high-throughput molecular approaches like ChIP-Seq and RNA-Seq we determined Ebf1 and Pax5 dependent changes in spatial organization of chromatin. Then, the 3D structures of regions, that exhibited a developmental switch between transcriptionally permissive and repressive compartments, across the cell stages, were generated at 5kb resolution using the spring 3D modeling approach (Kadlof, Rozycka, and Plewczynski 2020) In addition to genome-wide 3D modeling at 20kb resolution. The spring method that was used to model locus-specific 3D chromatin structures operates on context-dependent chromatin interactions to define constraints between beads that are initially stretched. The ensemble energy is then minimized, resulting in shortening of the spring and the fiber folds into a structure that satisfies the chosen contacts. The 3D structure analysis has confirmed the genetic switch and provided more insights leading to the conclusion that the number of restraints (loops) in a particular locus play important role in genetic switch. Further to understand the molecular basis for genetic switch, we mapped the differential chromatin interaction profile (DCI) calculated at each bin (40kb) by comparing the Hi-C contact maps between

two developmental stages and performed the TR (transcriptional regulator) enrichment analysis. The analysis involved data obtained from over 5000 ChIP-Seq data sets representing various transcription factor binding profiles in addition to comprehensive compendium of all the active regulatory elements discovered till date. It revealed the top TRs enriched at regions with increased/decreased interactions (Wang, Zhang, and Zang 2021) during B cell developmental transitions. Among the highly enriched transcriptional regulators, during the developmental transition from pre-pro-B to pro-B we found transcription factors like Ebf1, Ezh1 in addition to the chromatin modifiers like CTCF and Rad21. Next, by performing global transcriptome analysis we found that Ebf1 alone is able to induce lineage-specific genes and suppress alternate lineage genes independently, reiterating its importance during B cell fate commitment. The De novo transcription factor binding site analysis has revealed that both Ebf1 binds either alone or in combination with Pax5 to 57.2% of the promoters and 51.2% of cis-regulatory elements that are involved in long-range interactions captured by Hi-C. In addition, we identified the novel Ebf1 dependent active enhancer elements and their interacting promoter partners by overlaying the significant loop interactions obtained from Rag2^{-/-} Hi-C data with Ebf1-ER H3K27ac peak information. These studies not only unraveled the key Ebf1 dependent regulatory elements but also aided in assembling physical gene regulatory networks. The newly identified components within the regulatory networks will further be functionally examined for their contributions to B cell development. Ebf1 dependent regulatory elements that are important for inducing B cell gene expression program will set stage for future experiments of characterization and understanding of their function. These studies delineate for the first time: the architectural role of Ebf1, in regulating the order of molecular events that dictate the B cell fate choice of MPPs.

Lastly, to understand the gene usage patterns during Igh recombination, we performed the HTGTS analysis of three new mouse strains in which the RAG1 NTR has been mutated. The mutations eliminate RAG1 ubiquitin ligase activity (P326G, or PG), eliminate the major autoubiquitination site (K233R, or KR), or remove the first 215 amino acids of the NTR (D215). Critical analysis of antigen repertoire obtained for these mutants indicated that the first 215 amino acids of the NTR (D215) are important for "short-range" recombination events (defined as recombination between two gene segments within a

recombination center) as the d215 mice showed strong reduction in short-range recombination events. Collectively, this study delineates the mechanistic details involved in regulatory role of lineage determinants like Ebf1, Pax5 and RAG genes during B cell development.

Conclusions

- Differential chromatin re-localization predominantly occurs during pre-proB to pro-B cell transition
- Crosstalk between chromatin modifiers and transcription factors (transcriptional regulators) orchestrate the genetic switch between A and B compartments.
- Reinforcement of lineage specific gene expression is contingent upon specificity and frequency of interactions between promoters and their *cis*-regulatory elements
- Binding of Ebf1/Pax5 at *cis*-regulatory elements is positively correlated with expression of their target genes that are involved in chromatin interactions.
- The first 215 aa of RAG1 are required to maintain the proper balance between short-range, collisional recombination and long-range recombination mediated by RAG chromatin scanning.
- Our data indicate that the RAG1 P326G mutation decreases V(D)J recombinase activity during B cell development.

6. References

- Adolfsson, J., R. Mansson, N. Buza-Vidas, A. Hultquist, K. Liuba, C. T. Jensen, D. Bryder, L. Yang, O. J. Borge, L. A. Thoren, K. Anderson, E. Sitnicka, Y. Sasaki, M. Sigvardsson, and S. E. Jacobsen. 2005. 'Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment', *Cell*, 121: 295-306.
- Ba, Z., J. Lou, A. Y. Ye, H. Q. Dai, E. W. Dring, S. G. Lin, S. Jain, N. Kyritsis, K. R. Kieffer-Kwon, R. Casellas, and F. W. Alt. 2020. 'CTCF orchestrates long-range cohesin-driven V(D)J recombinational scanning', *Nature*, 586: 305-10.
- Bain, G., E. C. Maandag, D. J. Izon, D. Amsen, A. M. Kruisbeek, B. C. Weintraub, I. Krop, M. S. Schlissel, A. J. Feeney, M. van Roon, and et al. 1994. 'E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements', *Cell*, 79: 885-92.
- Bickmore, W. A., and B. van Steensel. 2013. 'Genome architecture: domain organization of interphase chromosomes', *Cell*, 152: 1270-84.
- Dai, H. Q., H. Hu, J. Lou, A. Y. Ye, Z. Ba, X. Zhang, Y. Zhang, L. Zhao, H. S. Yoon, A. M. Chapdelaine-Williams, N. Kyritsis, H. Chen, K. Johnson, S. Lin, A. Conte, R. Casellas, C. S. Lee, and F. W. Alt. 2021. 'Loop extrusion mediates physiological Igh locus contraction for RAG scanning', *Nature*, 590: 338-43.
- Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. 'Capturing chromosome conformation', *Science*, 295: 1306-11.
- Dixon, J. R., I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget, A. Y. Lee, Z. Ye, A. Kim, N. Rajagopal, W. Xie, Y. Diao, J. Liang, H. Zhao, V. V. Lobanenko, J. R. Ecker, J. A. Thomson, and B. Ren. 2015. 'Chromatin architecture reorganization during stem cell differentiation', *Nature*, 518: 331-6.
- Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. 2012. 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature*, 485: 376-80.
- Dobrev, G., J. Dambacher, and R. Grosschedl. 2003. 'SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin mu gene expression', *Genes Dev*, 17: 3048-61.
- Dostie, J., and J. Dekker. 2007. 'Mapping networks of physical interactions between genomic elements using 5C technology', *Nat Protoc*, 2: 988-1002.
- Dostie, J., T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, T. A. Honan, E. D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R. D. Green, and J. Dekker. 2006. 'Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements', *Genome Res*, 16: 1299-309.
- Dudley, D. D., J. Sekiguchi, C. Zhu, M. J. Sadofsky, S. Whitlow, J. DeVido, R. J. Monroe, C. H. Bassing, and F. W. Alt. 2003. 'Impaired V(D)J recombination and lymphocyte development in core RAG1-expressing mice', *J Exp Med*, 198: 1439-50.
- Fortin, J. P., and K. D. Hansen. 2015. 'Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data', *Genome Biol*, 16: 180.
- Gao, H., K. Lukin, J. Ramirez, S. Fields, D. Lopez, and J. Hagman. 2009. 'Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by Ebf1 and Pax5', *Proc Natl Acad Sci U S A*, 106: 11258-63.

- Georgopoulos, K., M. Bigby, J. H. Wang, A. Molnar, P. Wu, S. Winandy, and A. Sharpe. 1994. 'The Ikaros gene is required for the development of all lymphoid lineages', *Cell*, 79: 143-56.
- Gibcus, J. H., and J. Dekker. 2013. 'The hierarchy of the 3D genome', *Mol Cell*, 49: 773-82.
- Gisler, R., P. Akerblad, and M. Sigvardsson. 1999. 'A human early B-cell factor-like protein participates in the regulation of the human CD19 promoter', *Mol Immunol*, 36: 1067-77.
- Gorkin, D. U., D. Leung, and B. Ren. 2014. 'The 3D genome in transcriptional regulation and pluripotency', *Cell Stem Cell*, 14: 762-75.
- Guo, C., H. S. Yoon, A. Franklin, S. Jain, A. Ebert, H. L. Cheng, E. Hansen, O. Despo, C. Bossen, C. Vettermann, J. G. Bates, N. Richards, D. Myers, H. Patel, M. Gallagher, M. S. Schlissel, C. Murre, M. Busslinger, C. C. Giallourakis, and F. W. Alt. 2011. 'CTCF-binding elements mediate control of V(D)J recombination', *Nature*, 477: 424-30.
- Han, J., Z. Zhang, and K. Wang. 2018. '3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering', *Mol Cytogenet*, 11: 21.
- Hauke, Ralph J., and Stefano R. Tarantolo. 2000. 'Hematopoietic Growth Factors', *Laboratory Medicine*, 31: 613-15.
- Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38: 576-89.
- Hill, L., A. Ebert, M. Jaritz, G. Wutz, K. Nagasaka, H. Tagoh, D. Kostanova-Poliakova, K. Schindler, Q. Sun, P. Bonelt, M. Fischer, J. M. Peters, and M. Busslinger. 2020. 'Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion', *Nature*, 584: 142-47.
- Hu, J., Y. Zhang, L. Zhao, R. L. Frock, Z. Du, R. M. Meyers, F. L. Meng, D. G. Schatz, and F. W. Alt. 2015. 'Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes', *Cell*, 163: 947-59.
- Ikawa, T., H. Kawamoto, L. Y. Wright, and C. Murre. 2004. 'Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent', *Immunity*, 20: 349-60.
- Imakaev, M., G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker, and L. A. Mirny. 2012. 'Iterative correction of Hi-C data reveals hallmarks of chromosome organization', *Nat Methods*, 9: 999-1003.
- Jain, S., Z. Ba, Y. Zhang, H. Q. Dai, and F. W. Alt. 2018. 'CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning', *Cell*, 174: 102-16 e14.
- Ji, Y., W. Resch, E. Corbett, A. Yamane, R. Casellas, and D. G. Schatz. 2010. 'The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci', *Cell*, 141: 419-31.
- Johanson, T. M., A. T. L. Lun, H. D. Coughlan, T. Tan, G. K. Smyth, S. L. Nutt, and R. S. Allan. 2018. 'Transcription-factor-mediated supervision of global genome architecture maintains B cell identity', *Nat Immunol*, 19: 1257-64.
- Kadluf, M., J. Rozycka, and D. Plewczynski. 2020. 'Spring Model - Chromatin Modeling Tool Based on OpenMM', *Methods*, 181-182: 62-69.
- Kieffer-Kwon, K. R., Z. Tang, E. Mathe, J. Qian, M. H. Sung, G. Li, W. Resch, S. Baek, N. Pruett, L. Grontved, L. Vian, S. Nelson, H. Zare, O. Hakim, D. Reyon, A. Yamane, H. Nakahashi, A. L. Kovalchuk, J. Zou, J. K. Joung, V. Sartorelli, C. L. Wei, X. Ruan, G. L. Hager, Y. Ruan, and R. Casellas. 2013. 'Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation', *Cell*, 155: 1507-20.

- Kikuchi, K., A. Y. Lai, C. L. Hsu, and M. Kondo. 2005. 'IL-7 receptor signaling is necessary for stage transition in adult B cell development through up-regulation of Ebf1', *J Exp Med*, 201: 1197-203.
- Kimmel, M. 2014. 'Stochasticity and determinism in models of hematopoiesis', *Adv Exp Med Biol*, 844: 119-52.
- Kwon, K., C. Hutter, Q. Sun, I. Bilic, C. Cobaleda, S. Malin, and M. Busslinger. 2008. 'Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development', *Immunity*, 28: 751-62.
- Kim, K. D., H. Tanizawa, O. Iwasaki, and K. Noma. 2016. 'Transcription factors mediate condensin recruitment and global chromosomal organization in fission yeast', *Nat Genet*, 48: 1242-52.
- Stadhouders, R., E. Vidal, F. Serra, B. Di Stefano, F. Le Dily, J. Quilez, A. Gomez, S. Collombet, C. Berenguer, Y. Cuartero, J. Hecht, G. J. Fillion, M. Beato, M. A. Marti-Renom, and T. Graf. 2018. 'Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming', *Nat Genet*, 50: 238-49.
- van Steensel, B., and E. E. M. Furlong. 2019. 'The role of transcription in shaping the spatial organization of the genome', *Nat Rev Mol Cell Biol*, 20: 327-37.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, 1-10.
- Laurenti, E., S. Doulatov, S. Zandi, I. Plumb, J. Chen, C. April, J. B. Fan, and J. E. Dick. 2013. 'The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment', *Nat Immunol*, 14: 756-63.
- Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science*, 326: 289-93.
- Lin, H., and R. Grosschedl. 1995. 'Failure of B-cell differentiation in mice lacking the transcription factor Ebf1', *Nature*, 376: 263-7.
- Lin, S. G., C. Guo, A. Su, Y. Zhang, and F. W. Alt. 2015. 'CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination', *Proc Natl Acad Sci U S A*, 112: 1815-20.
- Lin, Y. C., C. Benner, R. Mansson, S. Heinz, K. Miyazaki, M. Miyazaki, V. Chandra, C. Bossen, C. K. Glass, and C. Murre. 2012. 'Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate', *Nat Immunol*, 13: 1196-204.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*, 17: 10–12.
- Medina, K. L., J. M. Pongubala, K. L. Reddy, D. W. Lancki, R. Dekoter, M. Kieslinger, R. Grosschedl, and H. Singh. 2004. 'Assembling a gene regulatory network for specification of the B cell fate', *Dev Cell*, 7: 607-17.
- Misteli, T. 2007. 'Beyond the sequence: cellular organization of genome function', *Cell*, 128: 787-800.

- Nagano, T., Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, and P. Fraser. 2013. 'Single-cell Hi-C reveals cell-to-cell variability in chromosome structure', *Nature*, 502: 59-64.
- Naumova, N., M. Imakaev, G. Fudenberg, Y. Zhan, B. R. Lajoie, L. A. Mirny, and J. Dekker. 2013. 'Organization of the mitotic chromosome', *Science*, 342: 948-53.
- Nutt, S. L., B. Heavey, A. G. Rolink, and M. Busslinger. 1999. 'Commitment to the B-lymphoid lineage depends on the transcription factor Pax5', *Nature*, 401: 556-62.
- O'Riordan, M., and R. Grosschedl. 1999. 'Coordinate regulation of B cell differentiation by the transcription factors Ebf1 and E2A', *Immunity*, 11: 21-31.
- Peschon, J. J., P. J. Morrissey, K. H. Grabstein, F. J. Ramsdell, E. Maraskovsky, B. C. Gliniak, L. S. Park, S. F. Ziegler, D. E. Williams, C. B. Ware, J. D. Meyer, and B. L. Davison. 1994. 'Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice', *J Exp Med*, 180: 1955-60.
- Pongubala, J. M., D. L. Northrup, D. W. Lancki, K. L. Medina, T. Treiber, E. Bertolino, M. Thomas, R. Grosschedl, D. Allman, and H. Singh. 2008. 'Transcription factor Ebf1 restricts alternative lineage options and promotes B cell fate commitment independently of Pax5', *Nat Immunol*, 9: 203-15.
- Pongubala, J. M. R., and C. Murre. 2021. 'Spatial Organization of Chromatin: Transcriptional Control of Adaptive Immune Cell Development', *Front Immunol*, 12: 633825.
- Reynaud, D., I. A. Demarco, K. L. Reddy, H. Schjerven, E. Bertolino, Z. Chen, S. T. Smale, S. Winandy, and H. Singh. 2008. 'Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros', *Nat Immunol*, 9: 927-36.
- Roessler, S., I. Gyory, S. Imhof, M. Spivakov, R. R. Williams, M. Busslinger, A. G. Fisher, and R. Grosschedl. 2007. 'Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5', *Mol Cell Biol*, 27: 579-94.
- Schatz, D. G., and Y. Ji. 2011. 'Recombination centres and the orchestration of V(D)J recombination', *Nat Rev Immunol*, 11: 251-63.
- Schebesta, M., B. Heavey, and M. Busslinger. 2002. 'Transcriptional control of B-cell development', *Curr Opin Immunol*, 14: 216-23.
- Sigvardsson, M., D. R. Clark, D. Fitzsimmons, M. Doyle, P. Akerblad, T. Breslin, S. Bilke, R. Li, C. Yeaman, G. Zhang, and J. Hagman. 2002. 'Early B-cell factor, E2A, and Pax-5 cooperate to activate the early B cell-specific mb-1 promoter', *Mol Cell Biol*, 22: 8539-51.
- Sigvardsson, M., M. O'Riordan, and R. Grosschedl. 1997. 'Ebf1 and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes', *Immunity*, 7: 25-36.
- Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, and W. de Laat. 2006. 'Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)', *Nat Genet*, 38: 1348-54.
- Singh, H., K. L. Medina, and J. M. Pongubala. 2005. 'Contingent gene regulatory networks and B cell fate specification', *Proc Natl Acad Sci U S A*, 102: 4949-53.
- Sivakumar, A., J. I. de Las Heras, and E. C. Schirmer. 2019. 'Spatial Genome Organization: From Development to Disease', *Front Cell Dev Biol*, 7: 18.
- Szydlowski, N. A., J. S. Go, and Y. S. Hu. 2019. 'Chromatin imaging and new technologies for imaging the nucleome', *Wiley Interdiscip Rev Syst Biol Med*, 11: e1442.

- Tamaru, H. 2010. 'Confining euchromatin/heterochromatin territory: jumonji crosses the line', *Genes Dev*, 24: 1465-78.
- Teng, G., Y. Maman, W. Resch, M. Kim, A. Yamane, J. Qian, K. R. Kieffer-Kwon, M. Mandal, Y. Ji, E. Meffre, M. R. Clark, L. G. Cowell, R. Casellas, and D. G. Schatz. 2015. 'RAG Represents a Widespread Threat to the Lymphocyte Genome', *Cell*, 162: 751-65.
- Thal, M. A., T. L. Carvalho, T. He, H. G. Kim, H. Gao, J. Hagman, and C. A. Klug. 2009. 'Ebf1-mediated down-regulation of Id2 and Id3 is essential for specification of the B cell lineage', *Proc Natl Acad Sci U S A*, 106: 552-7.
- Zhang, Y., X. Zhang, Z. Ba, Z. Liang, E. W. Dring, H. Hu, J. Lou, N. Kyritsis, J. Zurita, M. S. Shamim, A. Presser Aiden, E. Lieberman Aiden, and F. W. Alt. 2019. 'The fundamental role of chromatin loop extrusion in physiological V(D)J recombination', *Nature*, 573: 600-04.

Regulatory interplay between chromatin organization and lineage determinants during B cell specification

ORIGINALITY REPORT

23%

SIMILARITY INDEX

21%

INTERNET SOURCES

15%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

www.ncbi.nlm.nih.gov

Internet Source

2

academic.oup.com

Internet Source

3

Submitted to University of Hyderabad,
Hyderabad

Student Paper

4

Elizabeth M Mandel, Rudolf Grosschedl,

"Transcription control of early B cell
differentiation", Current Opinion in
Immunology, 2010

Publication

5

Kara Lukin, Scott Fields, Jacqueline Hartley,
James Hagman. "Early B cell factor: Regulator
of B lineage specification and commitment",
Seminars in Immunology, 2008

Publication

6

rupress.org

Internet Source

Anurupa is a
co-author in these
Publications
I HRY
20/06/2022

15%

2%

1%

<1%

<1%

<1%

Dr. Jagan Pongubala
Professor
Department of Animal Biology
School of Life Sciences
University of Hyderabad
H.O. Central University, Gachibowli,
Hyderabad-500 046. T.S. INDIA.

7	Julita Ramírez, Kara Lukin, James Hagman. "From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment", Current Opinion in Immunology, 2010 Publication	<1 %
8	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
9	www.cell.com Internet Source	<1 %
10	Laslo, P.. "Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system", Seminars in Immunology, 200808 Publication	<1 %
11	Valentina Agoni. "Electrical induction hypothesis to explain enhancer-promoter communication", PeerJ, 2015 Publication	<1 %
12	coek.info Internet Source	<1 %
13	www.pnas.org Internet Source	<1 %
14	Mandel, E.M.. "Transcription control of early B cell differentiation", Current Opinion in Immunology, 201004 Publication	<1 %

15	biologicalproceduresonline.biomedcentral.com Internet Source	<1 %
16	www.biostars.org Internet Source	<1 %
17	aacrjournals.org Internet Source	<1 %
18	www.hindawi.com Internet Source	<1 %
19	Kay L Medina, Harinder Singh. "Genetic networks that regulate B lymphopoiesis", Current Opinion in Hematology, 2005 Publication	<1 %
20	www.tandfonline.com Internet Source	<1 %
21	Peter Laslo, Jagan M.R. Pongubala, David W. Lancki, Harinder Singh. "Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system", Seminars in Immunology, 2008 Publication	<1 %
22	escholarship.umassmed.edu Internet Source	<1 %
23	www.nature.com Internet Source	<1 %
24	hdl.handle.net	

<1 %

25

s3.amazonaws.com

Internet Source

<1 %

26

Molecular Biology of B-Cell and T-Cell Development, 1998.

Publication

<1 %

27

pure.rug.nl

Internet Source

<1 %

28

Michal Kadlof, Julia Rozycka, Dariusz Plewczynski. "Spring Model – Chromatin Modeling Tool Based on OpenMM", Methods, 2019

Publication

<1 %

29

Belton, Jon-Matthew, Rachel Patton McCord, Johan Harmen Gibcus, Natalia Naumova, Ye Zhan, and Job Dekker. "Hi-C: A comprehensive technique to capture the conformation of genomes", Methods, 2012.

Publication

<1 %

30

www.jimmunol.org

Internet Source

<1 %

31

pubman.mpdl.mpg.de

Internet Source

<1 %

32

www.frontiersin.org

Internet Source

<1 %

33

"Epigenetic Regulation of Lymphocyte Development", Springer Science and Business Media LLC, 2012

Publication

<1 %

34

"Mechanisms of Lymphocyte Activation and Immune Regulation XI", Springer Science and Business Media LLC, 2007

Publication

<1 %

Exclude quotes On

Exclude matches < 14 words

Exclude bibliography On

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

Ravi Boya^{1,†}, Anurupa Devi Yadavalli^{1,†}, Sameena Nikhat¹, Sreenivasulu Kurukuti¹,
Dasaradhi Palakodeti² and Jagan M. R. Pongubala^{1,*}

¹Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad 500046, India and

²Institute 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.

ARTICLE

The RAG1 N-terminal region regulates the efficiency and pathways of synapsis for V(D)J recombination

Helen A. Beilinson¹, Rebecca A. Glynn^{2,3}, Anurupa Devi Yadavalli^{1,4}, Jianxiong Xiao¹, Elizabeth Corbett¹, Huseyin Saribasak¹, Rahul Arya³, Charline Miot³, Anamika Bhattacharyya⁵, Jessica M. Jones⁵, Jagan M.R. Pongubala⁴, Craig H. Bassing^{2,3}, and David G. Schatz^{1,6}

Immunoglobulin and T cell receptor gene assembly depends on V(D)J recombination initiated by the RAG1-RAG2 recombinase. The RAG1 N-terminal region (NTR; aa 1–383) has been implicated in regulatory functions whose influence on V(D)J recombination and lymphocyte development in vivo is poorly understood. We generated mice in which RAG1 lacks ubiquitin ligase activity (P326G), the major site of autoubiquitination (K233R), or its first 215 residues (Δ 215). While few abnormalities were detected in R1.K233R mice, R1.P326G mice exhibit multiple features indicative of reduced recombination efficiency, including an increased Igk⁺:Ig λ ⁺ B cell ratio and decreased recombination of *Igh*, *Igk*, *Ig λ* , and *Tcrb* loci. Previous studies indicate that synapsis of recombining partners during *Igh* recombination occurs through two pathways: long-range scanning and short-range collision. We find that R1 Δ 215 mice exhibit reduced short-range *Igh* and *Tcrb* D-to-J recombination. Our findings indicate that the RAG1 NTR regulates V(D)J recombination and lymphocyte development by multiple pathways, including control of the balance between short- and long-range recombination.

Introduction

The diversity of jawed vertebrate adaptive immune responses depends on programmed assembly and hypermutation of antigen receptor (AgR) genes (Cooper and Alder, 2006). The first AgR gene diversification process to occur in developing lymphocytes is V(D)J recombination, which assembles Ig and TCR genes from arrays of V, in some cases D, and J gene segments (Schatz and Swanson, 2011). V(D)J recombination is initiated by the endonuclease RAG, a heterotetramer made up of RAG1 and RAG2. RAG binds to and cleaves DNA at recombination signal sequences (RSSs) that flank rearranging gene segments. RAG-mediated DNA cleavage requires the synapsis of two RSSs and leads to simultaneous generation of DNA double-strand breaks (DSBs) at the two sites (Schatz and Swanson, 2011). The DNA ends are processed and ligated by nonhomologous end joining repair factors (Rooney et al., 2004).

RAG1 is the primary DNA-binding and cleaving component of RAG, while RAG2 is an essential accessory factor. The minimal portion of RAG1 required for RSS binding and cleavage is called the “core” domain (R1core), consisting of aa 384–1008 of the 1040-aa protein (numbering according to mouse RAG; Fig. 1 A; Sadofsky et al., 1993; Silver et al., 1993). RAG1 also contains two

noncore regions, one at the N-terminus (aa 1–383; hereafter RAG1 N-terminal region [NTR]) and a short C-terminal tail (aa 1009–1040) that modulates RAG cleavage activity in vitro (Grundy et al., 2010; Kim et al., 2015) and is not a focus of this study. Mice expressing R1core exhibit a partial block in early B and T cell development, reduced numbers of mature lymphocytes, an altered V β gene segment repertoire in *Tcrb* recombination events, and an increase in aberrant V(D)J recombination products, demonstrating the importance of RAG1 noncore regions (Dudley et al., 2003; Horowitz and Bassing, 2014; Talukder et al., 2004). The functional significance of the RAG1 NTR is highlighted by the wide array of atypical SCID-like phenotypes found in human patients with mutations in this region (Notarangelo et al., 2016). While multiple regulatory functions have been associated with the RAG1 NTR, largely through in vitro and cell line studies, little is known about its roles in the context of lymphocyte development and AgR repertoire formation in vivo (Jones and Simkus, 2009).

Distinct functions have been attributed to different portions of the RAG1 NTR. RAG1 aa 287–351 constitute a zinc-RING finger domain that binds four zinc atoms and possesses ubiquitin ligase

¹Department of Immunobiology, Yale School of Medicine, Yale University, New Haven, CT; ²Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ³Department of Pathology and Laboratory Medicine, Children’s Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ⁴Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad, India; ⁵Department of Biochemistry and Molecular & Cellular Biology, Georgetown University, Washington, DC; ⁶Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.

Correspondence to David G. Schatz: david.schatz@yale.edu.

© 2021 Beilinson et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

