Regulation of chromatin organizer SATB1 via TCR induced alternative promoter switch during T- cell development

A thesis

Submitted in partial fulfilment of the requirements

Of the degree of

Doctor of Philosophy

By

Indumathi Patta

IISER ID No. 20123156

Under the guidance of

Prof. Sanjeev Galande

At

Indian Institute of Science Education and Research, Pune

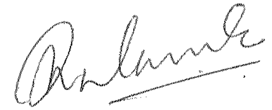
March, 2019

Dedicated To

“*My Parents*”

CERTIFICATE

Certified that the work incorporated in the thesis entitled "**Regulation of chromatin organizer SATB1 via TCR induced alternative promoter switch during T-cell development**" Submitted by **INDUMATHI PATTA** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

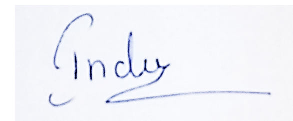


Prof Sanjeev Galande

Date March 27th, 2019 (Supervisor)

Declaration

I declare that this written submission represents my ideas in my own words and where others ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.



Date Indumathi Patta

27/3/2019 IISER ID: 20123156

Acknowledgements

Completion of my PhD research was possible with the support of several amazing people. I would like to express my sincere gratitude to all of them. First and foremost, I would like to express my special appreciation and thanks to my PhD advisor Professor Sanjeev Galande, who has been a tremendous mentor to me. I am grateful to him for enabling me to grow as a researcher. I thank him for supporting new ideas, and establishing timely collaborations for the completion of various aspects of my work. I deeply thank him for the support, constructive suggestions and patient guidance during the entire process of my Ph.D.

I would like to thank my thesis committee members, Dr. Girdhari Lal, Dr. Girish Ratnaparkhi, Dr. Vineeta Bal for their invaluable comments and critical suggestions during the annual meetings. I would like to thank our collaborators Dr. Girdhari Lal, Dr. Krishanpal Karmodiya, Dr. Vasudevan Seshadri, Dr. Jyoti Misra-Sen, Dr. Riitta Lahesmaa, and Dr. Manas Kumar Santra for providing their critical suggestions and support during the period of my doctorate study. I am extremely grateful, and would like to express my appreciation for Dr. Girdhari Lal for providing me resources to begin my research work at NCCS, Pune and for providing valuable guidance. I would like to extend my gratitude to Drs. Krishanpal Karmodiya and Satyajeet Khare for their invaluable help in discussion on data analysis. I am grateful to Dr. Vasudevan Seshadri for his guidance as well as allowing me to perform ribosome profiling experiments in his lab. I would like to express my sincere thanks to Dr. Deepa Subramanyam for help in providing the access to FACS during the course of my work. I would like to thank Dr. Manas Kumar Santra for the encouraging experience while performing the shRNA related experiments in his lab. My special thanks to Dr. Jyoti Misra-Sen for providing her support in the completion of experiments using TCF KO mice at NIH and Kamal for doing these experiments.

I am grateful to IISER Pune for giving me an opportunity to pursue my PhD. I would like to thank both past and present Directors, Prof. KN Ganesh and Prof. Jayant Udgaonkar

for their support. I would also like to thank Prof. LS Shashidhara, former Chair, Biology, for his support in my early days at IISER Pune. I thank Dr. Sachin Atole, Dr. Sagar Tarate, Dr. Mahesh Sahare, Mr. Kalyan, and all the staff members of the animal house facility at IISER Pune, for their timely help and animal availability as per my requirements. I would like to thank Dr. Santosh Podder, Mr. Vijay Vitthal, Ms. Aditi Gavande and all the staff members of the microscopy facility at IISER Pune. My special thanks to Mrinalini Virkar, Shabnam Patil, Rupali Jadhav, Kalpesh Thakare, Piyush Gadekar, Sandeep Bejjanki, Mahesh Rote, Sandeep Sankpal, and all the staff members of Biology management, as well as Tushar Kurulkar from the Academics office in IISER.

I would like to thank Dr. Shekhar C. Mande, ex-director of NCCS for granting me a special permission to work at NCCS. I would like to thank Dr. Ramanamurthy, Dr. Bankar, Mrs. Vaishali, Mr. Thorat, Mr. Shaikh, and all the staff members at the Experimental animal facility at NCCS for their timely help and support. My special thanks to Amit, Ashwini, Atul, Pratibha, Himangini, and technical support at FACS facility at NCCS. I am extremely thankful to Shilpi, Narayana, Santhosh, Abdul, Vishal, and Rajesh for helping me in various aspects during my work at NCCS Pune.

I would like to thank Prof. Riitta Lahesmaa and Dr. Zhi Chen at the University of Turku, Finland for their support in performing experiments in their labs. I would like to thank Ankitha Shetty and Imran for providing support during my brief stay at University of Turku, Finland. I would like to thank Subhash, Moin, Obaid, Omid, Santosh, Karoliina, Verna and Kartiek for providing friendly environment during my work in Finland.

I would like to thank all SG lab members for providing friendly environment in the lab. I thank the senior members of the lab - Drs. Praveena, Kamal, Sunita, Rafeeq, Rahul, Rini, Manjunath, and Vijay for providing support in the initial days of my PhD. I would especially like to thank Ayush, who worked with me for last 3 years, for constant support, scientific discussion and scientific motivation to thrive in science. I would like to express my gratitude and special thanks to Drs. Rutika and Sapana for their support and motivation. I would like thank Manu for being a good friend and his helpful nature. I

thank Akhila for providing assistance and reagents in some experiments. I would like to thank Dr. Smitha for her critical suggestions for the manuscript. I would like to express my special thanks to Nilam and Madumitha for helping me in the work during their rotation in the lab. I would like to thank Ankita, Sneha, Saurabh, Anirudh, Mouli, Ashwin, Satyajeet, Ameya, Rahul, Mrinmoy, Suyash, Mukul, Deepak, Abhshek, Ashish, Rashim, and all the current and past members of the SG lab. Lastly, but not any lesser, I am grateful to Nethra, Pratima, and Madhavi for supporting me in many ways in the initial days of my PhD and I am glad to have friends like them. I wish to express my gratitude to the Lab manager Mr. Vipradas for not only providing his services but for being an elder who extremely cared for students. I would like to thank the lab attendant Mr. Santosh for providing timely help.

I am extremely thankful to the ‘Council of Scientific and Industrial Research’ (CSIR), New Delhi for the financial support during my Ph.D. I would like to thank Department of Biotechnology, Ministry of Science & Technology (DBT) for providing the travel grant to allow my research in the laboratory at University of Turku, Finland.

The days would have passed far more slowly without the support of my friends Natasha, Ayush, Libi, Srilatha, and Satish who have made my days happy during my PhD life at IISER. I am always grateful to you guys.

I sincerely express my gratitude and appreciation to my respectful teachers JV Ramana, Satya Narayana, Prasad Rao, and Kiran who had inspired me in each stage of my education.

I am indebted to my parents, sister and brother for their unconditional support, love, encouragement, and motivation to pursue science. I couldn’t have reached this stage without their support. I deeply express my gratitude to all of them.

Indumathi Patta

Abbreviations

BAF	Barrier-to-Autointegration factor
BCA	Bicinchoninic acid Assay
BCL11B	B-Cell CLL/Lymphoma 11B (Zinc Finger Protein)
bHLH	Basic helix-loop-helix
BUR	Base Unpairing Region
BSA	Bovine Serum Albumin
CDS	Coding Sequence
CLP	Common Lymphoid Progenitor
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation with High-throughput Sequencing
CD8SP	CD8 Single Positive
CD4SP	CD4 Single Positive
cTECs	Cortical Thymic Epithelial Cells
CD	Cluster Determining Region
CTCF	CCCTC-Binding Factor
cDNA	Complementary DNA
4C	Circularized Chromosome Conformation Capture
DNase I	Deoxyribonuclease I
DAG	Diacylglycerol
DRE	Distal Enhancer Region
3D	Three Dimentional
DAPI	4′,6-Diamidino-2-Phenylindole
DP	Double Positive

DN	Double Negative
ECL	Enhanced Chemiluminescence
E2A	Transcription Factor E2-Alpha
EDTA	Ethylenediamine tetra acetic Acid
FPKM	Fragments Per Kilo base of transcript per Million mapped reads
FACS	Florescence Activated Cell Sorter
GATA3	GATA Binding Protein 3
GFP	Green Fluorescent Protein
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HES1	Hairy and enhancer of split-1
Hi-C	Chromosome Conformation Capture with high-throughput sequencing
HSCs	Hematopoietic Stem Cells
HDAC	Histone Deacetylase Complex
HP1	Heterochromatin Protein 1
HD	Homeo Domain
ITAM	Immuno-Receptor Tyrosine-based Activation Motif
IRES	Internal Ribosome Entry Site
IFNg	Interferon Gamma
Inr	Initiator
Kb	Kilo Base Pairs
KCl	Potassium Chloride
KO	Knock Out
Lck	Lymphocyte Cell-Specific Protein-Tyrosine Kinase
LEF1	Lymphoid Enhancer Binding Factor 1

MHC	Major Histocompatibility Complex
MAPK	Mitogen-Activated Protein Kinase
MPP	Multipotent Progenitor
MAR	Matrix Associated Region
mRNA	Messenger RNA
Mb	Mega Base Pairs
MACS	Model-based Analysis of ChIP-Seq
$MgCl_2$	Magnesium Chloride
NaCl	Sodium Chloride
Nur77	Nuclear Receptor Subfamily 4 group A member 1
NP40	Nonidet P-40
PMSF	Phenylmethylsulfonyl Fluoride
PDZ	Post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1)
PVDF	Polyvinylidene Fluoride or Polyvinylidene Difluoride
PCR	Polymerase Chain Reaction
qRT-PCR	Quantitative Real-time Polymerase Chain Reaction
RNA pol II	RNA Polymerase II
RNA-seq	RNA Sequencing
RACE	Rapid Amplification of cDNA Ends
RIPA	Radio Immunoprecipitation Assay Buffer
RAG	Recombination Activating Gene
RPMI	Roswell Park Memorial Institute medium
RUNX3	Runt Related Transcription Factor 3
SATB1	Special AT-rich binding protein 1

SDS	Sodium Dodecyl Sulfate
SMC	Structural Maintenance of Chromosomes
SP	Single Positive
TCR	T-Cell Receptor
TCF7	T-cell Specific Transcription Factor-7
TCF1	T-cell Specific Transcription Factor-1
TOX	Thymocyte Selection Associated High Mobility Group Box
ThPOK	Th-inducing POK
tTregs	Thymic Regulatory T-cells
Tris-HCl	2-Amino-2-(hydroxymethyl) propane-1, 3-diol hydrochloride
TSS	Transcription Start Site
TADs	Topologically Associated Domains
Th1	T helper Cell Type 1
TF	Transcription Factor
ULD	ubiquitin-like domain
UTR	Untranslatable Region
uORFs	Upstream Open Reading Frame

Contents

Contents **Page No.**

Synopsis 15

Chapter 1: Regulation of chromatin organizer SATB1 by alternative promoter switch during T-cell development 20

1.1 Introduction

1.1.1 Early T-cell development 20

1.1.2 Late T-cell development – CD4/CD8 T lineage choice 22

1.1.3 The transcriptional regulation of early T lineage commitment 26

1.1.4 Transcriptional regulation of CD4/CD8 T lineage choice 30

1.1.5 The role of lineage specific transcription factors in T-cell development 32

1.1.6 The role of chromatin modifiers in T-cell development 33

1.1.7 The role of T-cell enriched chromatin organizer SATB1 in T-cell development 34

1.1.8 Transcriptional regulation of gene expression 38

1.2 Results

1.2.1 Identification of putative alternative promoters of *Satb1* gene 43

1.2.2 Identification of *Satb1* transcript variants from mouse thymocytes by 5' RACE analysis 44

1.2.3 Detection of *Satb1* transcript variants by RNA-seq analysis on mouse thymocytes 47

1.2.4 Alternative first exons of *Satb1* transcript variants act as 5'UTR elements 49

1.2.5 The cell type specific expression of *Satb1* transcript variants during T-cell development 51

1.2.6 The SATB1 protein levels are higher in immature CD4SP than DP - A discrepancy between expression levels of SATB1 mRNA and protein 54

1.2.7 SATB1 protein is stable in DP thymocytes – Cycloheximde and MG132 chase assay 56

1.2.8 Monitoring translation efficiency of *Satb1* 5' UTRs by polysome profiling of thymocytes 57

1.2.9 *Satb1* transcript variants differ in translation efficiencies *in vitro* 59

1.3 Discussion 61

1.4 Methods

1.4.1 Mice 65

1.4.2 Flow cytometry 65

1.4.3 Data base analysis 65

1.4.4 5' Rapid amplification of cDNA ends (5' RACE) 66

1.4.5 cDNA synthesis and quantitative real-time PCR analysis (qRT-PCR) 66

1.4.6 Cycloheximide and MG132 chase assay 67
1.4.7 Western blotting 67
1.4.8 Polysome profiling analysis 68
1.4.9 *In vitro* transcription, *in vitro* translation, and Luciferase reporter assay 69
1.5 References 71

Chapter 2: Molecular players of *Satb1* alternative promoter switch during CD4SP T lineage differentiation in thymus

2.1 Introduction 84

2.2 Results
2.2.1 Expression pattern of *Satb1* transcript variants during TCR mediated positive selection process 92
2.2.2 Persistent TCR signaling induces *Satb1* promoter switch during T-cell development 94
2.2.3 Lineage specific chromatin accessibility of *Satb1* P2 promoter region during development from progenitors 96
2.2.4 Developmental stage specific chromatin dynamics of *Satb1* alternative promoters during T-cell development 97
2.2.5 TCF1 regulates the expression of *Satb1* by binding to P2 promoter in CD4SP thymocytes 99
2.2.6 CD4SP thymocytes from TCF1 KO mice show downregulation of SATB1 expression 101
2.2.7 TCR signaling induces *Satb1* promoter switch in the peripheral $CD4^+$ T-cells 103
2.3 Discussion 105
2.4 Methods
2.4.1 Mice 113
2.4.2 Flow cytometry 113
2.4.3 Isolation of Naive $CD4^+$ T-cells and cell culture 114
2.4.4 Data base availability 114
2.4.5 cDNA synthesis and Quantitative real-time PCR analysis (qRT-PCR) 114
2.4.6 Western Blotting 115
2.4.7 ChIP- qPCR 116
2.4.8 Transfections and luciferase activity Assay 117
2.5 References 118

Chapter 3: Identifying the mechanism of combinatorial expression of *Satb1* transcript variants during T-cell development
3.1 Introduction 125

3.2 Results
3.2.1 Interaction profiling of *Satb1* alternative promoters with distal enhancer region located upstream to *Satb1* TSS 133
3.2.2 Expression profiling of *Satb1* transcript variants in Th1 cell type 135
3.2.3 Occupancy of chromatin architectural proteins at *Satb1* gene locus 136
3.2.4 SATB1 interacts with cohesin and co-localizes in thymocyte nucleus 139
3.2.5 SATB1- and cohesin- bound DNA regions were depleted of CTCF occupancy at *Cd3* locus 141
3.2.6 SATB1 and cohesin bound DNA regions were are depleted of CTCF occupancy at the *Cd8* locus 143
3.3 Discussion 146
3.4 Methods
3.4.1 Mice 152
3.4.2 Isolation of T-cells and cell culture 152
3.4.3 cDNA synthesis and quantitative PCR analysis (qRT-PCR) 152
3.4.4 Data base (ChIP-seq and Hi-C) analysis 153
3.4.5 Immunostaining of thymocytes 153
3.4.6 Co-immunoprecipitation and Immunoblotting 154
3.5 References 155
Publications 162

Synopsis

The development of T-cells occurs in an organ called thymus (Miller, 1961). T-cell development in the thymus has been distinguished into discrete stages based on the surface expression of co-receptors CD4 and CD8. During early double negative (DN) stage of T-cell development, thymocytes are highly proliferative and start expressing RAG-1 and RAG-2, a recombination activating genes which then initiate the TCR gene rearrangement. Thymocyte stages with successful rearrangement of TCR γ, δ gene loci are directed towards the γδ T-cell lineage (Bonneville, M, 1989), whereas thymocytes with successful rearrangement of TCRβ gene locus at DN3 stage are directed towards the αβ T-cell lineage by the process of β-selection. Pre-TCR signaling is an important player in the differentiation of DN4 stage into double positive (DP) which exhibit the surface expression of both CD4 and CD8 co-receptor molecules. DP thymocytes undergo the rearrangement of TCRα chain which then assembles with the pre-existing TCRβ chain to form complete αβTCR on their surface (Kearse et al., 1995). During thymic T-cell development, DP thymocytes which exhibit the completely rearranged TCR, will be scanned for the expression of functional receptors by the process of positive and negative selection (Germain, 2002). During T-cell development, the positively selected double positive thymocytes differentiate into either CD4SP or CD8SP depending on their TCR interaction with the type of MHC-self-peptide complexes. If αβTCRs on DP thymocytes interact with peptide-MHC-I complexes, they develop into the $CD8^{+}$ T lineage, on the other hand if they interact with peptide-MHC-II complexes, then they are diverted towards the $CD4^{+}$ T lineage (Germain, 2002). Along with TCR signaling, an interplay between the lineage specific transcription factors such as ThPOK and RUNX3 plays an important role in the CD4/CD8 T lineage choice and single positive (SP) T-cell differentiation.

SATB1 is a T-lineage-enriched chromatin organizer and global gene regulator. SATB1 regulates the multitude of genes important for T-cell survival, selection, lineage specification, and differentiation during T-cell development. SATB1 binds to its target loci and recruits the chromatin modifying machinery onto them. Studies on both

constitutive and conditional knockout (KO) mice have revealed that in the absence of SATB1, thymocytes fail to develop beyond DP stage (Alvarez et al., 2000; Kondo et al., 2016). SATB1 is essential for the TCR mediated positive and negative selection of thymocytes, and for the establishment of immune tolerance (Kondo et al., 2016). It binds to the *cis* regulatory elements of the lineage specific major transcription factors and regulates their expression during T-cell development (Kakugawa et al., 2017). SATB1 plays an essential role in the development of tTregs in the thymus via activation of Treg specific super-enhancers (Kitagawa et al., 2017). Although SATB1 is crucial for the expression of essential genes important for the T-cell development and function, the transcriptional regulation of SATB1 was not addressed yet. Therefore, understanding the transcriptional regulation of SATB1 will be essential to understand its developmental stage specific expression.

My doctoral study focuses to investigate the transcriptional regulation of SATB1 and its implication during T-cell development.

The key findings of my study are divided into three chapters as follows:

1. Regulation of chromatin organizer SATB1 by alternative promoter switch during T-cell development.

Here, we demonstrate the mechanism of SATB1 regulation by alternative promoter switch during T-cell development in the thymus. Profiling of histone modifications at the *Satb1* gene locus in thymocytes led to the identification of putative alternative promoters of *Satb1.* The analyses of 5’ RACE and RNA-seq have identified the different transcript variants of *Satb1*. The *Satb1* transcript variants differ in the distinct first exon sequences. The alternative first exons of *Satb1* transcript variants do not contribute to SATB1 protein sequence and therefore we classified them as 5’ untranslatable regions (5’ UTRs) of *Satb1* mRNA. *Satb1* transcript variants express in a combinatorial manner in the different stages of T-cell development and their expression is cell type specific during T-cell development. *Satb1* transcript variants differ in the translation efficiencies both *in vivo* and *in vitro* as confirmed by the polysome profiling and *in vitro* translation assay, respectively. Also the 5’ UTR sequences of *Satb1* transcript variants exhibit

distinct secondary structures. Finally, we demonstrate that the differential translatability and the combinatorial expression of *Satb1* transcript variants regulate the levels of SATB1 protein in developing thymocytes.

2. Molecular players of *Satb1* alternative promoter switch during CD4SP T lineage differentiation in the thymus.

Our study demonstrates that the *Satb1* alternative promoter switch in various developmental stages of T-cell development is induced by persistent TCR signaling. We found that the changes in TCR signaling induce the switch in the expression of *Satb1* transcript variants during T-cell development. We show that *Satb1* P2 promoter switch, and increased P2 variant expression along with constitutively expressed predominant P3 variant during the differentiation of CD4SP thymocytes from DP thymocytes is induced by the persistent TCR signaling. Further, only the TCR engaged CD4 but not TCR engaged DP thymocytes specifically express the P2 variant. Our results confirm that the expression of P2 transcript variant is induced not just by TCR signaling but by the persistent or stronger TCR signaling. Therefore CD8SP thymocytes do not express the P2 transcript variant during their development, which requires a weaker or cessation of TCR signaling at the DP stage. Additionally, we show that not only during T-cell development, TCR activation of peripheral $CD4^+$ T-cells also switches the *Satb1* alternative promoters, thus resulting in the expression of P2 variant along with P3 suggesting that *Satb1* is regulated by TCR signal via selective alternative promoter usage. Furthermore, we found that the *Satb1* alternative promoters exhibit distinct lineage specific chromatin dynamics during T-cell development from the progenitors. Especially the chromatin dynamics at P2 promoter is developmental stage specific during T-cell development from the progenitors. Finally, we show that TCF1 regulates the *Satb1* alternative promoter switch by directly binding to the P2 promoter region during development of CD4SP from DP thymocytes. Our results demonstrate that the CD4 SP thymocytes specific P2 promoter switch is mediated by TCF1 during T-cell development.

3. Identification of the mechanism of combinatorial expression of *Satb1* transcript variants during T-cell development

Since the combination of *Satb1* transcript variants expresses in a cell type specific manner in T-cell development, this study focuses to identify the mechanisms which bring the combinatorial expression of *Satb1* variants. We found that the *Satb1* alternative promoters interact with the distal enhancer region as shown by Hi-C and virtual 4C analysis. We show that the chromatin organizing proteins such as cohesin and CTCF bind to the *Satb1* alternative promoters. However, cohesin binding was observed at the distal enhancer region wherein CTCF binding was absent. On the other hand, SATB1 binds to its alternative promoters as well as the distal enhancer region. We show that SATB1 and cohesin bind to the *Satb1* alternative promoters, and also to the distal enhancer region, and presumably cooperate to mediate the chromatin interactions at the *Satb1* gene locus. SATB1 interacts with cohesin (SMC1), and co-localizes in the thymocyte nuclei. SATB1 and cohesin also bind to the super-enhancers and other regulatory regions of key genes involved in the T-cell development. Our work suggests that SATB1 interacts with cohesin in the thymocyte nuclei and together might regulate the chromatin interactions at the *Satb1* gene locus as well as at several key genes of developing thymocytes, thus maintain the T lineage specificity.

Taken together, the results elaborated in the thesis demonstrate that the combinatorial alternative promoter usage drives *Satb*1 expression during various stages of thymocyte development and also during the activation of peripheral T-cells. The selective expression of combination of *Satb1* transcript variants during T-cell development has a crucial role on the regulation of SATB1 protein levels. Persistent TCR signaling induces *Satb1* alternative promoter switch in CD4SP thymocytes during their differentiation from DP thymocytes. Interestingly, we found that chromatin dynamics at the *Satb1* alternative promoters is lineage specific during T-cell development from their progenitors. We have further identified the molecular players of this *Satb1* alternative promoter switch in developing thymocytes and found that TCF1 is one of the players. Further studies are required to fully dissect the role of developmental stage specific transcription factors and the epigenetic mechanisms in regulating the expression of SATB1 via alternative

promoters in a cell type specific manner. Whether perturbation of the multitude of *Satb1* alternative promoters exerts any effect on the lineage specific expression of SATB1 is yet to be understood.

References

Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, Kohwi-Shigematsu T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev 14, 521-35.

Bonneville M, Ishida I, Mombaerts P, Katsuki M, Verbeek S, Berns A, Tonegawa S. (1989). Blockage of alpha beta T-cell development by TCR gamma delta transgenes. Nature *342,* 931-4.

Germain RN. (2002). T-cell development and the CD4-CD8 lineage decision. Nat Rev Immunol *2,* 309-22.

Kakugawa K, Kojo S, Tanaka H, Seo W, Endo TA, Kitagawa Y, Muroi S, Tenno M, Yasmin N, Kohwi Y, Sakaguchi S, Kowhi-Shigematsu T, Taniuchi I. (2017). Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. Cell Rep 19, 1176-1188.

Kearse KP, Roberts JL, Singer A. (1995). TCR alpha-CD3 delta epsilon association is the initial step in alpha beta dimer formation in murine T cells and is limiting in immature CD4+ CD8+ thymocytes. Immunity 2, 391-9.

Kitagawa Y, Ohkura N, Kidani Y, Vandenbon A, Hirota K, Kawakami R, Yasuda K, Motooka D, Nakamura S, Kondo M, Taniuchi I, Kohwi-Shigematsu T, Sakaguchi S. (2017). Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. Nat Immunol 18, 173-183.

Kondo M, Tanaka Y, Kuwabara T, Naito T, Kohwi-Shigematsu T, Watanabe A. (2016). SATB1 Plays a Critical Role in Establishment of Immune Tolerance. J Immunol *196,* 563-72.

Miller JF. (1961). Immunological function of the thymus. Lancet *2,* 748-9

Chapter 1

Regulation of chromatin organizer SATB1 by alternative promoter switch during T-cell development

1.1 Introduction

1.1.1 Early T-cell development

The development of T-cells occurs in an organ called thymus (Miller, 1961). The precursor hematopoietic stem cells (HSCs) in the bone marrow differentiate into MPP (multipotent progenitor) and then into CLP (common lymphoid progenitor). CLP is a bi-potent cell, which can be directed to differentiate into either a B-cell or a T-cell. Once CLPs migrate to the thymus, they are directed toward the T-cell lineage but still they have the potential to differentiate into the B-cell lineage as well (Izon et al., 2001; Koch et al., 2001; Wilson, MacDonald and Radtke, 2001). Notch-1 signaling in the thymic micro-environment directs the lymphoid precursor differentiation (Schmitt and Zúñiga-Pflücker, 2002) towards the T-cell lineage, thus in the absence of Notch-1 signals, the lymphoid precursors are directed towards the B-cell lineage in the thymus (Radtke et al., 1999; Izon et al., 2002; Izon, Punt and Pear, 2002).

The development of T-cells in the thymus has been distinguished into discrete stages based on the surface expression of the co-receptors CD4 and CD8 (Figure 1.1.1). The early T-cell developmental stages lack the surface expression of both CD4 and CD8, hence they are called double negative (DN) (Figure 1.1.1). DN thymocyte stages are named DN1 through DN4, in order of their appearance during the development. The earliest T lineage committed cells in the thymus are the DN1 thymocytes, which are characterized as $CD44^{+}25^{-}$; and then DN2 thymocytes as $CD44^{+}25^{+}$; DN3 thymocytes as $CD44^{-}25^{+}$; and finally DN4 thymocytes as $CD44^{-}25^{-}$ (Pearse et al., 1989; Godfrey et al., 1993). During this early DN stage of T-cell development, thymocytes are highly proliferative and start expressing RAG-1 and RAG-2, the recombination activating genes, which then initiate the TCR gene rearrangement.

The first proliferation events occur at DN1 and DN2 stage, before the expression of RAG proteins which then start expressing at the DN2 and DN3 stages resulting in the rearrangement of TCR-β, γ, δ gene loci (Kawamoto et al., 2003). The thymocyte stages with successful rearrangement of TCR γ, δ gene loci are directed towards the γδ T-cell lineage (Bonneville et al., 1989), whereas thymocytes with successful rearrangement of TCRβ gene locus at the DN3 stage of development, are directed towards the αβ T-cell lineage via the process of β-selection. The TCRα chains do not express at the DN3 stage. The DN3 stage expresses an invariant chain, pre-Tα which can assemble with TCRβ to form the pre-TCR signaling complex, wherein it transduces the survival signals during the development (Saint-Ruf *et al.*, 1994; Fehling *et al.*, 1995). The pre-TCR signaling is important for the differentiation of DN4 stage into double positive (DP) thymocytes, which exhibit the surface expression of both CD4 and CD8 co-receptor molecules (Figure 1.1.1). DP thymocytes undergo the rearrangement of TCRα chain, which then assembles with the pre-existing TCRβ chain to form the complete functional αβTCR receptor on their surface (Kearse, Roberts and Singer, 1995).

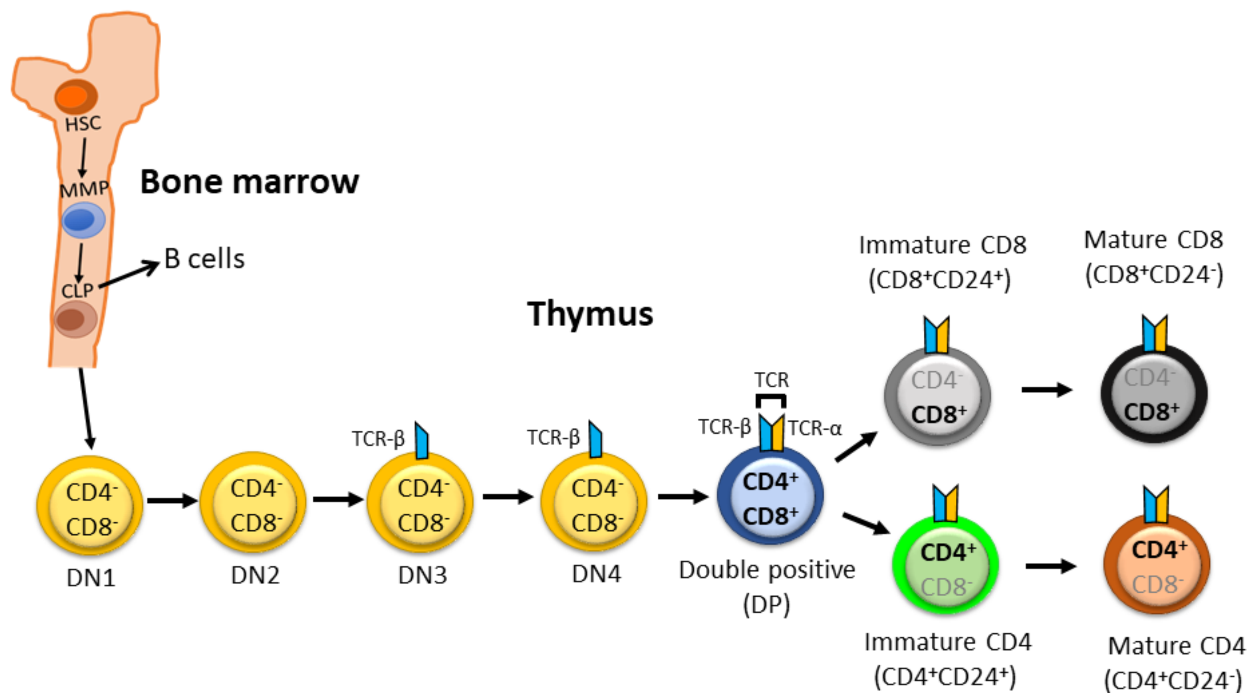


Figure 1.1.1. Schematic representation of the T-cell development in the thymus. A Common lymphoid precursor (CLP) from the bone marrow migrates to the thymus, and undergoes a series of distinct developmental programs to differentiate into a completely mature T-cell. The T-cell developmental

stages are characterized based on the surface expression of CD4 and CD8 co-receptors. The initial developmental stages are characterized as DN ($CD4^-CD8^-$), which then developed into DP ($CD4^+CD8^+$), where it express the completely rearranged TCR on their surface. Upon TCR stimuli, the double positive thymocytes undergo positive and negative selection process. The positively selected double positive thymocytes differentiate into either the CD4SP ($CD4^+CD8^-$) or the CD8SP ($CD4^-CD8^+$) T lineages, depending on their TCR self-peptide MHC interactions.

1.1.2 Late T-cell development – the CD4/CD8 T lineage choice

The αβTCRs of DP thymocytes start engage with the self-peptide MHC complexes expressed by the thymic stromal cells (Figure 1.1.2). As shown in the Figure 1.1.2, during the process of TCR-self-peptide-MHC interactions, the DP thymocytes which express αβTCRs with high affinity towards the self-peptide MHC complexes, undergo apoptosis by a process called negative selection (Kappler, Roehm and Marrack, 1987; Sha *et al.*, 1988). On the other hand, the DP thymocytes with αβTCR that exhibit very low affinity for the self-peptide MHC complexes undergo the process of death by neglect (Figure 1.1.2) (Mombaerts *et al.*, 1992). Now, only the DP thymocytes whose αβTCR have an intermediate affinity for the self-peptide MHC complexes differentiate into the $CD4^+$ SP or $CD8^+$ SP thymocytes by the process of positive selection (Figure 1.1.2) (Kisielow *et al.*, 1988; Kaye *et al.*, 1989; Hogquist *et al.*, 1994). Hence the process of positive selection not only identifies the relatively few DP thymocytes with useful TCRs; it further promotes the differentiation of DP into the single positive thymocytes (Figure 1.1.2). During this process, the DP thymocytes exhibit the transcriptional termination of one of the co-receptors such that those DP thymocytes with the transcriptional termination of CD8 differentiate into the $CD4^+$ T-cells, whereas the DP thymocytes that exhibit the transcriptional termination of CD4 differentiate into the $CD8^+$ T-cells (Figure 1.1.1) (Chan *et al.*, 1993; Davis *et al.*, 1993; Benveniste, Knowles and Cohen, 1996).

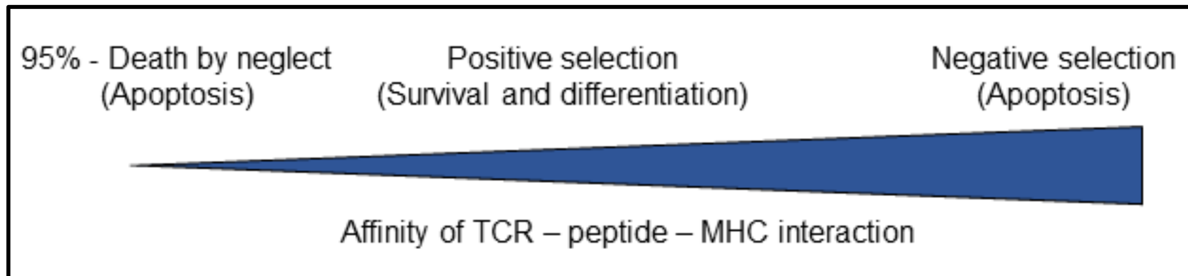


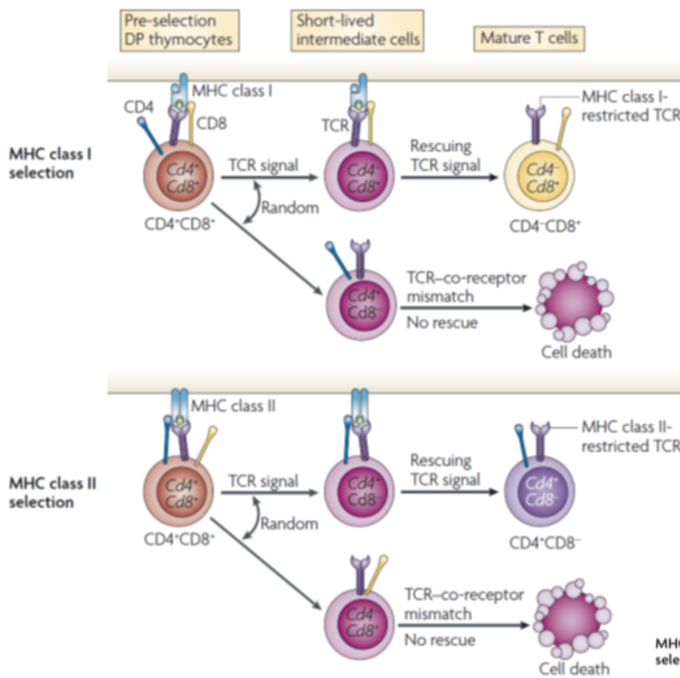
Figure 1.1.2. An instrumental role played by TCR signaling during T-cell development. The strength of TCR signaling received by the developing T-cells during thymic T-cell development plays an instrumental role in the selection of functional developing T-cells as shown in the above figure. During T-cell development, once DP thymocytes express the completely rearranged TCR receptor on their surface, they will be scanned for the expression of functional TCRs. During this process, approximately 90-95% of thymocytes which express nonfunctional TCRs undergo apoptosis by the process of death by neglect. Further, the thymocytes which express the TCRs of high affinity for the self-peptide MHC complexes undergo death by negative selection process. Only 3-5 % of the thymocytes, which express the TCRs of an intermediate affinity towards the self-peptide MHC complexes undergo the positive selection and survive.

The CD4/CD8 lineage choice of DP thymocytes is explained by the classical models of lineage commitment such as instruction, stochastic and an alternative kinetic signal model. According to the instruction model, also called as the strength of signal instructional model (Robey *et al.*, 1991), the quantitative differences in the TCR signal received by the $TCR^{+}CD4$ combination and TCR+CD8 combination, results in the termination of other co-receptor expression (Itano et al., 1996). This model suggests that the TCR+CD4 combination generates a strong TCR signal than the TCR+CD8 combination and thus the stronger TCR signals facilitate the differentiation of DP into the $CD4^{+}$ T-cell lineage by terminating the expression of CD8, whereas the weaker TCR signal allows the differentiation of DP towards the $CD8^{+}$ T lineage by downregulating the expression of CD4 (Figure 1.1.3 B). In contrast, the stochastic selection model explains that the CD4/CD8 T lineage choice is a random event during development (Davis *et al.*, 1993). This model proposes that the thymocytes which engage the self-peptide MHC complexes with the matched TCR+co-receptor combination, will receive the TCR mediated survival signal, whereas those with the mismatched TCR+co-receptor

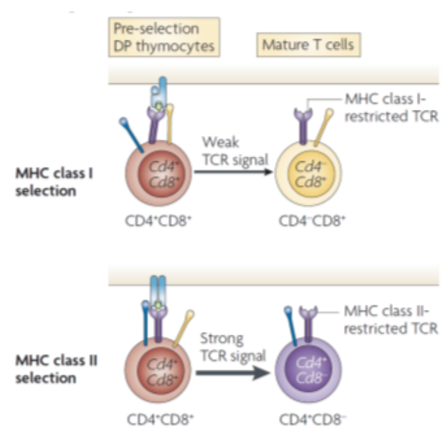
combination, such as CD4 in case of MHC-I engaged TCR, and CD8 in case of MHC-II engaged TCR, will undergo death (Figure 1.1.3 A). However, both the instruction and the stochastic selection models indicates that the CD4/CD8 lineage commitment occurs in the TCR signaled DP thymocytes simultaneously with the positive selection.

Other than these two classical models, the kinetic signaling model postulates that the duration of positively selecting TCR signaling in the DP thymocytes induces the lineage commitment (Yasutomo *et al.*, 2000; Singer, 2002). During the positive selection, the TCR mediated signals induce DP thymocytes to terminate the CD8 expression and hence they are converted to $CD4^+CD8^-$ intermediate cells, which are the lineage uncommitted cells and have potential to differentiate into either the $CD4^+$ T- or the $CD8^+$ SP T-cells (Brugnera *et al.*, 2000; Cibotti *et al.*, 2000). The persistent TCR signaling in the $CD4^+CD8^-$ intermediates results in the differentiation towards the $CD4^+$ T-cells, whereas the cessation of TCR signaling in the $CD4^+CD8^-$ intermediate cells leads to their differentiation towards the $CD8^+$ T-cells (Figure 1.1.3 C).

A) Stochastic selection model



B) Strength-of-signal instructional model



C) Kinetic signal instructional model

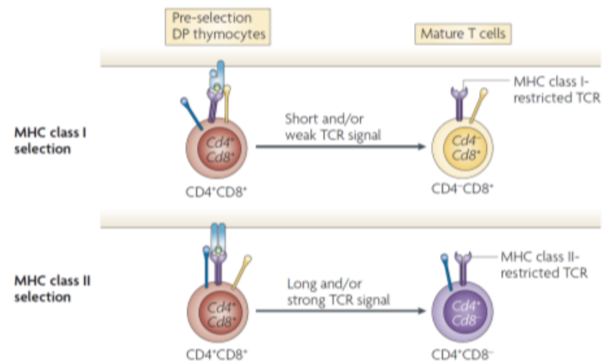


Figure 1.1.3. Different classical models of the CD4/CD8 T lineage choice during T-cell development.

A) The stochastic selection model proposes that the correct combination of TCR, co-receptor, MHC interactions allows the survival and further differentiation into the CD4/CD8 SP stage. According to this model, the CD4/CD8 lineage choice is a random event, wherein the MHC-I:TCR interactions lead to the $CD8^+$ T lineage and MHC-II:TCR interactions lead to the $CD4^+$ T lineage. B) The strength of signal instructional model postulates that a stronger TCR signal facilitates the CD4SP lineage choice, whereas a weaker TCR signal allows the CD8SP lineage choice. According to this model, the MHC-II TCR interactions generates the stronger TCR signaling resulting in the $CD4^+$ T lineage choice, whereas the MHC-I TCR interactions generates the weaker TCR signaling thereby resulting in the $CD8^+$ T lineage choice. C) The kinetic or the duration of signal instructional model suggests that the persistent TCR signaling allows the CD4SP T-cell differentiation, whereas the cessation of TCR signaling leads to the CD8SP T-cell differentiation (Reproduced from Singer, Adoro and Park, 2008).

The strength of signal model suggests that the TCR signal strength dictates the lineage choice of CD4/CD8 (Itano *et al.*, 1996) and also the MHC class II specific thymocytes became differentiated into the $CD8^+$ T-cells in the absence of CD4 expression (Matechak *et al.*, 1996). The observations suggest that the CD4 assists the TCR signaling and hence the weak TCR signaling in the absence of CD4, leads to the $CD8^+$ T-cell differentiation. When the pre-selection DP thymocytes were exposed to the varying concentrations of phorbol ester and ionomycin, the higher concentrations and the longer duration of exposure resulted in $CD4^+$ T-cell differentiation, whereas exposure to the lower concentrations led to $CD8^+$ T-cell differentiation (Ohoka *et al.*, 1997). Similar observations were proposed on the basis of the effect of levels of MAPK activity on the lineage choice decision of CD4/CD8, wherein the high MAPK activity promotes the $CD4^+$ T-cell development, whereas the low MAPK activity favors the $CD8^+$ T-cell development (Sharp *et al.*, 1997). The lineage choice decision was also controlled by the extent of LCK activation because of the effect of more LCK activation resulting in the $CD4^+$ T-cell fate, and low LCK activation leads to the $CD8^+$ T-cell fate (Basson *et al.*, 1998; Hernández-Hoyos *et al.*, 2000; Legname *et al.*, 2000). Further, in the DP thymocytes, the CD4 cytoplasmic tails are associated with more Lck molecules than the CD8 cytoplasmic tails, indicating that during the T-cell development, the engagement of CD4 and TCR with MHC class-II specific peptides leads to the increased association of LCK with the signaling complex than the CD8:TCR engagement with the MHC class-I specific peptides (Veillette *et al.*, 1989; Weist *et al*. 1993). Hence, stronger TCR signaling leads to $CD4^+$ T-cell development and weaker TCR signaling leads to $CD8^+$ T-cell development.

1.1.3 The transcriptional regulation of early T lineage commitment

The thymic environment provides cues, which initiate the programmed transcriptional events for the specification and commitment of the multipotent progenitors into the T lineage. Among the DN stages of thymocyte development, the DN1 stage shows more characteristics of the early thymic precursors (ETPs) and they develop into DN2. Signaling through Notch1, IL-7 receptor and IL-2 receptor at the DN2 stage of

development, prepares them to acquire the T-cell identity. During the transition from DN2 to DN3 stage, the thymocytes start expressing the recombinases RAG1/2, and also the proteins of TCR signaling such as CD3є and Zap70. Therefore at the DN3 stage, thymocytes undergo rearrangement of the TCRβ chain, resulting in the selection of αβ T-cells through the β-selection process, which distinguishes the αβ T-cell development from the γδ T-cell development. The pre-TCR signaling at the DN4 stage leads to their differentiation into the DP thymocytes. During this process of acquiring T-cell identity in the thymus, multipotent progenitors at the different stages of development tend to lose their potential to differentiate into alternative lineages such as the B-cells. Hence, before the development of DN4 stage, multiple signaling pathways play a pivotal role in the αβ T lineage commitment of the precursors (Yui and Rothenberg, 2014).

The role of Notch signaling

Different phases of transcriptional activation events result in the development of various stages of T-cell development. The Notch signaling initiates the transcriptional events of the target genes important during the initial T lineage commitment. The expression of T lineage specific transcription factors leads to the opening of TCR loci for the recombination events and also represses the progenitor signature gene expression. As a result of the TCR rearrangement, the TCR signaling leads to the activation of new regulatory genes essential for further T-cell developmental stages, which are mostly TCR-dependent and Notch-independent. The T-cell specification genes such as TCF7, GATA3 and BCL11B are induced by the Notch signaling and are essential for the establishment of transcriptional networks in the progenitor cells. Notch1 is highly expressed on the surface of the progenitors, which originate in the bone marrow and its expression is maintained through the development until the early DN3 stage. As thymic epithelial cells express the Notch ligand such as Dll4, thus Notch receptor and ligand interactions facilitate the T-cell fate by allowing the progenitors to lose their potential to differentiate into other lineages (Schmitt and Zúñiga-Pflücker, 2002; Abe *et al.*, 2010). The Notch1 signaling is essential for the transition of DN1 to DN2 stage (Pui *et al.*, 1999; Radtke *et al.*, 1999) and of DN2 to DN3 stage of development (Wolfer *et al.*, 2002). Loss of Notch1 leads to the developmental defect at the DN1 stage and thus

Lck-Cre mediated conditional deletion of Notch1 resulting in the impaired β-selection leads to the defective development of αβ T-cells (Pui *et al.*, 1999; Radtke *et al.*, 1999; Wolfer *et al.*, 2002). The downstream targets of Notch1 signaling that play a crucial role in the T-cell development are HES1, TCF1, CD25 and pre TCR-α (pTα) (Deftos *et al.*, 1998; Tomita *et al.*, 1999; Maillard *et al.*, 2006; Weber *et al.*, 2011). The next section describes the roles of key transcription factors involved in T lineage commitment and T-cell development.

HES1

HES, a bHLH family transcriptional repressor, is expressed by the T-cells as well as by the various immune cells such as B cells, mast cells and dendritic cells (Caton, Smith-Raska and Reizis, 2007; Sakata-Yanagimoto *et al.*, 2008). At the stage of T lineage commitment, HES1 is expressed in the ETPs and continues through the DN3a stage. HES1 deficiency leads to the defect in the T-cell development at the early DN stage, and resulting in the ectopic expression of myeloid lineage genes, indicating the forced hematopoiesis towards the myeloid lineage (Tomita *et al.*, 1999; Wendorff *et al.*, 2010; De Obaldia *et al.*, 2013).

TCF1

TCF1 is encoded by the *Tcf7* gene. Notch binds to the *Tcf7* enhancer located upstream of the transcriptional start site, thus resulting in the transcriptional activation of *Tcf7* (Germar *et al.*, 2011). Once expressed, TCF1 can act as either an activator or repressor of gene expression. In response to the Notch signaling, TCF1 is highly expressed in the ETPs, wherein it is important for the transition of DN1 to DN2 (Gifford and Meissner, 2012). TCF1 is important for the survival and proliferation of ETPs (Germar *et al.*, 2011). TCF1 is primarily responsible for the activation of T-cell specification genes in collaboration with the notch signaling through a feedback circuit. Thymus from TCF1 KO mouse shows significantly reduced number of DN1 thymocytes (Weber *et al.*, 2011), whereas the overexpression of TCF1 leads to the T lineage development *in vitro* in the absence of Notch signaling (Germar *et al.*, 2011). This indicates that TCF1 act as a T

lineage specifying transcription factor in response Notch signaling and is critical for early T lineage development in the thymus.

GATA3

GATA3 plays a crucial role throughout the development including the T lineage commitment, survival, T-cell specification, and also diversification of T lineage subset upon the TCR signaling (Weber *et al.*, 2011) (Ho, Tai and Pai, 2009; Hosoya, Maillard and Engel, 2010). It plays a major role in the commitment of progenitors toward the T lineage by excluding the B cell fate (Garćia-Ojeda *et al.*, 2013). It regulates the expression of Bcl11b at the DN2 stage and also prepares the DN3 cells for β-selection.

BCL11B

BCL11B is a six zinc finger containing transcription factor and has co-repressor activity due to its interaction with the transcription repressor complexes (Cismasiu *et al.*, 2005). The temporal expression of BCL11B is essential for the lineage commitment (Ikawa *et al.*, 2010; Li *et al.*, 2010; Li, Leid and Rothenberg, 2010). BCL11B is very important for T-cell development at the DN3 stage, at which BCL11B is important to maintain the silenced state of id2, which antagonizes Notch signaling (Li *et al.*, 2010; Li, Leid and Rothenberg, 2010). Deficiency of BCL11B leads to the defect in the development of αβ T-cell lineage (Wakabayashi *et al.*, 2003; Inoue *et al.*, 2006) and then leads to the reprogramming of T-cell lineage towards NKT cells (Li *et al.*, 2010).

E2A and HEB

After transcriptional activation of TCRβ gene, the successful VDJ recombination and the assembly of pre-TCRα with β-chain were monitored at the DN3 thymocytes via a developmental checkpoint known as the β-selection. E2A is required for the activation of VDJ recombination at the TCRβ locus (Agata *et al.*, 2007). The impairment of E2A and HEB function leads to the defective VDJ recombination and then the development is blocked at the DN3 stage (Barndt, Dai and Zhuang, 2000). E2A acts together with Notch to activate the genes important for the early stages of T-cell development (Ikawa *et al.*, 2006).

1.1.4 Transcriptional regulation of CD4/CD8 T lineage choice

The role of TCR signaling in CD4/CD8 T lineage choice

TCRs on thymocytes were tested for their ability to recognize the self-peptide MHC complexes on the cortical epithelial cells. Poor TCR signaling leads to the apoptosis of thymocytes by the process of death by neglect. Almost 90-95% of the developing thymocytes undergo this process as they fail to express the functional TCRs on their surface and hence they die in the thymus. The strong TCR signals result in the apoptosis of thymocytes by the process of negative selection (Palmer, 2003; Stritesky, Jameson and Hogquist, 2012). TCR signals above a certain threshold allow the survival and differentiation of thymocytes by the process of positive selection (Starr, Jameson and Hogquist, 2003). During this process, the initial TCR signaling results in the partial downregulation of CD8 and upregulation of TCR molecules on the thymocyte surface, leading to the generation of the $CD4^{+}CD8^{int}$ population. At this stage, due to the reduced levels of CD8, the TCR signaling is interrupted in case of TCR encountering with the MHC class-I peptides, resulting in the commitment of thymocytes towards the $CD8^{+}$ T lineage, with further downregulation of CD4 and upregulation of CD8. If the TCR encounters the MHC class-II peptides, the loss of CD8 does not interfere with the TCR signaling and hence leads to the development of the $CD4^{+}$ T lineage. Hence the cessation of TCR signaling allows development of the $CD8^{+}$ T lineage, whereas sustained TCR signaling leads to the $CD4^{+}$ T lineage differentiation (Singer, Adoro and Park, 2008).

Molecular pathways downstream of the TCR signaling

TCR recognition of the peptide MHC complexes along with the co-receptor recognition of MHCs induces the activation of Lck, which then phosphorylate the immune receptor tyrosine based activation motifs (ITAMs) of CD3 molecules (Figure 1.1.4). This enables the binding of Zap70 with the phosphorylated ITAMs of CD3 and then phosphorylated by Lck. The phosphorylated Zap70 is activated and then phosphorylates the LAT complexes, which in turn activate the further downstream signaling events. The LATs associate with SLP76, Itk, PLCγ and form a signalosome complex. PLCγ acts on the

membrane-bound PIP2 and catalyzes it into the membrane bound DAG and IP3. IP3 diffuses into the cytoplasm and is then bound by the IP3 receptors on the endoplasmic reticulum (ER), resulting in the release of $Ca2^+$ from the ER. The released $Ca2^+$ then facilitates the dephosphorylation, and the nuclear localization of NFAT, a transcription factor (Figure 1.1.4). The membrane-bound DAG activates Protein kinase C (PKCθ and PKCη), which then activates and allows the nuclear translocation of NFkB, another transcription factor. DAG is also bound by RasGRP, which activates the Ras-MAPK pathway leading to the activation of its downstream effector proteins ERK1 and ERK2, resulting in the further activation and nuclear entry of the AP1 transcription factor complex consisting of Jun and Fos heterodimer. All of these nuclear translocated transcription factors regulate the expression of specific set of target genes orchestrating the survival and the differentiation events of thymocytes during T-cell development (Figure 1.1.4).

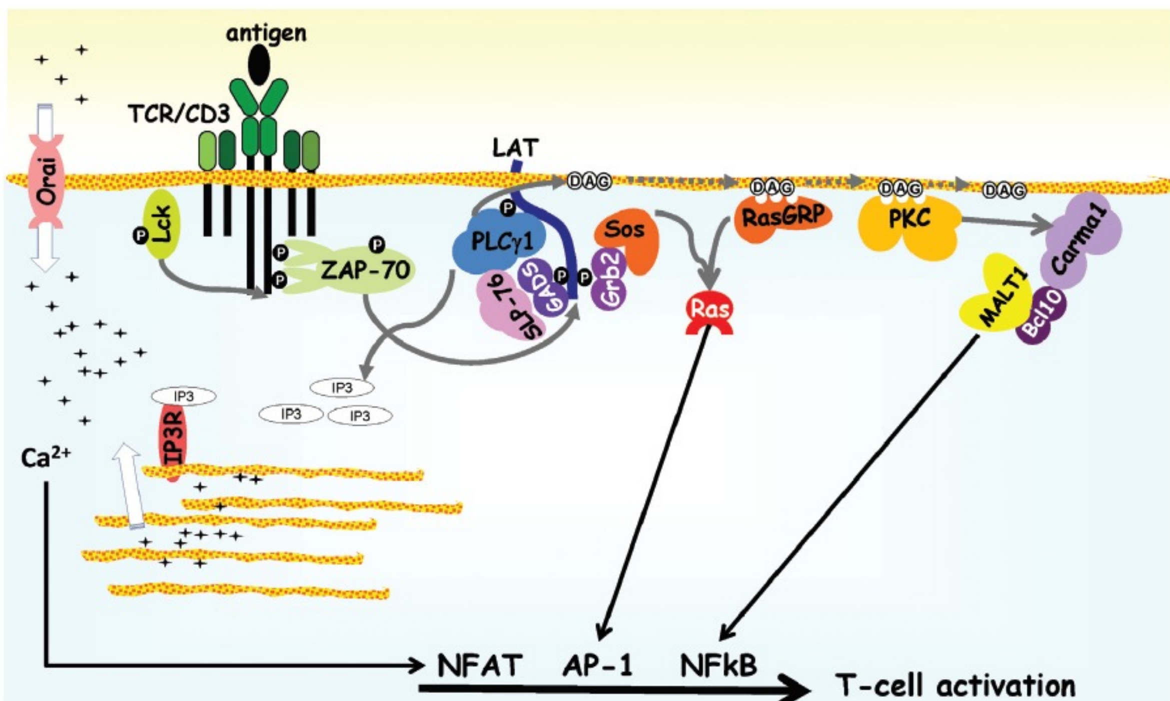


Figure 1.1.4. Schematic representation of the TCR signaling. The interaction of TCR with the self-peptide–MHC complexes leads to the activation of several TCR proximity kinases such as Lck, ZAP70, LAT, and SLP-76, which then mediate several downstream phosphorylation events, and activates the

Ras MAPK signaling, $Ca2^+$ release from the endoplasmic reticulum, also increases the $Ca2^+$ influx, PLCγ mediated PIP2 breakdown into IP3 and DAG, and DAG mediated activation of PKCθ. These signaling pathways lead to the activation of major transcription factors such as NFκB, NFAT and AP-1. These canonical transcription factors then translocate into the nucleus and then regulate the expression of genes important for the T-cell activation and function (Reproduced from Simeoni and Bogeski, 2015).

1.1.5 The role of lineage specific transcription factors during T-cell development

During T-cell development, the lineage specific transcription factors play a crucial role in the CD4/CD8 T lineage choice mediated by the MHC restricted TCR signal. During this process of lineage commitment, the lineage specific transcription factors downregulate the expression of other co-receptor molecules. The major CD4/$CD8^+$ T lineage specific major transcription factors are ThPOK in case of the $CD4^+$ T lineage and RUNX3 in case of the $CD8^+$ T lineage.

The role of Runx family proteins in the $CD8^+$ T lineage choice

Runx family proteins are known to play major roles in the $CD8^+$ T lineage commitment. Runx proteins bind to the enhancer region of the *Cd8* locus and thus lead to the activation of CD8 gene expression (Sato *et al.*, 2005). Distinct Runx proteins such as Runx1 at the DN stage, and Runx3 at the CD8 SP stage, bind to the silencer region of the *Cd4* locus and then downregulate CD4 gene expression (Sawada *et al.*, 1994; Siu *et al.*, 1994; Taniuchi *et al.*, 2002). Runx proteins also regulate the expression of major transcription factors of the $CD4^+$ T lineage such as ThPOK. Runx proteins bind to the regulatory regions of the *Zbtb7b* gene, which encodes ThPOK, thus resulting in the repression of Zbtb7b expression (Setoguchi *et al.*, 2008). The deficiency of Runx1 and Runx3 leads to the ectopic expression of ThPOK, which then leads to the re-direction of MHC class-I selected thymocytes towards the $CD4^+$ T lineage (Setoguchi *et al.*, 2008). Taking these observations together, the Runx proteins play a crucial role in the $CD8^+$ T lineage development by shutting down the important gene networks specific to the $CD4^+$ T lineage.

The role of ThPOK in the CD4$^+$ T lineage choice

ThPOK is encoded by a gene *Zbtb7b* that plays an essential role in directing the MHC class-II selected thymocytes towards the CD4$^+$ T lineage differentiation. ThPOK antagonizes the silencer activities at the distal enhancer region (DRE) of its gene and also the silencer element at CD4 locus thus allowing their expression in the developing CD4$^+$ T-cells (Muroi *et al.*, 2008). *Zbtb7b* gene expression is activated by GATA3 in the CD4$^+$ T lineage committed thymocytes. GATA3 binds to the two regions at *zbtb7b* gene locus and activates its expression, hence Zbtb7b expression was significantly reduced in the GATA3 deficient thymocytes (Wang *et al.*, 2008). The conditional deletion of GATA3 in DP thymocytes leads to a remarkable impairment of CD4$^+$ T-cell development (Hernández-Hoyos *et al.*, 2003; Pai *et al.*, 2003).

Along with the above-mentioned canonical transcription factors downstream of the TCR signaling and the lineage specific major transcription factors, several chromatin modifying mechanisms play an important role in the regulation of lineage specific gene expression during T-cell development.

1.1.6 The role of chromatin modifiers in T-cell development

The conditional depletion of Dnmt1 in the early T-cell developmental stages leads to the demethylation of DNA, resulting in the specific ablation of αβ T-cell development but no change in the γδ T-cell number. The silencing of CD4 gene expression during the transition of DP into CD8$^+$ T lineage requires HDAC and HP1 chromatin repressor proteins (Chi, 2004). On the other hand, Mi-2β mediated recruitment of p300 on the *Cd4* proximal enhancer region is required for the CD4 gene expression. Also, the BAF chromatin remodeling complex binds to the enhancer region of *Cd8*, thus facilitating the CD8 gene expression (Kioussis and Ellmeier, 2002; Sato *et al.*, 2005). A chromatin organizer which is highly enriched specifically in the thymocytes and is critical for the T-cell development is SATB1 (Alvarez *et al.*, 2000; Kondo *et al.*, 2016).

1.1.7 The role of T-cell enriched chromatin organizer SATB1 in T-cell development

SATB1 (special AT-rich binding protein 1), originally identified due to its propensity to bind the DNA sequences with the high degree of base-unpairing called BURs (base-unpairing regions), is a higher-order chromatin organizer and a lineage-specific transcription factor (Dickinson *et al*., 1992; De Belle, Cai and Kohwi-Shigematsu, 1998; Cai, Han and Kohwi-Shigematsu, 2003). SATB1 forms an unusual 'cage-like' three-dimensional structure in the mouse thymocytes, where it is predominantly found, and presumably circumscribes the heterochromatin (Figure 1.1.5. A) (Dickinson *et al.*, 1992; Cai, Han and Kohwi-Shigematsu, 2003). SATB1 is essential for the looping of chromatin and tethering of the special AT-rich DNA regions to the nuclear matrix (De Belle, Cai and Kohwi-Shigematsu, 1998; Cai, Lee and Kohwi-Shigematsu, 2006), thus forming the chromatin 'loopscape' (Pavan Kumar *et al.*, 2007). SATB1 regulates its target gene expression by acting as a docking site for a number of chromatin modifiers and nucleosome remodelers (Yasui *et al.*, 2002; Kumar *et al.*, 2005). SATB1 interacts with the SWI/SNF complexes and recruit these complexes on the target DNA, thus regulates the expression of vast number of genes (Alvarez *et al.*, 2000; Yasui *et al.*, 2002; Cai, Lee and Kohwi-Shigematsu, 2006). SATB1 organizes the MHC class-I locus and Th2 cytokine genes by tethering the MARs to the nuclear matrix (Cai, Lee and Kohwi-Shigematsu, 2006; Pavan Kumar *et al.*, 2007). In thymocyte nucleus, SATB1 seems to involve in the dynamic organization of open chromatin (Galande *et al.*, 2007). The N-terminal PDZ-like domain of SATB1 is important for its homodimerization and also mediates its interactions with the interaction partners (Figure 1.1.5. B) (Galande *et al.*, 2001; Purbey *et al.*, 2008). The MAR-binding property of SATB1 resides within the C-terminal cut domain and in the homeodomain (HD) (Dickinson, Dickinson and Kohwi-Shigematsu, 1997; Purbey *et al.*, 2008). SATB1 regulates the expression of genes such as myc and Bcl2, which are the targets of Wnt/β-catenin pathway (Cai, Han and Kohwi-Shigematsu, 2003; Ma *et al.*, 2007). SATB1 interacts with β-catenin and thus recruits β-catenin and p300 on the essential genes of Th2 cells, therefore SATB1 mediates Th2 differentiation specifically by mediating the Wnt/ β-catenin signaling pathway (Notani *et al.*, 2010).

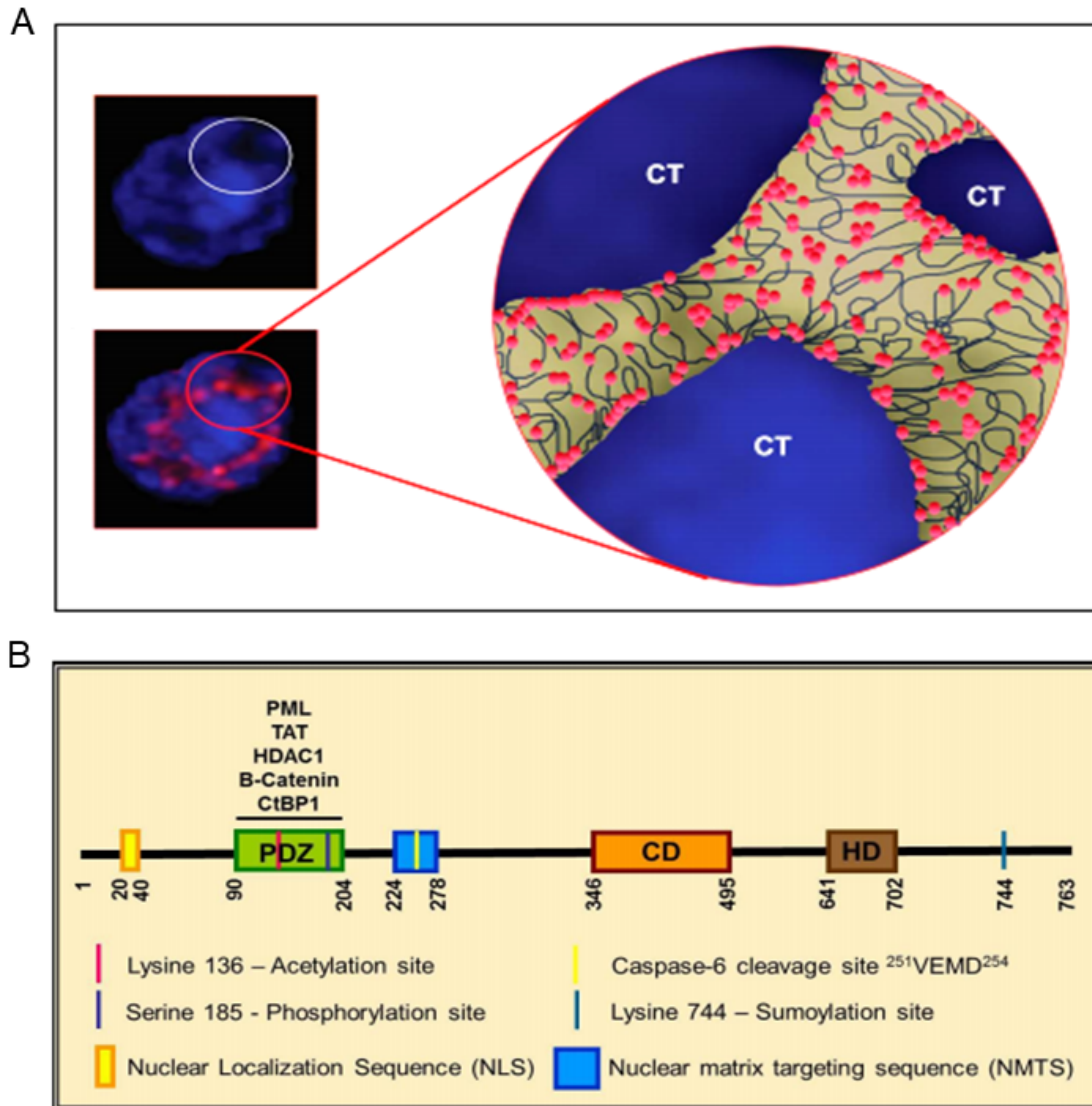


Figure 1.1.5. SATB1 is a T-cell enriched chromatin organizer. (A) The mouse thymocytes were subjected to the immunostaining with DAPI and anti-SATB1. The overlay of SATB1 and DAPI immunostaining suggests that SATB1 forms a cage like three-dimensional structure in thymocyte nuclei. SATB1 binds to its genomic targets in the transcriptionally poised regions and maintains the dynamic organization of chromatin in the thymocyte nuclei. (Reproduced from Galande *et al.,* 2007). (B) Schematic representation of the domain structure of SATB1. SATB1 is a 763 amino acid containing protein. Its N-terminal PDZ-like domain is important for interactions with the protein partners and also for the homo and hetero dimerization. Recently this domain has been shown to fold into an ubiquitin-like domain (ULD) (Wang et al., 2012; Wang *et al.,* 2014). The cut domain (CD) and the homeodomains (HD) are important for the DNA-binding activity of SATB1. (Reproduced from Burute, Gottimukkala and Galande, 2012).

Since its discovery, numerous functions have been ascribed to the SATB1 namely; genome organizer, chromatin modifier, transcription factor, and repressor (Galande *et al.*, 2007; Purbey *et al.*, 2009; Stephen *et al.*, 2017). SATB1 is not a tissue or cell type specific protein; it is highly expressed in the mature neurons. Specifically, in the postnatal cerebral cortex, SATB1 binds to the genomic loci of multiple immediate early genes, temporally regulating their expression, thus implicated in the synaptic plasticity (Balamotis *et al.*, 2012). It has been shown that SATB1 regulates the embryonic stem cell differentiation, early embryonic lineage segregation (Goolam and Zernicka-Goetz, 2017), also regulates the self-renewal of hematopoietic stem cells (Will *et al.*, 2013) and directs their differentiation towards the lymphoid lineages (Satoh *et al.*, 2013). SATB1 acts as an oncogenic regulator and its higher expression levels correlate with the tumor progression and poor prognosis (Han *et al.*, 2008; Mir *et al.*, 2016; Frömberg, Engeland and Aigner, 2018; Naik and Galande, 2019).

SATB1 is essential for the TCR mediated positive and negative selection of thymocytes, and for the establishment of immune tolerance (Kondo *et al.*, 2016). SATB1 is required for the lineage specification of Tregs in the thymus via the activation of Treg specific super-enhancers (Kitagawa *et al.*, 2017). SATB1 is important in specifying the thymic T-cell pool by activating the genes encoding the lineage specifying transcription factors (Figure 1.1.6) (Kakugawa *et al.*, 2017). SATB1 also plays a role in T-cell activation and differentiation (Pavan Kumar *et al.*, 2006; Notani *et al.*, 2010).

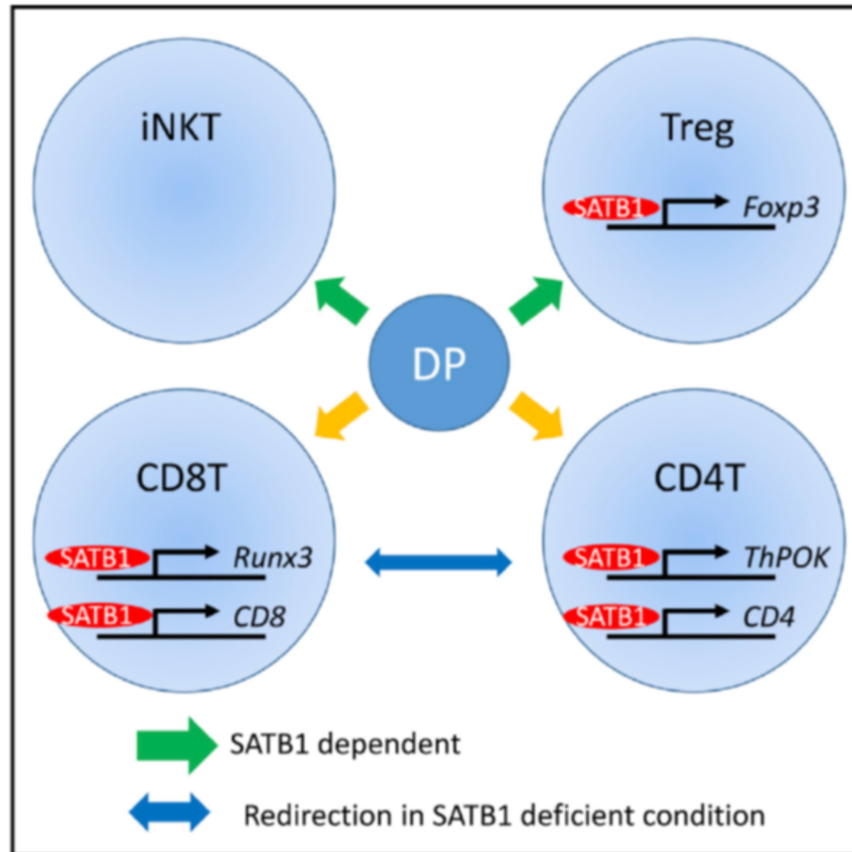


Figure 1.1.6. SATB1 regulates key lineage specifying genes of developing T-cells. The studies using SATB1 knockout mice revealed that the T-cell development was blocked at the DP stage, leading to further reduction in the number of SP thymocytes. During T-cell development, SATB1 binds to the regulatory elements of genes encoding the major transcription factors of developing thymocytes such as ThPOK in case of CD4SP, Runx3 in case of CD8SP, FOXP3 in case of Tregs, and also the genes encoding CD4, CD8 co-receptors, and regulates their expression. Loss of SATB1 leads to the partial redirection of MHC-I and MHC-II selected thymocytes towards their opposite lineage during T-cell development (Reproduced from Kakugawa *et al.*, 2017).

Considering the diversity of functions assigned to SATB1, studying its regulation is crucial for understanding the underlying molecular mechanisms. Post-translational modifications, such as phosphorylation and acetylation of SATB1, have contrasting effects on the transcriptional activity of SATB1 and also on the recruitment of its interaction partners (Pavan Kumar *et al.*, 2006; Notani *et al.*, 2010). SATB1 is regulated by the FOXP3 induced micro-RNAs miR-7 and miR-155, which specifically target the 3’

UTR of *Satb1* mRNA (Beyer *et al.*, 2011; McInnes *et al.*, 2012). Interestingly, in T-cells the levels of SATB1 protein is under the control of TCR signaling and exhibits the differential expression pattern in different stages of T-cell development (Gottimukkala *et al.*, 2016).

However, the molecular mechanism of transcriptional regulation of SATB1 is not yet studied. Studying the transcriptional regulation of SATB1 would be important to delineate its developmental stage specific expression and thus understanding its role in lineage specific manner during T-cell development.

1.1.8 Transcriptional regulation of gene expression

Transcription constitutes the major regulatory step of gene expression. Deregulation of transcription is associated with several developmental disorders and diseases such as cancer (Arrowsmith *et al.*, 2012; Dawson and Kouzarides, 2012). The transcription in eukaryotes is preceded by a sequence of events, which includes the chromatin de-condensation, histone modifications, and binding of basal transcription factors and RNA Pol II on to the special DNA cis regulatory elements called promoters (Figure 1.1.7. A). Promoters are short sequences of nearly 100 bp in length and are located near the transcription start site (TSS) of a gene (Roeder, 1996). The RNA Pol II is recruited at these promoters, which dictate the accurate position of the transcription initiation complex, and then assemble the pre-initiation complex (Roeder, 1996). The most commonly used core promoter elements by the majority of protein encoding genes are TATAA box, the first eukaryotic protein coding gene promoter identified, and the initiator (Inr) (Smale *et al.*, 1998; Smale and Kadonaga, 2003; Frith *et al.*, 2008). The TATA box is found to be located nearly 20-30 bp upstream to the TSS of a gene and acts as a binding site for the basal transcription factor TFIID (Mathis and Chambon, 1981). On the other hand, the Inr elements span the TSS of genes (Smale and Baltimore, 1989; Smale *et al.*, 1998). Early reports suggest that the majority of gene promoters contain the TATA box elements, however the recent reports revealed that only 5-7% of the eukaryotic promoters harbor the TATA box elements (Smale and Kadonaga, 2003; Frith *et al.*, 2008). The TATA box elements are present in the core promoters of tissue

specific genes, and whereas the Inr elements are present in the core promoters of housekeeping or the ubiquitously expressed genes (Gershenzon and Ioshikhes, 2005; Sandelin *et al.*, 2007; Lenhard, Sandelin and Carninci, 2012).

The cell type specific expression of a gene is governed by another cis-regulatory element called enhancer (Spitz and Furlong, 2012; Shlyueva, Stampfel and Stark, 2014). Enhancers are few hundred to thousand base pairs in length and consist of binding sites for key transcription factors. Enhancers can act on the promoters which are located in the long distance, by direct association with the promoters and allow the initiation of transcription at the promoter region (Figure 1.1.7. B).

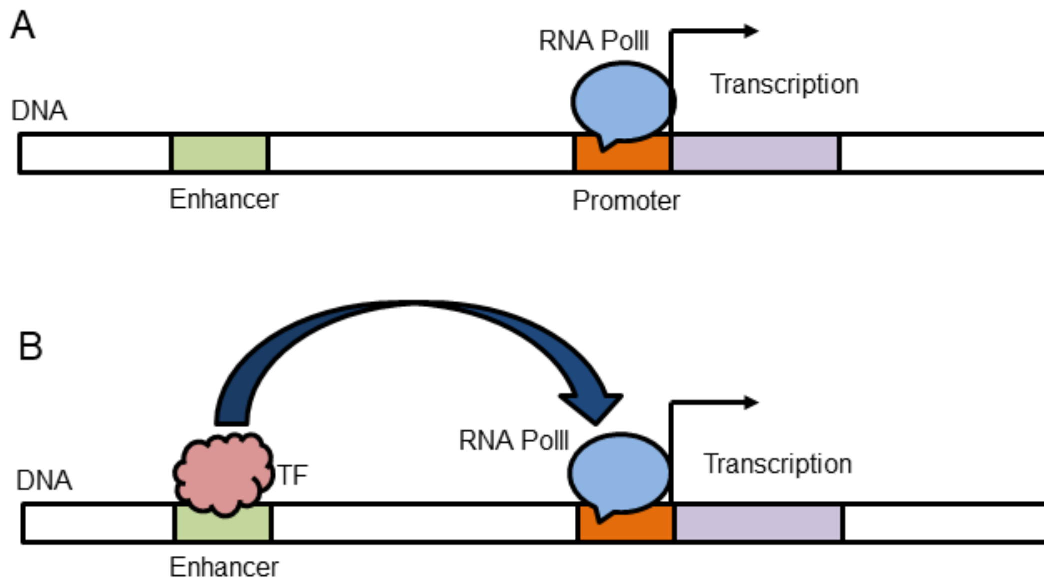


Figure 1.1.7. The role of DNA cis regulatory elements in transcription. (A) In eukaryotes, the transcription is mediated by DNA cis regulatory elements such as the promoter and the enhancer. The RNA PolII binds to the promoter region, and then initiates the transcription. Promoter is a DNA sequence of nearly 100 bp in length, and located near the TSS of a gene. The transcription factors binds to the promoter region of a gene and then allow the initiation of transcription by the RNA PolII. (B) The cell type specific expression of a gene is achieved by the binding of active transcription factors to another DNA cis regulatory element called enhancer. Upon binding of the specific transcription factors, enhancer interacts with the promoter region, which is located at a long distance and then form enhancer–promoter loop which facilitates transcription from the promoters of cell type specific genes.

Multiple genome-wide studies identified that the transcription initiation occurs not only from the core promoter regions around the TSS of a protein coding genes but also occurs at the enhancers (de Santa *et al.*, 2010; Kim *et al.*, 2010). The recent findings also suggest that the majority of mammalian genes show the divergent transcription, a phenomenon of transcription initiation in both the directions (Wu and Sharp, 2013). Thus the divergent transcription from the promoters and also from the enhancers acts as a source of the regulatory RNAs such as long non-coding RNA and enhancer RNA (Wu and Sharp, 2013). The specific motif enriched or the cell type or the developmental stage specific transcription factors (TF) plays a major role in the initiation of transcription from the promoters and enhancers.

In the eukaryotic nuclei, the DNA is wrapped around the histone octamer to form the nucleosome, which is a basic unit of the chromatin. The post-translational modifications of the histone tails play decisive roles towards the regulation and organization of the chromatin in the interphase of nuclei. Thus along with the transcription factors, the chromatin modifications also impart a major role in the transcription initiation from these regulatory elements and serve as a platform by which we can distinguish the transcription from the promoters and enhancers. The widely studied chromatin marks to distinguish the transcriptionally active promoter and enhancer are the post-translational modifications of histone tails (Figure 1.1.8). H3K4me3 mark is associated with the transcriptionally active promoters, whereas H3K27me3 is associated with the repressed promoters (Shilatifard, 2012). The enrichment of both H3K4me3 and H3K27me3 is associated with the poised transcription of a gene (Kouzarides, 2007; Voigt, Tee and Reinberg, 2013). The deposition of H3K36 methylation towards the 3' end of a gene, mediated by the Set2 histone methyltransferase is indicative of transcription elongation (Pokholok *et al.*, 2005; Rao *et al.*, 2005; Rando and Chang, 2009), whereas H3K79me3 is deposited towards the 5' end of actively transcribing genes (Krogan *et al.*, 2003). H3K4me1 and H3K27ac chromatin marks are typically known to associate with the transcriptionally active enhancer region (Heintzman *et al.*, 2007; Calo and Wysocka, 2013). The presence of H3K27 acetylation mark distinguishes the state of enhancers such as poised versus

active (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011; Zentner, Tesar and Scacheri, 2011). The poised state of enhancers shows the enrichment of H3K4me1 but lack H3K27ac occupancy and also show the enrichment of H3K27me3 mark (Rada-Iglesias *et al.*, 2011; Zentner, Tesar and Scacheri, 2011). H3K9me3 mark was also detected at the poised enhancer (Zentner, Tesar and Scacheri, 2011). Additionally, the ratio of H3K4/H3K4me1 is widely used to identify the promoters, which characteristically display higher enrichment of H3K4me3, whereas the enhancers associate with higher H3K4me1 (Figure 1.1.8). H3K9ac and H3K18ac marks were also detected at the putative enhancer regions (Ernst *et al.*, 2011; Zentner, Tesar and Scacheri, 2011). These epigenetic marks play an essential role in the transcription, as they are essential for the RNA Pol II recruitment, transcription initiation, and elongation.

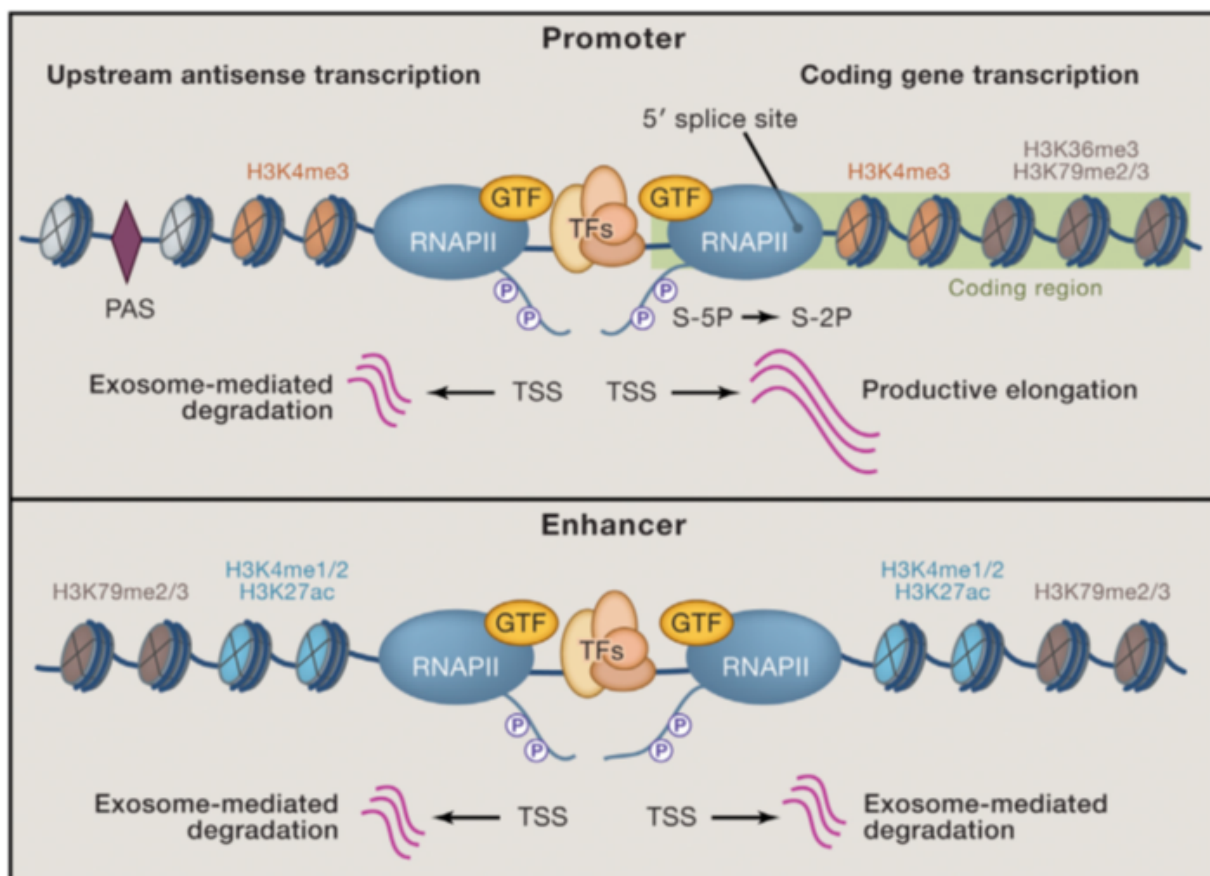


Figure 1.1.8. The profiling of histone marks at the transcriptionally active promoter and enhancer regions. In eukaryotes, transcription takes place at specific DNA elements called promoters, wherein the

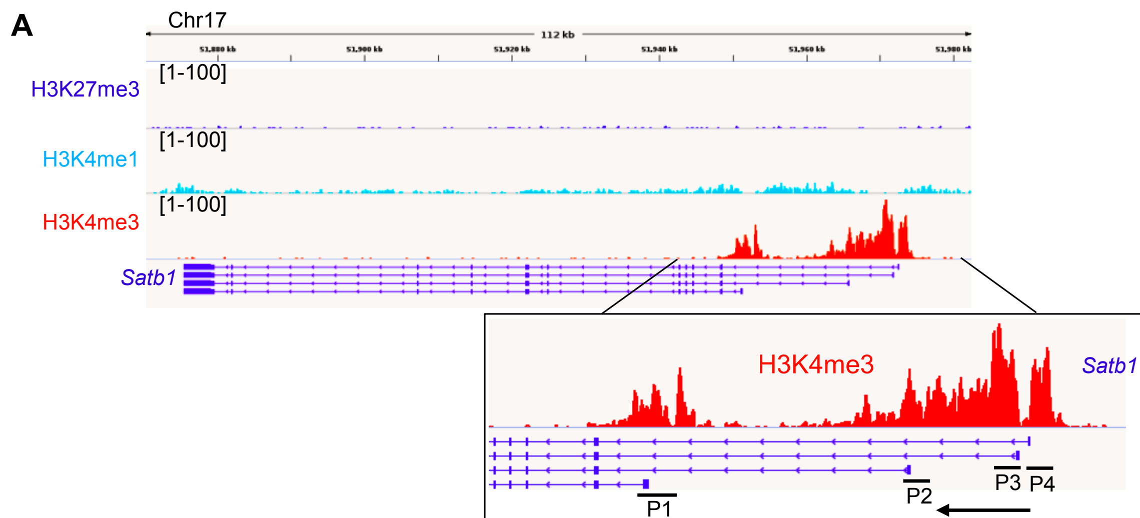
RNA PolII binds and initiates the transcription. The cell type specific gene expression is achieved by the interaction of enhancer with the promoter region and this process is mediated by specific transcription factors. The post-translational modification of histone tails have crucial importance in the positioning of these cis regulatory elements, thereby playing a role in the transcription. H3K4me3 mark is highly enriched at the promoter region, and H3K36me3, H3K79me3 marks are enriched on the gene body of transcriptionally active genes. The enhancer region is marked by H3K4me1/2 and H3K27ac. The transcriptionally active enhancers show the characteristic H3K79me2/me3 mark. In case of poised enhancers, the nucleosomes are still marked by H3K79me3 but they lack H3K27ac (Reproduced from Kim and Shiekhattar, 2015).

The protein complexes which play a role in writing, reading and erasing these chromatin marks have crucial importance in the regulation of gene expression in a context-dependent manner (Tarakhovsky, 2010). Therefore, perturbation in the expression of chromatin modifiers affects the global gene expression and is therefore associated with several diseases and developmental defects. In case of T-cell development, the expression levels of T-cell enriched global chromatin organizer SATB1 play a critical role. As mentioned above, disturbance in the expression of SATB1 leads to abortive T-cell development. Thus, investigating the molecular mechanisms that regulate the transcription of SATB1 would be crucial to understand its developmental stage specific expression and its role. Therefore, the current study focuses to first delineate and characterize the transcriptional regulation of SATB1.

1.2 Results

1.2.1 Identification of putative alternative promoters of *Satb1* gene

To understand the transcriptional regulation of *Satb*1 gene during the development of T-cells, we profiled the histone modification marks such as H3K4me3, H3K4me1 and H3K27me3 at *Satb1* gene locus (Figure 1.2.1 A) using the publicly available ChIP-seq datasets from mouse DP thymocytes (Wei *et al*., 2011). Analysis of these histone marks revealed that multiple regions upstream of *Satb*1 TSS were enriched with the H3K4me3 mark, with no H3K4me1 (Figure 1.2.1 A); a characteristic feature of a promoter (Heintzman *et al*., 2007; Djebali *et al*., 2012). ChIP-seq analyses were overlaid with UCSC annotated ref-seq (NCBI36) for mouse *Satb*1 (Figure 1.2.1 A) that predicted the four different *Satb1* transcript variants with distinct first exons. Interestingly, unlike S*atb*1, another SATB family transcription factor S*atb*2 has a single region upstream of its TSS with the occupancy of H3K4me3, with no H3K4me1 (Figure 1.2.1 B), suggesting that the presence of alternative promoters like regulatory elements is exclusive to S*atb*1.



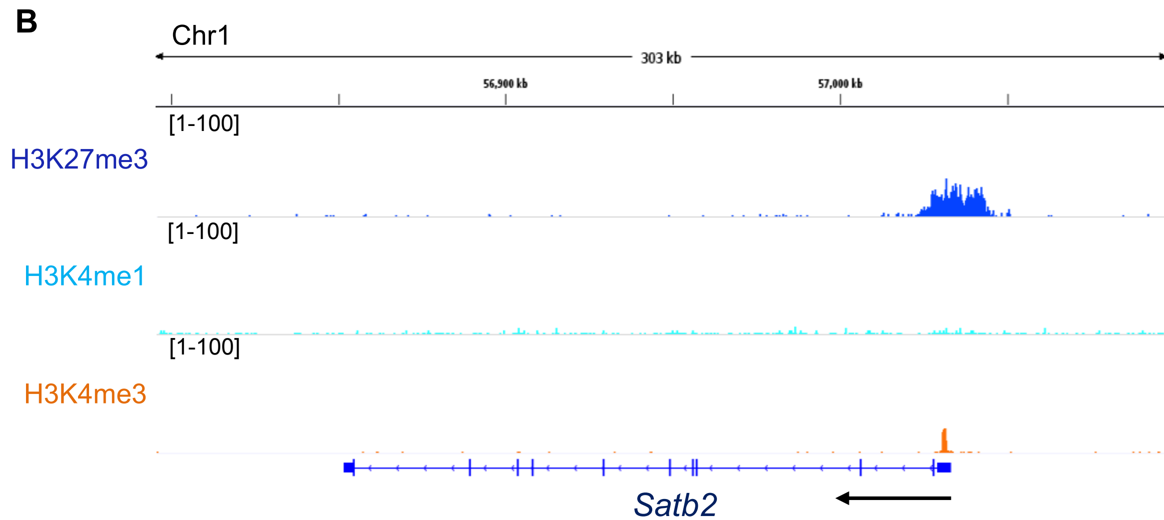
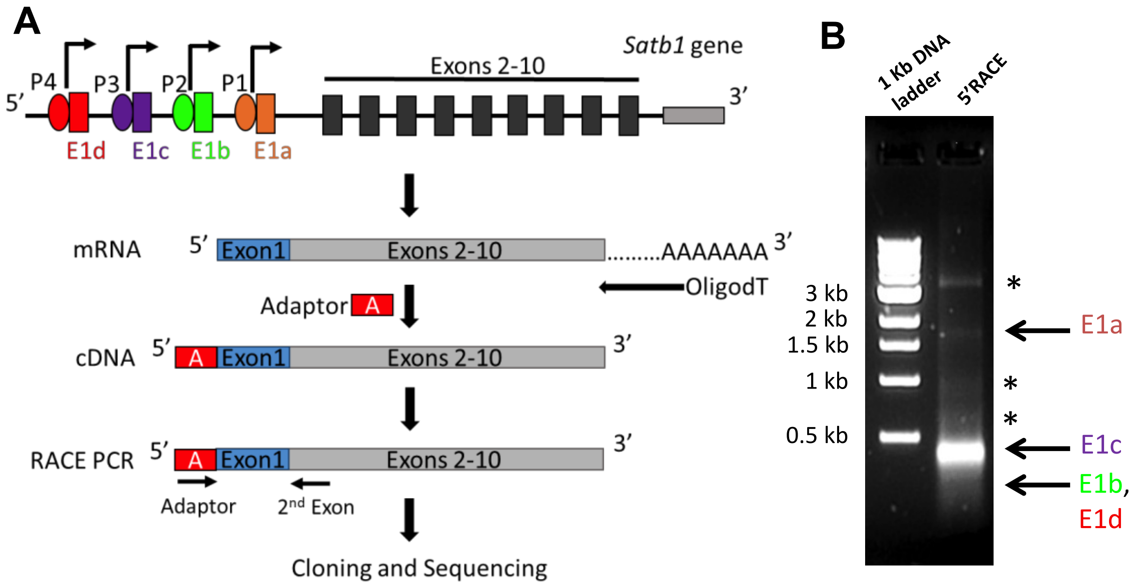


Figure 1.2.1. The ChIP-seq analysis of histone marks at *Satb1* and *Satb2* gene loci in the mouse thymocytes. (A) Publicly available ChIP-seq data sets of histone modifications such as H3K4me3, H3K4me1 and H3K27me3 performed in the double positive (DP) thymocytes were analyzed and used for the histone mark profiling at *Satb1* gene locus. The ChIP-seq peaks were mapped to the Ref-seq for mouse *Satb1* from the UCSC genome browser. The identified putative alternative promoter regions of *Satb1* were labeled as P1, P2, P3 and P4. The black arrow indicates the direction of the transcription at *Satb1* gene locus. Inset shows the magnified image of the *Satb1* alternative promoters. (B). ChIP-seq analysis of histone marks such as H3K4me3, H3K4me1 and H3K27me3 were analyzed at the *Satb2* gene locus in the DP thymocytes. The ChIP-seq analyses of histone marks were mapped to the ref-seq for mouse *Satb2* from the UCSC genome browser. The black arrow indicates the direction of transcription at the *Satb2* gene locus.

1.2.2 Identification of *Satb1* transcript variants in mouse thymocytes by 5' RACE analysis

To evaluate the usage and the transcriptional activity of predicted *Satb*1 alternative promoters, we performed 5' Rapid Amplification of cDNA Ends (5' RACE) (Frohman, 1993) using the RNA extracted from mouse thymocytes. Briefly, first strand cDNA synthesis was performed with the total RNA isolated from the mouse thymocytes using a modified oligo (dT) primer (SMARTer RACE kit, Clontech), followed by RACE-PCR using the universal forward primer provided in the kit, and the reverse primer spanning the exon-2 of *Satb*1 gene as shown in the schematic (Figure 1.2.2 A). The amplified sequences were cloned and sequenced using Sanger sequencing (Figure 1.2.2 B). The

results of 5’ RACE led to the identification of different S*atb*1 transcript variants with the alternative first exons, named E1a, E1b, E1c and E1d (Figure 1.2.2 C). The E1a exon is located approximately 2.48 Kb upstream of the E2 exon and similarly, E1b, E1c and E1d exons are located 17.14 Kb, 23.18 Kb and 23.875 Kb upstream, respectively from the E2 exon. Hence these alternative first exons are alternatively spliced to the E2 exon that resulting in the generation of *Satb1* transcript variants with different 5’ end sequences (Figure 1.2.2 C).



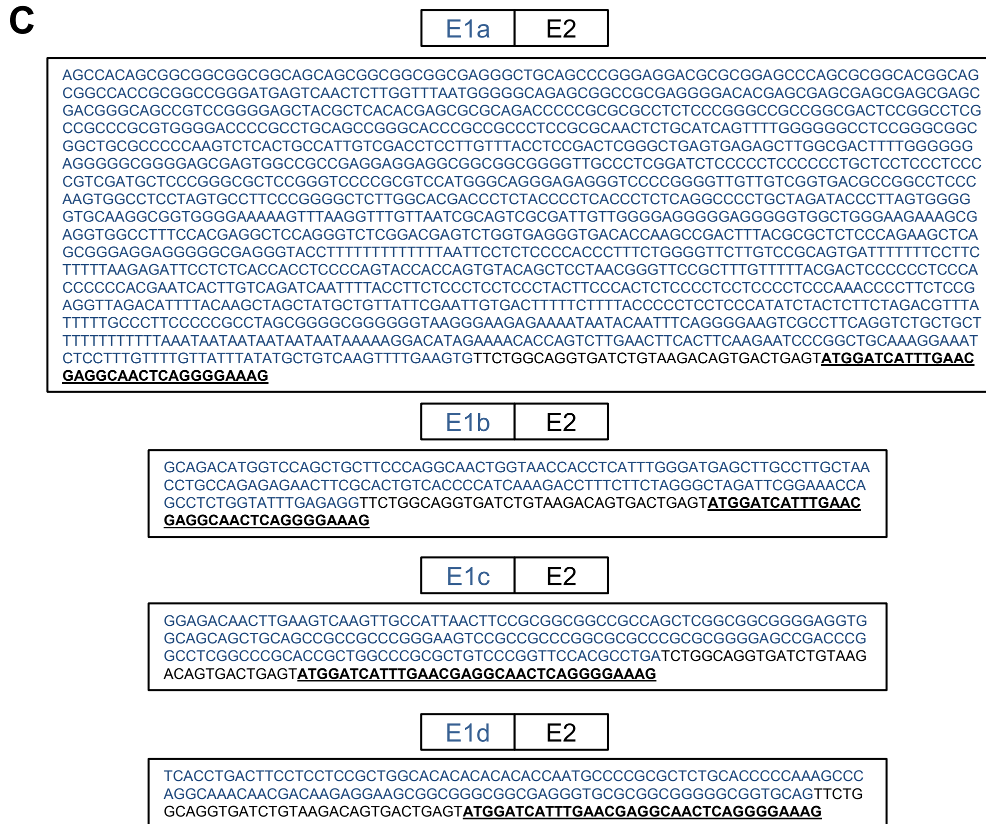
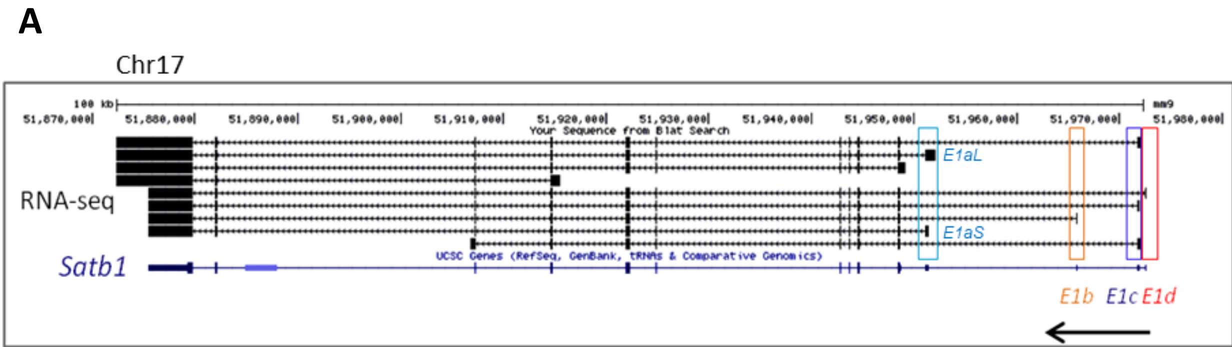


Figure 1.2.2. 5' RACE analysis of the mouse *Satb1* transcript variants. A) Schematic depicting the strategy of 5' RACE analysis of the mouse *Satb1.* In the schematic, the position of *Satb1* alternative promoters P1, P2, P3 and P4 were indicated (The genomic distances are not drawn to scale). To identify the transcription from these alternative promoters and thus, the generation of *Satb1* transcript variants with the alternative promoter specific exons, the total RNA was used for cDNA synthesis. Once the cDNA

synthesis is performed, an adapter (oligo DNA) is added to the 5’ end of cDNA. The adapter specific forward primer and the *Satb1* exon-2 specific reverse primer were used for the amplification of *Satb1* alternative transcripts with the alternative first exons. (B) Single cell suspension of thymocytes were prepared from 3 week old mice and used for RNA extraction using the Trizol method. 5’ RACE analyses were performed using the total RNA from mouse thymocytes. The RACE-PCR products were resolved on 1% agarose gel, purified and cloned into the pRACE vector, and then sequenced. (C) The identified distinct first exon sequences of *Satb*1 transcript variants such as E1a, E1b, E1c, and E1d are shown. The alternative 5’ UTR sequences of *Satb1* transcript variants, which constitutes the first exon, are indicated by sequences in blue. SATB1 protein coding sequence begins in the second exon, which is indicated by the underlined black colored sequence.

1.2.3 Detection of *Satb1* transcript variants by RNA-seq analysis of mouse thymocytes

To confirm the existence of alternative first exons of *Satb1*, we analyzed the publicly available data of RNA-seq performed in the mouse thymocytes (DP and $CD4^{+}$SP) (Hu *et al.*, 2013) and confirmed that indeed *Satb1* gene shows the expression of alternative transcript variants in mouse thymocytes (Figure 1.2.3). The RNA-seq analysis identified the *Satb1* transcript variants with E1a, E1b, E1c, and E1d exons (Figure 1.2.3.A). The E1a containing transcript variant has shown two isoforms, one with the short (E1aS) and the other with the long (E1aL) alternative first exons, the latter was detected in the 5’ RACE analysis. The expression of *Satb1* transcript variants in the DP thymocytes were indicated as FPKM values (Figure 1.2.3.B). All together both 5’ RACE and RNA-seq analyses identified the *satb1* transcript variants with the following alternative first exons: E1aS, E1aL, E1b, E1c, and E1d.



B

Transcripts	FPKM
E1a L	26.9731
E1a S	45.8507
E1b	116.115
E1c	619.175
E1d	194.549

Figure 1.2.3. The RNA-seq analysis of mouse *Satb1* transcript variants. (A) The RNA-seq analysis was performed by using the publicly available data sets of mouse thymocyte subpopulations - CD3^{lo}DP and CD4SP. The RNA-seq reads were mapped to the mouse genome mm9 using Bowtie2 and Tophat2, details of which are provided in the 'Methods' section. The identified *Satb1* transcript variants were shown and named as E1aL, E1aS, E1b, E1c, and E1d. The arrow indicates the directionality of the transcription at *Satb1* gene locus. (B) The expression values of *Satb1* transcript variants in DP thymocytes were indicated as FPKM values.

1.2.4 The alternative first exons of *Satb1* transcript variants act as 5' UTR elements

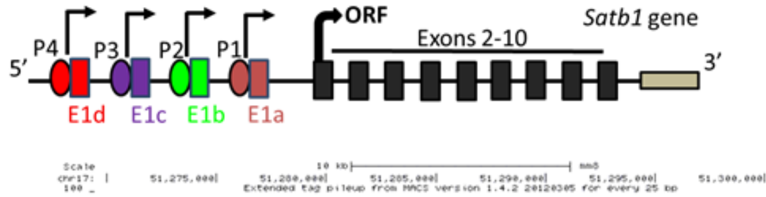
SATB1 is a 764 amino acid containing protein, which is encoded by the *Satb1* gene located on the chromosome 17 in mouse and chromosome 3 in human. We found that the alternative first exons of *Satb1* mRNA variants do not contribute in encoding the SATB1 protein and the protein coding sequence (CDS) of *Satb*1 begins in the exon-2 (Figure 1.2.4.A). Therefore the alternative first exon sequence along with a part of second exon sequence upstream to the protein coding mRNA sequence of *Satb1*, act as 5' untranslatable region (5' UTR) of the *Satb1* mRNA. Finally, mapping the sequences of *Satb1* alternative first exons to the mouse genome revealed that their TSS reside within the above identified distinct H3K4me3 ChIP-seq peaks (Figure 1.2.4.B), indicating that distinct alternative promoters (hereafter named P1, P2, P3 and P4) were used for the generation of *Satb*1 transcript variants with alternative 5' UTRs, named P1 (E1a) - both P1S (E1aS) and P1L (E1aL), P2 (E1b), P3 (E1c), and P4 (E1d), respectively (Figure 1.2.4.C).

A **Amino acid sequence of SATB1**

MDHLNEATQGKEHSEMSNNVSDPKGPPAKIARLEQNGSPLGRGRLGSTGGKMQGVPLKHSGHLMKTNLRKGTMLPVFCVVEHYENAIEYDCKEEHAEFVLVRKDMLFN
QLIEMALLSLGYSHSSAAQAKGLIQVGKWNPVPLSYVTDAPDATVADMLQDVYHVVTLKIQLHSCPKLEDLPPEQWSHTTVRNALKDLLKDMNQSSLAKECPLSQSMISSI
VNSTYYANVSAAKCQEFGRWYKHFKKTKDMMVEMDSLSELSQQGANHVNFGQQPVPGNTAEQPPSPAQLSHGSQPSVRTPLPNLHPGLVSTPISPQLVNQQLVMAQL
LNQQYAVNRLLAQQSLNQQYLNHPPPVSRSMNKPLEQQVSTNTEVSSEIYQWVRDELKRAGISQAVFARVAFNRTQGLLSEILRKEEDPKTASQSLLVNLRAMQNFLQLPE
AERDRIYQDERERSLNAASAMGPAPLLSTPPSRPPQVKTATLATERNGKPENNTMNINASIYDEIQQEMKRAKVSQALFAKVAATKSQGWLCELLRWKEDPSPENRTLWE
NLSMIRRFLSLPQPERDAIYEQESNAVHHHGDRPPHIIHVPAEQIQQQQQQQQQQQQQQQPPPPPPQPQPQPQAGPRLPPRQPTVASSAESDEENRQKTRPRTKISVEA
LGILQSFIQDVGLYPDEEAIQTLSAQLDLPKYTIIKFFQNQRYYLKHHGKLKDNSGLEVDVAEYKDEELLKDLEESVQDKNANTLFSVKLEEELSVEGSTDVNADLKD

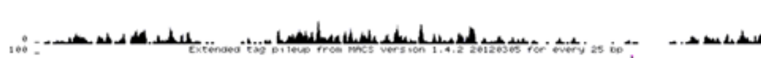
ORF of *Satb1*

ATG GAT CAT TTG AAC GAG GCA ACT CAG GGG AAA GAA CAT TCA GAA ATG TCT AAC AAT GTG AGT GAT CCG AAG GGT CCA CCC GCC AAG ATT GCC
CGC CTG GAG CAG AAC GGG AGC CCT CTA GGA AGA GGA AGG CTT GGG AGC ACA……………….TAA



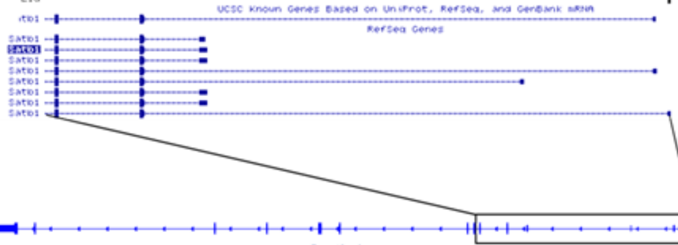
B

H3K4me1_DP



H3K4me3_DP

5′RACE



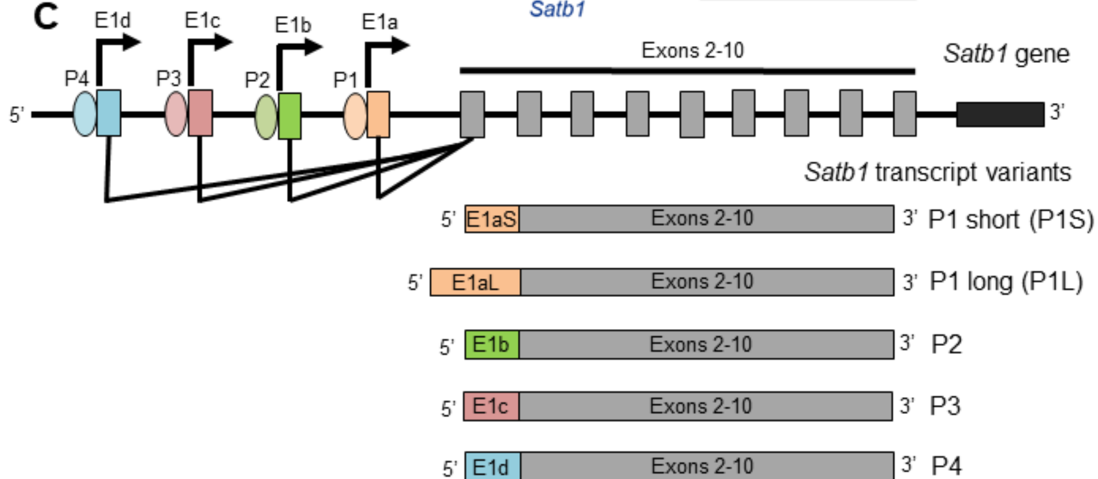


Figure 1.2.4. The alternative first exons of *Satb1* transcript variants act as 5'UTR elements. (A) The primary amino acid sequence of SATB1 protein and the open reading frame (ORF) of SATB1 coding sequence is shown. The coding sequence or ORF, which encodes the SATB1 protein begins in the exon-

2 but not in any of the alternative first exons as indicated in the diagram. (B) The sequences of *Satb1* transcript variants obtained from the 5' RACE analysis of mouse thymocytes, were overlaid with the ChIP-seq peaks of H3K4me3 and H3K4me1 modifications in the mouse thymocytes, and were then mapped to the ref-seq of mouse *Satb1* from the UCSC genome browser. The sequences of *Satb1* transcript variants from 5' RACE analysis are indicated as E1a, E1b, E1c, and E1d. (C) Schematic depiction of the genomic locations of the alternative promoters such as P1, P2, P3, and P4 at *Satb1* gene locus. The *Satb1* alternative promoter usage leads to alternative splicing of *Satb1* mRNA, resulting in the generation of *Satb1* transcript variants with alternative first exons such as E1aS, E1aL, E1b, E1c, and E1d. (Genomic distances are not drawn to scale).

1.2.5 The cell type specific expression of *Satb1* transcript variants during T-cell development

To explore the importance of expression of various *Satb1* transcript variants*,* we first characterized their expression pattern during T-cell development (Figure 1.2.5.A). We performed the quantitative RT-PCR analysis of *Satb1* transcript variants in various developmental stages of thymocytes such as $CD4^-CD8^-$ double negative (DN), $CD4^+CD8^+$ double positive (DP), total $CD4^-CD8^+$SP (CD8SP), total $CD4^+CD8^-$SP (CD4SP), $CD4^+CD24^+$SP (immature CD4SP), and $CD4^+CD24^-$SP (mature CD4SP) thymocytes.

The expression analyses revealed the differential expression pattern of *Satb*1 transcript variants in a cell type specific manner during T-cell development. DP and immature CD4SP thymocytes have shown the higher expression levels of *Satb1* mRNA compared to other thymocyte developmental stages such as DN, mature CD4SP, and CD8SP (Figure 1.2.5.B and 1.2.5.C). We then further evaluated the expression pattern of *Satb*1 transcript variants in these developmental stages. The expression of P1 and P4 transcript variants was higher in the DP thymocytes compared to other cell types (Figure 1.2.5.B). The P3 transcript variant expression was higher in the DP as well as in the immature CDSP compared to DN, mature CD4SP, and CD8SP thymocytes (Figure 1.2.5.B and 1.2.5.C). Interestingly unlike other transcript variants, The P2 transcript variant has shown distinct expression pattern, specific to the $CD4^+$ T lineage. During T-cell development, the expression of P2 transcript variant begins during the transition

from DN to DP and their expression levels further increased in the CD4 SP thymocytes, but reduces in the CD8 SP (Figure 1.2.5.B and 1.2.5.C). Specifically in the CD4 SP thymocytes, P2 is highly expressed in the immature CD4SP compared to the mature CD4SP (Figure 1.2.5.B).

Altogether, these analyses indicated that the *Satb1* mRNA of DP stage majorly consists of the higher expression levels of P1, P3 and P4 variants, but less of P2 variant compared to that of CD4SP. The P2 was predominantly expressed in the immature CD4SP thymocytes along with the P3 transcript variant. These findings revealed the cell type specific expression of a combination of *Satb*1 transcript variants during T-cell development.

Next, we analyzed the RNA-seq performed in $CD3^{lo}$ DP and CD4SP thymocytes (Hu *et al.*, 2013) and the differential splicing of *Satb1* alternative first exons in these two thymocyte subpopulations are shown by Sashimi plot analysis. The expression levels of total *Satb1* and its variants are indicated by FPKM values (Figure 1.2.5.D). Altogether, the expression analysis along with RNA-seq analysis (Figure 1.2.5.D) revealed that P3 is the predominant variant among all *Satb1* transcript variants and was expressed constitutively, in combination with the cell type specific transcript variants in the various stages of thymocyte development.

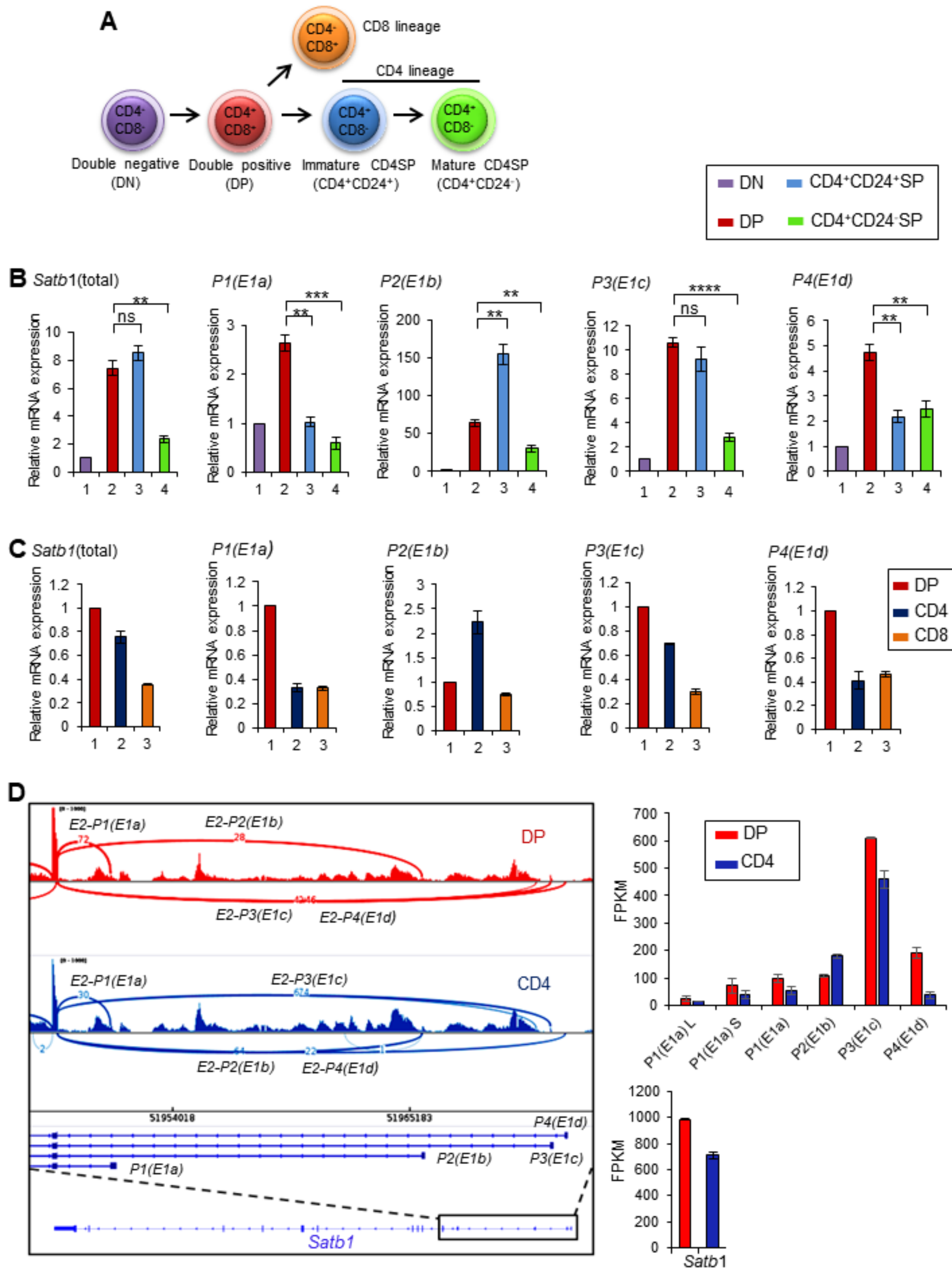


Figure 1.2.5. Cell type specific expression of the *Satb*1 transcript variants during thymocyte development. (A) Schematic diagram depicting the various developmental stages of thymocytes. The

initial developmental stages are $CD4^-CD8^-$ double negative (DN), which subsequently develops into $CD4^+CD8^+$ double positive (DP thymocytes. DP cells then develop into either $CD8^+$ T lineage or $CD4^+$ T lineage. (B and C) Three week old C57/BL6 mice were used for isolation of thymus which was further used for preparation of single cell suspension. The thymocytes were subjected to the surface staining with anti-CD4, anti-CD8, and anti-CD24 antibodies. The $CD4^-CD8^-$ DN, $CD4^+CD8^+$DP, $CD4^+CD24^+$ immature CD4SP, $CD4^+CD24^-$ mature CD4SP, $CD4^+CD8^-$ total CD4SP and $CD4^-CD8^+$ total CD8SP thymocyte subpopulations were FACS sorted. The FACS sorted thymocyte sub populations were used for the RNA isolation and cDNA synthesis. The quantitative RT-PCR analyses were performed to measure the expression levels of total *Satb*1 mRNA and the *Satb*1 transcript variants in mentioned developmental stages of thymocytes. The presented data is from three independent experiments and was shown as means of ± SEM. p values were calculated using the two-tailed Student's t-test. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. (D) The RNA-seq analysis of $CD3^{lo}$DP and CD4SP thymocytes were performed using the publicly available datasets. The Sashimi plot analysis represents the alternative splicing events of the *Satb1* alternative first exons in the DP and CD4SP thymocytes. The RNA-seq based gene expression values (FPKM) of *Satb1* expression in DP and CD4SP thymocytes are presented.

1.2.6 The SATB1 protein levels are higher in immature CD4SP than DP - A discrepancy between the expression levels of SATB1 mRNA and protein

Since we observed the presence of distinct combination of *Satb1* transcript variants in DP and CD4SP thymocytes, simultaneously, we quantitated the SATB1 protein levels in these thymocyte developmental stages. We observed that SATB1 protein levels were higher in the immature CD4SP thymocytes compared to DP and mature CD4SP thymocytes (Figure 1.2.6), confirming the previous observation from our group (Gottimukkala *et al.*, 2016). During the development from DP to immature CD4SP, interestingly, we observed that the *Satb1* transcript levels barely differ, but a stark difference was observed in the SATB1 protein levels between these two populations. We hypothesized that the combinatorial expression of the *Satb*1 transcript variants in DP and immature CD4SP thymocytes might play a role in the differential SATB1 protein expression in these developmental stages of thymocytes. As shown in the Figure 1.2.5 B, compared to the immature CD4SP the majority of *Satb1* transcripts present in the DP consists of all *Satb1* transcript variants but less of P2 variant, which is only highly expressed in the immature CD4SP. We then asked whether the combinatorial expression of transcript variants plays any role in regulating the levels of SATB1 protein.

Before further dissecting the role of transcript variants in maintaining the different protein levels, we also checked the stability of SATB1 protein in DP thymocytes. As DP cells exhibit higher levels of expression of *Satb*1 mRNA but significantly lower levels of SATB1 protein compared to the protein levels of immature CD4SP, we wished to ascertain whether the phenomenon we see is due to differential translatability of these transcript variants or due to the active degradation of SATB1 protein in DP thymocytes.

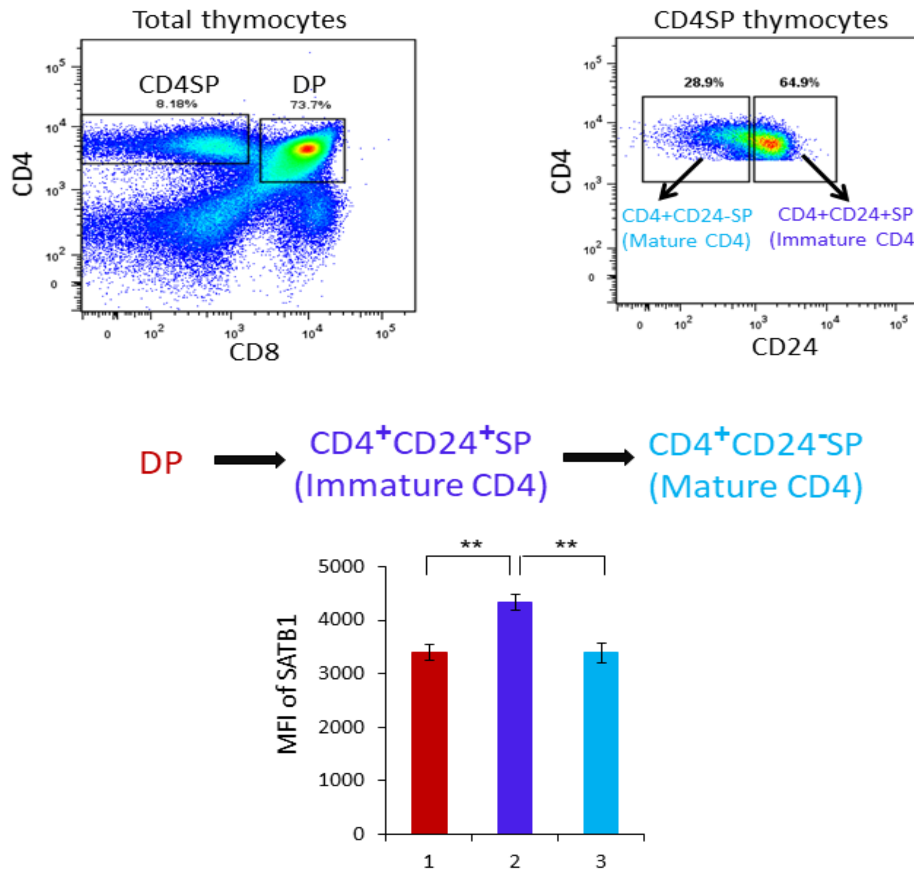


Figure 1.2.6. The expression profiling of SATB1 protein levels during the CD4^{+} T lineage differentiation. Three week old C57/BL6 mice were used for the isolation of thymus, which was used for the preparation of single cell suspension of thymocytes. The thymocytes were subjected to surface staining with fluorophore conjugated anti-CD4 and anti-CD8. After the surface staining, thymocytes were permeabilized and subjected to the intracellular staining of SATB1 with conjugated anti-SATB1 antibody by using the FOXP3/Transcription factor intracellular staining kit as described in 'Materials and Methods'. The FACS analysis were performed using BD FACS canto II (BD Biosciences). Expression of SATB1 protein in CD4^{+}CD8^{+} DP, immature CD4^{+}CD24^{+}SP and mature CD4^{+}CD24^{-}SP thymocytes was analyzed using flow cytometry and the mean florescence intensities (MFI) of SATB1 in these three populations were shown.

.

1.2.7 SATB1 protein is stable in double positive (DP) thymocytes –cycloheximide and MG132 chase assays

To evaluate the reason for significantly lower levels of SATB1 protein in DP thymocytes compared to the immature, we performed the MG132 and cycloheximide chase assay in DP thymocytes. DP thymocytes were FACS sorted by surface staining the thymocytes with fluorophore conjugated anti-CD4 and anti-CD8. We treated the sorted DP thymocytes with MG132, a proteasome inhibitor and cycloheximide, a translation inhibitor, for 4 hrs. We observed no change in the SATB1 protein levels upon MG132 treatment in the presence or absence of cycloheximide (Figure 1.2.7 A). However, only slight increase was observed in the SATB1 protein levels in total thymocytes treated with MG132 and cycloheximide (Figure 1.2.7 B) but not in the DP population. Our results confirm that the low SATB1 protein levels in DP thymocytes compared to the immature CD4SP thymocytes are not due to the active degradation of SATB1 protein in DP thymocytes.

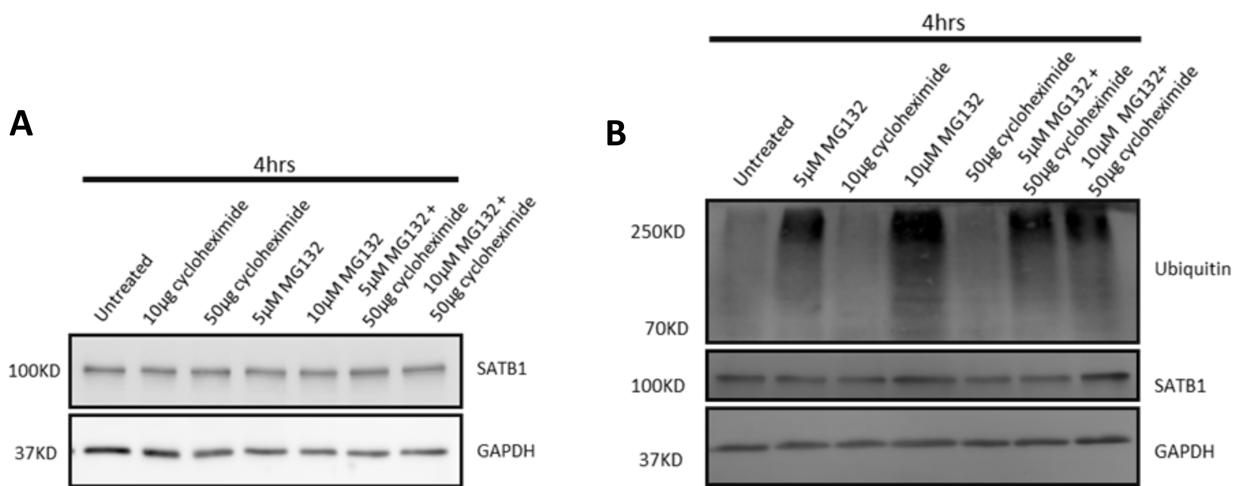


Figure 1.2.7. Cycloheximide and MG132 chase assay in DP thymocytes. (A) Three week old mice were used for the preparation of single cell suspension of thymocytes. The thymocytes were surface stained with fluorophore conjugated anti-CD4 and anti-CD8 antibodies, and FACS sorting of $CD4^{+}CD8^{+}$ DP thymocytes was performed. DP thymocytes were cultured in the presence or absence of the indicated concentrations of MG132 and cycloheximide for 4 hrs. The cells were harvested and used for

immunoblotting with anti-GAPDH and anti-SATB1 antibodies (N=3). (B) Three week old mice were used for the preparation of single cell suspension of thymocytes. Total thymocytes were cultured in the presence or absence of the indicated concentrations of MG132 and cycloheximide for 4 hrs. After incubation, cells were harvested and subjected to the immunoblotting analysis by using anti-SATB1, anti-GAPDH, and anti-ubiquitin antibodies.

1.2.8 Monitoring translation efficiency of Satb1 5' UTRs by polysome profiling of thymocytes

To further evaluate the functional significance of *Satb1* transcript variants in regulating the levels of SATB1 protein, we performed polysome profiling. The polysome profiling of total thymocytes were performed to monitor the presence of *Satb1* mRNA in the ribosome free and ribosome bound states (Figure 1.2.8 A) and thus to confirm whether all the *Satb1* transcript variants can be translatable or not. Thymii from three old week mice were snap frozen and lysed in the hypotonic buffer containing cycloheximide. The nuclei were separated and the cytoplasmic fraction was loaded onto a 10-50% sucrose gradient and subjected to ultracentrifugation. After separation, the fractions were collected and used for RNA extraction and for the gene expression analysis. Actin mRNA was used as a positive control which shows its abundance only in the highly translatable or polysome fractions (Figure 1.2.8 B), indicating its high translatability. Interestingly, we found that all the *Satb1* transcript variants were present in the subunit, monosome and polysome bound states (Figure 1.2.8 B). However, we did not observe the exclusive association of any of the *Satb1* transcript variants in the non-translatable fractions, indicating that all the *Satb1* transcript variants are translatable. As shown in the Figure 1.2.8 B, the P1 and P2 transcript variants of *Satb1* were present comparatively more in the translatable fractions than the non-translatable fractions. However, P3 and P4 transcript variants of *Satb1* were observed both in the translatable fractions as well as non-translatable fractions (Figure 1.2.8 B). The ratio of polysome to monosome bound states (mentioned above each peak) suggests that the translatability of the P1 and P2 transcript variants is comparatively higher than that of the P3 and P4 variants of *Satb1* (Figure 1.2.8 B).

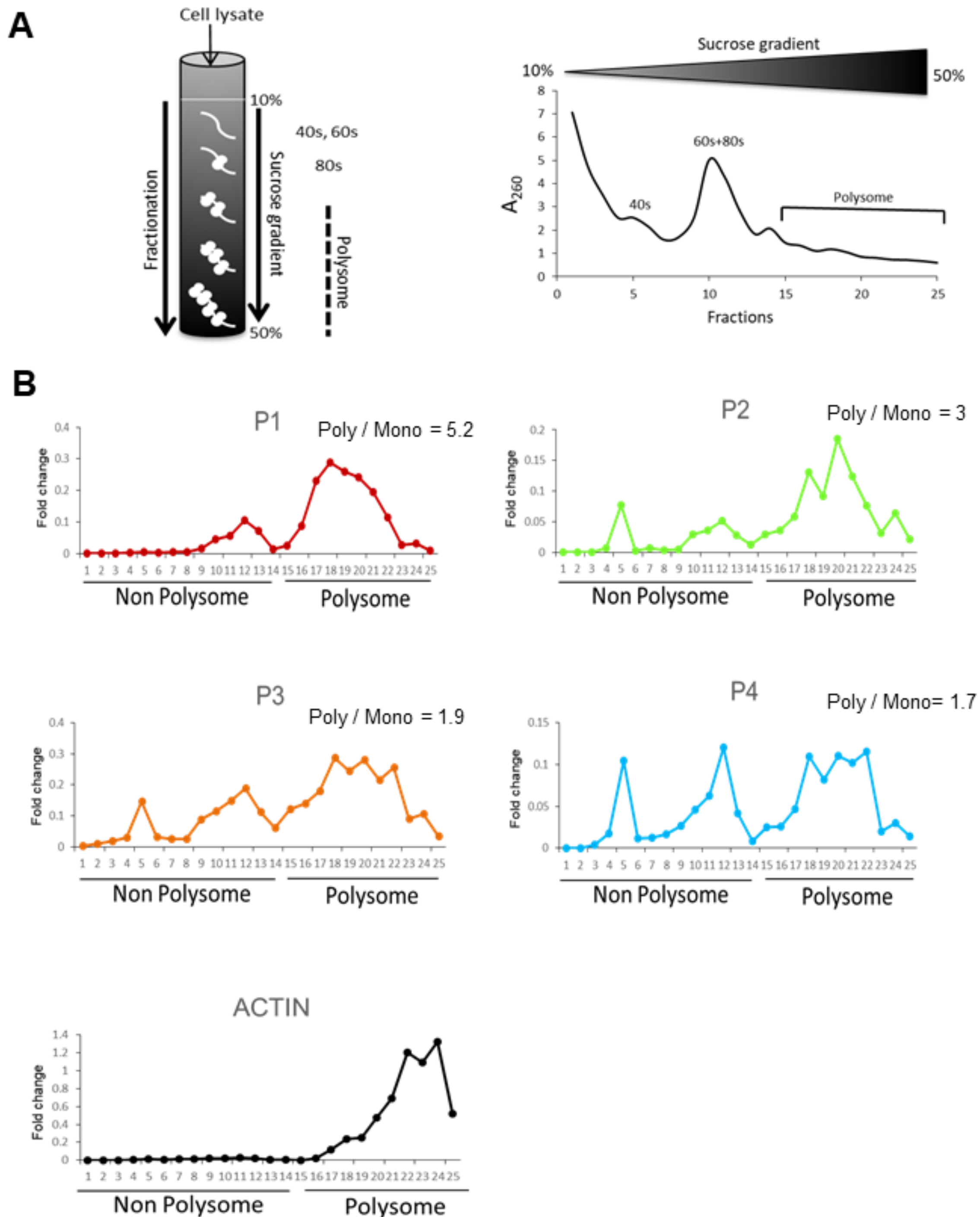


Figure 1.2.8. Polysome profiling in total thymocytes. (A) Three week old C57/BL6 mice were used for the isolation of thymus. The thymii were snap frozen as soon as they were isolated from the mice. Thymii were pulverized under the liquid nitrogen conditions and the lysate was prepared by using a hypotonic lysis buffer (see methods) supplemented with 100 µg/mL cycloheximide. The lysate was loaded onto the 10-50% sucrose gradient and subjected to the ultracentrifugation. The fractions were collected as

indicated in the figure and OD at 260 nm was measured from the collected fractions. (B) The fractions were used for the extraction of RNA by using Trizol method. The isolated RNAs from the collected fractions were used for the cDNA synthesis which was used for the quantitative RT-PCR analysis using Sybr green master mix. The qRT-PCR analyses of *Satb1* transcript variants such as P1, P2, P3 and P4 along with actin were performed. The levels of these transcripts present in the non-polysome fraction which includes both subunits and monosome fractions, and the polysome fractions were plotted.

1.2.9 *Satb*1 transcript variants differ in the translation efficiencies *in vitro*

To further probe into the functional significance of the *Satb1* alternative transcripts, we performed *in vitro* transcription and translation assays. The *Satb*1 alternative 5' UTRs were cloned upstream of firefly luciferase coding sequence in the pGL3 basic vector (Promega), and used for *in vitro* transcription (Figure 1.2.9 A). Equimolar concentrations of the *in vitro* transcribed RNAs normalized to their base pair lengths of respective UTRs were used for *in vitro* translation. The translated product was subjected to luciferase activity measurement. *In vitro* luciferase assay results revealed that the P2 and P1S 5' UTR sequences were most efficiently translatable (Figure 1.2.9 B). The luciferase activity was reduced in case of P1L, P3 and P4 5' UTR sequences, indicating that these *Satb1* 5' UTRs have less translation capacity *in vitro* (Figure 1.2.9 B). We then performed the secondary structure analysis of these 5' UTR sequences by using the mfold web server (Zuker, 2003), and found that the P2 and P4 5' UTR sequences form the less stable secondary structures followed by P1S. Whereas, the P1L and P3 5 'UTRs were shown to form the more stable secondary structures, as indicated by their higher -ΔG values (Figure 1.2.9 C). This might be one of the reasons for the poor translatability of P1L and P3 5'UTRs *in vitro*. Although, the P4 5' UTR sequence forms less stable secondary structure, we observed the low translatability of P4 5' UTR *in vitro,* indicating the regulatory elements other than the secondary structures might affect the translatability of P4 5' UTR. Hence the majority of *Satb1* transcript variants present in the DP thymocytes have 5' UTR sequences which can affect the translation rate, compared to the 5 'UTRs of *Satb1* transcript variants present in the immature CD4SP thymocytes, presumably resulting in the lower SATB1 protein levels in DP compared to immature CD4SP.

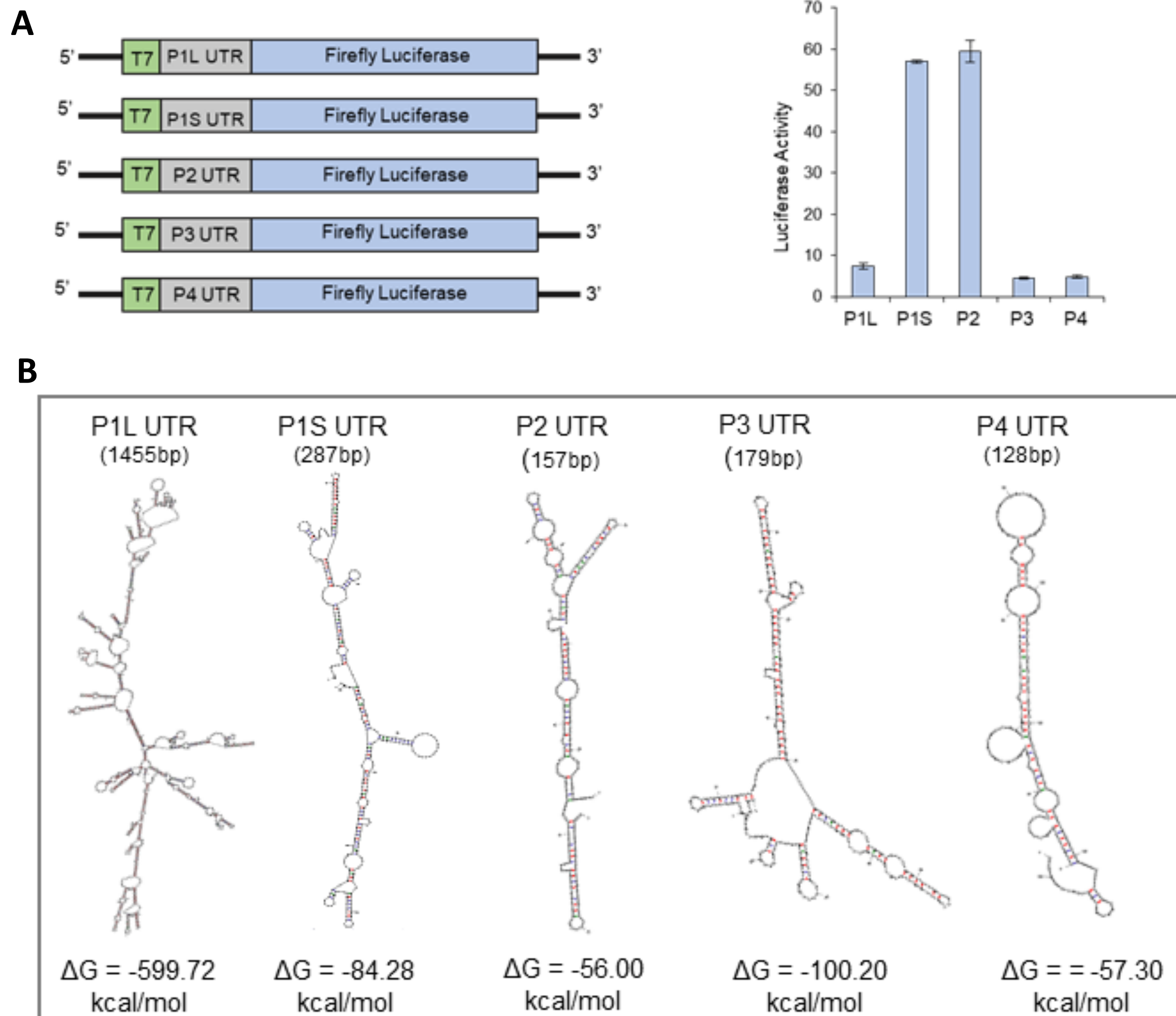


Figure 1.2.9. ***Satb1* transcript variants differ in the translatability *in vitro***. (A) The distinct 5' UTR sequences of *Satb1* transcript variants were amplified from the cDNA of mouse thymocytes and cloned upstream of firefly luciferase coding sequence in pGL3 basic vector (Promega). As shown in the figure, these constructs were used as templates for *in vitro* transcription experiment by including the T7 promoter sequence in the sense primer. *In vitro* transcription was performed using the T7 polymerase. The equimolar concentrations of *in vitro* transcribed RNAs normalized to their base pair lengths were used for *in vitro* translation. In vitro translation was performed using the rabbit reticulocyte lysate and luciferase activity assay was performed with *in vitro* translated products (N=3). (B) The secondary structure analysis of *Satb1* 5' UTR sequences was performed using the mfold web server. –ΔG values, which indicate the stability of secondary structures, were calculated for each 5' UTR sequence of the *Satb*1 transcript variants.

1.3 Discussion

SATB1 is a global chromatin organizer and is predominantly expressed in thymus, the development niche for T-cells (Dickinson *et al.*, 1992). SATB1 plays a pivotal role during thymocyte development by regulating the expression of multiple genes encoding T-cell specific cytokines and surface receptors (Alvarez *et al.*, 2000; Kondo *et al.*, 2016). During T-cell development, SATB1 plays an essential role in the selection of thymocytes (Kondo *et al.*, 2016). The absence of SATB1 results in the defective T-cell development (Alvarez *et al.*, 2000). During T-cell development, SATB1 is under the control of TCR signaling and is differentially expressed in various stages of T-cell development (Gottimukkala *et al.*, 2016). Although SATB1 plays a pivotal role in the T-cell development, how SATB1 is regulated with such a stringency of development was addressed in this study. In the current study, we have identified the putative alternative promoters of *Satb1* gene in mouse thymocytes. The alternative promoters were widely studied to regulate the gene expression in a tissue- or developmental stage-specific manner. The genome-wide studies provide evidence that the mammalian genome uses the alternative promoters to increase the frequency of diversity in the transcriptome and proteome, there by regulating the complexity of an organism (Landry, Mager and Wilhelm, 2003).

We show that the *Satb1* gene locus exhibits the putative alternative promoter like regulatory elements as revealed by the analysis of its epigenetic landscape through the profiling of signature histone modification marks. The post-translational modifications of histone tails have been widely used to study the transcriptional *cis* regulatory elements. It is well reported that the post-translational modifications of histone H3 is of great interest in understanding the transcription of a gene (Krogan et al., 2003, Pokholok et al., 2005; Rao et al., 2005; Kouzarides, 2007; Shilatifard, 2012; Voigt, Tee and Reinberg, 2013). The major *cis* regulatory elements of transcription include the promoter and an enhancer which are characterized by the enrichment of H3K4me3, and H3K4me1 marks respectively. We found that multiple regions upstream of the *Satb1* TSS, exhibit higher enrichment of H3K4me3 and low enrichment of H3K4me1 marks, which is indicative of multiple promoters like signature. In contrast, analysis at the gene

locus of another SATB family protein *Satb2* revealed only a single region with more H3K4me3 and low H3K4me1 mark. The *Satb2* expression is repressed in the mouse thymocytes as indicated by the enrichment of H3K27me3 mark. We found that the *Satb1* alternative promoter usage leads to the generation of multiple transcript variants, wherein the alternative promoter specific first exon was alternatively spliced to the common second exon. We have identified different *Satb1* transcripts with distinct first exons named as P1(E1a), P2(E1b), P3(E1c), and P4(E1d). Interestingly, the P1 transcripts harbor the long and the short first exon sequences, therefore we characterized them as long P1L and short P1S. Thus, we have identified four different alternative promoters and five distinct alternative transcript variants of *Satb1* in mouse. Moreover, the *Satb1* gene locus exhibits three alternative promoters in case of human T-cells (Khare *et al.,* 2019), some of which are similar to that of mouse indicating evolutionary conservation. The alternative splice variants of *Satb1* differ in the non-overlapping first exon sequences. These alternative first exons and the initial sequence of second exon contribute to the 5' UTR region (5' untranslated region) of *Satb1* mRNA, since the protein coding sequence begins in the exon-2.

SATB1 is essential for T-cell development. During T-cell development in thymus, SATB1 protein exhibits differential expression pattern (Gottimukkala *et al.*, 2016). Here we have shown that SATB1 protein is abundantly present in the immature CD4SP thymocytes compared to other T-cell developmental stages. How SATB1 is regulated with such stringency during T-cell development was addressed in this study. Among the thymocyte developmental stages, the double positive (DP) and immature CD4SP thymocytes have shown higher expression levels of *Satb1* mRNA. However, the transcript variants of *Satb1* show cell type specific expression pattern in a combinatorial manner during T-cell development. We observed that the P3 transcript variant is predominant and shows constitutive expression among the *satb1* transcript variants. The DP thymocytes show the higher expression of P1, P4 transcript variants along with the constitutively expressed P3 transcript variant. Strikingly, The P2 transcript variant was highly expressed specifically in the immature CD4SP thymocytes compared to DP or CD8SP thymocytes, and is selectively regulated during the stage transition from DP to immature CD4SP. Along with P2, the immature CD4SP thymocytes also exhibit the

expression of constitutive P3 transcript variant. Our analyses identified the cell type specific expression of a combination of *Satb1* mRNA variants during thymocyte development.

Although total *Satb1* mRNA levels differed marginally between DP and immature CD4, a significant difference was observed in the SATB1 protein levels between these two populations. During T-cell development, the immature CD4SP exhibits higher expression of SATB1 protein levels compared to the DP thymocytes. Our results confirm that the stability of SATB1 protein does not have any effect on the levels of SATB1 protein in DP during T-cell development. Since the SATB1 protein in DP cells does not undergo active degradation, we hypothesized that the significant difference in SATB1 protein levels in DP and immature CD4SP might be because of the expression of different combinations of *Satb1* transcript variants in DP and immature CD4SP thymocytes. We have shown that the combination of *Satb1* transcript variants with distinct 5' UTR sequences expressed in the immature CD4SP thymocytes contribute to the higher levels of SATB1 protein during T-cell development.

The regulatory elements within the 5' UTRs, such as an upstream open reading frames (uORFs), secondary structures, internal ribosome entry sites (IRES) and sequences for RNA binding proteins play an instrumental role in the regulation of translation initiation (Davuluri *et al.*, 2008). We identified the mechanism of differential translatability of *Satb*1 transcript variants *in vivo* and *in vitro* that is dependent on their 5' UTR sequences. Amongst the *Satb*1 alternative 5' UTRs, the P2, P1S and P4 5' UTR sequences form less stable secondary structures. The P1L 5' UTR sequence form more stable secondary structure followed by the P3 5' UTR sequence. These observations indicate that the presence of type of *Satb1* 5' UTR sequence clearly exert strong effect on the translation of *Satb1* mRNA.

When we assessed the translation efficiency of the *Satb1* transcript variants *in vivo* and *in vitro* by performing the polysome profiling and *in vitro* translation assays, we found that P1S and P2 5' UTRs show the higher translation efficacies as compared to the P1L, P3 and P4. Although the P4 5' UTR sequence forms less stable secondary structure, we observed that the translation efficiency of P4 is low. This observation

indicates that the P4 5’ UTR might harbor regulatory features other than the secondary structures that affect its translation.

Other than P1S, the DP thymocytes exhibit the expression of majority of *Satb1* transcript variants such as the P1L, P3, and P4 that have less translatability, which might affect the levels of SATB1 protein in DP. In contrast, the immature CD4SP thymocytes revealed the expression of only P2 and P3, wherein P2 contributes to the higher expression levels of SATB1 protein as was also observed in case of human Th2 differentiation (Khare *et al.,* 2019). These results suggest that the combinatorial expression of *Satb1* transcript variants in a cell type specific manner plays a key role in the maintenance of SATB1 protein levels. Altogether, the results indicate the importance of the P2 transcript variant expression in immature CD4SP thymocytes in maintaining the SATB1 protein levels than that of DP.

In conclusion, we show that expression of *Satb1* is regulated by multiple alternative promoters during T-cell development. The usage of alternative promoters leads to the expression of multiple *Satb1* transcript variants with distinct 5’ UTR sequences. The *Satb1* transcript variants expresse in a combinatorial manner during T-cell development. The expression of combination of *Satb1* transcript variants is cell type specific during T-cell development. *Satb1* transcript variants differ in the translatability both *in vitro* and *in vivo*. The combinatorial expression and the differential translatability of *Satb1* transcript variants play an essential role in maintaining the SATB1 protein levels during various stages of T-cell development. Thus, our study demonstrates that SATB1 expression is tightly regulated by alternative promoter switch during T-cell development.

1.4 Methods

1.4.1 Mice

Three week old C57BL/6 mice were used to prepare the single cell suspension of thymocytes for 5' RACE analysis and for the sorting of subpopulation of thymocytes which are in different stages of T-cell development. All mice were bred and maintained under pathogen free environment and the experiment procedures were performed according to the guidelines of the animal house facility at IISER Pune and NCCS Pune.

1.4.2 Flow cytometry

The single cell suspension of thymocytes were prepared using thymii from 3 week old C57BL/6 mice and were used for surface staining. Before the surface staining of total thymocytes, Fc receptor blocking was performed using the purified anti-CD16.32 antibody (BD Biosciences). Then, thymocytes were subjected to surface staining using the following the flurochrome tagged antibodies: eFluor 450 anti-mouse CD4 (Clone GK1.5, eBioscience); APC anti-mouse CD4 (Clone GK1.5, eBioscience); FITC anti-mouse CD4 (Clone GK1.5, BD Biosciences); PE anti-mouse CD8a (Clone 53-6.7, BD Biosciences); eFluor 450 anti-mouse CD24 (Clone M1/69, eBioscience); PerCP Cy5.5 anti-mouse CD24 (Clone M1/69, eBioscience). The thymocyte sub populations such as $CD4^{-}CD8^{-}$ double negative (DN), $CD4^{+}CD8^{+}$ double positive (DP), immature CD4SP ($CD4^{+}CD24^{+}$), mature CD4SP ($CD4^{+}CD24^{-}$), $CD4^{+}$ SP thymocytes (total CD4SP), and $CD8^{+}$ SP thymocytes (total CD8SP) were FACS sorted using FACS Aria III SORP (BD biosciences). In case of intracellular staining, immediately after the surface staining, the thymocytes were permeabilized and then intracellular staining was performed using anti-SATB1 (BD Biosciences) by using the Foxp3/Transcription Factor intracellular staining kit (eBiosciences). The flow cytometry analyses were performed using FACS Canto II (BD Biosciences).

1.4.3 Data base analysis

GSE20898 (Wei et al., 2011) data set was used for ChIP-seq analyses of genome-wide occupancy of H3K4me3, H3K27me3, and H3K4me1 performed in mouse DP thymocytes. Raw reads were aligned using Bowtie2 and peak calling was performed

using MACS2. MACS generated peaks were visualized by using IGV genome browser. RNA-seq analysis of DP and CD4SP thymocytes were performed using the dataset GSE48138 (Hu et al., 2013). RNA-seq read alignment was performed using Bowtie 2 and TopHat2. Cuffdiff was used for further differential gene expression analysis.

1.4.4 5’ Rapid amplification of cDNA ends (5’ RACE)

Single cell suspension of thymocytes were prepared using thymii from 3 week old C57BL/6 mice. Total thymocyte were used for the RNA extraction using Trizol (Invitrogen) method. Total RNA was used for cDNA synthesis which was performed using SMARTer RACE 5’/3’ kit (Clontech) according to the manufacturer’s protocol. 5’ RACE PCR was carried out using the forward universal primer mix provided with the kit, and the reverse primer 5’-TGCTCCCAAGCCTTCCTCTTCCTAGAG-3’, specific to the exon-2 of Satb1. The resulting PCR products were subjected to gel electrophoresis and DNA bands were gel purified using Nucleospin gel extraction kit (MACHEREY-NAGEL). Nested PCR was performed for few sequences as indicated in the Figure 1.2.1 B, using the following *Satb1* exon-2 specific reverse primer 5’-CTGTCTTACAGATCACCTGCCAG-3’. The amplified DNA fragments were cloned into linearized pRACE vector provided with the kit, and then transformed into DH5α strain of *E. coli* (Promega). Recombinant Plasmid DNAs were isolated from an individual bacterial clones by alkaline lysis method and were subjected to sequencing by Sanger sequencing method.

1.4.5 cDNA synthesis and Quantitative real-time PCR analysis (qRT-PCR)

Isolation of total RNA from the sorted thymocyte subpopulations was performed using the RNeasy mini kit (Qiagen). Following DNase I (Promega) digestion, the RNA was subjected to cDNA synthesis using High capacity cDNA synthesis kit (Applied Biosystems). The quantitative PCR analyses were performed using Sybr green qPCR master mix (Roche) at the following PCR conditions: step 1, 95°C-5 min; step 2, 95°C-45 sec, 60°C-45 sec, 72°C-1 min for 40 cycles. The following qPCR primers were used:

Satb1 (total)-F: 5’-TGATAGAGATGGCGTTGCTG-3’
Satb1 (total)-R: 5’-TTTTGAGGGTGACCACATGA-3’
P1(E1a)-F: 5’-CAAGAATCCCGGCTGCAAAG-3’
P1(E1a)-R: 5’-CCCTGAGTTGCCTCGTTCAA-3’
P2(E1b)-F: 5’-AGATTCGGAAACCAGCCTCTG-3’
P2(E1b)-R: 5’-GGACCCTTCGGATCACTCAC-3’
P3(E1c)-F: 5’-CGGTTCCACGCCTGATTCT-3’
P3(E1c)-R: 5’-GTGGACCCTTCGGATCACTC-3’
P4(E1d)-F: 5’-CCAAAGCCCAGGCAAACAAC-3’
P4(E1d)-R: 5’-CCCTGAGTTGCCTCGTTCAA-3’
m18s-F: 5’-GTAACCCGTTGAACCCCATT-3’
m18s-R: 5’-CCATCCAATCGGTAGTAGCG-3’

1.4.6 Cycloheximide and MG132 chase assay

Three week old C57BL/6 mice were used for isolation of thymus. Thymii were used for the preparation of single cell suspension. Single cell suspension of thymocytes were subjected to Fc receptor blocking using the purified anti-CD16/CD32 (Clone 2.4G2, BD Biosciences). Then thymocytes were surface stained using the following flurochrome tagged antibodies: FITC anti-mouse CD4 (Clone GK1.5, BD Biosciences); PE anti-mouse CD8a (Clone 53-6.7, BD Biosciences). $CD4^+CD8^+$ double positive (DP) thymocytes were FACS sorted using FACS Aria III SORP (BD biosciences). Sorted DP thymocytes were cultured in the presence or absence of different concentrations of MG132 (Sigma-Aldrich), and Cycloheximide (Sigma-Aldrich) in RPMI-1640 media supplemented with 10% FBS and penicillin-streptomycin for 4hrs. After incubation, cells were harvested and used for western blotting experiment.

1.4.7 Western blotting

Cells were lysed using RIPA lysis buffer (50 mM Tris-HCl at PH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.25% sodium deoxycholate, 1 mM PMSF, and 1X protease inhibitor cocktail (Roche) and the protein quantification was performed using BCA method. The total protein was separated on 10-12% SDS polyacrylamide gel and then

transferred to PVDF membrane (Millipore). The PVDF membrane was then blocked with 5% milk and probed with the following antibodies; anti-SATB1 (1:1000, BD Biosciences), anti-GAPDH (1:4000, ABM), and anti-ubiquitin (1:1000, Millipore). The signals were visualized using ECL luminescence detection reagent (BIO-RAD) on ImageQuant LAS 4000 system (GE Healthcare Life Sciences).

1.4.8 Polysome profiling analysis

Thymii from six 3 week old C57BL/6 mice were immediately snap frozen as soon as mice were dissected and were pulverized under liquid nitrogen. The powder obtained was then lysed in the lysis buffer (5 mM Tris-HCl at pH 7.5, 2.5 mM MgCl2, 1.5 mM KCl, 100 µg/mL cycloheximide (Sigma), 0.1% Triton-X-100, 2mM DTT, 500 U/mL RNase inhibitor (ABI), 1X EDTA free protease inhibitor cocktail (Roche). After lysis, the nuclei were pelleted down by centrifugation at 14,000 rpm for 15 min, at 4 ^{o}C, with no brake during deceleration using a bench top centrifuge (Rotor IL-053, Eppendorf). The supernatant was collected into pre-chilled 1.5 ml tubes and optical density (OD) at 260 nm was measured using spectrophotometer (NanoDrop 2000c, Thermo Scientific). The supernatant with OD at 260nm of 30-40 was layered onto the 10mL linear sucrose gradient (10-50% sucrose (w/v), which was made in 1X gradient buffer (20 mM HEPES at pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 100 µg/mL cycloheximide (Sigma), 500 U/mL RNase inhibitor (ABI), 1X EDTA free protease inhibitor cocktail (Roche) and centrifuged in a SW40Ti rotor (Beckman) for 3 hrs at 35,000 rpm at 4 ^{o}C, with no brake applied during deceleration. Sucrose gradient was then subjected to the fractionations using the gradient fractionation system (ISCO Model 160 gradient former). Fractions were used for RNA extraction using Trizol (Invitrogen) method. Following DNaseI (Promega) digestion, RNA was subjected to cDNA synthesis which was performed using the iScript cDNA synthesis kit (Biorad). Quantitative real time PCR (qRT-PCR) analyses were performed using SYBR green qPCR master mix (Roche) and by using the *Satb1* alternative transcript specific amplification primers.

The following qPCR primers were used for qPCR analyses:

P1(E1a)-F: 5’-CAAGAATCCCGGCTGCAAAG-3’
P1(E1a)-R: 5’-CCCTGAGTTGCCTCGTTCAA-3’
P2(E1b)-F: 5’-AGATTCGGAAACCAGCCTCTG-3’
P2(E1b)-R: 5’-GGACCCTTCGGATCACTCAC-3’
P3(E1c)-F: 5’-CGGTTCCACGCCTGATTCT-3’
P3(E1c)-R: 5’-GTGGACCCTTCGGATCACTC-3’
P4(E1d)-F: 5’-CCAAAGCCCAGGCAAACAAC-3’
P4(E1d)-R: 5’-CCCTGAGTTGCCTCGTTCAA-3’
β-Actin-F: 5’-GCCTTCCTTCTTGGGTATGG-3’
β-Actin-R: 5’-GCACTGTGTTGGCATAGAGG-3’

1.4.9 *In vitro* transcription, *in vitro* translation, and Luciferase reporter assay

Full length *Satb1* 5’UTRs such as P1L, P1S, P2, P3, and P4 5’UTRs were amplified from thymocyte cDNA and then cloned upstream of firefly luciferase gene in pGL3 promoter vector (Promega). The sequences of an individual clones were confirmed by Sanger sequencing method. A template for *in vitro* transcription was prepared by PCR amplification of firefly luciferase coding sequences along with the cloned *Satb1* 5’ UTRs by using a pair of primers which include T7 promoter in the sense primer specific to the 5’UTR, and an antisense primer specific to the firefly luciferase gene. The resulting PCR amplified products were gel extracted and quantified using spectrophotometer (NanoDrop 2000c, Thermo Scientific). Equimolar concentrations of DNA normalized to their base pair lengths were taken for *in vitro* transcription reaction using T7 RNA Polymerase, according to the manufacturer’s protocol of mMESSAGE mMACHINE T7 kit (Invitrogen). *In vitro* transcribed RNA was purified by $LiCl_2$ precipitation method and then equimolar concentrations of RNA normalized to their base pair sizes were taken for *in vitro* translation reaction using rabbit reticulocyte lysate (Invitrogen). Luciferase activity was measured by using 5 μL of *in vitro* translated product and using luciferase assay kit (Promega).

The following primer sequences were used for cloning of the 5' UTRs of *Satb1*:

P1L 5'UTR-F: 5'-GCAAGCTTAGCCACAGCGGCGGCGGCGGCAG-3'
P1L 5'UTR-R: 5'-GCAAGCTTACTCAGTCACTGTCTTACAGATCA-3'
P1S 5'UTR-F: 5'-GCAAGCTTGTTATTCGAATTGTGACTT-3'
P1S 5'UTR-R: 5'-GCACCATGGACTCAGTCACTGTCTTACAG-3'
P2 5'UTR-F: 5'-GCAAGCTTGCAGACATGGTCCAGCTGCT-3'
P2 5'UTR-R: 5'-GCCCATGGACTCAGTCACTGTCTTACAGATC-3'
P3 5'UTR-F: 5'-GCAAGCTTGGAGACAACTTGAAGTCAA-3'
P3 5'UTR-R: 5'-GCCCATGGACTCAGTCACTGTCTTACAG-3'
P4 5'UTR-F; 5'-GCAAGCTTTCACCTGACTTCCTCCTCCGCT-3'
P4 5'UTR-R: 5'-GCCCATGGACTCAGTCACTGTCTTACA-3'

1.5 References

Abe N, Hozumi K, Hirano K, Yagita H, Habu S. (2010). Notch ligands transduce different magnitudes of signaling critical for determination of T-cell fate. Eur J Immunol *40,* 2608-2617.

Agata Y, Tamaki N, Sakamoto S, Ikawa T, Masuda K, Kawamoto H, Murre C. (2007). Regulation of T cell receptor beta gene rearrangements and allelic exclusion by the helix-loop-helix protein, E47. Immunity *27,* 871-84.

Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, Kohwi-Shigematsu T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev *14,* 521-35.

Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. (2012). Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov *11,* 384-400.

Balamotis MA, Tamberg N, Woo YJ, Li J, Davy B, Kohwi-Shigematsu T, Kohwi Y. (2012). Mol Cell Biol *32,* 333-47

Barndt RJ, Dai M, Zhuang Y. (2000). Functions of E2A-HEB heterodimers in T-cell development revealed by a dominant negative mutation of HEB. Mol Cell Biol *20,* 6677-85.

Basson MA, Bommhardt U, Mee PJ, Tybulewicz VL, Zamoyska R. (1998). Molecular requirements for lineage commitment in the thymus--antibody-mediated receptor engagements reveal a central role for lck in lineage decisions. Immunol Rev *165,*181-94.

de Belle I, Cai S, Kohwi-Shigematsu T. (1998). The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. J Cell Biol *141,* 335-48.

Benveniste P, Knowles G, Cohen A. (1996). CD8/CD4 lineage commitment occurs by an instructional/default process followed by positive selection. Eur J Immunol *26,* 461-71.

Beyer M, Thabet Y, Müller RU, Sadlon T, Classen S, Lahl K, Basu S, Zhou X, Bailey-Bucktrout SL, Krebs W, et al. (2011). Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. Nat Immunol *12,* 898-907.

Bonneville M, Ishida I, Mombaerts P, Katsuki M, Verbeek S, Berns A, Tonegawa S. (1989). Blockage of alpha beta T-cell development by TCR gamma delta transgenes. Nature *342,* 931-4.

Brugnera E, Bhandoola A, Cibotti R, Yu Q, Guinter TI, Yamashita Y, Sharrow SO, Singer A. (2000). Coreceptor reversal in the thymus: signaled CD4+8+ thymocytes initially terminate CD8 transcription even when differentiating into CD8+ T cells. Immunity *13,* 59-71.

Burute M, Gottimukkala K, Galande S. (2012). Chromatin organizer SATB1 is an important determinant of T-cell differentiation. Immunol Cell Biol *90,* 852-9.

Cai S, Han HJ, Kohwi-Shigematsu T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nat Genet *34,* 42-51.

Cai S, Lee CC, Kohwi-Shigematsu T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat Genet *38,* 1278-88.

Calo E, Wysocka J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell. *49,* 825-37.

Caton ML, Smith-Raska MR, Reizis B. (2007). Notch-RBP-J signaling controls the homeostasis of CD8- dendritic cells in the spleen. J Exp Med *204,* 1653-64.

Chan SH, Cosgrove D, Waltzinger C, Benoist C, Mathis D. (1993). Another view of the selective model of thymocyte selection. Cell *73,* 225-36.

Chi T. (2004). A BAF-centred view of the immune system. Nat Rev Immunol *4,* 965-77.

Cibotti R, Bhandoola A, Guinter TI, Sharrow SO, Singer A. (2000). CD8 coreceptor extinction in signaled CD4(+)CD8(+) thymocytes: coordinate roles for both transcriptional and posttranscriptional regulatory mechanisms in developing thymocytes. Mol Cell Biol *20,* 3852-9.

Cismasiu VB, Adamo K, Gecewicz J, Duque J, Lin Q, Avram D. (2005). BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. Oncogene *24,* 6753-64.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107,* 21931-6.

Davis CB, Killeen N, Crooks ME, Raulet D, Littman DR. (1993). Evidence for a stochastic mechanism in the differentiation of mature subsets of T lymphocytes. Cell *73,* 237-47.

Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. (2008). The functional consequences of alternative promoter use in mammalian genomes. Trends Genet *24,* 167-77.

Dawson MA, Kouzarides T. (2012). Cancer epigenetics: from mechanism to therapy. Cell *150,* 12-27.

Deftos ML, He YW, Ojala EW, Bevan MJ. (1998). Correlating notch signaling with thymocyte maturation. Immunity *9,* 777-86.

Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T. (1992). A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. Cell *70,* 631-45.

Dickinson LA, Dickinson CD, Kohwi-Shigematsu T. (1997). An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. J Biol Chem *272,* 11463-70.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, et al. (2012). Landscape of transcription in human cells. Nature *489,* 101-8.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473,* 43-9.

Fehling HJ, Krotkova A, Saint-Ruf C, von Boehmer H. (1995). Crucial role of the pre-T-cell receptor alpha gene in development of alpha beta but not gamma delta T cells. Nature *375,* 795-8.

Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. (2008). A code for transcription initiation in mammalian genomes. Genome Res *18,* 1-12.

Frohman MA. (19930. Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. Methods Enzymol *218,* 340-56.

Frömberg A, Engeland K, Aigner A. (2018). The Special AT-rich Sequence Binding Protein 1 (SATB1) and its role in solid tumors. Cancer Lett *417,* 96-111.

Galande S, Dickinson LA, Mian IS, Sikorska M, Kohwi-Shigematsu T. (2001). SATB1 cleavage by caspase 6 disrupts PDZ domain-mediated dimerization, causing detachment from chromatin early in T-cell apoptosis. Mol Cell Biol *21,* 5591-604.

Galande S, Purbey PK, Notani D, Kumar PP. (2007). The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. Curr Opin Genet Dev *17,* 408-14.

García-Ojeda ME, Klein Wolterink RG, Lemaître F, Richard-Le Goff O, Hasan M, Hendriks RW, Cumano A, Di Santo JP. (2013). GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice. Blood *121,* 1749-59.

Germar K, Dose M, Konstantinou T, Zhang J, Wang H, Lobry C, Arnett KL, Blacklow SC, Aifantis I, Aster JC, Gounari F. (2011). T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. Proc Natl Acad Sci U S A *108,* 20060-5.

Gershenzon NI, Ioshikhes IP. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics *21,* 1295-300.

Gifford CA, Meissner A. (2012). Epigenetic obstacles encountered by transcription factors: reprogramming against all odds. Curr Opin Genet Dev *22,* 409-15.

Godfrey DI, Kennedy J, Suda T, Zlotnik A. (1993). A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8- triple-negative adult mouse thymocytes defined by CD44 and CD25 expression. J Immunol *150,* 4244-52.

Goolam M, Zernicka-Goetz M. (2017). The chromatin modifier Satb1 regulates cell fate through Fgf signalling in the early mouse embryo. Development *144,* 1450-1461.

Gottimukkala KP, Jangid R, Patta I, Sultana DA, Sharma A, Misra-Sen J, Galande S. (2016). Regulation of SATB1 during thymocyte development by TCR signaling. Mol Immunol *77,* 34-43.

Han HJ, Russo J, Kohwi Y, Kohwi-Shigematsu T. (2008). SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. Nature *452,* 187-93.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39,* 311-8.

Hernández-Hoyos G, Sohn SJ, Rothenberg EV, Alberola-Ila J. (2000). Lck activity controls CD4/CD8 T cell lineage commitment. Immunity *12,* 313-22.

Hernández-Hoyos G, Anderson MK, Wang C, Rothenberg EV, Alberola-Ila J. (2003). GATA-3 expression is controlled by TCR signals and regulates CD4/CD8 differentiation. Immunity *19,* 83-94.

Ho IC, Tai TS, Pai SY. (2009). GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. Nat Rev Immunol *9,* 125-35.

Hogquist KA, Jameson SC, Heath WR, Howard JL, Bevan MJ, Carbone FR. (1994). T cell receptor antagonist peptides induce positive selection. Cell *76,* 17-27.

Hosoya T, Maillard I, Engel JD. (2010). From the cradle to the grave: activities of GATA-3 throughout T-cell development and differentiation. Immunol Rev *238,* 110-25.

Hu G, Tang Q, Sharma S, Yu F, Escobar TM, Muljo SA, Zhu J, Zhao K. (2013). Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. Nat Immunol *14,* 1190-8.

Ikawa T, Kawamoto H, Goldrath AW, Murre C. (2006). E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. J Exp Med *203,* 1329-42.

Ikawa T, Hirose S, Masuda K, Kakugawa K, Satoh R, Shibano-Satoh A, Kominami R, Katsura Y, Kawamoto H. (2010). An essential developmental checkpoint for production of the T cell lineage. Science *329,* 93-6.

Inoue J, Kanefuji T, Okazuka K, Watanabe H, Mishima Y, Kominami R. (2006). Expression of TCR alpha beta partly rescues developmental arrest and apoptosis of alpha beta T cells in Bcl11b-/- mice. J Immunol *176,* 5871-9.

Itano A, Salmon P, Kioussis D, Tolaini M, Corbella P, Robey E. (1996). The cytoplasmic domain of CD4 promotes the development of CD4 lineage T cells. J Exp Med *183,* 731-41.

Izon DJ, Punt JA, Xu L, Karnell FG, Allman D, Myung PS, Boerth NJ, Pui JC, Koretzky GA, Pear WS. (2001). Notch1 regulates maturation of CD4+ and CD8+ thymocytes by modulating TCR signal strength. Immunity *14,* 253-64.

Izon DJ, Aster JC, He Y, Weng A, Karnell FG, Patriub V, Xu L, Bakkour S, Rodriguez C, Allman D, Pear WS. (2002). Deltex1 redirects lymphoid progenitors to the B cell lineage by antagonizing Notch1. Immunity *16,* 231-43.

Izon DJ, Punt JA, Pear WS. (2002). Deciphering the role of Notch signaling in lymphopoiesis. Curr Opin Immunol *14,* 192-9.

Kakugawa K, Kojo S, Tanaka H, Seo W, Endo TA, Kitagawa Y, Muroi S, Tenno M, Yasmin N, Kohwi Y, Sakaguchi S, Kowhi-Shigematsu T, Taniuchi I. (2017). Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. Cell Rep 19, 1176-1188.

Kappler JW, Roehm N, Marrack P. (1987). T cell tolerance by clonal elimination in the thymus. Cell *49,* 273-80.

Kawamoto H, Ohmura K, Fujimoto S, Lu M, Ikawa T, Katsura Y. (2003). Extensive proliferation of T cell lineage-restricted progenitors in the thymus: an essential process for clonal expression of diverse T cell receptor beta chains. Eur J Immunol *33,* 606-15.

Kaye J, Hsu ML, Sauron ME, Jameson SC, Gascoigne NR, Hedrick SM. (1989). Selective development of CD4+ T cells in transgenic mice expressing a class II MHC-restricted antigen receptor. Nature *341,* 746-9.

Kearse KP, Roberts JL, Singer A. (1995). TCR alpha-CD3 delta epsilon association is the initial step in alpha beta dimer formation in murine T cells and is limiting in immature CD4+ CD8+ thymocytes. Immunity *2,* 391-9.

Khare S, Shetty A, Biradar R, Patta I, Sathe A, Reddy PC, Chen Z, Lahesmaa R and Galande S. (2019). NF-kB signaling and IL-4 signaling regulate SATB1 expression via alternative promoter usage during Th2 differentiation. Front. Immunol. doi: 10.3389/fimmu.2019.00667. (In press).

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465,* 182-7.

Kim TK, Shiekhattar R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. Cell *162,* 948-59.

Kioussis D, Ellmeier W. (2002). Chromatin and CD4, CD8A and CD8B gene expression during thymic differentiation. Nat Rev Immunol *2,* 909-19.

Kisielow P, Teh HS, Blüthmann H, von Boehmer H. (1988). Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. Nature *335*, 730-3.

Kitagawa Y, Ohkura N, Kidani Y, Vandenbon A, Hirota K, Kawakami R, Yasuda K, Motooka D, Nakamura S, Kondo M, Taniuchi I, Kohwi-Shigematsu T, Sakaguchi S. (2017). Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. Nat Immunol 18, 173-183.

Koch U, Lacombe TA, Holland D, Bowman JL, Cohen BL, Egan SE, Guidos CJ. (2001). Subversion of the T/B lineage decision in the thymus by lunatic fringe-mediated inhibition of Notch-1. Immunity *15,* 225-36.

Kondo M, Tanaka Y, Kuwabara T, Naito T, Kohwi-Shigematsu T, Watanabe A. (2016). SATB1 Plays a Critical Role in Establishment of Immune Tolerance. J Immunol *196,* 563-72.

Kouzarides T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, Greenblatt JF, Shilatifard A. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell *11,* 721-9.

Kumar PP, Purbey PK, Ravi DS, Mitra D, Galande S. (2005). Displacement of SATB1-bound histone deacetylase 1 corepressor by the human immunodeficiency virus type 1 transactivator induces expression of interleukin-2 and its receptor in T cells. Mol Cell Biol *25,* 1620-33.

Kumar PP, Bischof O, Purbey PK, Notani D, Urlaub H, Dejean A, Galande S. (2007). Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. Nat Cell Biol *9*, 45-56.

Landry JR, Mager DL, Wilhelm BT. (2003). Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet *19,* 640-8.

Legname G, Seddon B, Lovatt M, Tomlinson P, Sarner N, Tolaini M, Williams K, Norton T, Kioussis D, Zamoyska R. (2000). Inducible expression of a p56Lck transgene reveals a central role for Lck in the differentiation of CD4 SP thymocytes. Immunity *12,* 537-46.

Lenhard B, Sandelin A, Carninci P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet *13,* 233-45.

Li L, Leid M, Rothenberg EV. (2010). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science *329,* 89-93.

Li P, Burke S, Wang J, Chen X, Ortiz M, Lee SC, Lu D, Campos L, Goulding D, Ng BL, Dougan G, Huntly B, Gottgens B, Jenkins NA, Copeland NG, Colucci F, Liu P. (2010). Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science *329,* 85-9.

Ma C, Zhang J, Durrin LK, Lv J, Zhu D, Han X, Sun Y. (2007). The BCL2 major breakpoint region (mbr) regulates gene expression. Oncogene *26,* 2649-57.

Maillard I, Tu L, Sambandam A, Yashiro-Ohtani Y, Millholland J, Keeshan K, Shestova O, Xu L, Bhandoola A, Pear WS. (2006). The requirement for Notch signaling at the beta-selection checkpoint in vivo is absolute and independent of the pre-T cell receptor. J Exp Med *203*, 2239-45.

Matechak EO, Killeen N, Hedrick SM, Fowlkes BJ. (1996). MHC class II-specific T cells can develop in the CD8 lineage when CD4 is absent. Immunity *4,* 337-47.

Mathis DJ, Chambon P. The SV40 early region TATA box is required for accurate in vitro initiation of transcription. (1981). Nature *290,* 310-5.

McInnes N, Sadlon TJ, Brown CY, Pederson S, Beyer M, Schultze JL, McColl S, Goodall GJ, Barry SC. (2012). FOXP3 and FOXP3-regulated microRNAs suppress SATB1 in breast cancer cells. Oncogene *31,* 1045-54.

MILLER JF. (1961). Immunological function of the thymus. Lancet *2,* 748-9.

Mir R, Pradhan SJ, Patil P, Mulherkar R, Galande S. (2016). Wnt/β-catenin signaling regulated SATB1 promotes colorectal cancer tumorigenesis and progression. Oncogene *35,* 1679-91.

Mombaerts P, Clarke AR, Rudnicki MA, Iacomini J, Itohara S, Lafaille JJ, Wang L, Ichikawa Y, Jaenisch R, Hooper ML, et al. (1992). Mutations in T-cell antigen receptor genes alpha and beta block thymocyte development at different stages. Nature *360,* 225-31.

Muroi S, Naoe Y, Miyamoto C, Akiyama K, Ikawa T, Masuda K, Kawamoto H, Taniuchi I. (2008). Cascading suppression of transcriptional silencers by ThPOK seals helper T cell fate. Nat Immunol *9,* 1113-21.

Naik R, Galande S. (2019). SATB family chromatin organizers as master regulators of tumor progression. Oncogene *38,* 1989-2004.

Notani D, Gottimukkala KP, Jayani RS, Limaye AS, Damle MV, Mehta S, Purbey PK, Joseph J, Galande S. (2010). Global regulator SATB1 recruits beta-catenin and regulates T(H)2 differentiation in Wnt-dependent manner. PLoS Biol *8,* e1000296.

De Obaldia ME, Bell JJ, Wang X, Harly C, Yashiro-Ohtani Y, DeLong JH, Zlotoff DA, Sultana DA, Pear WS, Bhandoola A. (2013). T cell development requires constraint of the myeloid regulator C/EBP-α by the Notch target and transcriptional repressor Hes1. Nat Immunol *14,* 1277-84.

Ohoka Y, Kuwata T, Asada A, Zhao Y, Mukai M, Iwata M. (1997). Regulation of thymocyte lineage commitment by the level of classical protein kinase C activity. J Immunol *158,* 5707-16.

Pavan Kumar P, Purbey PK, Sinha CK, Notani D, Limaye A, Jayani RS, Galande S. (2006). Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. Mol Cell *22*, 231-43.

Pai SY, Truitt ML, Ting CN, Leiden JM, Glimcher LH, Ho IC. (2003). Critical roles for transcription factor GATA-3 in thymocyte development. Immunity *19,* 863-75.

Palmer E. (2003). Negative selection--clearing out the bad apples from the T-cell repertoire. Nat Rev Immunol *3,* 383-91.

Pearse M, Wu L, Egerton M, Wilson A, Shortman K, Scollay R. (1989). A murine early thymocyte developmental sequence is marked by transient expression of the interleukin 2 receptor. Proc Natl Acad Sci U S A *86,* 1614-8.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122,* 517-27.

Pui JC, Allman D, Xu L, DeRocco S, Karnell FG, Bakkour S, Lee JY, Kadesch T, Hardy RR, Aster JC, Pear WS. Notch1 expression in early lymphopoiesis influences B versus T lineage determination. (1999). Immunity *11,* 299-308.

Purbey PK, Singh S, Kumar PP, Mehta S, Ganesh KN, Mitra D, Galande S. (2008). PDZ domain-mediated dimerization and homeodomain-directed specificity are required for high-affinity DNA binding by SATB1. Nucleic Acids Res *36*, 2107-22.

Purbey PK, Singh S, Notani D, Kumar PP, Limaye AS, Galande S. (2009). Acetylation-dependent interaction of SATB1 and CtBP1 mediates transcriptional repression by SATB1. Mol Cell Biol *29,* 1321-37.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470,* 279-83.

Radtke F, Wilson A, Stark G, Bauer M, van Meerwijk J, MacDonald HR, Aguet M. (1999). Deficient T cell fate specification in mice with an induced inactivation of Notch1. Immunity *10,* 547-58.

Rando OJ, Chang HY. (2009). Genome-wide views of chromatin structure. Annu Rev Biochem *78*, 245-71.

Rao B, Shibata Y, Strahl BD, Lieb JD. (2005). Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. Mol Cell Biol *25,* 9447-59.

Robey EA, Fowlkes BJ, Gordon JW, Kioussis D, von Boehmer H, Ramsdell F, Axel R. (1991). Thymic selection in CD8 transgenic mice supports an instructive model for commitment to a CD4 or CD8 lineage. Cell *64,* 99-107.

Roeder RG. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci 21, 327-35.

Saint-Ruf C, Ungewiss K, Groettrup M, Bruno L, Fehling HJ, von Boehmer H. (1994). Analysis and expression of a cloned pre-T cell receptor gene. Science *266,* 1208-12.

Sakata-Yanagimoto M, Nakagami-Yamaguchi E, Saito T, Kumano K, Yasutomo K, Ogawa S, Kurokawa M, Chiba S. (2008). Coordinated regulation of transcription factors through Notch2 is an important mediator of mast cell fate. Proc Natl Acad Sci U S A 105, 7839-44.

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. (20070. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet *8,* 424-36.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

Sato T, Ohno S, Hayashi T, Sato C, Kohu K, Satake M, Habu S. (2005). Dual functions of Runx proteins for reactivating CD8 and silencing CD4 at the commitment process into CD8 thymocytes. Immunity *22,* 317-28.

Satoh Y, Yokota T, Sudo T, Kondo M, Lai A, Kincade PW, Kouro T, Iida R, Kokame K, Miyata T, Habuchi Y, Matsui K, Tanaka H, Matsumura I, Oritani K, Kohwi-Shigematsu T, Kanakura Y. (2013). The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. Immunity 38, 1105-15.

Sawada S, Scarborough JD, Killeen N, Littman DR. (1994). A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. Cell *77*, 917-29.

Schmitt TM, Zúñiga-Pflücker JC. (2002). Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. Immunity *17,* 749-56.

Setoguchi R, Tachibana M, Naoe Y, Muroi S, Akiyama K, Tezuka C, Okuda T, Taniuchi I. (2008). Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. Science *319,* 822-5.

Sha WC, Nelson CA, Newberry RD, Kranz DM, Russell JH, Loh DY. (1988). Positive and negative selection of an antigen receptor on T cells in transgenic mice. Nature *336*, 73-6.

Sharp LL, Schwarz DA, Bott CM, Marshall CJ, Hedrick SM. (1997). The influence of the MAPK pathway on T cell lineage commitment. Immunity *7,* 609-18.

Shilatifard A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem *81*, 65-95.

Shlyueva D, Stampfel G, Stark A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-86.

Simeoni L, Bogeski I. (2015). Redox regulation of T-cell receptor signaling. Biol Chem *396*, 555-68.

Singer A. (2002). New perspectives on a developmental dilemma: the kinetic signaling model and the importance of signal duration for the CD4/CD8 lineage decision. Curr Opin Immunol *14*, 207-15.

Singer A, Adoro S, Park JH. (2008). Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. Nat Rev Immunol *8,* 788-801.

Siu G1, Wurster AL, Duncan DD, Soliman TM, Hedrick SM. (1994). A transcriptional silencer controls the developmental expression of the CD4 gene. EMBO J *13,* 3570-9.

Smale ST, Jain A, Kaufmann J, Emami KH, Lo K, Garraway IP. (1998). The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. Cold Spring Harb Symp Quant Biol *63,* 21-31.

Smale ST, Baltimore D. (1989). The "initiator" as a transcription control element. Cell *57,* 103-13.

Smale ST, Kadonaga JT. (2003). The RNA polymerase II core promoter. Annu Rev Biochem *72,* 449-79.

Spitz F, Furlong EE. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13,* 613-26.

Starr TK, Jameson SC, Hogquist KA. (2003). Positive and negative selection of T cells. Annu Rev Immunol *21,* 139-76.

Stephen TL, Payne KK, Chaurio RA, Allegrezza MJ, Zhu H, Perez-Sanz J, Perales-Puchalt A, Nguyen JM, Vara-Ailor AE, Eruslanov EB, Borowsky ME, Zhang R, Laufer TM, Conejo-Garcia JR. (2017). SATB1 Expression Governs Epigenetic Repression of PD-1 in Tumor-Reactive T Cells. Immunity *46,* 51-64.

Stritesky GL, Jameson SC, Hogquist KA. (2012). Selection of self-reactive T cells in the thymus. Annu Rev Immunol *30*, 95-114.

Tarakhovsky A. (2010). Tools and landscapes of epigenetics. Nat Immunol *11*, 565-8.

Taniuchi I, Osato M, Egawa T, Sunshine MJ, Bae SC, Komori T, Ito Y, Littman DR. (2002). Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. Cell *111,* 621-33.

Tomita K, Hattori M, Nakamura E, Nakanishi S, Minato N, Kageyama R. (1999). The bHLH gene Hes1 is essential for expansion of early T cell precursors. Genes Dev *13,* 1203-10.

Veillette A, Zúñiga-Pflücker JC, Bolen JB, Kruisbeek AM. (1989). Engagement of CD4 and CD8 expressed on immature thymocytes induces activation of intracellular tyrosine phosphorylation pathways. J Exp Med *170,* 1671-80.

Voigt P1, Tee WW, Reinberg D. (2013). A double take on bivalent promoters. Genes Dev 27, 1318-38.

Wakabayashi Y, Watanabe H, Inoue J, Takeda N, Sakata J, Mishima Y, Hitomi J, Yamamoto T, Utsuyama M, Niwa O, Aizawa S, Kominami R. (2003). Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. Nat Immunol *4,* 533-9.

Wang L, Wildt KF, Zhu J, Zhang X, Feigenbaum L, Tessarollo L, Paul WE, Fowlkes BJ, Bosselut R. (2008). Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of CD4(+) T cells. Nat Immunol *9*, 1122-30.

Wang Z, Yang X, Chu X, Zhang J, Zhou H, Shen Y, Long J. (2012). The structural basis for the oligomerization of the N-terminal domain of SATB1. Nucleic Acids Res *40*, 4193-202.

Wang Z, Yang X, Guo S, Yang Y, Su XC, Shen Y, Long J. (2014). Crystal structure of the ubiquitin-like domain-CUT repeat-like tandem of special AT-rich sequence binding protein 1 (SATB1) reveals a coordinating DNA-binding mechanism. J Biol Chem *289,* 27376-85.

Weber BN, Chi AW, Chavez A, Yashiro-Ohtani Y, Yang Q, Shestova O, Bhandoola A. (2011). A critical role for TCF-1 in T-lineage specification and differentiation. Nature *476,* 63-8.

Wei G, Abraham BJ, Yagi R, Jothi R, Cui K, Sharma S, Narlikar L, Northrup DL, Tang Q, Paul WE, Zhu J, Zhao K. (2011). Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity *35*, 299-311.

Wiest DL, Yuan L, Jefferson J, Benveniste P, Tsokos M, Klausner RD, Glimcher LH, Samelson LE, Singer A. (1993). Regulation of T cell receptor expression in immature CD4+CD8+ thymocytes by p56lck tyrosine kinase: basis for differential signaling by CD4 and CD8 in immature thymocytes expressing both coreceptor molecules. J Exp Med *178,* 1701-12.

Wendorff AA, Koch U, Wunderlich FT, Wirth S, Dubey C, Brüning JC, MacDonald HR, Radtke F. (2010). Hes1 is a critical but context-dependent mediator of canonical Notch signaling in lymphocyte development and transformation. Immunity *33,* 671-84.

Will B, Vogler TO, Bartholdy B, Garrett-Bakelman F, Mayer J, Barreyro L, Pandolfi A, Todorova TI, Okoye-Okafor UC, Stanley RF, et al. (2013). Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. Nat Immunol *14,* 437-45.

Wilson A, MacDonald HR, Radtke F. (2001). Notch 1-deficient common lymphoid precursors adopt a B cell fate in the thymus. J Exp Med *194,* 1003-12.

Wolfer A, Wilson A, Nemir M, MacDonald HR, Radtke F. (2002). Inactivation of Notch1 impairs VDJbeta rearrangement and allows pre-TCR-independent survival of early alpha beta Lineage Thymocytes. Immunity *16,* 869-79.

Wu X, Sharp PA. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990-6.

Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. Nature *419,* 641-5.

Yasutomo K, Doyle C, Miele L, Fuchs C, Germain RN. (2000). The duration of antigen receptor signalling determines CD4+ versus CD8+ T-cell lineage fate. Nature *404*, 506-10.

Yui MA, Rothenberg EV. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat Rev Immunol 14, 529-45.

Zentner GE, Tesar PJ, Scacheri PC. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res *21,* 1273-83.

Zuker M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31,* 3406-15.

Yasutomo K, Doyle C, Miele L, Fuchs C, Germain RN. (2000). The duration of antigen receptor signalling determines CD4+ versus CD8+ T-cell lineage fate. Nature *404,* 506-10.

Chapter 2

Molecular players of *Satb1* alternative promoter switch during CD4SP T lineage differentiation in the thymus

2.1 Introduction

The two major subtypes of T lymphocytes in the periphery are $CD4^+$ helper and $CD8^+$ cytotoxic T lymphocytes. Proper development of these two T lymphocytes is essential in mounting a functional immune response, wherein the $CD4^+$ T-cells are responsible for the immune response against the extracellular pathogens and the parasites, whereas the $CD8^+$ T-cells mount immune response to the intracellular pathogens such as viruses and tumor antigens. During T-cell development in thymus, the $CD4^+$ and $CD8^+$ T lymphocytes develop from the double positive (DP) thymocytes which express both CD4 and CD8 co-receptors on their surface. During early T-cell development in the thymus, the double negative (DN) thymocytes which lack the surface expression of both CD4 and CD8 co-receptors, exhibit a partially rearranged TCR receptor and subsequently develop into $CD4^+CD8^+$ double positive (DP) thymocytes. At the DP stage, thymocytes undergo a complete rearrangement of T-cell receptors (TCR) and are scanned for the expression of proper functional TCRs on their surface by the process of positive and negative selection (Germain, 2002). As shown in the Figure 2.1.1, the TCRs of DP thymocytes are in frequently engagement with the self-peptide MHC complexes expressed by the thymic stromal cells such as the cortical thymic epithelial cells (cTECs) and the dendritic cells (Bousso *et al.*, 2002). During the process of TCR-self peptide MHC interactions, DP thymocytes which express TCRs with high affinity towards self-peptide MHC complexes, undergo apoptosis by the process of negative selection (Kappler, Roehm and Marrack, 1987; Sha *et al.*, 1988). Whereas the DP thymocytes with αβTCR with too low affinity for self-peptide MHC complexes undergo death by neglect (Mombaerts *et al.*, 1992). Only those DP thymocytes whose αβTCR have intermediate affinity for self-peptide MHC complexes differentiate into $CD4^+$ or $CD8^+$ SP thymocytes by the process of positive selection (Kisielow *et al.*, 1988; Kaye *et*

al., 1989; Hogquist *et al.*, 1994). By these processes only 3-5% of the cortical DP thymocytes survive (Egerton *et al.*, 1990; Goldrath and Bevan, 1999) and then migrate to the medullary region of the thymus (Witt *et al.*, 2005), wherein they undergo the differentiation towards the CD4/CD8 SP lineages.

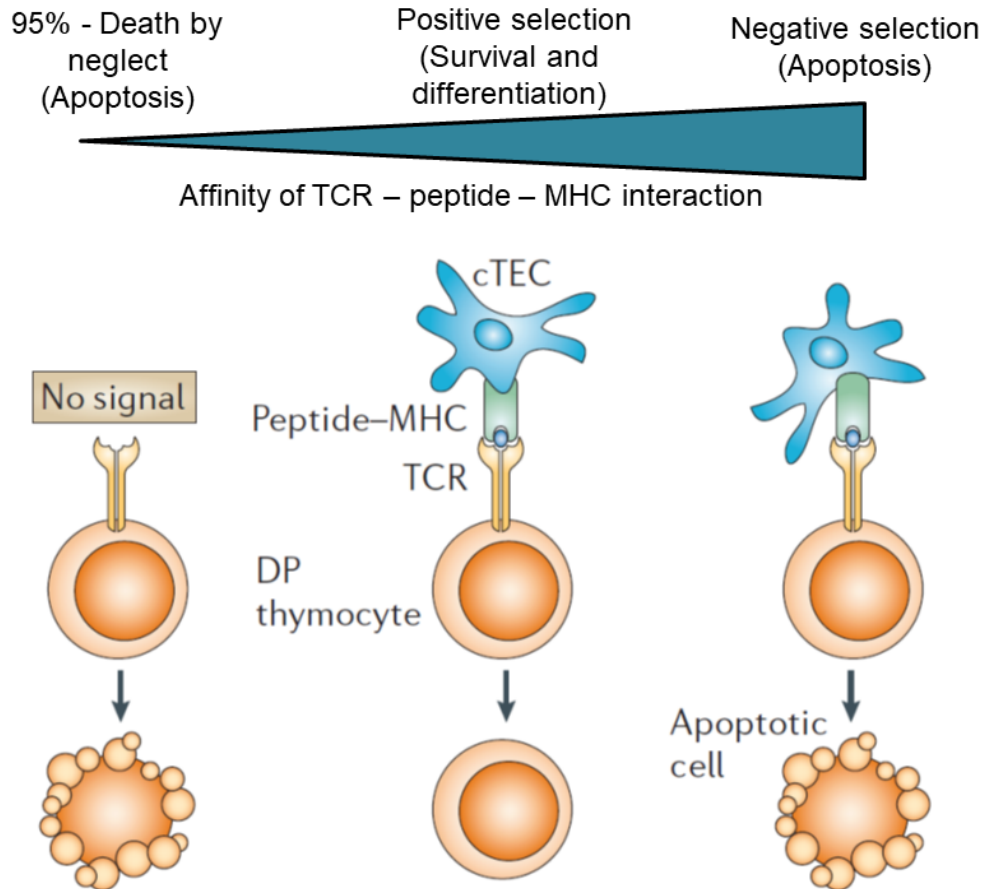


Figure 2.1.1. The fate of DP thymocytes during T-cell development is decided by the TCR signal strength. Upon TCRα chain rearrangement, the double positive (DP) thymocytes are scanned for the expression of a functional TCR receptor. DP thymocytes whose TCRs have very low affinity or no affinity for the self-peptide MHC complexes undergo apoptosis by the process of death by neglect. About 90 to 95% of the developing DP thymocytes undergo this process. DP thymocytes whose TCRs have the high affinity towards the self-peptide MHC complexes also undergo death by the negative selection, thereby preventing the auto-immune disorders. Now only those DP thymocytes whose TCRs display an intermediate affinity towards the self-peptide MHC complexes survive by the process of positive selection. (Adapted from Takahama, 2006).

During T-cell development, the positively selected double positive thymocytes differentiate into either CD4SP or CD8SP thymocytes depending on the interaction of their TCRs with the type of self-peptide MHC complexes. If the αβTCRs on DP thymocytes interact with the peptide-MHC-I complexes, they develop into $CD8^+$ T lineage. On the other hand, if they interact with the peptide-MHC-II complexes, they get diverted towards the $CD4^+$ T lineage (Figure 2.1.2) (Germain, 2002). The CD4SP development from DP thymocytes requires a stronger and longer duration of TCR signal, whereas the CD8SP T-cell development requires a weaker and shorter duration of the TCR signaling (Singer, Adoro and Park, 2008). The positively selected DP thymocytes are guided to differentiate into the SP thymocytes by means of the above-mentioned mechanisms and then they relocate from the cortical region to the medulla (Witt *et al.*, 2005). The newly generated SP thymocytes are called immature SP cells, which are characterized by the surface expression of CD69 and CD24 molecules (Figure 2.1.2). These immature SP thymocytes subsequently differentiate into completely mature SP T-cells and exhibit the downregulation of both CD69 and CD24 molecules on their surface (Figure 2.1.2).

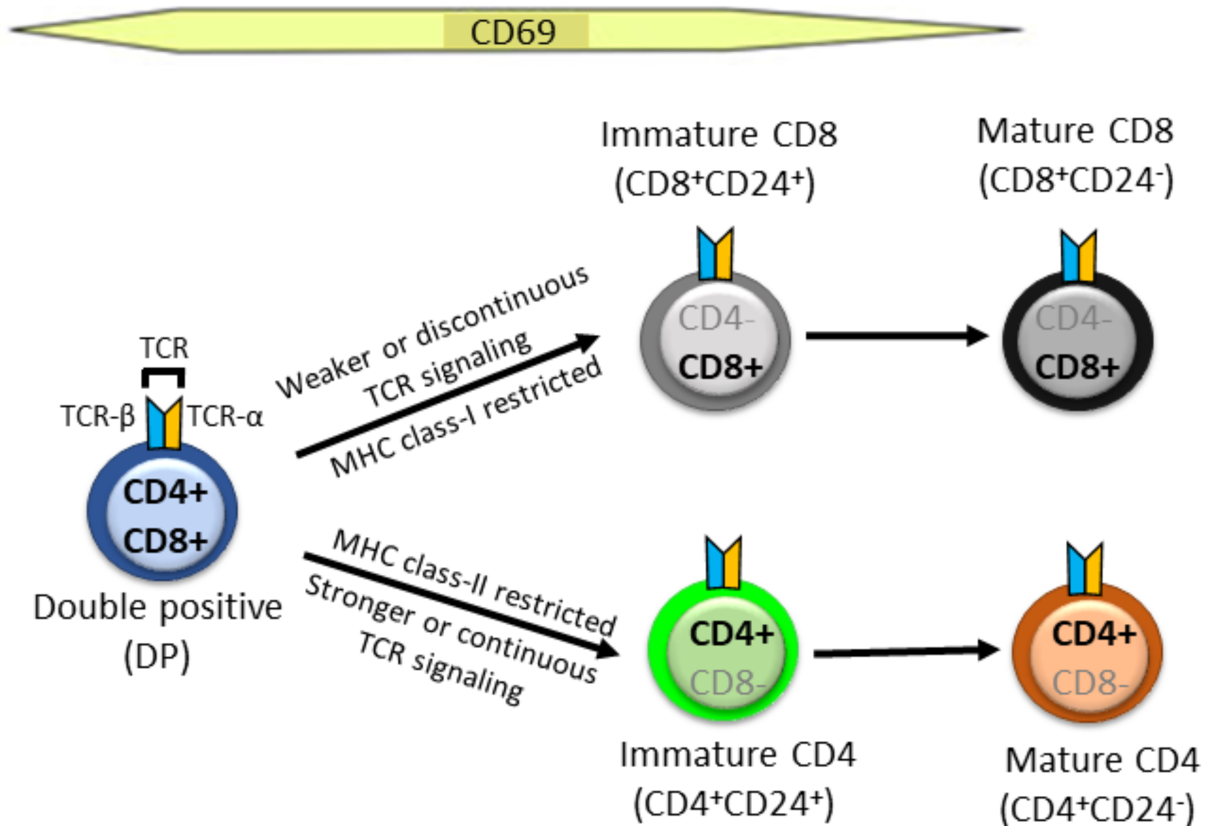


Figure 2.1.2. Differentiation of double positive thymocytes into CD4/CD8 SP T-cells. During T-cell development in the thymus, the TCR signaling plays an instrumental role in the survival, lineage commitment and differentiation of developing thymocytes. DP thymocytes upon their interaction with the MHC self-peptide complexes differentiate into either CD4 or CD8 SP T-cells. The interaction of TCR MHC-I self-peptide complexes leads to the commitment of DP thymocytes into the $CD8^+$ lineage. During this process, DP thymocytes receive discontinuous or a weaker TCR signaling. On the other hand, the interaction of DP thymocytes with MHC-II self-peptide complexes leads to their commitment towards the $CD4^+$ T lineage, which requires continuous or a stronger TCR signaling.

Along with the TCR signal, an interplay of the lineage specific transcription factors such as ThPOK and RUNX3 plays an important role in the lineage choice, and in the single positive (SP) T-cell differentiation (Figure 2.1.3). ThPOK is essential for the $CD4^+$ T lineage commitment, whereas RUNX3 is important for the $CD8^+$ T lineage commitment (Figure 2.1.3) (Taniuchi *et al.*, 2002; He *et al.*, 2005; Sato *et al.*, 2005). The lineage

specific transcription factors have crucial role not only in directing the SP T lineage differentiation but also in repressing the essential genes characteristic of the other lineage. The studies in the 1990s had revealed that the *CD4* gene is repressed by an intronic transcriptional silencer region (*CD4* silencer), on which Runx3 binds and represses the expression of CD4 in $CD8^+$ T-cells (Sawada *et al.*, 1994; Siu *et al.*, 1994; Taniuchi *et al.*, 2002). In contrast, Runx3 binds to an enhancer region at the *CD8* locus resulting in the transcriptional activation of the *CD8* gene during $CD8^+$ T lineage differentiation (Sato *et al.*, 2005). Runx3 is important for the reactivation of CD8 and silencing of CD4 during $CD8^+$ T lineage differentiation (Sato *et al.*, 2005). Thus, Runx3 plays a crucial role in balancing the expression of CD4 and CD8 during $CD8^+$ T-cell choice. On the other hand, ThPOK expression is limited to the $CD4^+$ T lineage and perturbation in the expression of ThPOK leads to severe reduction in the $CD4^+$ T lineage differentiation and resulting in the redirection of the $CD4^+$ T lineage into the $CD8^+$ T lineage (He *et al.*, 2005; Sun *et al.*, 2005). Ectopic overexpression of ThPOK leads to the forced redirection of the $CD8^+$ T lineage into the $CD4^+$ T lineage. These T lineage specific master regulators not only dictate the expression of lineage specific genes but also silence the key factors of the other lineage. Runx3 binds to the silencer region of ThPOK and represses its expression in CD8 cells (Figure 2.1.3) (He *et al.*, 2008; Setoguchi *et al.*, 2008). Therefore in case of Runx3 deficiency, ThPOK is derepressed, leading to the redirection of CD8 into $CD4^+$ T lineage (He *et al.*, 2008; Setoguchi *et al.*, 2008). ThPOK also represses the expression of Runx3, and loss of ThPOK leads to the redirection of CD4 into $CD8^+$ T lineage (Egawa and Littman, 2008). In addition to ThPOK, several other key transcription factors such as GATA3, Myb, TOX, and TCF1/LEF1 play an important role in the differentiation of $CD4^+$ T-cells in the thymus. These transcription factors act upstream to ThPOK in mediating the transcriptional activation of ThPOK (Figure 2.1.3). One of these transcription factors is GATA3 which directly binds to the *ThPOK* locus, and critical for the expression of ThPOK (Wang *et al.*, 2008). Since ThPOK depletion does not affect the GATA3 expression, GATA3 functions upstream of ThPOK during $CD4^+$ T-cell specification (Wang *et al.*, 2008). Depletion of GATA3 in DP thymocytes leads to an impairment of CD4 SP T-cell development (Hernández-Hoyos *et al.*, 2003; Pai *et al.*, 2003). The

expression of GATA3 is upregulated especially in the CD4SP T-cells but not in the CD8 T-cells during TCR mediated positive selection (Hernández-Hoyos *et al.*, 2003). Since ThPOK expression did not rescue the impairment in CD4 T-cell development in the absence of GATA3, indicating the importance of GATA3 not only for ThPOK expression but also for the commitment of the $CD4^+$ T lineage. ThPOK expression is regulated by another protein TOX, depletion of which leads to an impairment in the expression of ThPOK, resulting in the reduced $CD4^+$ T-cell number (Aliahmad and Kaye, 2008). However, the absence of TOX does not affect the expression of GATA3 (Aliahmad and Kaye, 2008). Myb is another transcription factor playing a role in the $CD4^+$ T lineage commitment (Maurice *et al.*, 2007). T lineage specific depletion of Myb resulted in the impairment of $CD4^+$ T lineage differentiation, with minimal effect on the $CD8^+$ T lineage (Bender *et al.*, 2004; Lieu *et al.*, 2004). Myb binds to the GATA3 promoter and activates its transcription, indicating that Myb acts upstream of GATA3 and ThPOK during $CD4^+$ T-cell differentiation (Maurice *et al.*, 2007). A recent report elucidated the importance of TCF1/LEF1 in $CD4^+$ T-cell differentiation (Steinke *et al.*, 2014). Depletion of both TCF1 and LEF1 in DP thymocytes resulted in an impairment of both ThPOK expression and $CD4^+$ T-cell differentiation, thus TCF1/LEF1 transcription factors act upstream of ThPOK. Although TCF1/LEF1 proteins are not important for the $CD8^+$ T-cell commitment from DP, these transcription factors repress the $CD4^+$ T lineage specific genes in the committed $CD8^+$ T-cells, therefore the conditional depletion of TCF1/LEF1 leads to the depression of $CD4^+$ T-cell specific genes in CD8 T-cells (Steinke *et al.*, 2014; Xing *et al.*, 2016).

The T lineage enriched global chromatin organizer SATB1 plays an important role in the development of SP T-cells from DP, thus SATB1 knockout (KO) results in the blockade of T-cell development at the DP stage. Furthermore, SATB1 plays an important role in specifying the thymic T-cell pool by transcriptional activation of the lineage specifying transcription factors (Kakugawa *et al.*, 2017). SATB1 is essential for the regulation of CD4 and ThPOK expression during CD4 commitment of the positively selected thymocytes, but is dispensable for the maintenance of ThPOK in the committed $CD4^+$ T-cells (Figure 2.1.3) (Kakugawa *et al.*, 2017). These observations confirm the essential role of SATB1 in $CD4^+$ SP T-cell development from DP.

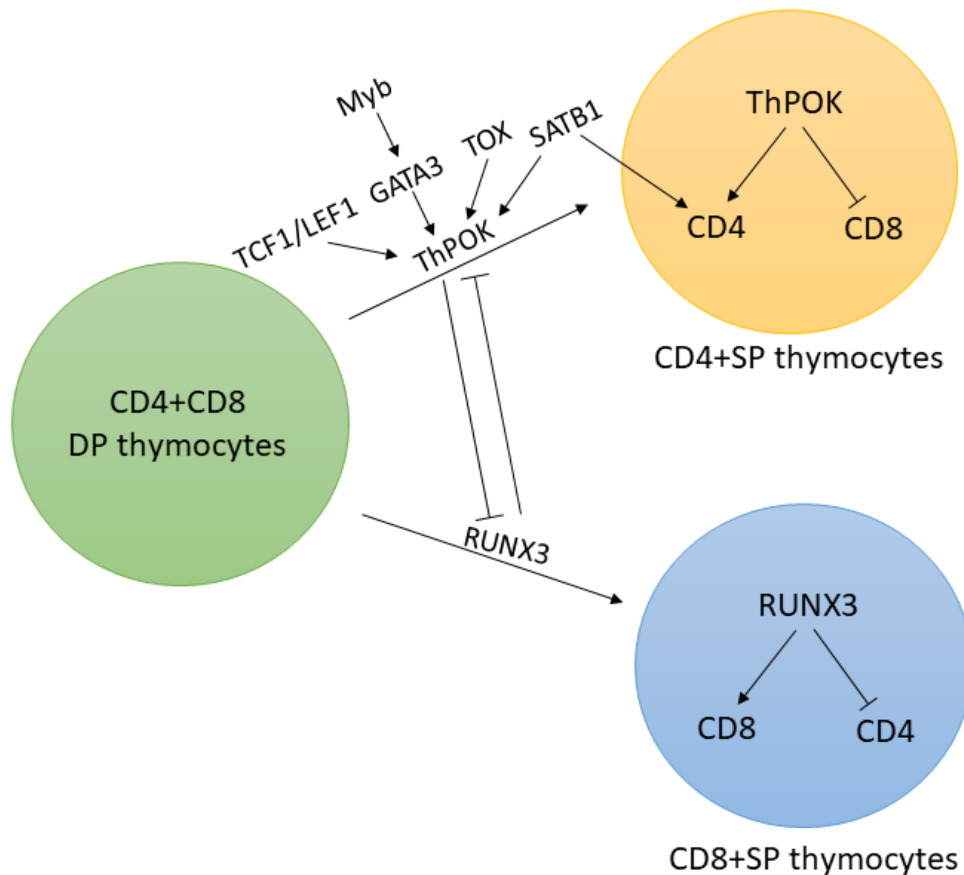


Figure 2.1.3. The role of transcription factors in CD4$^+$ T lineage commitment. ThPOK is the major transcription factor involved in the CD4SP T lineage commitment and differentiation, whereas RUNX3 is the master regulator of CD8$^+$ T lineage differentiation. While favoring the differentiation of lineage-committed cells, the major transcription factors simultaneously repress the signature genes of the opposite lineage. ThPOK represses the expression of CD8$^+$ T lineage genes such as Runx3 and CD8, whereas Runx3 is involved in the downregulation of CD4$^+$ T lineage specifying genes including ThPOK and CD4. As shown in the figure, the multitude of transcription factors such as GATA3, TOX, TCF1/LEF1, and SATB1 play a role in the transcriptional activation of ThPOK during CD4$^+$ T lineage commitment. Therefore, dysregulation of any of these factors results in the defective CD4$^+$ SP T-cell development.

Additionally, during T-cell development, we have observed that the SATB1 protein levels were increased in the immature CD4SP T-cells compared to DP (Chapter 1, Figure 1.2.6). As shown in Chapter 1, we found that the expression of a combination of the *Satb1* transcript variants plays an important role in maintaining the SATB1 protein

levels during the transition of DP into immature CD4SP T-cells. However, the mechanisms by which *Satb1* alternative promoter switch is mediated during the transition of DP into CDSP development is not understood. Therefore, the current study focuses on identifying the molecular players involved in the *Satb1* alternative promoter switch during these processes.

Here, we show that the expression of majority of the *Satb1* transcript variants is increased during the TCR mediated positive selection process and their levels were comparatively higher in the positively selected $CD69^{+}TCR^{hi}$ thymocyte population which includes the positively selected DP, CD4, and CD8SP. However, we show that the expression of the P2 transcript variant in CD4SP thymocytes was induced by the persistent TCR signaling during the development of T-cells. We also observe that the peripheral TCR signaling also induces the P2 promoter switch along with the predominant P3 variant. TCR signaling in the peripheral T-cells also leads to the downregulation of P1 transcript variant. The P4 variant expression was marginal in the TCR activated and naïve $CD4^{+}$ T-cells. These results suggest that *Satb1* is regulated by the TCR signal via selective alternative promoter usage. Furthermore, we observed that the *Satb1* P2 promoter exhibits the characteristic lineage specific chromatin accessibility during T-cell development from the progenitors. Finally, we identify that TCF-1 acts as a direct regulator of SATB1 expression, potentially by regulating the P2 promoter switch during CD4SP development from DP thymocytes.

2.2 Results

2.2.1 Expression pattern of *Satb1* transcript variants during TCR mediated positive selection process

Since the TCR signaling plays a pivotal role in the selection, CD4/CD8 T lineage commitment and differentiation of developing thymocytes, we first monitored the expression pattern of SATB1 protein and its transcript variants during TCR mediated positive selection process. Thymocytes that are undergoing the positive selection process were characterized based on the surface expression of CD69 and TCRβ as $CD69^{-}TCR^{-}$ which have not received the TCR signaling, $CD69^{+}TCR^{int}$, $CD69^{+}TCR^{hi}$ which are TCR engaged, positively selected thymocytes, and $CD69^{-}TCR^{hi}$ which represent the post positive selection thymocytes. The SATB1 protein levels, as shown previously (Gottimukkala et al., 2016), were elevated during TCR mediated positive selection and their levels were abundantly present in the positively selected $CD69^{+}TCR^{hi}$ population (Figure 2.2.1 A). We then FACS sorted thymocyte populations that are in different stages of positive selection and monitored the expression pattern of *Satb1* transcript variants using quantitative RT-PCR analysis. The expression profiling of *Satb1* transcript variants in the different stages of thymic selection process such as $CD69^{+}TCR^{int}$, $CD69^{+}TCR^{hi}$ and $CD69^{-}TCR^{hi}$, revealed that only $CD69^{+}TCR^{hi}$ population which includes TCR engaged DP, $CD4^{+}SP$ and $CD8^{+}SP$ shows the higher expression of majority of the *Satb1* transcript variants (Figure 2.2.1 B). These results suggest that total *Satb1* and also its transcript variants were expressed in the positively selected developing thymocytes, which include DP, $CD4^{+}SP$, and $CD8^{+}SP$. However, we reported earlier in Chapter 1 that the P2 transcript variant was expressed specifically in the $CD4^{+}SP$ thymocytes (Figure 1.2.5 B and C). Therefore we asked the question what confers the developmental stage specific or lineage specific expression of *Satb1* transcript variants, especially $CD4^{+}$ T lineage specific expression of P2 transcript variant along with P3 during T-cell development.

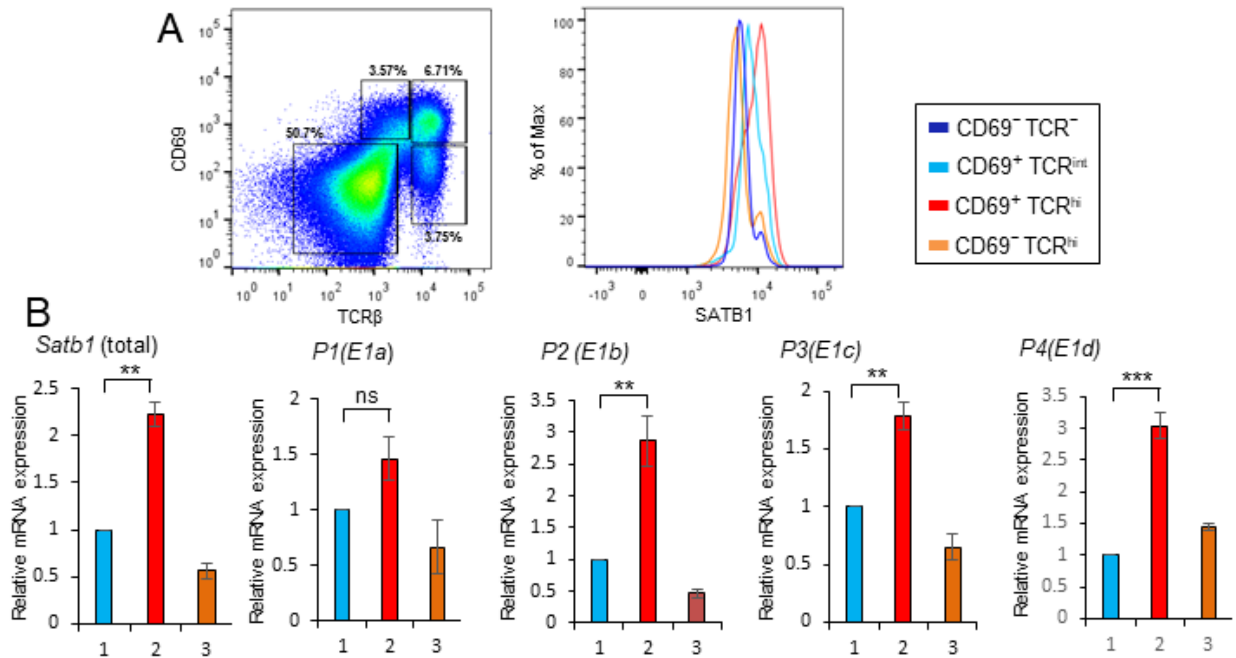


Figure 2.2.1. The expression pattern of *Satb1* transcript variants during positive selection of thymocytes. (A) The flow cytometry analysis of SATB1 expression in thymocytes undergoing positive selection. Single cell suspension was prepared from thymii of three week old C57BL/6 mice and subjected to the surface staining using the following antibodies: anti-CD4, anti-CD8, anti-CD69, and anti-TCRβ. After surface staining, thymocytes were permeabilized, and subjected to the intracellular staining with anti-SATB1 antibody using the FOXP3 intracellular staining kit. Thymocytes were analyzed using the BD FACS canto II flow cytometer. The results shown are from three independent experiments (N=3). **(B)** Thymocytes were subjected to the surface staining with anti-CD4, anti-CD8, anti-CD69, and anti-TCRβ. $CD69^{+}TCR^{int}$, $CD69^{+}TCR^{hi}$, and $CD69^{-}TCR^{hi}$ thymocyte subpopulations were FACS sorted and used for RNA isolation and cDNA synthesis. Quantitative RT-PCR analysis was performed to monitor the expression pattern of total *Satb1* mRNA and *Satb1* transcript variants (N=3) in these thymocyte subpopulations. Sequences of the primers used for the qRT-PCRs are listed in the methods section 2.4.5. The results shown are from three independent experiments (N=3).

2.2.2 Persistent TCR signaling induces *Satb1* alternative promoter switch during T-cell development

During development of thymocytes, the pre-TCR signaling plays an important role in the differentiation of DN4 stage into DP. DP thymocytes undergo the rearrangement of TCRα chain, which then assembles with the pre-existing TCRβ chain to form the complete functional αβTCR receptor on their surface (Kearse, Roberts and Singer, 1995). Persistent TCR signaling of DP thymocytes results in the differentiation of DP thymocytes towards the $CD4^+$ T lineage, whereas cessation of TCR signaling leads to their differentiation towards the $CD8^+$ T lineage (Yasutomo *et al.*, 2000; Singer, 2002; Singer, Adoro and Park, 2008). Additionally, during development the TCR+CD4 combination generates the stronger TCR signal than the TCR+CD8 combination. Therefore comparatively stronger TCR signals facilitate the differentiation of DP into the $CD4^+$ T lineage, whereas weaker TCR signals allow the differentiation of DP toward the $CD8^+$ T lineage (Itano et al., 1996; Basson et al., 1998). Furthermore, the cortical DP thymocytes are very sensitive and respond rapidly to the low affinity TCR ligands than SP thymocytes which respond to the high affinity TCR ligands presented in the medullary region of the thymus (Davey et al., 1998; Hogquist and Jameson, 2014). Therefore we evaluated whether these differences in the TCR signal received during various stages of thymocyte development might exert any effect on the *Satb1* alternative promoter switch. We first assessed the expression levels of SATB1 protein in the TCR engaged DP and $CD4^+$SP thymocyte populations that express CD69 on their surface (Figure 2.2.2 A). We observed that the SATB1 protein levels were significantly higher in the $CD69^+$CD4SP population than $CD69^+$DP, and their levels were further downregulated in the $CD69^-$CD4SP population (Figure 2.2.2 A). The $CD69^+$CD4SP thymocytes exhibit higher expression of P2 transcript variant along with P3 compared to the $CD69^+$DP, and $CD69^-$CD4SP thymocytes (Figure 2.2.2 B), indicating that the P2 promoter switch was observed only in the TCR engaged CD4SP thymocytes but not in the TCR engaged DP. These observations suggest that during development, the persistent and comparatively high affinity TCR signaling induce a switch in the *Satb*1 alternative promoter usage, and this promoter-switch leads to the higher SATB1 protein levels, which in turn might enable SATB1 to regulate its target genes to a higher degree.

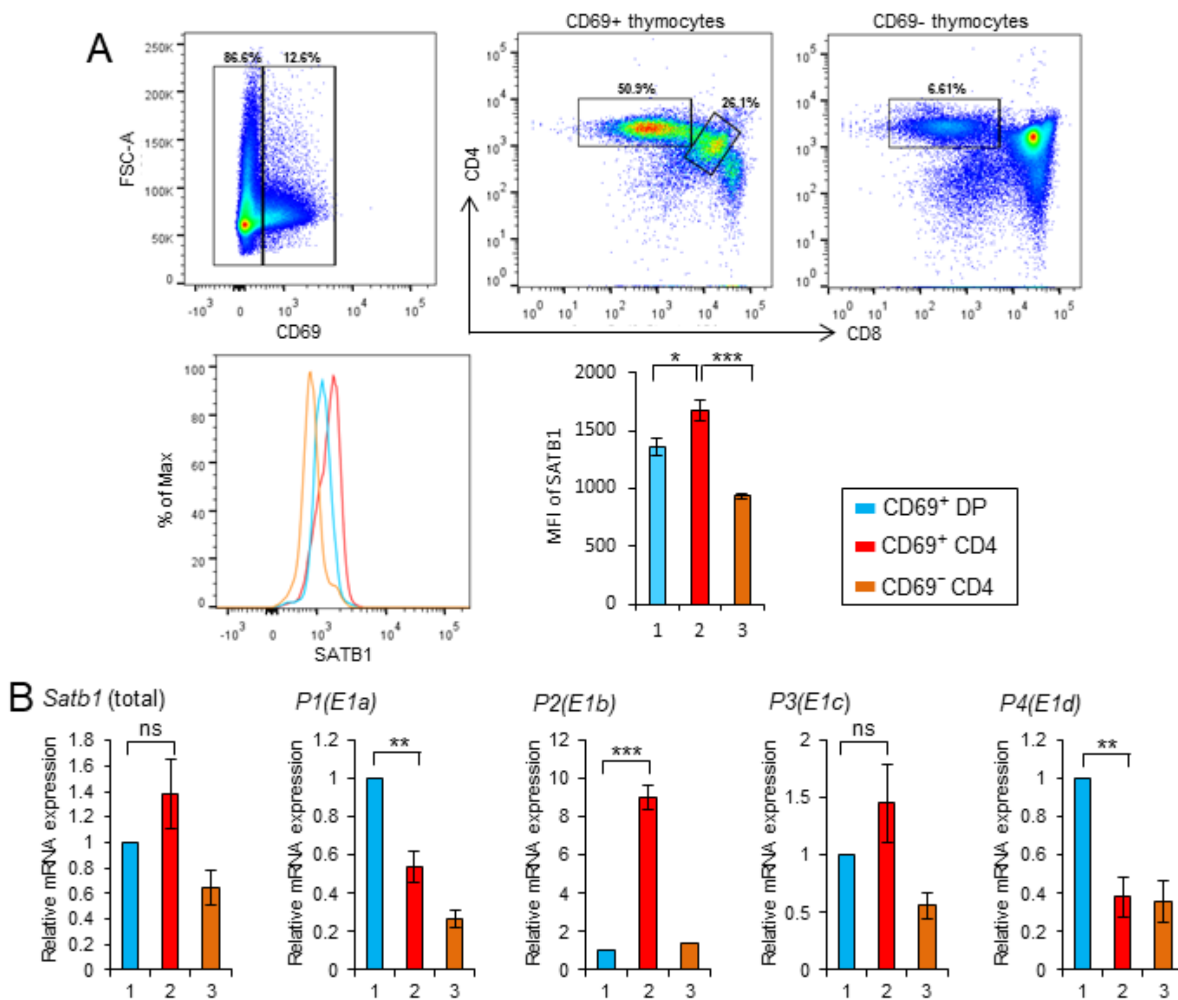


Figure 2.2.2. Persistent TCR signals induce the *Satb1* promoter switch in the CD4SP thymocytes. (A) The flow cytometry analysis of SATB1 in $CD69^{+}$DP, $CD69^{+}$CD4SP, and $CD69^{-}$CD4SP thymocytes. Three week old C57BL/6 mice were used for isolation of thymus. Single cell suspension was prepared from thymii and subjected to surface staining with the following antibodies: anti-CD4, anti-CD8, and anti-CD69. After surface staining, thymocytes were permeabilized, and subjected to the intracellular staining with anti-SATB1 antibody using the FOXP3 intracellular staining kit. Thymocytes were subjected to FACS analysis using FACS canto II analyzer. The results shown are from three independent experiments (N=3). P-values were calculated using student's t-test. **(B)** Thymocytes were subjected to the surface staining with anti-CD4, anti-CD8, and anti-CD69. The $CD69^{+}$DP, $CD69^{+}$CD4SP, and $CD69^{-}$CD4SP populations were FACS sorted and used for RNA extraction and cDNA synthesis. The qRT-PCR analysis was performed to monitor the expression pattern of total *Satb1* mRNA and its variants in these thymocyte subpopulations. Sequences of oligonucleotide primers used for the qRT-PCRs are listed in the methods section 2.4.5. The results shown are from three independent experiments (N=3).

2.2.3 Lineage specific chromatin accessibility of *Satb1* P2 promoter region during T-cell development from progenitors

Since the expression of *Satb1* transcript variants was found to be cell type specific during T-cell development, we assessed the chromatin dynamics of *Satb1* alternative promoters during their development from the progenitors by using publicly available ATAC-seq data. ATAC-seq is an assay of chromatin accessibility using a transposase coupled with high-throughput sequencing (Heng *et al.*, 2008; Shih *et al.*, 2016). During development, a precursor hematopoietic stem cell (HSC) is differentiated into a multipotent progenitor (MPP), which then differentiates into a common lymphoid precursor (CLP) (Figure 2.2.3 A). CLP has bi-potential to differentiate into the B-cell lineage in the bone marrow or the T-cell lineage once it enters the thymus (Figure 2.2.3 A). We observed that the chromatin region around the P1, P3, and P4 promoters of *Satb1* exists in an open chromatin configuration as indicated by the peaks of accessibility at these alternative promoter regions of *Satb1* locus. The peak height which corresponds to the openness of chromatin varies throughout the development from the precursor HSCs to MPPs, CLPs, B-cells and T-cells (Figure 2.2.3 B). However, as shown in the Figure 2.2.3 B, the chromatin region of the P2 promoter was found to be in the closed state, as indicated by negligible peaks at the P2 region in the precursors (Figure 2.2.3 B, boxed region around P2). Unlike other promoters (Figure 2.2.3 B, boxed regions around P1 and P3 promoters), the P2 promoter region was not accessible in the B-cell lineage as well, indicating its specificity only towards the T-cell lineage (Figure 2.2.3 B).

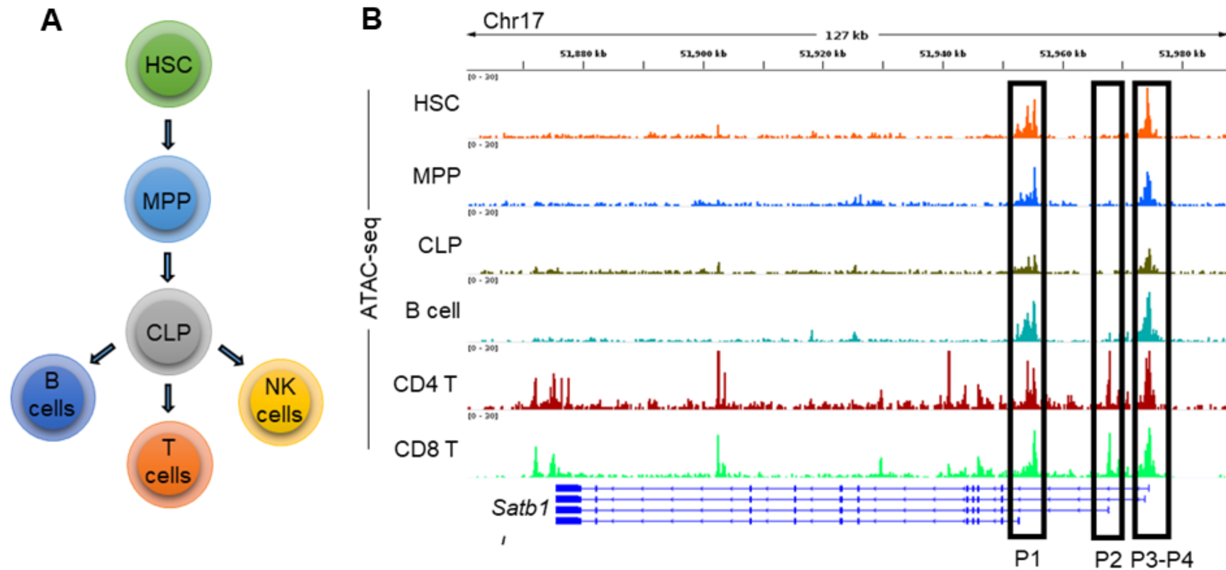


Figure 2.2.3. Lineage-specific chromatin accessibility at the *Satb1* alternative promoters during development from progenitors. During the development of the immune cells, a precursor hematopoietic stem cell (HCS) in the bone marrow differentiates into a multipotent progenitor cell (MPP), which then develops into a common lymphoid precursor (CLP) that can further differentiate either into a T-cell lineage or a B-cell lineage. We used publicly available ATAC-seq datasets performed in HSCs, MPPs, CLPs, B-cells and T-cells for analyzing the state of chromatin accessibility at the *Satb1* gene locus on chromosome 17. The *Satb1* alternative promoter regions (P1, P2, P3, and P4) are marked with rectangular boxes at the *Satb1* gene locus. ATAC-seq peaks at the *Satb1* gene locus in various cell types during the development in this compartment are depicted using different colors.

2.2.4 Developmental stage specific chromatin dynamics of the *Satb1* alternative promoters during T-cell development

As mentioned above the chromatin accessibility of *Satb1* alternative promoters exhibit lineage specificity during development from the precursor HSCs, we then asked what happens to the chromatin dynamics of *Satb1* alternative promoters during T-cell development in the thymus (Heng *et al.*, 2008; Shih *et al.*, 2016). Among *Satb1* alternative promoters, the P1, P3 and P4 promoter regions exhibit the chromatin accessibility throughout the development as indicated by the peaks at these regions (boxed regions around P1, and P3 in different developing stages of T-cells). However, the chromatin accessibility at the P2 promoter region is observed from the DN3 stage

onwards based on the appearance of accessibility peak at the P2 promoter (see boxed region around P2) at DN3, correlating with the occurrence of pre-TCR signaling. Further, the P2 promoter accessibility increases in the later developmental stages, especially during the SP lineages as indicated by an increase in the height of the peak at this promoter region (Figure 2.2.4, boxed region around P2). Among the various T-cell developmental stages, only the CD4SP stage exhibits comparatively higher P2 promoter accessibility. Although we also observe that the P2 promoter region is accessible in the CD8SP cells, the expression of P2 transcripts was found predominantly in the CD4^{+} T lineage but not in the CD8^{+} T lineage, presumably suggesting that repressor proteins might bind around or downstream to the TSS of P2 in the CD8 T-cells, thereby inhibiting the P2 transcription. These results confirm that the chromatin accessibility at the *Satb1* alternative promoters is developmental stage specific during T-cell development. Therefore, it would be interesting to further study the role of development stage specific transcription factors in the regulation of *Satb1* alternative promoters during T-cell development.

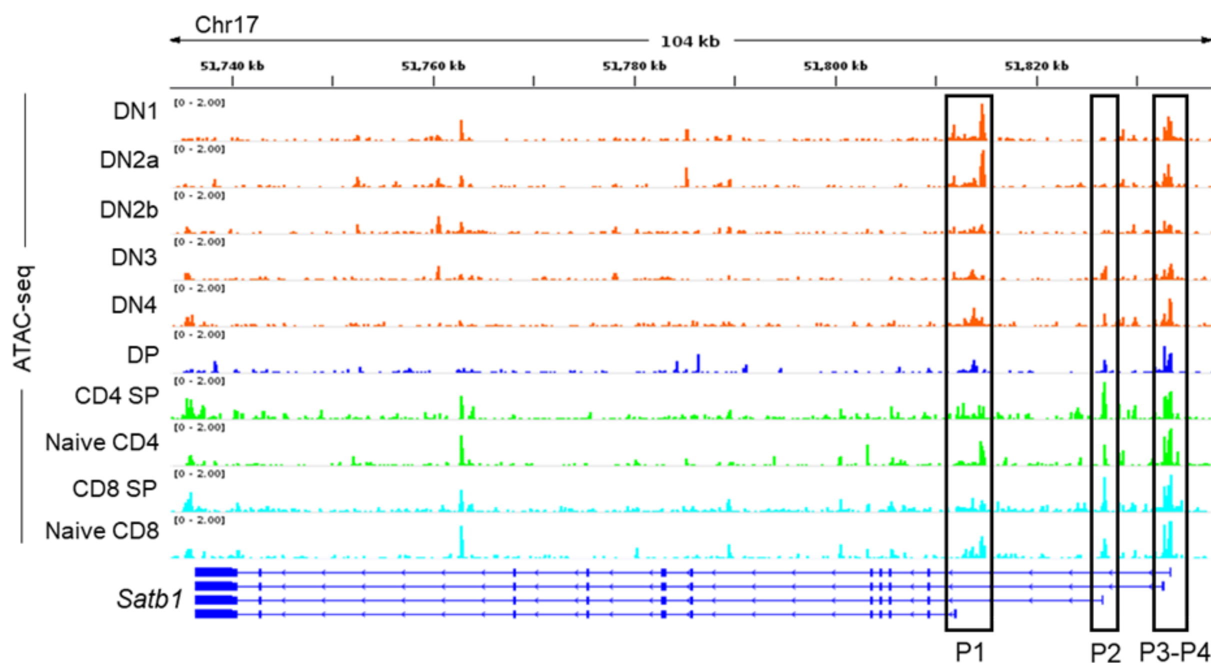


Figure 2.2.4. Lineage specific chromatin dynamics at the *Satb1* alternative promoters during T-cell development. During T-cell development, a progenitor CLP migrates to the thymus and in the thymic

microenvironment it is directed towards the T-cell lineage. The T-cell development is a multistep process which includes the initial double negative stages (DN-DN4). At DN3 stage, the pre-TCR signaling and thus the β-selection leads to the development of DN4 which then be developed into the double positive stage (DP). Upon TCR mediated signaling, the DP thymocytes become differentiated into either the CD4SP T-lineage or the CD8SP T-cells. To address the chromatin dynamics of *Satb1* alternative promoters, we analyzed the publicly available ATAC-seq datasets performed in the DN1, DN2a, DN2b, DN3, DN4, DP, CD4SP, naïve CD4SP, CD8SP and naïve CD8SP populations. The *Satb1* alternative promoter regions (P1, P2, P3, and P4) were marked by rectangular boxes at the *Satb1* gene locus. ATAC seq peaks at the *Satb1* gene locus in the different-cell types of T-cell development were shown by indicated colors.

2.2.5 TCF1 regulates the expression of SATB1 by binding to the P2 promoter in CD4SP thymocytes

Since the P2 promoter switch was identified during the differentiation of DP into the CD4SP lineage, we then asked whether a specific transcription factor might be involved in regulating this P2 promoter switch in the developing thymocytes. The development of $CD4^+$ T-lineage from DP thymocytes is influenced by various transcription factors such as Th-POK, GATA3, c-Myb, and Tox (He *et al.*, 2005; Maurice *et al.*, 2007; Aliahmad and Kaye, 2008; Wang *et al.*, 2008). Recent reports also suggest that the loss of TCF1 and LEF1 results in the impairment of CD4SP from DP, and causes the redirection of CD4SP into CD8SP (Steinke et al., 2014). We hypothesized that either of these $CD4^+$ T-lineage specifying transcription factors might regulate the *Satb1* P2 promoter switch, and thus resulting in the expression of CD4SP specific *Satb1* transcript variants such as P2 along with constitutively expressed predominant P3 variant. Interestingly, using in-silico analysis we identified a canonical TCF1-LEF1 DNA-binding motif within the P2 promoter of *Satb1* (Figure 2.2.5 A). To find whether TCF1 binds to this region in vivo, we FACS sorted CD4SP thymocytes and performed TCF1-ChIP using the chromatin isolated from CD4SP thymocytes. We found that indeed TCF1 binds to the P2 promoter of *Satb1* in CD4SP thymocytes (Figure 2.2.5 B). We also analyzed the publicly available data of TCF1 ChIP-seq performed using total thymocytes (Dose et al., 2014) and

observed the occupancy TCF1 on the P2 promoter, wherein the TCF1-LEF1-binding motif was identified (Figure 2.2.5 C).

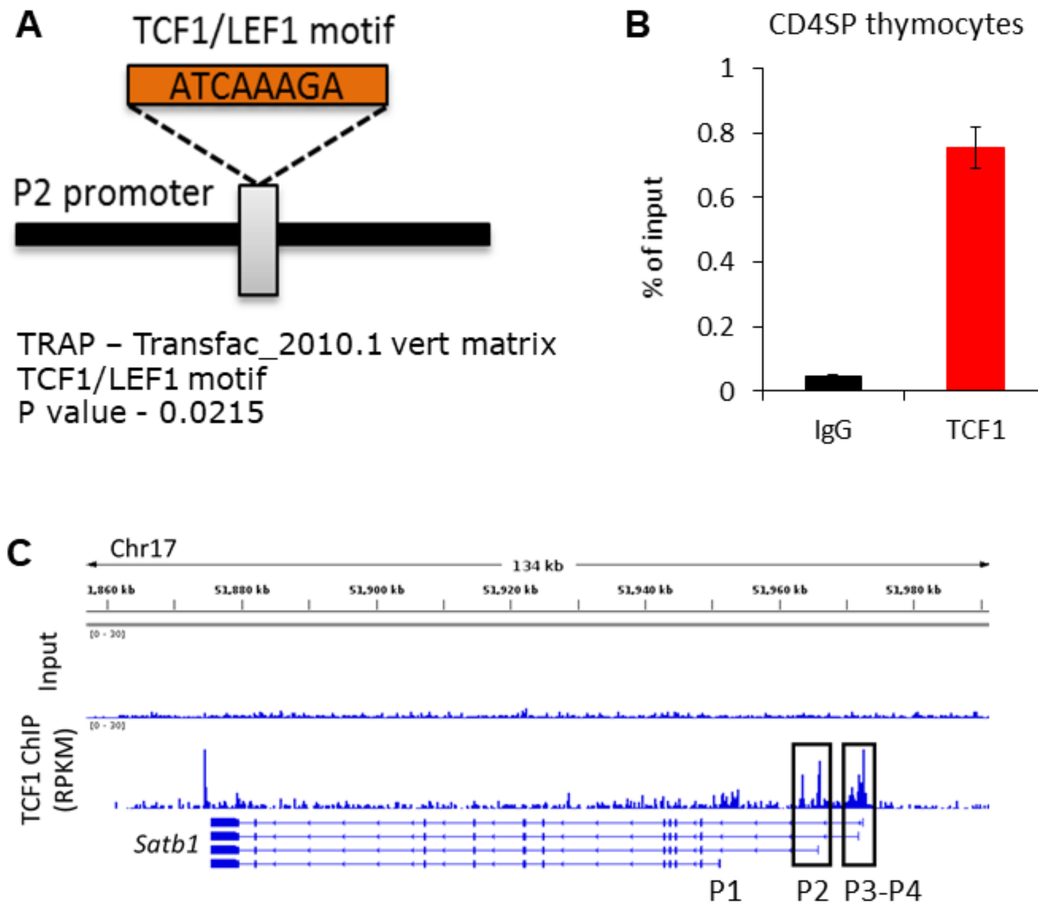


Figure 2.2.5. Occupancy of TCF1 on the P2 promoter of *Satb1* in the CD4SP thymocytes. (A) TCF1-LEF1 canonical DNA-binding motif was identified in the promoter region of P2 by using TRAP, a Transfac based transcription factor binding site finding tool. **(B)** Three week old C57BL/6 mice were used to prepare the single cell suspension of thymocytes and subjected to the surface staining with anti-CD4 and anti-CD8. The CD4SP thymocyte subpopulation was FACS sorted and used for crosslinking with 1% formaldehyde. The crosslinked nuclei were used for the chromatin immunoprecipitation (ChIP) with anti-TCF1 antibody. The TCF1-bound DNA was enriched and purified and used for qRT-PCR analyses. qRT-PCR analysis was performed using primers spanning the canonical TCF1-LEF1 binding site within the P2 promoter. Analysis of TCF1-bound P2 promoter region was plotted as % of input as indicated. **(C)** The data from publicly available TCF1 ChIP-seq performed in the mouse thymocytes was analyzed at the alternative promoter regions P1, P2, P3 and P4 of *Satb1* gene locus. TCF1-bound regions are indicated by the black colored rectangular boxes.

2.2.6 CD4SP thymocytes from TCF1 null mice exhibit the downregulation of *Satb1* expression

Since we observed the binding of TCF1 on *Satb1* P2 promoter region, we then asked whether TCF1 is required for the expression of the P2 transcript variant in the CD4SP thymocytes and thereby maintaining the SATB1 protein levels. Towards this, we FACS sorted DN, DP, CD4, and CD8SP thymocytes from WT and TCF1-knockout (KO) mice and used for the gene expression profiling of *Satb1* transcript variants using qRT-PCR. We observed that the expression of both the P2 transcript variant as well as the total *Satb1* transcript levels were significantly downregulated in the CD4SP thymocytes from TCF1-KO mice compared to the wild type (Figure 2.2.6 A and B). We found that the expression of P2 transcript variant and also total *Satb1* transcript level were upregulated in the DP thymocytes from TCF1-KO mice (Figure 2.2.6 A and B). Further, we observed a stark reduction in the SATB1 protein levels in the CD4SP T-cells from TCF1-KO mice compared to that of the wild type mice (Figure 2.2.6 C). Surprisingly, SATB1 protein levels were increased in case of DP thymocytes from TCF1-KO mice (Figure 2.2.6 C). These results indicate that TCF1 might differentially regulate the activity of the P2 promoter in a developmental stage-dependent manner by directly binding to the P2 promoter or via interaction with different protein partners. We then further assessed the importance of TCF1 in regulating the transcriptional activity of P2 promoter. To test this, we cloned the P2 promoter of *Satb1* in the pGL3 basic vector and then overexpressed this reporter construct in the presence or absence of TCF1 overexpression in HEK 293T cells. We then monitored the P2 promoter activity by measuring the luciferase activity in a dual luciferase assay with Renilla luciferase as an internal control for the normalization. We observed 2.3-fold increase in the promoter activity of P2 in the presence of TCF1 as evidenced by the luciferase activity assay (Figure 2.2.6 D). Taken together, these experiments unequivocally confirm that TCF1 is important for P2 promoter switch in CD4SP thymocytes, thereby regulating the expression of *Satb1* during T-cell development. Therefore our results demonstrate that TCF1 might act upstream of SATB1 during the development of CD4SP thymocytes (Figure 2.2.6 E).

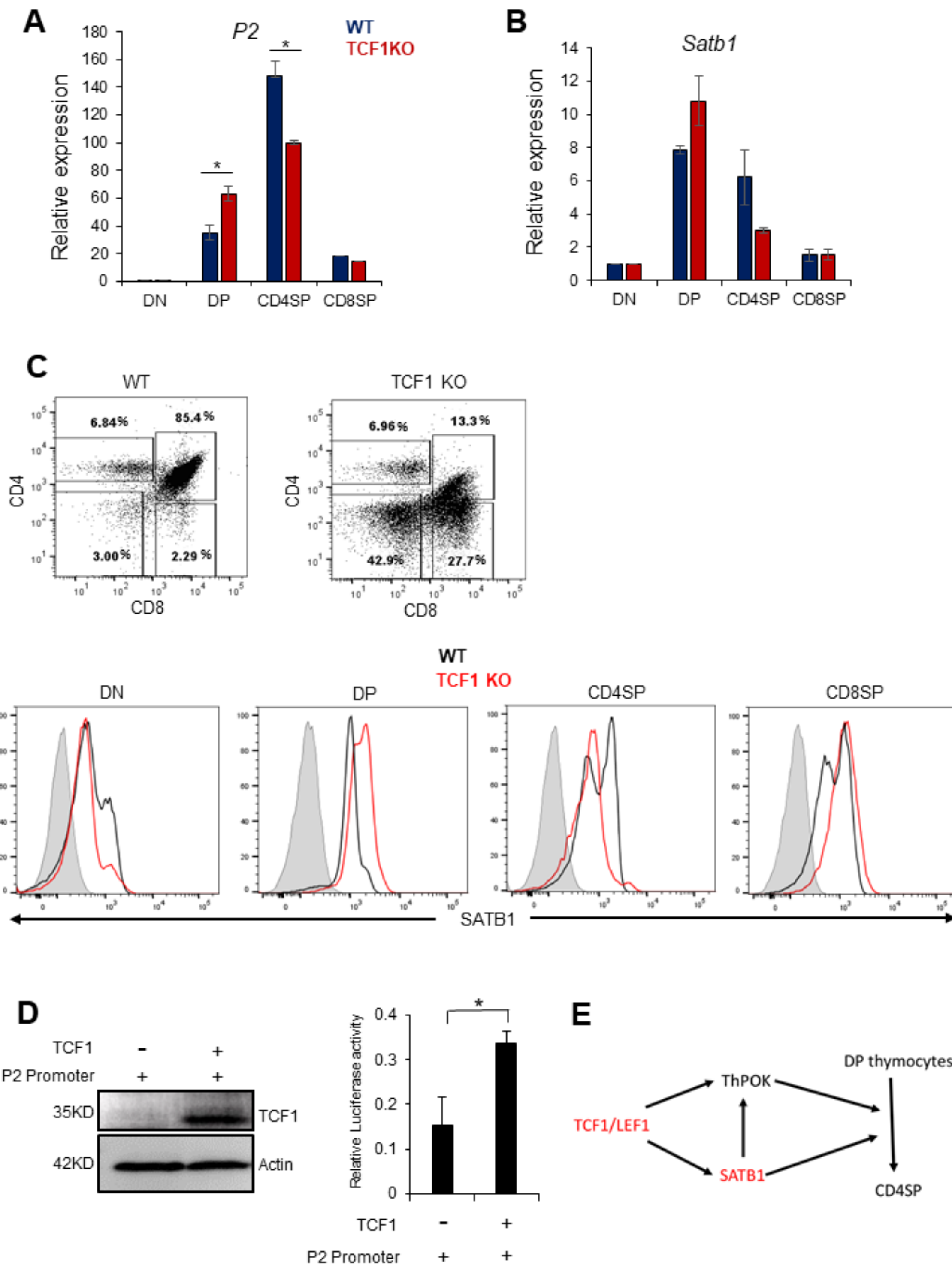


Figure 2.2.6. TCF1 regulates the expression of SATB1 in CD4SP thymocytes during T-cell development. (A and B) Three week old TCF1-KO and wild type mice were used for the preparation of

ingle cell suspension of thymocytes. Thymocytes were used for the surface staining with anti-CD4 and anti-CD8 antibodies. DN, DP, CD4SP, and CD8SP thymocytes were FACS sorted from TCF1-KO as well as wild type (WT) mice. Sorted thymocytes subpopulations were used for RNA extraction and cDNA synthesis. Further qRT-PCR analyses were performed to monitor the expression profile of P2 transcript variant and the total *Satb1* mRNA in CD4SP thymocytes along with other developmental stages from WT and TCF1-KO. The expression of Hprt is used as an endogenous control. **(C)** Single cell suspension of thymocytes was used for the surface staining with anti-CD4 and anti-CD8 antibodies. Surface stained thymocytes were permeabilized and used for the intracellular staining with anti-SATB1 antibody using the FOXP3 intracellular staining kit. The flow cytometry analyses of SATB1 protein in thymocyte developmental stages such as DN, DP, CDSP, and D8SP from wild type and TCF1-KO mice is depicted. **(D)** *Satb1* P2 promoter was cloned into the pGL3 basic vector and expressed in HEK 293T cells in the background of TCF1 overexpression. Renilla was used as transfection control. The data is from three independent experiments and is shown as means of ± SEM. *p<0.05. **(E)** The schematic depiction of TCF1 mediated SATB1 regulation during CD4SP differentiation during T-cell development in the thymus.

2.2.7 TCR signaling induces *Satb1* alternative promoter switch in the peripheral $CD4^+$ T-cells

Since the expression *Satb*1 P2 transcript variant was found to be TCR signal-dependent during T-cell development, we further evaluated whether similar regulation occurs in the peripheral $CD4^+$ T-cells. We therefore performed TCR stimulation of naïve $CD4^+$ T-cells from the periphery (spleen) using plate-bound anti-CD3 and soluble anti-CD28 for 48 hrs. SATB1 protein levels were increased upon TCR activation, as shown by the western blotting and flow cytometry analysis (Figure 2.2.7 A). Total *Satb*1 mRNA levels were also increased in the TCR activated $CD4^+$ T-cells compared to the naïve $CD4^+$ T-cells (Figure 2.2.7 A). When tested for the expression of *Satb1* transcript variants upon activation, the expression of P2 transcripts along with the P3 was higher compared to the naïve $CD4^+$ T-cells (Figure 2.2.7 B). In contrast, the expression of the P1 transcript variant was downregulated and a very slight increase was observed in case of P4 transcript variant upon TCR activation (Figure 2.2.7 B). These results again confirm that the expression of P2 transcript variant along with P3 is required to maintain the SATB1 protein levels during peripheral T-cell activation as well.

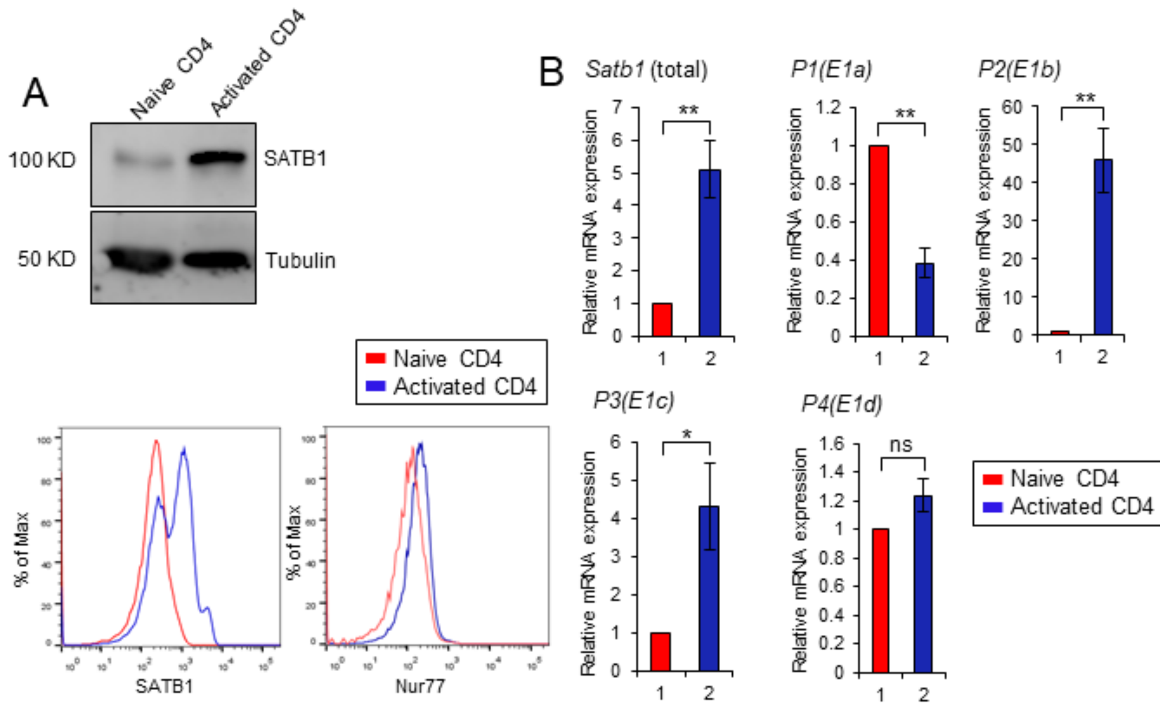


Figure 2.2.7. TCR signaling mediates *Satb1* alternative promoter switch in the peripheral CD4$^+$ T-cell. (A) Single cell suspension of spleenocytes were prepared using spleen from 6 week old mice and subjected to the magnetic sorting of CD4$^+$ T-cells by using mouse CD4$^+$ T-cell isolation kit (BD Biosciences). Splenic T-cells were subjected to the surface staining with anti-CD4, anti-CD8, and anti-CD25 and subjected to FACS sorted to isolate naïve CD4$^+$ T-cells. Naïve CD4$^+$ T-cells were cultured in the presence of plate-bound anti-CD3 and soluble anti-CD28 for 48 hrs. SATB1 protein levels were monitored in both naïve and activated CD4$^+$ T-cells by immunoblotting and by flow cytometry analysis. Simultaneously, CD4$^+$ T-cells were also stained for Nur77 and flow cytometry analysis was performed. **(B)** Naïve and TCR activated cells were used for RNA extraction and cDNA synthesis. qRT-PCR analyses of total *Satb1* and its transcript variants were performed on naïve and TCR activated CD4$^+$ T-cells. The presented data is from three independent experiments (N=3). The primers used for qRT-PCR analysis were listed in the methods section 2.4.5.

2.3 Discussion

T-cells play an essential role in the cell mediated immune response. The development of T-cells is a dynamic and multistep process during which a lymphoid progenitor undergoes a series of development to generate the functional mature CD4 and CD8 T-cells in thymus. It has been reported previously that lack of a T-cell enriched chromatin organizer SATB1 resulting in the blockade of T-cell development at the double positive (DP) stage (Alvarez *et al.*, 2000; Kondo *et al.*, 2016). During T-cell development, SATB1 is essential for TCR mediated T-cell selection, and immune tolerance (Kondo *et al.*, 2016). SATB1 regulates the expression of genes essential for T-cell development and function (Alvarez *et al.*, 2000; Hao *et al.*, 2015; Kakugawa *et al.,* 2017). Previously we have shown that SATB1 protein is induced by TCR signaling during T-cell development and in the peripheral T-cell activation (Gottimukkala et *al.,* 2016). Therefore understanding the transcriptional regulation of SATB1 during T-cell development is crucial. In Chapter-1, we demonstrated that SATB1 is regulated by an alternative promoter switch in a developmental stage specific manner during T-cell development, wherein we found that the *Satb1* transcript variants express in a combinatorial manner in a cell type specific manner to maintain the levels of SATB1 protein. It was reported earlier that approximately 30-50% of mouse and human genes are regulated by the usage of alternative promoters (Baek et al., 2007; Kimura et al, 2006; Sun et al., 2006). Alternative promoter usage plays an important role in the transcriptional regulation of gene expression in a cell or developmental stage specific manner, thus contributing to the complexity of gene expression regulation. The usage of alternative promoters by immune system plays a crucial role in the physiological function. Especially in case of T-cells, a lymphocyte specific gene Lck is regulated in a lineage specific manner by the usage of two alternative promoters (Voronova, 1987), where in the proximal promoter of Lck is used only in the developing thymocytes and is silenced in the peripheral T-cells, whereas the distal promoter was used by both the developing T-cells and peripheral T-cells (Wildin, 1991; Reynolds, 1990). Further, dys-regulation in the usage of Lck alternative promoters was observed in several malignancies (Garvin et al., 1988; Sartor et al., 1989). Similarly, the usage of alternative promoters at the *Notch* gene locus results in the activation of ligand independent unconventional Notch signaling during T-

cell development (Gómez-del Arco, 2010). Tissue or lineage specific usage of the alternative promoters at the *GATA3* gene locus was observed and it was also found that GATA3 alternative promoters are used selectively in the thymus and in the differentiated Th2 cells (Asnagli, 2002). Developmental stage specific regulation of alternative promoter switch plays an important role in the transcriptional regulation of Runx1 during hematopoiesis and also increases the complexity to the Runx1 functions (Pozner, 2007; Bee et al., 2010).

Since *Satb1* is regulated in a developmental stage specific manner by the alternative promoter switch, thus investigating the signaling and the molecular players involved in the cell type specific expression of *Satb1* transcript variants during T-cell development is essential to understand the physiological significance of this effect. Therefore, the current study focuses to elucidate the molecular players of *Satb1* alternative promoter switch in the developing thymocytes. As shown in the Figure 2.3, we have demonstrated the mechanism of *Satb1* alternative promoter switch during double positive (DP) to CD4SP T-cell differentiation in the thymus. T-cell development in the thymus is a highly dynamic process, and is dependent on the TCR signaling. The pre-TCR signaling at the late DN stage leads to the development of DP thymocytes. At DP stage of development, the TCRα chain rearrangement takes place and then TCRα chain assembles with TCRβ chain to form complete TCR on the surface of DP thymocytes (Kearse, Roberts and Singer, 1995). Now, DP thymocytes will be scanned for the expression of functional TCRs on their surface by interacting with the self-peptide MHC complexes expressed by the thymic stromal cells. DP thymocytes, whose TCRs have strong affinity for the self-peptide MHC complexes and receive stronger TCR signaling will undergo negative selection (Kappler, Roehm and Marrack, 1987; Sha *et al.*, 1988), whereas those with TCRs which don't recognize the self-peptide MHC complexes will undergo apoptosis by the process of death by neglect (Mombaerts *et al.*, 1992). Now, the DP thymocytes with TCRs that recognize the self-peptide MHC complexes with intermediate affinity will survive and undergo positive selection (Kisielow *et al.*, 1988; Kaye *et al.*, 1989; Hogquist *et al.*, 1994). These positively selected DP thymocytes are directed either towards the $CD4^{+}$ T lineage or $CD8^{+}$ T lineage, depending on the kinetics of TCR signaling that they receive. Persistent or stronger

TCR signaling leads to the $CD4^+$ T lineage differentiation, whereas cessation or weaker TCR signaling results in the $CD8^+$ T lineage differentiation (Yasutomo *et al.*, 2000; Singer, 2002).

During T-cell development, the *Satb1* alternative promoter switch was observed especially during the differentiation of DP thymocytes into CD4SP, which in turn depends on the kinetics of TCR signaling. Therefore we asked whether the TCR signaling plays a role in the *Satb1* alternative promoter switch during the T-cell development. When we assessed the expression pattern of *Satb1* transcript variants in the thymocytes undergoing various stages of TCR mediated positive selection process such as $CD69^+TCR^{int}$, $CD69^+TCR^{hi}$ and $CD69^-TCR^{hi}$, interestingly we observed that only $CD69^+TCR^{hi}$ population which indicates thymocyte population undergoing positive selection and includes the TCR engaged DP, $CD4^+SP$ and $CD8^+SP$ cells, exhibits higher expression levels of majority of the *Satb1* transcript variants. Also, $CD69^+TCR^{hi}$ population displays higher expression of total *Satb1* mRNA. However, as shown in the Chapter 1, the P2 transcript variant expression was highly expressed specifically in the $CD4^+SP$ thymocytes during the development. The development of CD4SP thymocytes from DP requires a stronger or persistent TCR signaling. As we have shown in Chapter 1 that DP thymocytes show the higher expression levels of P1, P4 along with the constitutively expressed predominant form of P3 transcript variant, whereas CD4SP thymocytes show the higher expression levels of P2 transcript variant expression along with the P3 variant. We found that the P2 promoter switch, and the increased P2 variant expression in CD4SP thymocytes from DP thymocytes is induced by the persistent TCR signaling. We observed that only TCR engaged CD4 ($CD69^+CD4$) but not DP ($CD69^+CD4^+CD8^+$) thymocytes show the expression of the P2 variant. Our results indicate that the P2 transcript variant expression in T-cell development is induced just not by TCR signaling but by the persistent TCR signaling or comparatively stronger TCR signaling which resulting in the CD4SP differentiation. This might be the reason that the CD8SP thymocytes are not expressing the P2 transcript variant during T-cell development, since their development requires a weaker or cessation of TCR signaling at the DP stage. Along with the expression of P2 and P3 transcript variants, we observe that $CD69^+CD4SP$ thymocytes show the expression of significantly increased levels of

SATB1 protein compared to the $CD69^+$DP. These results confirm the importance of the expression of combination of P2 and P3 transcript variants in maintaining higher levels of SATB1 protein. Additionally, we observed that not only during T-cell development, anti-CD3 and anti-CD28 mediated TCR activation of the peripheral $CD4^+$ T-cells also switches *Satb1* alternative promoters, thus resulting in the expression of the P2 transcript variant along with P3 in TCR activated CD4 T-cells compared to naïve CD4. Also we observed that the P1 transcript variant levels were downregulated and a slight increase was observed in the P4 transcript variant expression in TCR activated CD4 T-cells compared to the naïve CD4, suggesting that SATB1 is regulated by TCR signal via selective alternative promoter usage. Thus upon TCR activation of peripheral $CD4^+$ T-cells, SATB1 protein levels were increased in the TCR activated CD4 T-cells compared to the naïve $CD4^+$ T-cells. Our results corroborating with the previous observation that SATB1 protein levels were elevated upon the TCR activation of peripheral $CD4^+$ T-cells (Gottimukkala *et al.,* 2016; Stephen *et al.,* 2017).

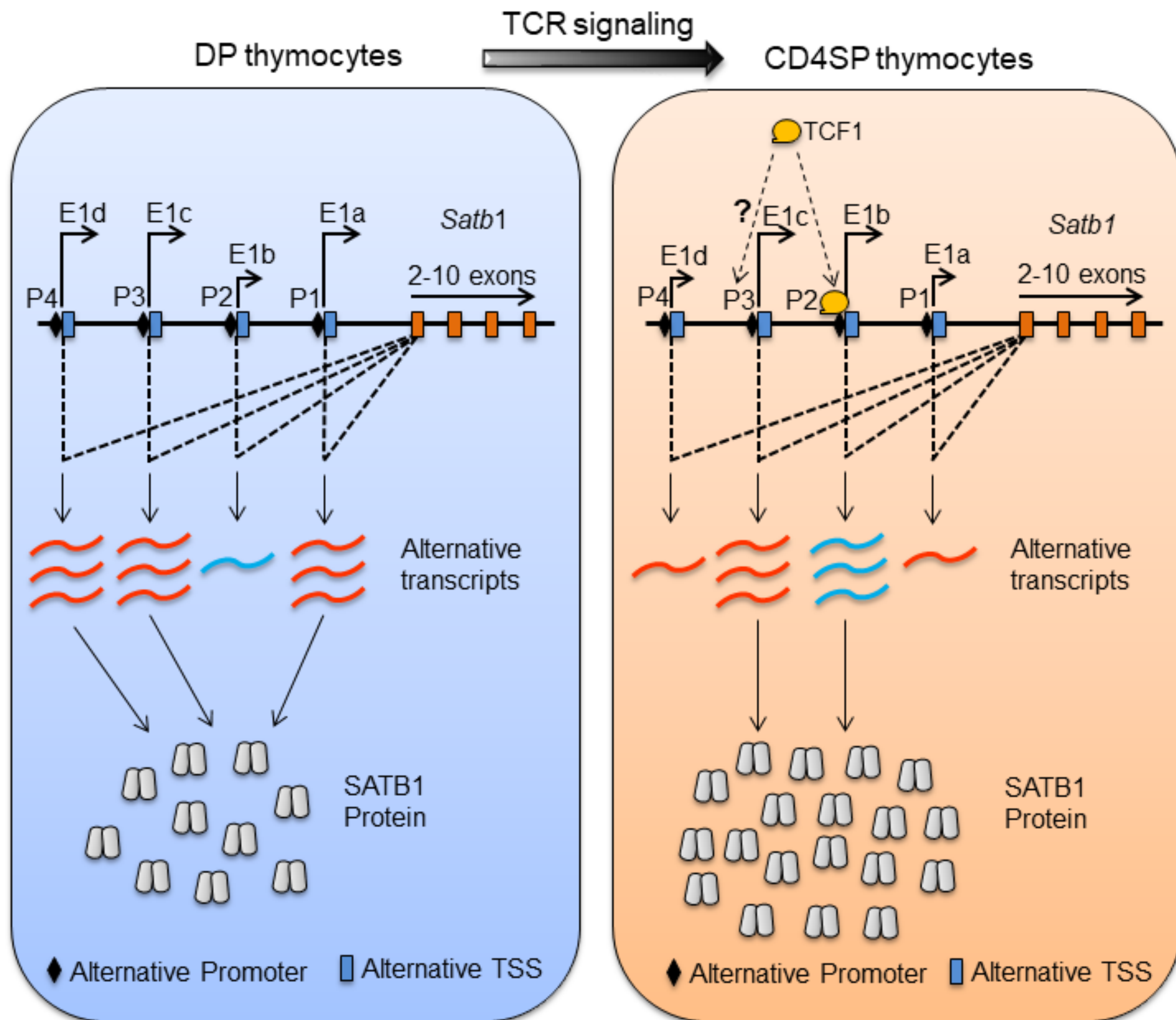


Figure 2.3. A schematic model depicting TCR signal mediated *Satb1* alternative promoter switch during T-cell development. During T-cell development in the thymus, double positive (DP) thymocytes are generated as a result of pre-TCR signaling at DN4 stage. At DP stage of development, TCRα gene rearrangement takes place and DP thymocytes receive TCR signaling. DP thymocytes are very sensitive and respond to the low affinity TCR ligands than SP thymocytes, whereas SP thymocytes respond to the high affinity TCR ligands presented in the medullary region of the thymus. Persistent or stronger TCR signaling at DP stage leads to their differentiation towards the $CD4^+$ T lineage, whereas cessation or weaker TCR signaling leads to $CD8^+$ T lineage differentiation. During T-cell development, DP thymocytes exhibit expression of P1 (P1S and P1L), P4 transcript variants along with the constitutively active P3 variant. Upon persistent TCR signaling mediated differentiation of DP into CD4SP, the $CD4^+$ T-cells especially immature CD4SP cells switch to the expression of the P2 transcript variant along with the P3

variant presumably due to the direct binding of transcription factors such as TCF1. TCF1 directly binds to the P2 promoter in CD4SP thymocytes and regulates the expression of the P2 transcript variant. However, whether TCF1 also regulates the expression of P3 in CD4SP is yet to be investigated. Since *Satb1* transcript variants exhibit the differential translatability, the combinatorial expression of *Satb1* transcript variants plays a critical role in maintaining the SATB1 protein levels. The combination of P2 along with P3 transcript variants contributes to the significantly higher SATB1 protein levels in immature CD4SP during T-cell development.

Interestingly, we also found that the *Satb1* alternative promoters show distinct lineage specific chromatin dynamics during T-cell development from the progenitors. Especially, the *Satb1* P2 promoter shows characteristic lineage specific chromatin accessibility during T-cell development from the progenitors. Unlike other promoters, the P2 promoter region is not accessible in the progenitors such as HSCs, MPPs, CLPs and also in the B cells. P2 promoter region was present in the open chromatin state only in the T-cell lineage. Also, during T-cell development in thymus, the accessibility of P2 promoter region begins at the DN3 stage onwards corroborating with the pre-TCR signaling received at this developmental stage onwards, and is increased in SP T-cells especially in the $CD4^+$ SP T-cells. Although P2 promoter accessibility is observed in the $CD8^+$ SP T-cells, the P2 transcripts were not detected in these T-cells, indicating that $CD8^+$ SP T-cells specific repressor protein complexes might bind around the TSS or downstream of P2 promoter region, thereby inhibiting the transcription of P2 in the $CD8^+$SP T-cells in order to maintain the lower levels of SATB1 protein in CD8SP compared to CD4SP (Gottimukkala *et al.,* 2016).

We further investigated the molecular players involved in the *Satb1* alternative promoter switch during CD4SP T-cells differentiation from DP. We found that TCF-1 acts upstream of SATB1 in the CD4 SP T-cells, potentially by regulating the P2 promoter switch during CD4SP development from DP. We identified presence of a canonical TCF1/LEF1 binding motif at the *Satb1* P2 promoter region. Furthermore, TCF1 occupancy was observed at the *Satb1* P2 promoter region in the CD4SP thymocytes. We observed that *Satb1* P2 transcript variant expression levels as well as total *Satb1* transcript levels were downregulated in CD4SP thymocytes from TCF1 KO mice

compared to the wild type mice. The CD4SP thymocytes from TCF1 KO mice also display reduced levels of SATB1 protein, indicating an important role played by TCF1 in regulating the SATB1 expression in CD4SP T-cells. Interestingly, we found that the DP thymocytes from the TCF1 KO mice reveal increased expression of total *Satb1* and also the P2 transcript variant. SATB1 protein levels in the DP thymocytes from TCF1 KO mice were increased compared to the wild type DP thymocytes. Currently we have not identified the reason for this phenomenon, but we hypothesize that TCF1 might regulate the expression of SATB1 in a developmental stage specific manner by directly binding to *Satb1* alternative promoters or by its gene activation and repression functions mediated by interactions with different protein partners.

In conclusion, we demonstrate that the combinatorial alternative promoter usage drives *Satb*1 expression during various stages of thymocyte development and also during the activation of peripheral T-cells. The selective expression of combination of *Satb1* transcript variants during T-cell development has a crucial role on the regulation of SATB1 protein levels as shown in the schematic (Figure 2.3). Persistent TCR signaling induces *Satb1* alternative promoter switch in CD4SP thymocytes during their differentiation from DP thymocytes. Our study demonstrates that not just TCR signaling, but a stronger or continuous TCR signaling during SP T lineage differentiation induces the *Satb1* alternative promoter switch. Not only during T-cell development, TCR activation of the peripheral $CD4^+$ T-cells also shows the switch of P2 and P3 promoters, thereby increased SATB1 protein levels. Interestingly, we found that chromatin dynamics at the *Satb1* alternative promoters is lineage specific during T-cell development from their progenitors. We have further identified the molecular players of this *Satb1* alternative promoter switch in developing thymocytes and found that TCF1 is one of the players. We provide evidence that TCF1 directly binds to the *Satb1* P2 promoter region and activates the P2 transcription, thereby maintaining the higher levels of SATB1 protein in CD4SP thymocytes. However, SATB1 is an important regulator of T-cell development, and its expression is regulated differentially in a developmental stage specific manner during T-cell development. Thus, further studies are required to fully dissect the role of developmental stage specific transcription factors and the epigenetic mechanisms in regulating the expression of SATB1 via alternative promoters

in a cell type specific manner. Whether perturbation of the multitude of *Satb1* alternative promoters exerts any effect on the lineage specific expression of SATB1 is yet to be understood.

2.4 Methods

2.4.1 Mice

Three weeks old C57BL/6 mice were used to prepare the single cell suspension of thymocytes and used for sorting of subpopulation of thymocytes which are in the different developmental stages. 6-8 week old C57BL/6 mice were used for isolation of naïve $CD4^{+}$ T-cells. TCF-1 knock out (TCF1 KO) mice were obtained by H. Clevers. All mice were bred and maintained under specific pathogen free environment and the experiment procedures were performed according to the guidelines of animal house facilities at IISER Pune, NCCS Pune and National Institute of Aging, Baltimore, USA.

2.4.2 Flow cytometry

Single cell suspension of thymocytes was used for the surface staining of thymocytes. Prior to the surface staining of thymocytes, FC receptor blocking was performed using the purified anti-CD16.32 antibody (BD Biosciences). Thymocytes were subjected to the surface staining with the following fluorochrome tagged antibodies: eFluor 450 anti-mouse CD4 (Clone GK1.5, eBioscience); APC anti-mouse CD4 (Clone GK1.5, eBioscience); FITC anti-mouse CD4 (Clone GK1.5, BD Biosciences); PE anti-mouse CD8a (Clone 53-6.7, BD Biosciences); PE-Cy7 anti-mouse TCRβ (Clone H57-597, eBioscience); PE-Cy7 anti-mouse CD69 (Clone H1.2F3, BD Biosciences); PE anti-Mouse CD25 (Clone PC61, BD Biosciences); BV510 anti-mouse CD25 (Clone PC61, BD Biosciences). The thymocyte subpopulations such as $CD69^{-}TCR^{-}$, $CD69^{+}TCR^{int}$, $CD69^{+}TCR^{hi}$, $CD69^{+}TCR^{-}$, $CD69^{+}DP$, $CD69^{+}CD4SP$, and $CD69^{-}CD4SP$ thymocytes were FACS sorted using FACS Aria III SORP (BD biosciences). Intracellular staining of SATB1 was performed using anti-SATB1 antibody (BD Biosciences) by using Foxp3/Transcription Factor staining kit (eBiosciences). Flow cytometry analyses were performed using FACS Canto II (BD Biosciences).

2.4.3 Isolation of Naïve CD4$^+$ T-cells and cell culture

For isolation of naïve CD4$^+$ T-cells, single cell suspension of spleen was prepared and RBC were lysed using RBC lysis buffer. Cells were subjected to magnetic sorting by using mouse naïve CD4$^+$ T-cell isolation kit (BD Biosciences). After magnetic sorting, cells were surface stained with eFluor 450 anti-mouse CD4 (Clone GK1.5, eBioscience) or FITC anti-mouse CD4 (Clone GK1.5, BD Biosciences); PE anti-mouse CD8a (Clone 53-6.7, BD Biosciences); BV510 anti-mouse CD25 (Clone PC61, BD Biosciences), and naïve CD4$^+$ T-cells were FACS sorted using FACS sorter Aria II (BD Biosciences). Naïve CD4$^+$ T-cells were cultured in the presence of 0.5 μg/mL of plate bound anti-mouse CD3 (Clone 17A2, eBioscience) and 1.5 μg/mL of soluble anti-mouse CD28 (Clone 37.51, eBioscience) for 48 hrs. Cells were harvested for RNA extraction and for intracellular staining of SATB1 and Nur77 using anti-SATB1 antibody (BD Biosciences) and anti-Nur77PE (eBiosciences) by using the Foxp3/Transcription Factor staining kit (eBiosciences). Flow cytometry analyses were performed using FACS Canto II (BD Biosciences).

2.4.4 Data base availability

GSE77695 data set was used for the analysis of ATAC-seq performed in HSCs, MPPs, CLPs, B-cells, CD4 T-cells, and CD8 T-cells. GSE100738 data set was used for ATAC seq analysis performed in DN1, DN2a, DN2b, DN3, DN4, DP, CD4SP, CD4 naïve, CD8SP and CD8 naïve T-cells. Publicly available GSE46662 (Dose et al., 2014) data set was used for TCF1 ChIP-seq analysis performed in DP. The raw reads were aligned using Bowtie2 and peak calling was performed using MACS2. The peaks were visualized using IGV genome browser.

2.4.5 cDNA synthesis and Quantitative real-time PCR analysis (qPCR)

Isolation of total RNA from the sorted thymocyte subpopulations were performed using Qiagen RNeasy mini kit (Qiagen). Following DNase I (Promega) digestion, the RNA was subjected to cDNA synthesis using High capacity cDNA synthesis kit (Applied Biosystems). Quantitative RT-PCR analyses were performed using Sybr green qPCR

master mix (Roche) at the following PCR conditions: step 1, 95°C-5 min; step 2, 95°C-45 sec, 60°C-45 sec, 72°C-1 min for 40 cycles. The following qPCR primers were used:

Satb1 (total)-F: 5'-TGATAGAGATGGCGTTGCTG-3'
Satb1 (total)-R: 5'-TTTTGAGGGTGACCACATGA-3'
P1(E1a)-F: 5'-CAAGAATCCCGGCTGCAAAG-3'
P1(E1a)-R: 5'-CCCTGAGTTGCCTCGTTCAA-3'
P2(E1b)-F: 5'-AGATTCGGAAACCAGCCTCTG-3'
P2(E1b)-R: 5'-GGACCCTTCGGATCACTCAC-3'
P3(E1c)-F: 5'-CGGTTCCACGCCTGATTCT-3'
P3(E1c)-R: 5'-GTGGACCCTTCGGATCACTC-3'
P4(E1d)-F: 5'-CCAAAGCCCAGGCAAACAAC-3'
P4(E1d)-R: 5'-CCCTGAGTTGCCTCGTTCAA-3'
m18s-F: 5'-GTAACCCGTTGAACCCCATT-3'
m18s-R: 5'-CCATCCAATCGGTAGTAGCG-3'
Hprt-F: 5'-CTGGTGAAAAGGACCTCTCG-3'
Hprt-R: 5'-TGAAGTACTCATTATAGTCAAGGGCA-3'

2.4.6 Western Blotting

Cells were lysed using RIPA lysis buffer (50 mM Tris-HCl at PH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.25% sodium deoxycholate, 1 mM PMSF, and 1X protease inhibitor cocktail (Roche) and the protein quantification was performed using BCA method. The total protein was separated on 10-12% SDS polyacrylamide gel and then transferred to PVDF membrane. The PVDF membrane was blocked with 5% milk and probed with the following antibodies; Anti-SATB1 (1:1000, BD Biosciences), anti-actin (1:4000, sigma), anti-TCF1 (1:500, Santa Cruz), and Anti-tubulin (1:4000, sigma). The signals were visualized using ECL luminescence detection reagent (BIO-RAD) on ImageQuant LAS 4000 system (GE Healthcare Life Sciences).

2.4.7 ChIP-qPCR

Chromatin immunoprecipitation (ChIP) was performed using CD4 SP thymocytes as previously described (Jayani et al., 2010). Approximately 50×10^6 CD4SP thymocytes were cross linked with 1% formaldehyde at room temperature for 10 minutes. After formaldehyde cross linking, 125 mM glycine was used for the neutralization. The cross linked cells were subjected to nuclei isolation using the hypotonic buffer (25 mM Tris-HCl pH 7.9, 1.5 mM MgCl2, 10mM KCl, 0.1% NP40, 1mM DTT, 0.5 mM PMSF, and 1x protease inhibitor cocktail (Roche). Isolated nuclei were lysed using the lysis buffer (50 mM Tris-HCl pH 7.9, 140 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Sodium deoxycholate, 0.1% SDS, 0.5 mM PMSF and 1x protease inhibitor cocktail (Roche) and the chromatin was subjected to sonication using Bioruptor sonication system (Diagenode, Belgium) for 15-30 cycles with 30 sec 'on' and 30 sec 'off 'to obtain the chromatin fragment size of 200–500 base pairs. Preclearing of the sonicated chromatin was performed using a cocktail containing 50% protein A/G beads slurry (Thermo scientific) and precleared chromatin was subjected to the immune precipitation with anti-mouse TCF1 antibody (Santa Cruz Biotech) overnight at 4 ^{o}C. Similarly, anti-mouse IgG was used as an isotype control. The immune precipitated complexes were pull down by adding Protein A/G beads cocktail which was pre-saturated by incubating with 10mg/mL tRNA and 1% BSA and at 4 ^{o}C for 4hrs. The immune precipitated bead bound chromatin was washed though roughly, and subjected to the elution by using elution buffer (1% SDS, 0.1 M $NaHCO_3$). The eluted chromatin was de-crosslinked and the protein was removed by treating with the proteinase K, and RNA was removed by treating with the RNase A. Further the immune-precipitated chromatin was purified and subjected to quantitative RT-PCR (qRT-PCR) analysis using the Sybr green qRT-PCR master mix (Roche).

The following primer pair was used for qRT-PCR analysis:

P2 promoter - F: 5'-CATGGAACCGAGTGCCTGTA-3'

P2 promoter - R: 5'-GGGATGAGCTTGCCTTGCTA-3'

2.4.8 Transfections and luciferase activity Assay

Satb1 P2 promoter region was amplified from the genomic DNA of mouse thymocytes. The amplified P2 promoter region was cloned upstream to the firefly luciferase gene in pGL3 basic vector (Promega). Similarly TCF1 gene was amplified from the cDNA of mouse thymocytes, and then cloned into p3X-CMV-FLAG10 vector (Sigma). The following cloning primers were used for the amplification of P2 promoter and TCF1:

P2 promoter - F: 5'-GCCGAGCTCACACGCTGTCTTTTCTTTAAATA-3'

P2 promoter - F: 5'-GCCCTCGAGCGTAATTCTAACCACCTCCCCC-3'

TCF1 - F: 5'-GCGAATTCATGTACAAAGAGACTGTCTACTCT-3'

TCF1 - R: 5'-GCGGATCCCTAGAGCACTGTCATCGGAAGGAA-3'

The P2 promoter containing pGL3 basic vector was overexpressed in HEK293T cells in the presence or absence of TCF1 over expression using Lipofectamine 2000 transfection reagent (Invitrogen). After 48 hrs of the transfection, cells were harvested for the western blotting to monitor TCF1 expression, and also used for the luciferase activity assay to monitor the P2 promoter activity. The lysates were subjected to the luciferase activity assay using dual luciferase assay kit (Promega). Renilla luciferase activity was used as endogenous control.

2.5 References

Aliahmad P, Kaye J. (2008). Development of all CD4 T lineages requires nuclear factor TOX. J Exp Med *205,* 245-56.

Asnagli H, Afkarian M, Murphy KM. (2002). Cutting edge: Identification of an alternative GATA-3 promoter directing tissue-specific gene expression in mouse and human. J Immunol *168,* 4268-71.

Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, Kohwi-Shigematsu T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev *14,* 521-35.

Baek D, Davis C, Ewing B, Gordon D, Green P. (2007). Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. Genome Res *17,* 145-55.

Basson MA, Bommhardt U, Mee PJ, Tybulewicz VL, Zamoyska R. (1998). Molecular requirements for lineage commitment in the thymus--antibody-mediated receptor engagements reveal a central role for lck in lineage decisions. Immunol Rev *165,*181-94.

Bee T, Swiers G, Muroi S, Pozner A, Nottingham W, Santos AC, Li PS, Taniuchi I, de Bruijn MF. (2010). Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. Blood *115,* 3042-50.

Bender TP, Kremer CS, Kraus M, Buch T, Rajewsky K. (2004). Critical functions for c-Myb at three checkpoints during thymocyte development. Nat Immunol *5*, 721-9.

Bousso P, Bhakta NR, Lewis RS, Robey E. (2002). Dynamics of thymocyte-stromal cell interactions visualized by two-photon microscopy. Science *296,* 1876-80.

Davey GM, Schober SL, Endrizzi BT, Dutcher AK, Jameson SC, Hogquist KA. (1998). Preselection thymocytes are more sensitive to T cell receptor stimulation than mature T cells. J Exp Med *188,* 1867-74.

Dose M, Emmanuel AO, Chaumeil J, Zhang J, Sun T, Germar K, Aghajani K, Davis EM, Keerthivasan S, Bredemeyer AL, et al. (2014). β-Catenin induces T-cell transformation by promoting genomic instability. Proc Natl Acad Sci U S A *111,* 391-6.

Egawa T, Littman DR. (2008). ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediated commitment to the cytotoxic T cell lineage. Nat Immunol *9*, 1131-9.

Egerton M, Scollay R, Shortman K. (1990). Kinetics of mature T-cell development in the thymus. Proc Natl Acad Sci U S A *87,* 2579-82.

Garvin AM, Pawar S, Marth JD, Perlmutter RM. (1988). Structure of the murine lck gene and its rearrangement in a murine lymphoma cell line. Mol Cell Biol *8*, 3058-64.

Germain RN. (2002). T-cell development and the CD4-CD8 lineage decision. Nat Rev Immunol *2,* 309-22.

Goldrath AW, Bevan MJ. (1999). Selecting and maintaining a diverse T-cell repertoire. Nature *402,* 255-62.

Gómez-del Arco P, Kashiwagi M, Jackson AF, Naito T, Zhang J, Liu F, Kee B, Vooijs M, Radtke F, Redondo JM, Georgopoulos K. (2010). Alternative promoter usage at the Notch1 locus supports ligand-independent signaling in T cell development and leukemogenesis. Immunity *33,* 685-98.

Gottimukkala KP, Jangid R, Patta I, Sultana DA, Sharma A, Misra-Sen J, Galande S. (2016). Regulation of SATB1 during thymocyte development by TCR signaling. Mol Immunol *77,* 34-43.

Hao B, Naik AK, Watanabe A, Tanaka H, Chen L, Richards HW, Kondo M, Taniuchi I, Kohwi Y, Kohwi-Shigematsu T, Krangel MS. (2015). An anti-silencer- and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. J Exp Med 212, 809-24.

He X, He X, Dave VP, Zhang Y, Hua X, Nicolas E, Xu W, Roe BA, Kappes DJ. (2005). The zinc finger transcription factor Th-POK regulates CD4 versus CD8 T-cell lineage commitment. Nature *433,* 826-33.

He X, Park K, Wang H, He X, Zhang Y, Hua X, Li Y, Kappes DJ. (2008). CD4-CD8 lineage commitment is regulated by a silencer element at the ThPOK transcription-factor locus. Immunity *28,* 346-58.

Heng TS, Painter MW; Immunological Genome Project Consortium. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol *9,* 1091-4.

Hernández-Hoyos G, Anderson MK, Wang C, Rothenberg EV, Alberola-Ila J. (2003). GATA-3 expression is controlled by TCR signals and regulates CD4/CD8 differentiation. Immunity *19,* 83-94.

Hogquist KA, Jameson SC, Heath WR, Howard JL, Bevan MJ, Carbone FR. (1994). T cell receptor antagonist peptides induce positive selection. Cell *76,* 17-27.

Hogquist KA, Jameson SC. (2014). The self-obsession of T cells: how TCR signaling thresholds affect fate 'decisions' and effector function. Nat Immunol *15,* 815-23.

Itano A, Salmon P, Kioussis D, Tolaini M, Corbella P, Robey E. (1996). The cytoplasmic domain of CD4 promotes the development of CD4 lineage T cells. J Exp Med *183,* 731-41.

Jayani RS, Ramanujam PL, Galande S. (2010). Studying histone modifications and their genomic functions by employing chromatin immunoprecipitation and immunoblotting. Methods Cell Biol *98,* 35-56.

Kakugawa K, Kojo S, Tanaka H, Seo W, Endo TA, Kitagawa Y, Muroi S, Tenno M, Yasmin N, Kohwi Y, Sakaguchi S, Kowhi-Shigematsu T, Taniuchi I. (2017). Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. Cell Rep 19, 1176-1188.

Kappler JW, Roehm N, Marrack P. (1987). T cell tolerance by clonal elimination in the thymus. Cell *49,* 273-80.

Kaye J, Hsu ML, Sauron ME, Jameson SC, Gascoigne NR, Hedrick SM. (1989). Selective development of CD4+ T cells in transgenic mice expressing a class II MHC-restricted antigen receptor. Nature *341,* 746-9.

Kearse KP, Roberts JL, Singer A. (1995). TCR alpha-CD3 delta epsilon association is the initial step in alpha beta dimer formation in murine T cells and is limiting in immature CD4+ CD8+ thymocytes. Immunity *2,* 391-9.

Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, et al. (2006). Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res *16,* 55-65.

Kisielow P, Teh HS, Blüthmann H, von Boehmer H. (1988). Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. Nature *335*, 730-3.

Kondo M, Tanaka Y, Kuwabara T, Naito T, Kohwi-Shigematsu T, Watanabe A. (2016). SATB1 Plays a Critical Role in Establishment of Immune Tolerance. J Immunol *196,* 563-72.

Lieu YK, Kumar A, Pajerowski AG, Rogers TJ, Reddy EP. (2004). Requirement of c-myb in T cell development and in mature T cell function. Proc Natl Acad Sci U S A *101*, 14853-8.

Maurice D, Hooper J, Lang G, Weston K. (2007). c-Myb regulates lineage choice in developing thymocytes via its target gene Gata3. EMBO J *26,* 3629-40.

Mombaerts P, Clarke AR, Rudnicki MA, Iacomini J, Itohara S, Lafaille JJ, Wang L, Ichikawa Y, Jaenisch R, Hooper ML, et al. (1992). Mutations in T-cell antigen receptor genes alpha and beta block thymocyte development at different stages. Nature *360,* 225-31.

Pai SY, Truitt ML, Ting CN, Leiden JM, Glimcher LH, Ho IC. (2003). Critical roles for transcription factor GATA-3 in thymocyte development. Immunity *19,* 863-75.

Pozner A, Lotem J, Xiao C, Goldenberg D, Brenner O, Negreanu V, Levanon D, Groner Y. (2007). Developmentally regulated promoter-switch transcriptionally controls Runx1 function during embryonic hematopoiesis. BMC Dev Biol *12,* 7:84.

Reynolds PJ, Lesley J, Trotter J, Schulte R, Hyman R, Sefton BM. (1990). Changes in the relative abundance of type I and type II lck mRNA transcripts suggest differential promoter usage during T-cell development. Mol Cell Biol *10,* 4266-70.

Sartor O, Gregory FS, Templeton NS, Pawar S, Perlmutter RM, Rosen N. (1989). Selective expression of alternative lck mRNAs in human malignant cell lines. Mol Cell Biol *9,* 2983-8.

Sato T, Ohno S, Hayashi T, Sato C, Kohu K, Satake M, Habu S. (2005). Dual functions of Runx proteins for reactivating CD8 and silencing CD4 at the commitment process into CD8 thymocytes. Immunity *22,* 317-28.

Sawada S, Scarborough JD, Killeen N, Littman DR. (1994). A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. Cell *77,* 917-29.

Setoguchi R, Tachibana M, Naoe Y, Muroi S, Akiyama K, Tezuka C, Okuda T, Taniuchi I. (2008). Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. Science *319,* 822-5.

Sha WC, Nelson CA, Newberry RD, Kranz DM, Russell JH, Loh DY. (1988). Positive and negative selection of an antigen receptor on T cells in transgenic mice. Nature *336,* 73-6.

Shih HY, Sciumè G, Mikami Y, Guo L, Sun HW, Brooks SR, Urban JF Jr, Davis FP, Kanno Y, O'Shea JJ. (2016). Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality. Cell *165,* 1120-1133.

Singer A. (2002). New perspectives on a developmental dilemma: the kinetic signaling model and the importance of signal duration for the CD4/CD8 lineage decision. Curr Opin Immunol *14*, 207-15.

Singer A, Adoro S, Park JH. (2008). Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. Nat Rev Immunol *8,* 788-801.

Siu G1, Wurster AL, Duncan DD, Soliman TM, Hedrick SM. (1994). A transcriptional silencer controls the developmental expression of the CD4 gene. EMBO J *13,* 3570-9.

Stephen TL, Payne KK, Chaurio RA, Allegrezza MJ, Zhu H, Perez-Sanz J, Perales-Puchalt A, Nguyen JM, Vara-Ailor AE, Eruslanov EB, Borowsky ME, Zhang R, Laufer TM, Conejo-Garcia JR. (2017). SATB1 Expression Governs Epigenetic Repression of PD-1 in Tumor-Reactive T Cells. Immunity *46,* 51-64.

Steinke FC, Yu S, Zhou X, He B, Yang W, Zhou B, Kawamoto H, Zhu J, Tan K, Xue HH. (2014). TCF-1 and LEF-1 act upstream of Th-POK to promote the CD4(+) T cell fate and interact with Runx3 to silence Cd4 in CD8(+) T cells. Nat Immunol *15,* 646-656.

Sun G, Liu X, Mercado P, Jenkinson SR, Kypriotou M, Feigenbaum L, Galéra P, Bosselut R. (2005). The zinc finger protein cKrox directs CD4 lineage differentiation during intrathymic T cell positive selection. Nat Immunol *6,* 373-81.

Sun H, Palaniswamy SK, Pohar TT, Jin VX, Huang TH, Davuluri RV. (2006). MPromDb: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-chip experimental data. Nucleic Acids Res *34*, D98-103.

Takahama Y. (2006). Journey through the thymus: stromal guides for T-cell development and selection. Nat Rev Immunol *6,* 127-35.

Taniuchi I, Osato M, Egawa T, Sunshine MJ, Bae SC, Komori T, Ito Y, Littman DR. (2002). Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. Cell *111,* 621-33.

Voronova AF, Adler HT, Sefton BM. (1987). Two lck transcripts containing different 5' untranslated regions are present in T cells. Mol Cell Biol *7,* 4407-13.

Wang L, Wildt KF, Zhu J, Zhang X, Feigenbaum L, Tessarollo L, Paul WE, Fowlkes BJ, Bosselut R. (2008). Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of CD4(+) T cells. Nat Immunol *9*, 1122-30.

Wildin RS, Garvin AM, Pawar S, Lewis DB, Abraham KM, Forbush KA, Ziegler SF, Allen JM, Perlmutter RM. (1991). Developmental regulation of lck gene expression in T lymphocytes. J Exp Med *173*, 383-93.

Witt CM, Raychaudhuri S, Schaefer B, Chakraborty AK, Robey EA. (2005). Directed migration of positively selected thymocytes visualized in real time. PLoS Biol *3,* e160.

Xing S, Li F, Zeng Z, Zhao Y, Yu S, Shan Q, Li Y, Phillips FC, Maina PK, Qi HH, et al. (2016). Tcf1 and Lef1 transcription factors establish CD8(+) T cell identity through intrinsic HDAC activity. Nat Immunol *17*, 695-703.

Yasutomo K, Doyle C, Miele L, Fuchs C, Germain RN. (2000). The duration of antigen receptor signalling determines CD4+ versus CD8+ T-cell lineage fate. Nature *404,* 506-10.

Chapter 3

Identifying the mechanism of the combinatorial expression of *Satb1* transcript variants during T-cell development

3.1 Introduction

Transcription is a major regulatory step during gene expression and is dependent on the transcription factors and RNA polymerase II (RNA pol II), which binds to the promoter region of a gene and initiates the transcription. Promoters are nearly 100 bp in length and located near the transcription start site (TSS) of a gene. The most commonly used core promoter elements of the protein encoding genes are the TATA box, and the initiator (Inr) (Smale and Baltimore, 1989; Smale and Kadonaga, 2003; Frith *et al.*, 2008). The RNA pol II binds to these promoter elements, and initiates the transcription, thus these elements decide the directionality of transcription (Figure 3.1.1 A). The cell type specific expression of a gene is governed by another *cis*-regulatory element called the enhancer (Spitz and Furlong, 2012; Shlyueva, Stampfel and Stark, 2014). Enhancers are typically few hundred to thousand base pairs in length and consist of binding sites for the key transcription factors. Enhancers can act on the promoters which are located at a long distance by direct association with the promoters and allow the initiation of transcription at the promoter region of cell type specific genes (Figure 3.1.1 B).

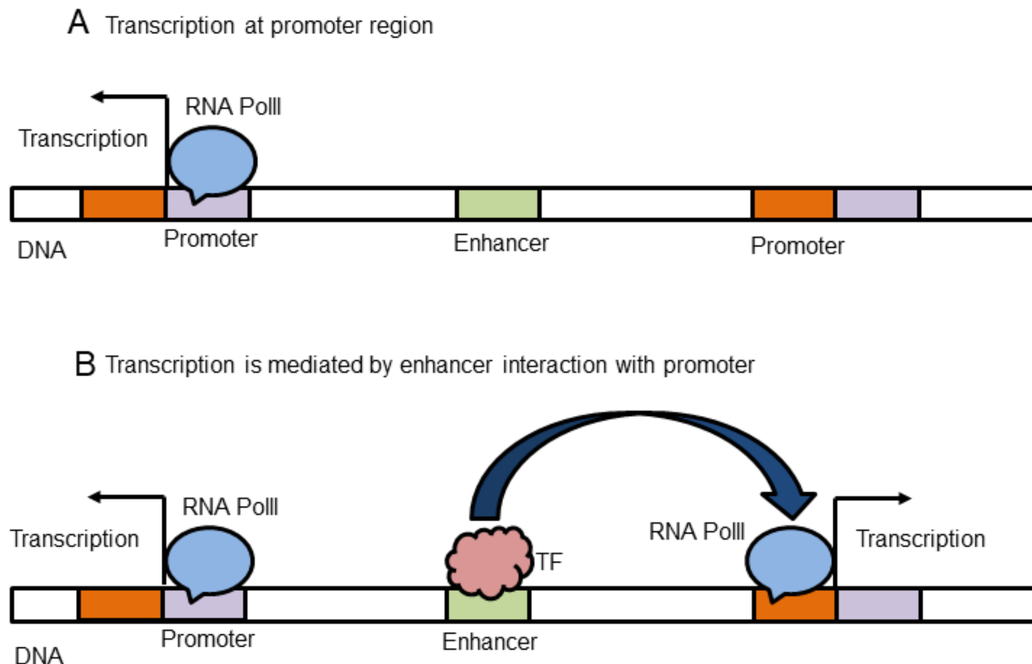


Figure 3.1.1. The initiation of the transcription by an interplay between DNA *cis* regulatory elements of a gene. (A) In general the regulation of transcription of a protein coding gene occurs at the promoter elements. Promoter is nearly 50-100 bp in length and is located near the TSS of a gene. The transcription factors bind to the promoter region and then recruit the RNA polymerase II onto the promoter and then the RNA polymerase II initiates transcription at the promoter. (B) In case of the cell type specific expression of a gene, another *cis* regulatory element such as an enhancer plays a crucial role. The active transcription factors and co-activator complexes bind to the enhancer and then mediate the interaction of the enhancer with the promoter region which is located at a long distance and then forms the enhancer–promoter loop which enables the RNA polymerase II to bind to the promoter region and initiates the transcription from the promoters of a cell type specific gene.

The long-range chromatin interactions between these *cis* regulatory DNA elements such as enhancers and promoters further result in the spatial organization of the genome into compartments called as "topologically associated domains" (TADs) that can be several Kb or Mb in length (Figure 3.1.2) (Dixon *et al.*, 2012; Nora *et al.*, 2012; Sexton *et al.*, 2012). These chromatin interactions are mediated by the chromatin architectural proteins such as CTCF and cohesin (Phillips-Cremins *et al.*, 2013; Wutz *et al.*, 2017). Several reports suggest that the cohesin complex, which consists of SMC1, SMC3 and

RAD21, play an essential role in the TAD formation (Kagey *et al.*, 2010; Phillips-Cremins *et al.*, 2013; Seitan *et al.*, 2013; Sofueva *et al.*, 2013).

Very few reports suggest that TADS vary between the different cell types of a given organism, indicating the invariant feature of genome organization (Dixon *et al.*, 2012; Nora *et al.*, 2012). However, the chromatin interactions within the TADs called “sub TADs” which include the long-range enhancer-promoter interactions seem to vary among the different cell types of an organism (Phillips-Cremins *et al.*, 2013; Dowen *et al.*, 2014; Rao *et al.*, 2014; Ji *et al.*, 2016), indicating that the chromatin interactions between the genome regulatory elements are cell type specific, thereby allowing the regulation of cell type specific gene expression. These cell type specific enhancer-promoter interactions are very important, since many of the disease associated mutations occur in the vicinity of the enhancers (Ernst *et al.*, 2011; Maurano *et al.*, 2012; Hnisz *et al.*, 2013; Farh *et al.*, 2015). Some of these enhancer-promoter interactions are mediated by the binding of specific transcription factors at the promoter regions (Choi and Engel, 1988; Ohtsuki, Levine and Cai, 1998; Butler and Kadonaga, 2001). Another important mechanism of maintaining the specificity of enhancer-promoter interactions is an insulation function of the boundary regions which prevent the enhancer action on other genes (Kellum and Schedl, 1991; Chung, Whiteley and Felsenfeld, 1993). The binding of CTCF to the boundary elements is very important in maintaining the insulator functions (Figure 3.1.2).

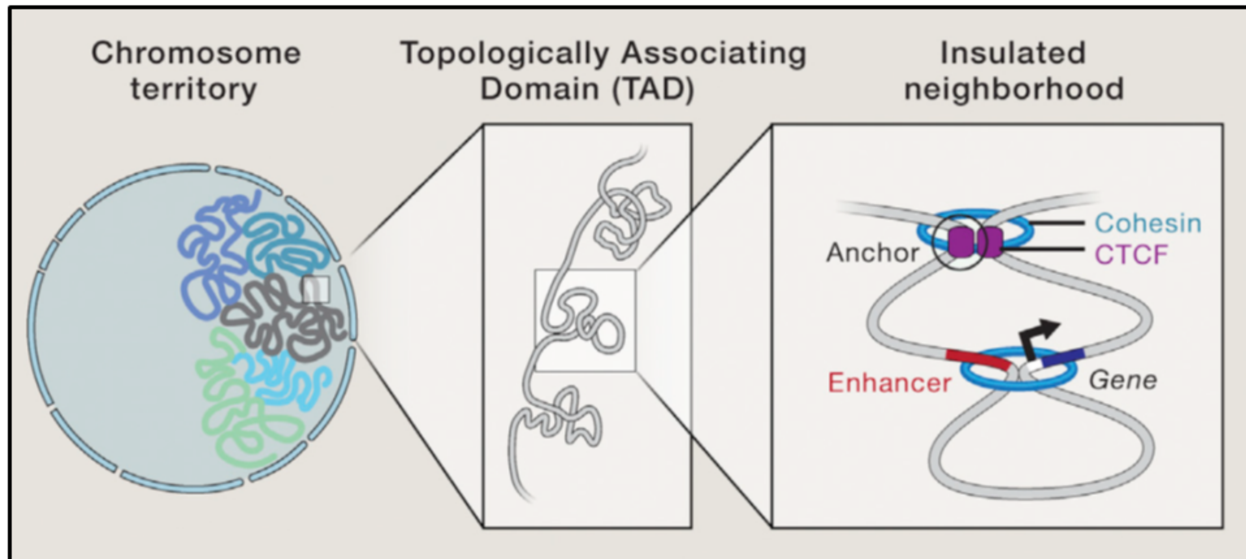


Figure 3.1.2.The higher-order chromatin interactions in the interphase of nucleus. In the interphase nucleus, the genome is highly folded, thus maintaining the minimal space of the nucleus. The chromatin interactions in the genome such as the specific promoter-enhancer interactions lead to the expression of cell type specific genes. These enhancer-promoter interactions occur in the spatial domains called topologically associated domains (TADs). The spatial organization of these domain interactions are mediated by insulator regions to which CTCF binds, thereby maintaining the domain structures and prevent the aberrant chromatin interactions (Reproduced from Hnisz, Day and Young, 2016).

Boundary activity between the TADs restricts interaction of an enhancer with the genes in the neighboring TADS, thus prevent the co-regulation of genes and deletion of the boundary regions lead to the ectopic expression of genes in the neighboring TADs (Lupiáñez *et al.*, 2015; Narendra *et al.*, 2015). TADs are well conserved among the species during evolution (Dixon *et al.*, 2012; Vietri Rudan *et al.*, 2015). Though TADs appear in human, mouse, dogs, and Drosophila, however, the TADs appearing in C. elegans seems to be different indicating the different pattern of chromosome organization, showing the TAD appearance only on the X chromosome, whereas their less frequent appearance on the somatic chromosomes. This observation was also supported by another finding that both CTCF, a boundary activity maintaining protein and CTCF enrichment motifs were lost in the genome of C. elegans and in some nematodes during evolution (Heger *et al.*, 2012). The above observations indicate that

the origin of CTCF and its enrichment in the boundary regions coincide with the evolution of TADs. However, even during the development TADs are cell type invariant (Dixon *et al.*, 2012; Nora *et al.*, 2012) but the sub domains highly cell type specific (Lieberman-Aiden *et al.*, 2009). Additionally, during development the chromatin interactions between the promoters and distal enhancers are also highly cell type specific (Dixon *et al.*, 2012; Li *et al.*, 2012; Sanyal *et al.*, 2012; Kieffer-Kwon *et al.*, 2013; Rao *et al.*, 2014). Hence, the chromatin interactions between the *cis* regulatory DNA elements such as promoter and enhancer within the subdomains of TADs maintain the cell type specific expression of genes (Figure 3.1.3).

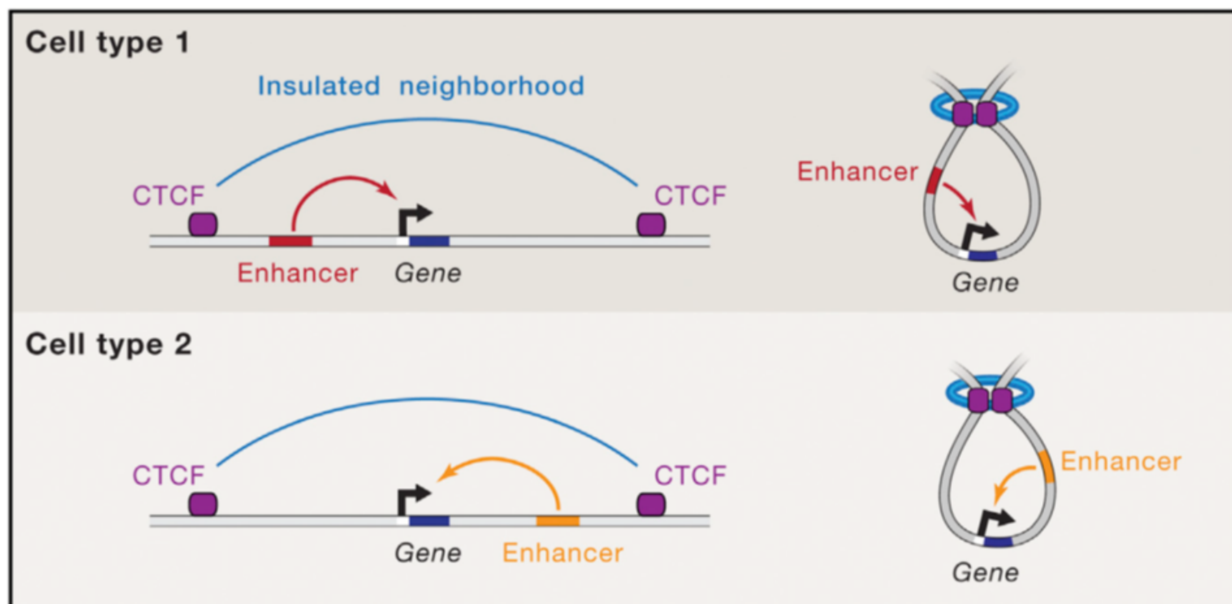


Figure. 3.1.3. The cell type specific expression of genes is governed by specific enhancer-promoter interactions. During development, the cell type specific gene expression is achieved by interactions between an enhancer and promoter. In the genome, these specific enhancer–promoters interactions is mediated by the boundary elements to which the proteins such as CTCF bind, thereby maintaining the insulation function. This insulation prevents the enhancer action on genes in the neighboring domain. Therefore an enhancer interacts with genes in the same domain and thus achieve the cell type specific gene expression (Reproduced from Hnisz, Day and Young, 2016).

Genome-wide analyses revealed that cohesin co-localizes with CTCF at majority of the sites in the mammalian genome (Parelho et al., 2008; Wendt et al., 2008). CTCF is a ubiquitously expressed 11 zinc finger containing protein involved in many cellular functions by regulating the gene expression as well as insulating the genome (Gazner and Felsenfeld, 2006; Filippova, 2008). It is a highly conserved protein and contains an N-terminal domain, a central domain with highly conserved 11 zinc fingers, and a C-terminal domain (Figure 3.1.5 A) (Ohlsson, 2001; Zlatanova and Caiafa, 2009). These three domains of CTCF are important for the dimerization as well as the interaction with multiple transcription factors (Figure 3.1.5 A) (Yusufzai and Felsenfeld, 2004; Zlatanova and Caiafa, 2009). CTCF interacts with a large number of transcription factors such as cohesin, YY1, PARP1, nucleophosmin and RNA pol II (Zlatanova and Caiafa, 2009). The dimerization of CTCF allows the DNA looping and the organization of genome along with cohesin (Yusufzai *et al.,* 2004; Parelho *et al.,* 2008, Rubio *et al.,* 2008; Zlatanova and Caiafa, 2009).

Cohesin is a multi-subunit complex which consists of two SMC (structural maintenance of chromosomes) family proteins such as SMC1 and SMC3, Scc1 (RAD21) and Scc3 (SA1/SA2) (Figure 3.1.5 B). The two SMC proteins form the core of the cohesin complex. The main structural characteristics of SMC proteins are the head domain with ATPase activity and the hinge domain (Strunnikov *et al*., 1993). The ATP binding head domain is formed as a result of interactions between the C- and N- termini of the SMC proteins (Figure 3.1.5 B). The hinge domain of cohesin allows the dimerization of SMC proteins. However, the hinge and head domains of SMC proteins are connected by a coiled coil region. The other two subunits of SMC complexes such as scc1 and Scc3 bind to the head domain allowing the stabilization of the ring structure formed by the dimerization of SMC proteins. The DNA binding activity of cohesin is a highly dynamic process and requires the chromatin associating factors (Wendt *et al*., 2008; Parelho *et al*., 2008; Rubio *et al*., 2008).

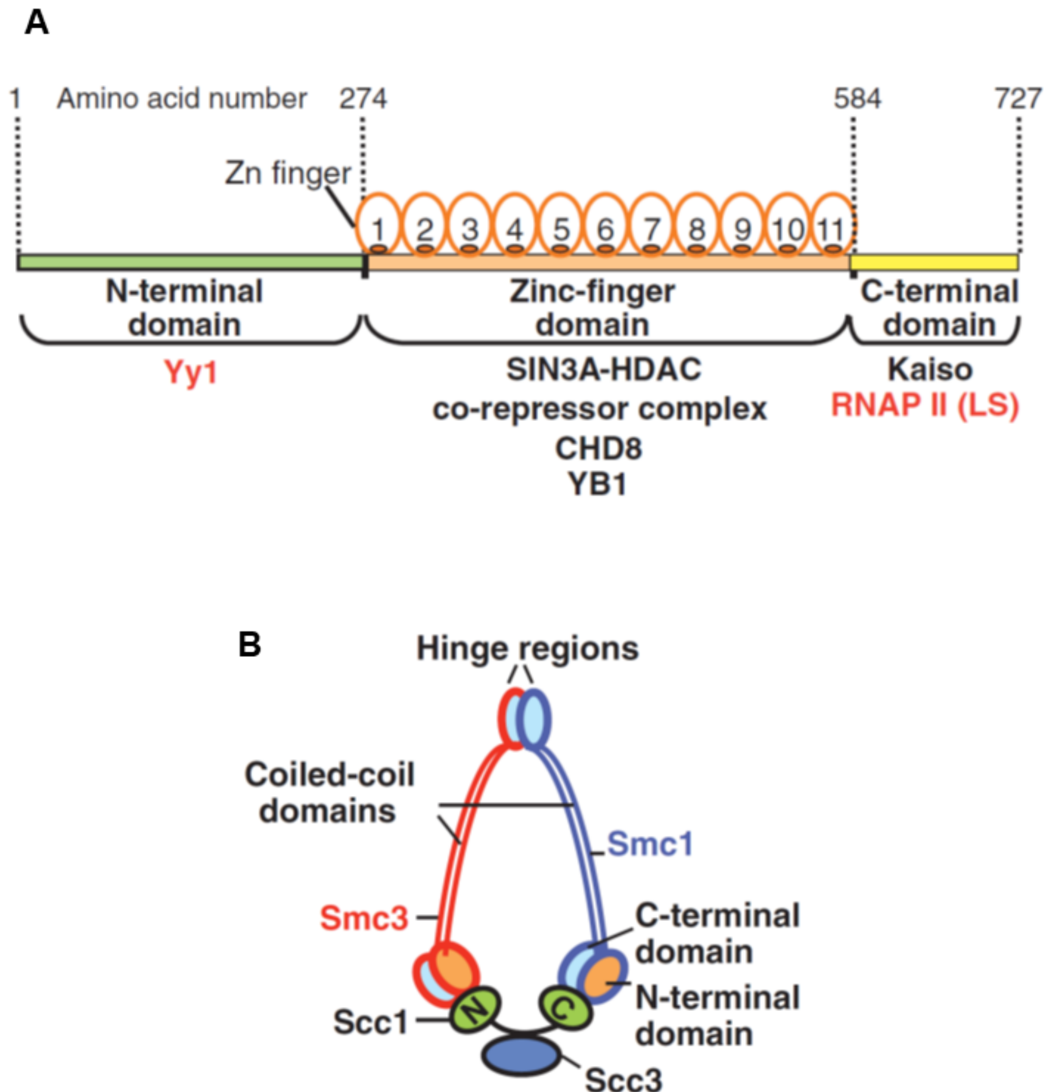


Figure 3.1.4. The Structure of CTCF and cohesin complex. A) The domain structure of CTCF. CTCF is a 727 amino acid containing protein with the following domain structure: N-terminal domain, central domain and C-terminal domain. The 11 zinc fingers in the central domain are highly conserved across species. As shown in the figure these independent domains play a crucial role in the interaction with the different protein partners thus regulating the genes differentially. **B) A schematic depiction indicates the structural composition of the cohesin complex.** Cohesin is a multimeric complex containing SMC1, SMC2, Scc1 and Scc3. SMC1 and SMC2 form the dimers through their hinge region upon interaction between the N and C-termini of SMC. The dimer is bound and stabilized by Scc1 and Scc3 upon their interaction with the N- and C-termini of SMC (Reproduced from Zlatanova and Caiafa, 2009).

Earlier, we identified the combinatorial expression of *Satb1* alternative transcripts in a cell-type specific manner during T-cell development (Chapter 1). Whether perturbing the activity of any of the promoters of *Satb1* have any effect on the combinatorial expression of transcripts is yet to be studied. Also, whether the interaction of any other *cis* regulatory DNA element with the *Satb1* alternative promoters mediate the combinatorial expression of *Satb1* transcript variants is not clear. Therefore, in the current study, we probed to investigate the molecular players which bring the combinatorial expression of *Satb1* transcript variants in the developing T-cells.

We identified the interaction of *Satb1* alternative promoters with the distal enhancer region. We show that the chromatin organizing proteins such as cohesin and CTCF bind to the *Satb1* alternative promoters. However, cohesin also binds to the distal enhancer region, wherein CTCF binding is absent. SATB1 binds to its alternative promoters as well as the distal enhancer region. Majority of the cohesion-bound and CTCF depleted regions were occupied by SATB1. We show that SATB1 physically interacts with cohesin and they co-localize in the thymocyte nuclei. SATB1 and cohesin bind to the *cis* regulatory elements including the promoters and enhancers of key genes of T-cells in the absence of CTCF binding. Our data suggests that the interaction of cohesin with the T lineage enriched chromatin organizer SATB1 presumably maintains the higher-order chromatin architecture in the developing T-cells and mediates the interactions between the *Satb1* alternative promoters and the distal enhancer region.

3.2 Results

3.2.1 The interaction profiling of *Satb1* alternative promoters with the distal enhancer region located upstream to the *Satb1* TSS

To study the regulation of combinatorial expression of *Satb1* transcript variants in a developmental stage specific manner during T-cell development, the understanding of chromatin interactions at the *Satb1* gene locus will be essential. Since the P3 promoter activity is constitutive throughout the development, we hypothesized as follows: 1) the P3 promoter could possibly interact with the differentially regulated alternative promoters in a cell type-dependent manner, 2) the distal regulatory elements such an enhancers may interact with the particular combination of promoters in a cell type-dependent manner, 3) the independent transcriptional events might occur at the alternative promoters of *Satb1*. However, all of these processes depend on the presence of cell type specific transcription factors during T-cell development. To understand the chromatin interactions at the *Satb1* gene locus, we analyzed the publicly available data sets of single cell Hi-C analysis (Nagano *et al.*, 2013) performed in the mouse Th1 cell type ($CD4^+$ T helper 1) using the 3D genome browser (Wang *et al.*, 2018). We analyzed the Hi-C analysis and plotted the results of virtual 4C analysis at the *Satb1* gene locus in nearly 40 Kb window (Figure 3.2.1 A). As shown by the virtual 4C analysis in the Figure 3.2.1 A, the intensity of peak at the *Satb1* locus indicates the interaction frequencies between the selected *Satb1* alternative promoters region (as shown in the rectangular box) with the chromatin in the downstream and upstream region of the selected promoters region. We found that the *Satb1* promoters interact with chromatin at the downstream region near the 3' end of *Satb1* gene and the upstream region of *Satb1* TSS (Figure 3.2.1 A). The upstream interacting region exhibits enhancer like signature as indicated by the enrichment of H3K27ac and lack of H3K4me3 peak (Figure 3.2.2 B). From this upstream region, the transcription of a long non-coding RNA takes place in a direction opposite to that of *Satb1* gene transcription (Figure 3.2.2 B). Since the downstream interacting region doesn't show any feature of containing the regulatory elements, we hypothesized that the interaction between the

upstream region and the *Satb1* alternative promoters might have a role in dictating the combinatorial expression of *Satb1* alternative promoters during T-cell development.

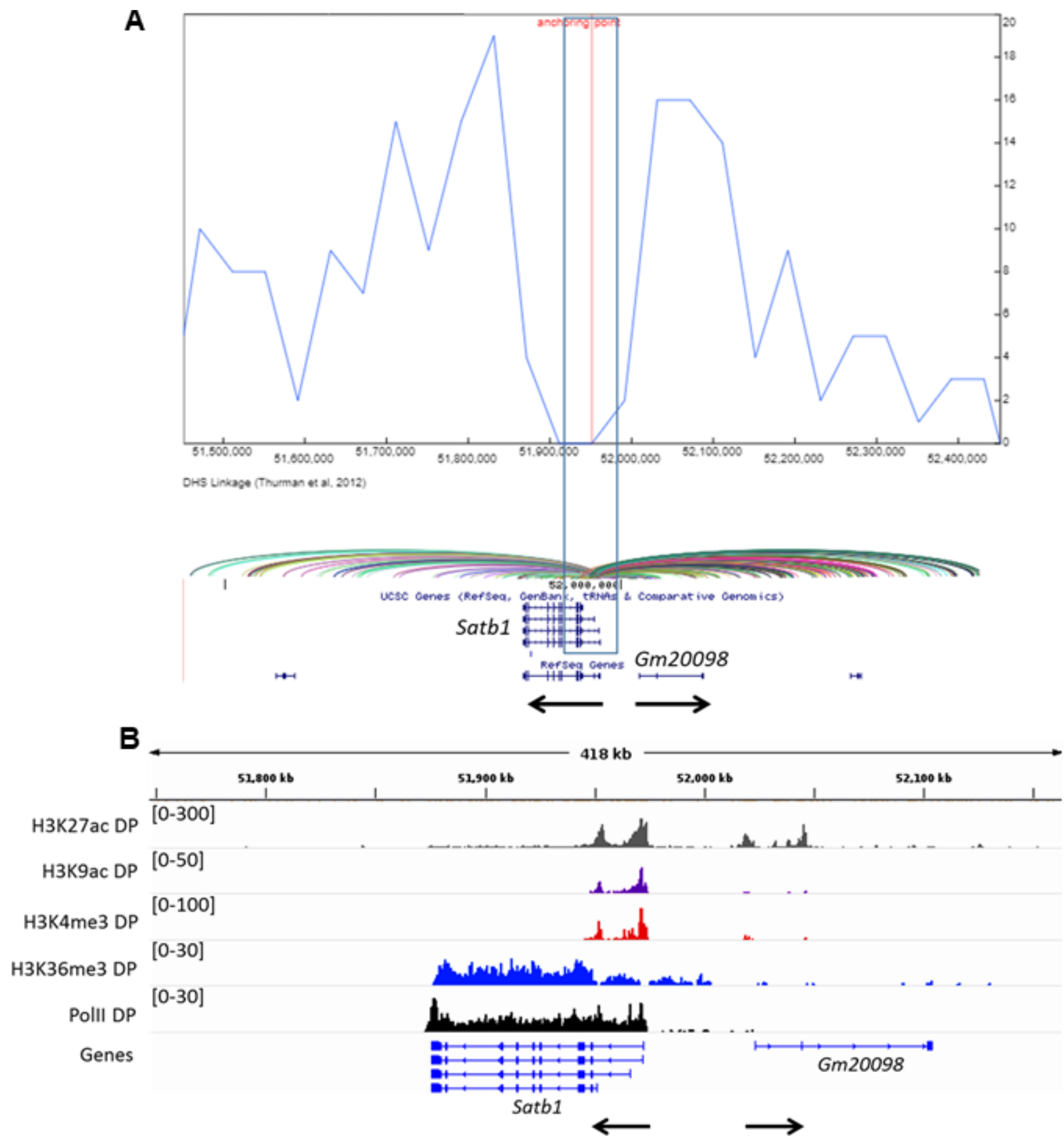


Figure3.2.1. Chromatin interactions at the *Satb1* gene locus. (A) The publicly available single cell Hi-C analyses performed in the mouse differentiated Th1 cell were used for the virtual 4C analysis which was performed using the 3D genome browser and was used for analyzing the chromatin interactions at the *Satb1* gene locus. The intensity of peak indicates the interactions of *Satb1* promoter regions with the

DNA regions located downstream and upstream of *Satb1* gene as mentioned with the rectangular box. The black color arrows indicate the transcriptional directionality of *Satb1* and the noncoding RNA Gm20098. (B) The ChIP-seq analyses of publicly available data sets of the histone marks such as H3K27ac, H3K4me3, H3K9ac, H3K36me3, along with RNA pol II ChIP-seq performed in the mouse double positive thymocytes were used for analysis at the *Satb1* gene locus to distinguish the regulatory DNA elements. The occurrence of signature histone modifications H3K4me3, H3K9ac, H3K4me3 at *Satb1* gene locus indicates the transcriptionally active promoters of *Satb1*. The presence of only H3K27ac and the lack of H3K4me3 at a region upstream of *Satb1* TSS indicates the putative enhancer element. The transcription of noncoding RNA occurs in the direction opposite to the *Satb1* gene transcription as indicated by arrows in the opposite direction.

3.2.2 The expression profiling of *Satb1* transcript variants in the Th1 cell type

Since we observed the interaction between *Satb1* alternative promoters and the distal enhancer region in Th1 cells as revealed by single cell Hi-C analysis (Figure 3.2.1), we further studied whether these chromatin interactions play an essential role in the combinatorial expression of the *Satb1* transcript variants in developing thymocytes. To address this, we performed the expression profiling of the *Satb1* transcript variants in Th1 cells to monitor whether these cells also exhibit the switching of the *Satb1* alternative promoters usage in a manner similar to the developing thymocytes. We differentiated naïve CD4$^+$ T-cells isolated from the mouse spleen into Th1 *in vitro* and performed the gene expression analysis by quantitative RT-PCR (qRT-PCR). Interestingly, we identified a switch in the *Satb1* alternative promoter usage even during Th1 differentiation. As shown in the Figure 3.2.2 A, we observed an increase in the expression of total *Satb1* transcripts in the differentiated Th1 cells compared to the naïve CD4$^+$ T-cells. The naïve CD4$^+$ T-cells exhibit the higher levels of P1 transcripts (Figure 3.2.2 A). Upon Th1 differentiation, the Th1 cells exhibit the expression of higher levels of P2 compared to the naïve (Figure 3.2.2 B). These results confirm that the *in vitro* differentiated Th1 cells also exhibit *Satb1* alternative promoter switch similar to the developing thymocyte stages.

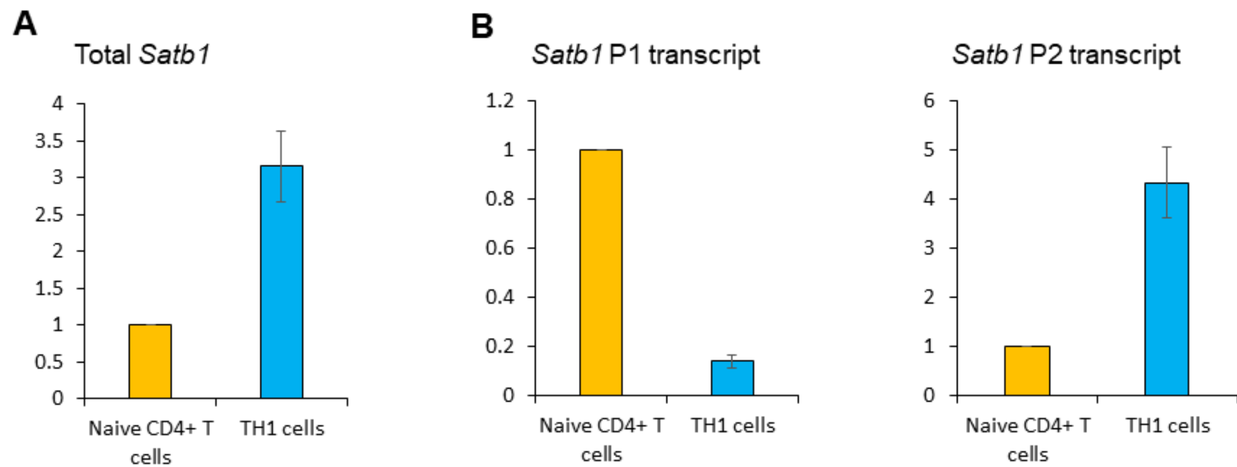


Figure 3.2.2.The expression profile of *Satb1* transcript variants in the differentiated Th1 cells. (A and B) Six week old C57BL/6 mice were sacrificed and spleen was isolated. Spleen was processed for the preparation of single cell supension and the RBCs were lysed using the lysis buffer as described in the 'Methods' section 3.4.2. The cells were subjected to the magnetic sorting using the mouse naïve CD4$^+$ T-cell isolation kit (MACS, Miltenyi Biotech) and naïve CD4$^+$ T-cells were isolated by negative the selection process. The iIsolated naïve CD4$^+$T-cells were differentiated to Th1 in the presence of IL12, IFNγ, anti-IL4, anti-CD3 and anti-CD28 for 72 hrs. After incubation, cells were harvested for RNA extraction using RNeasy Mini Kit (QIAGEN). The isolated RNA was subjected to cDNA synthesis and the quantitative gene expression analysis of *Satb1* transcript variants was performed. The expression of total *Satb1* and the *Satb1*transcript variants such as P1, P2 in the naïve and differentiated Th1 cell are shown.

3.2.3 The occupancy of chromatin architectural proteins at *Satb1* gene locus

Since we identified the interaction of *Satb1* alternative promoters with the upstream distal enhancer region, we then probed to investigate the protein complexes which could presumably mediate these chromatin interactions. To accomplish this, we used a strategy of ChIP-seq analysis and monitored for the co-occupancy of chromatin architectural proteins along with the signature histone modifications. Towards this, we used the publicly available data sets of ChIP-seq analysis of SMC1, CTCF along with H3K4me3, H3K27ac, H3K36me3, and H3K9ac histone marks (Shih *et al.*, 2012; Zhang *et al.*, 2012; Ing-Simmons *et al.*, 2015) to distinguish the promoter and the enhancer

elements. We identified the occupancy of chromatin architectural proteins such as CTCF and cohesion (SMC1 subunit) on the *Satb1* alternative promoters as well as at the distal enhancer region (Figure 3.2.3). Interestingly, the occupancy of CTCF was found only at the *Satb1* alternative promoters but not at the enhancer region. In contrast, cohesin (SMC1) occupancy was observed on both the alternative promoters and on the enhancer region (Figure 3.2.3). However, cohesin can't bind to the DNA efficiently and it requires another DNA-binding factor for its loading on to DNA (Wendt *et al*., 2008; Parelho *et al*., 2008; Rubio *et al*., 2008). Since CTCF does not bind the distal enhancer region of *Satb1*, we hypothesized that a T-cell-enriched DNA-binding factor might mediate the loading of cohesin on to the DNA. Interestingly, we identified the occupancy of SATB1 at the *Satb1* alternative promoters and also at the distal enhancer region. The overlay of ChIP-seq analysis of SATB1 (Kakugawa *et al.*, 2017) and cohesin together at the *Satb1* gene locus indicated that SATB1 and cohesin occupy at the same region of *Satb1* locus, wherein CTCF binding is absent (Figure 3.2.3). This result indicates that SATB1 might facilitate the loading of cohesin on to DNA in the absence of CTCF binding and facilitates the long-distance chromatin interactions.

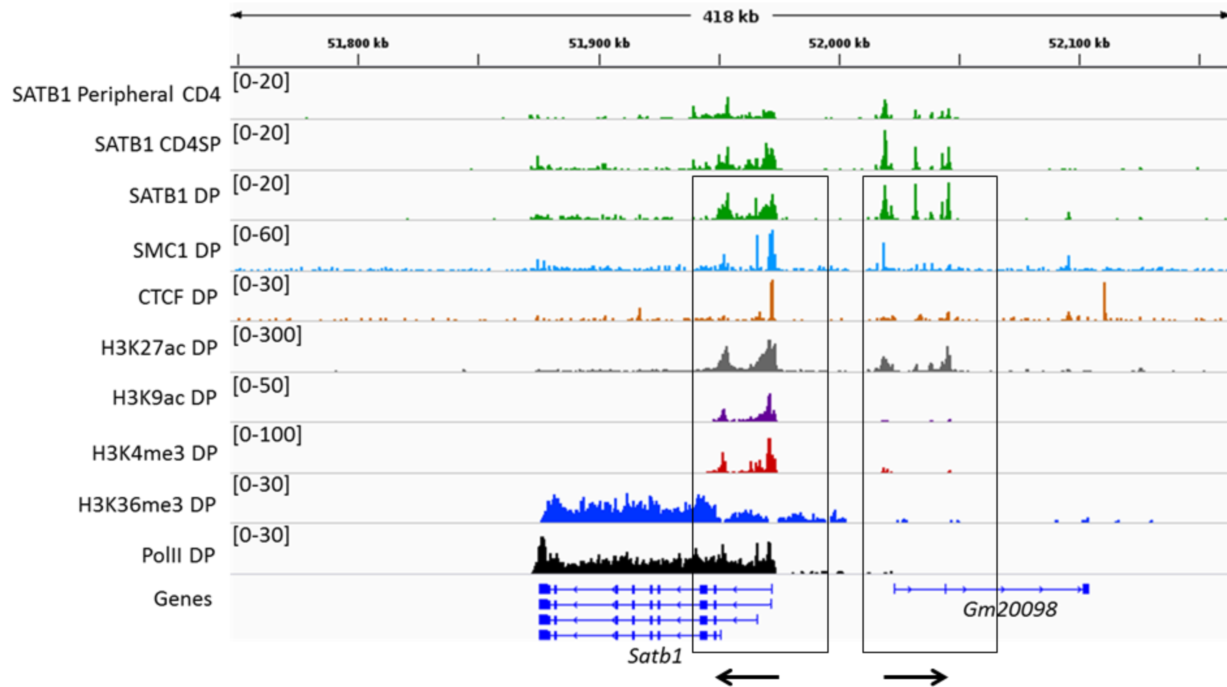


Figure 3.2.3. The occupancy of chromatin architectural proteins at the *Satb1* gene locus.The publicly available ChIP-seq data sets for cohesin (SMC1), CTCF, SATB1, H3K4me3, H3K27ac, H3K9ac, H3K36me3, and RNA pol II performed in the DP thymocytes were analyzed for the occupancy analysis at the *Satb1* gene locus. The presence of histone modifications H3K4me3, H3K9ac, H3K4me3 at the *Satb1* gene locus indicates the transcriptionally active promoters of *Satb1* which also exhibit the high enrichment of transcription elongation mark H3K36me3 throughout the *Satb1* gene body especially higher increase at the 3' end. We also observed higher enrichment of RNA pol II at the *Satb1* gene locus indicating an active transcription at this locus. The presence of only H3K27ac and lack of H3K4me3 enrichment at the region upstream of *Satb1* TSS indicates a putative enhancer like element, wherein the noncoding RNA transcription occurs in the direction opposite to the *Satb1* gene transcription as indicated by the arrows in the opposite direction. The overlay of SATB1, CTCF, and cohesin (SMC1) ChIP-seq analysis performed in DP thymocytes along with the histone modification marks indicating the co-occupancy of SATB1 and SMC1 at the *Satb1* alternative promoters and also at the distal enhancer region as depicted by the rectangular boxes. The CTCF occupancy was observed only at the *Satb1* alternative promoter region near the P3-P4 promoter.

3.2.4 SATB1 interacts with cohesin and co-localizes in the thymocyte nuclei

Since SATB1 and cohesin occupancy was observed at the same regulatory DNA regions in the DP thymocytes, we then asked whether these two proteins interact with each other. Additionally, since SATB1 is a DNA-binding protein, we asked whether the interaction of SATB1 with cohesion facilitates the loading of cohesion on DNA. We therefore performed co-immunoprecipitation of SATB1 and cohesion using thymocyte lysate and found that indeed SATB1 directly interacts with cohesin (Figure 3.2.4). We used anti-SATB1 for pulling down the protein complexes from thymocyte lysate and then performed western blotting using anti-SMC1 (cohesin). The immunoblot revealed the physical interaction between SATB1 and cohesin in thymocytes (Figure 3.2.4 A). Since SATB1 and cohesin interact with each other, we then monitored the subcellular compartment in which these two proteins interact. To address this, we first performed the immunostaining of thymocytes to monitor the localization of SATB1 and cohesin. The immunostaining analysis revealed their co-localization in the thymocyte nucleus (Figure 3.2.4 B), but not in the cytoplasm. The overlay of DAPI with SATB1 and cohesin immunostaining confirm that the interaction of SATB1 and cohesin is on the DNA i.e. within the nucleus and not outside of the nucleus (Figure 3.2.4 B). Therefore, the interaction of cohesin with SATB1 might enable the binding of cohesin onto the DNA in thymocyte nuclei. Further biochemical experiments are required to confirm this observation.

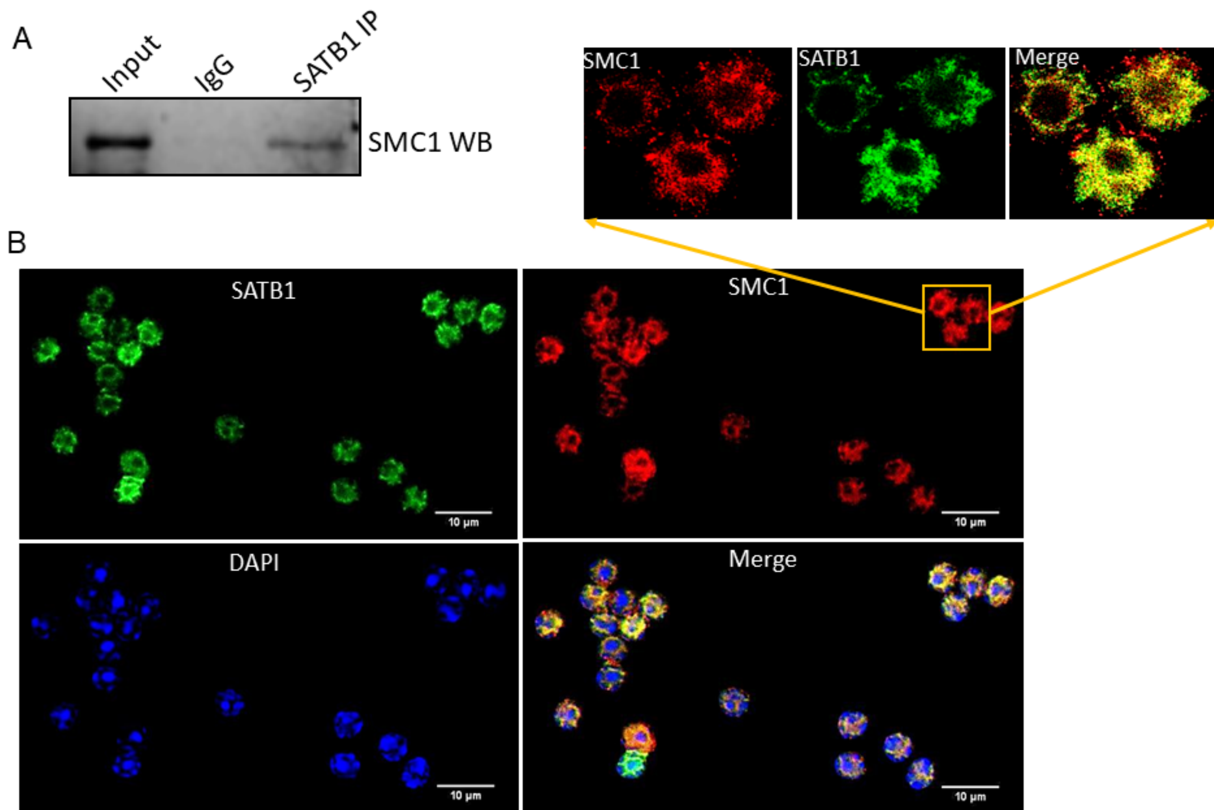


Figure 3.2.4. SATB1 and cohesin interact and co-localize in the thymocyte nuclei. (A) Three week old C57BL/6 mice were used for the isolation of thymii. The single cell suspension of thymocytes was prepared and thymocytes were lysed using the NP40 based lysis buffer as described in the 'Methods' section 3.4.6. The immunoprecipitation of SATB1-bound protein complexes from the thymocyte lysate was performed using anti-SATB1 antibody. Anti-SATB1 pulldown protein complexes were washed and subjected to the western blotting and immunoblots were probed using anti-SMC1 antibody. (B) The single cell suspension of thymocytes was used for the immunostaining using mouse anti-SATB1, rabbit anti-SMC1 and DAPI. The following secondary antibodies were used: anti-mouse GFP, anti-rabbit alexafluor 594. The individual and overlay images of anti-SATB1, anti-SMC1 and DAPI signals are shown. Inset shows the cells that were selected for the magnified image.

3.2.5 The SATB1- and cohesion-bound DNA regions are depleted of CTCF occupancy at the *Cd3* locus in mouse thymocytes

Since we observed that SATB1 and cohesin are highly co-localized in the thymocyte nucleus, we asked whether these two factors together regulate the expression of multiple T lineage specific genes other than *Satb1*. To understand this we analyzed the publicly available ChIP-seq data sets of SATB1, cohesin (SMC1) along with the active histone modification marks such as H3K4me3, H3K27ac, H3K9ac, H3K36me3, and also RNA pol II performed in the DP thymocytes as well as the CD4SP thymocytes (Shih *et al.*, 2012; Zhang *et al.*, 2012; Ing-Simmons *et al.*, 2015; Kakugawa *et al.*, 2017). Interestingly, we observed that SATB1 and cohesin occupancy was observed at the promoters and also at the previously reported super-enhancer regions of key genes in the T-cells, wherein the CTCF occupancy was characteristically reduced (Figure 3.2.5). We found that one of the key genes of T-cell development that encodes a component of TCR such as the *Cd3* locus was occupied by SATB1 and active histone marks in DP and CD4SP thymocytes, indicating that the active transcription occurs at the *Cd3* locus. The SATB1 bound regulatory DNA regions were also bound by cohesion and but not by CTCF. This result indicated that the T lineage enriched chromatin organizer SATB1 might be important for governing the DNA-binding profile of cohesin in the absence of CTCF occupancy and thereby maintaining the higher-order chromatin architecture of T-cells (Figure 3.2.5). We also found that the SATB1 and cohesion bound DNA regions were occupied by the active histone marks, indicating that these regulatory regions are transcriptionally active (Figure 3.2.5).

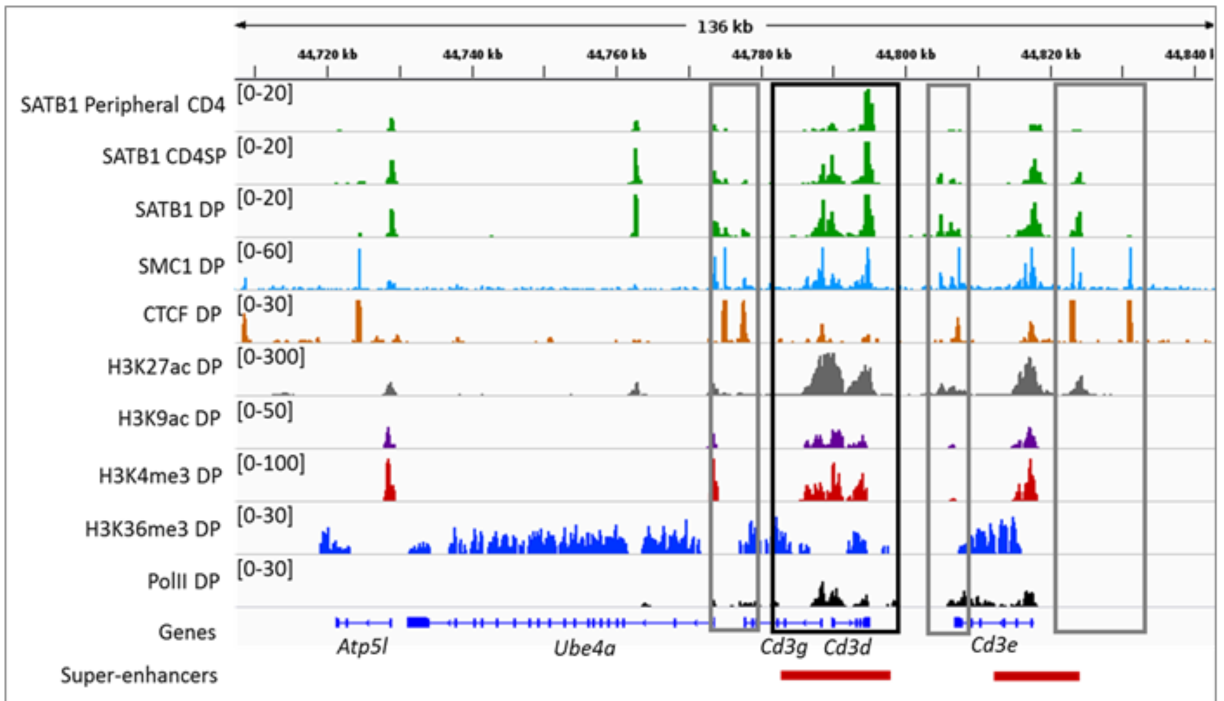


Figure 3.2.5. The occupancy of SATB1 and cohesin but not CTCF at the regulatory elements of *Cd3* locus in developing thymocytes. The publicly available data sets of ChIP-seq analysis of SATB1, cohesin (SMC1), CTCF, H3K27ac, H3K9ac, H3K4me3, H3K36me3 and RNA pol II performed in the double positive and CD4SP thymocytes were used for the analysis of their occupancies at the *Satb1* gene locus. The previously reported super-enhancer regions at the *Cd3* locus are indicated by the solid red lines at the bottom. The occupancy profile of SATB1, CTCF, and SMC1 at the *Cd3* locus has been shown. The regulatory regions marked by the black line box indicate the DNA regions occupied by both SATB1 and cohesin, but are depleted or less occupied by CTCF. The regulatory DNA regions in the grey color box lines depict the regions bound by both CTCF and cohesin (SMC1), but poorly occupied by SATB1.

3.2.6 The SATB1 and cohesin bound DNA regions are depleted of CTCF occupancy at the *Cd8* locus

Since we identified the binding of SATB1 and cohesin at key genes that were expressed throughout the development, we then studied whether the occupancy of SATB1 is affected at the regulatory elements of lineage specific genes. We analyzed the publicly available ChIP-seq of SATB1, and cohesin (SMC1) along with the active histone marks such as H3K4me3, H3K27ac, H3K9ac, H3K36me3, and also RNA pol II data sets performed in the double positive thymocytes as wells as in the CD4SP thymocytes (Shih *et al.*, 2012; Zhang *et al.*, 2012; Ing-Simmons *et al.*, 2015; Kakugawa *et al.*, 2017). We studied the binding profile of SATB1 at the CD8 T lineage specifying genes in the CD4SP thymocytes to address whether SATB1 can regulate the lineage specific gene expression during T-cell development. We then performed analysis of the ChIP-seq data of SATB1 and cohesin along with the active histone marks at the $CD8^+$ T lineage specific genes such as the *Cd8* locus, which includes both *Cd8a* as well as *Cd8b1.* During T-cell development, the *Cd8* gene is expressed at the DP stage and is maintained in the CD8SP T-cells and specifically downregulated in the CD4SP T-cells. We found that SATB1 binds to the regulatory regions of the *Cd8* gene locus including both the *Cd8a* and *Cd8b1* in DP thymocytes (Figure 3.2.6). The ChIP-seq analysis of SATB1 in DP thymocytes identified the occupancy of SATB1 at the promoter regions of *Cd8a* and *Cd8b1* gene loci as well as the *Cd8* super-enhancer region. Interestingly, SATB1 bound DNA regions were co-occupied by cohesin at the *Cd8* locus in DP thymocytes (Figure 3.2.6). SATB1 and cohesin bound DNA regions are devoid of CTCF at the *Cd8* locus in the DP thymocytes and are occupied by active histone marks and RNA pol II, an indicative of active transcription occurring at the *Cd8* locus in DP thymocytes (Figure 3.2.6). Interestingly, SATB1 occupancy was reduced at the *Cd8* locus in the CD4SP thymocytes, corroborating with the downregulation of *Cd8* in the CD4SP thymocyte subpopulation during T-cell development. However, SATB1 is dispensable for the expression of *Cd8* in DP thymocytes because the $CD4^+CD8^+$ double positive thymocyte development was not hampered in the SATB1 knockout mice (Alvarez *et al.*, 2000). This observation indicates that SATB1 is not the essential factor but it might cooperate with the other protein complexes to regulate the expression of

Cd8 in DP thymocytes. However, our analysis revealed that SATB1 binds to the *Cd8a* locus and work from others have also shown that it plays an essential role in the reinitiating the expression of *Cd8* during the $CD8^{+}$T lineage commitment from DP thymocytes. Thus, SATB1 plays an important role in the maturation of CD8 T-cells during the development (Banan, 1997; Nie *et al*., 2005; Nie*et al.*, 2008). Interestingly, when we compared the occupancy of SATB1 at the *Cd8* locus in both CD4 SP and in the peripheral CD4 T-cells, we observed reduced occupancy of SATB1 at the *Cd8* locus (Figure 3.2.6). We observed that instead of clustering of the SATB1 binding at the *Cd8* locus in DP thymocytes, SATB1 binding is reduced and only a single peak was observed near the 3' end of the *Cd8b1* locus, but no SATB1 peak was observed at the *Cd8a* locus (Figure 3.2.6). Collectively, these results suggest that SATB1 might play an important role in regulating the expression of the *Cd8* gene in a lineage specific manner during the CD4/CD8 T lineage commitment from DP and the reduced occupancy of SATB1 might be important for the repression of the *Cd8* gene in CD4 SP thymocytes, thereby maintaining the lineage specificity during development. This observation correlates with the previous finding that dysregulation in the expression of SATB1 leads to the partial redirection of MHCI/II selected thymocytes during T-cell development (Kakugawa *et al*., 2017), indicating that SATB1 might regulate target genes in a lineage specific manner during T-cell development.

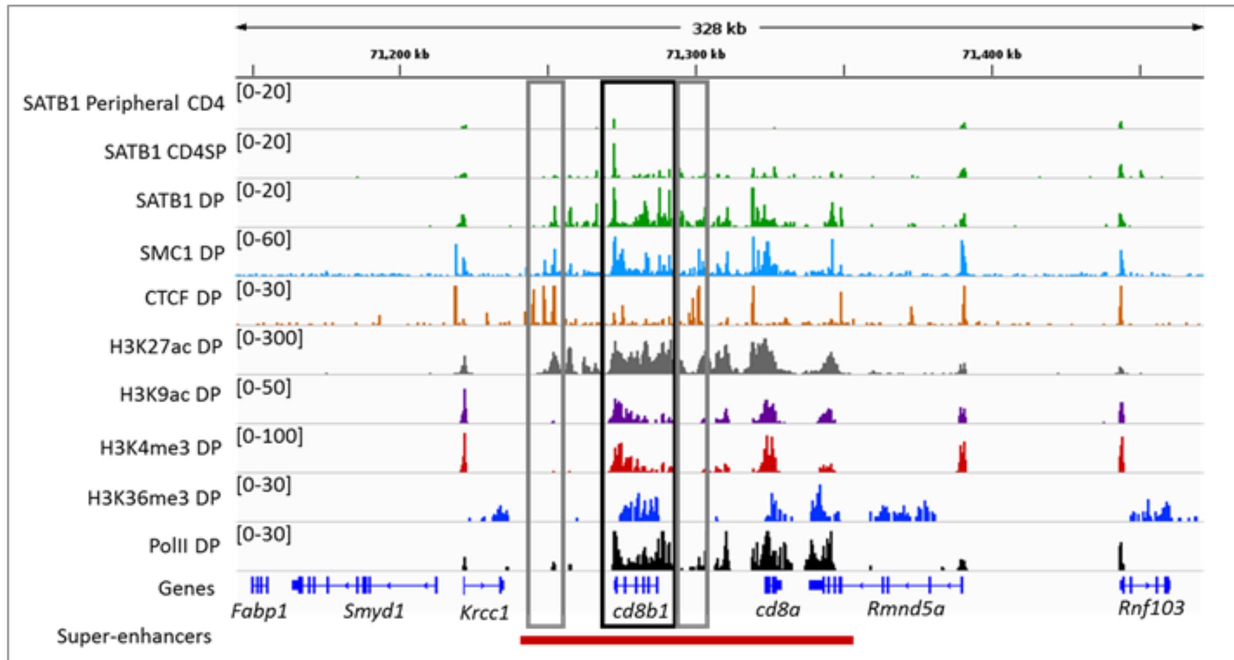


Figure 3.2.6. The occupancy of SATB1 and cohesin but not CTCF at the regulatory elements of the *Cd8* gene locus in developing thymocytes. (A) The publicly available data sets of ChIP-seq of SATB1, cohesin (SMC1), H3K4me3, H3K27ac, H3K9ac and RNA pol II performed in the DP thymocytes were analyzed at the *Cd8* gene locus including both the *Cd8a* and *Cd8b1* in double positive and also in the CD4SP thymocytes. The previously reported super-enhancer regions at the *Cd8* locus are indicated by the solid red line at the bottom. The occupancy profile of SATB1, CTCF, and SMC1 at the *Cd8* locus is shown. The regulatory regions marked by the black line box indicate the DNA regions occupied by both SATB1 and cohesin, but depleted or less occupied by CTCF. The regulatory DNA region in the grey line box indicates the CTCF and cohesin (SMC1) enriched regions that are poorly occupied by SATB1.

3.3 Discussion

In this chapter we describe the characterization of chromatin interactions at the *Satb1* gene locus. In the Chapters 1 and 2, we have shown that the *Satb1* transcript variants exhibit the combinatorial expression in a cell type dependent manner during T-cell development. However, it was not clear how the cell type dependent combinatorial activity of the *Satb1* alternative promoters was achieved. Therefore in the current study, we set out to characterize and understand the combinatorial regulation of *Satb1* alternative promoters. By analyzing the chromatin interaction frequencies at the *Satb1* gene locus using the 3D genome browser, we found that the *Satb1* alternative promoters interact with the upstream distal enhancer and also with the downstream DNA region at the *Satb1* gene locus as revealed by the Hi-C and virtual 4C analysis. To further characterize the regulatory elements in the interacting chromatin regions around the *Satb1* promoters, we have profiled the histone marks at the *Satb1* gene locus by using publicly available data sets. The analysis of occupancy of histone modification marks at the *Satb1* gene locus indicates that only the upstream distal enhancer region displays the enhancer specific histone marks such as higher enrichment of H3K27ac and low enrichment or lack of H3K4me3. However, no histone marks were found at the downstream interacting region of the *Satb1* gene locus. We observed that from the upstream enhancer region, the transcription of a noncoding RNA *Gm20098* takes place in an opposite direction to the transcription of *Satb1*. The role of this noncoding RNA is not known and further work is required to understand its function as well as role in the regulation of *Satb1* transcription, if any. We hypothesize that the interactions between the distal enhancer region and the *Satb1* alternative promoters might play an important role in regulating the combinatorial expression of *Satb1* alternative promoters during T-cell development.

The chromatin interactions in the genome are mediated by the protein complexes such as cohesin (SMC1) and CTCF. These two proteins mediate the chromatin interactions in spatial domains called topologically associated domains (TADs), thus maintaining the higher-order chromatin organization. These TAD structures are very important and are invariant in the nature across cell types during the development.

However, the chromatin interactions within these TADs are called sub TADs and are cell type specific. The boundary regions which maintains the TADs and the proteins complexes such as CTCF which bind to the boundary elements play an essential role in the organization of the TADs. The cell type specific expression of genes is achieved by an insulation function of the boundary regions and the binding of CTCF to these regions. These insulator regions prevent the action of enhancer on the genes in the neighboring TADs, thus allowing the specific enhancer-promoter interactions within the same domain, resulting in the cell type specific gene expression. The majority of CTCF bound sites in the genome are occupied by the cohesin complex which plays an important role in the organization of genome and in maintaining the cell type specific promoter-enhancer interactions.

When we assessed the protein complexes which bring about the chromatin interactions at the *Satb1* gene locus, we found that the chromatin organizing proteins cohesin and CTCF bind to the *Satb1* gene locus. Interestingly, we found the higher occupancy of cohesin and CTCF at the *Satb1* alternative promoters but not at the distal enhancer regions. Since cohesin alone doesn't bind to the DNA efficiently (Wendt *et al*., 2008; Parelho *et al*., 2008; Rubio *et al*., 2008), we then investigated the factors which might affect the chromatin interactions at the *Satb1* gene locus, in association with cohesin. Interestingly. we identified that the T-cell enriched chromatin organizer SATB1 binds to the *Satb1* gene regulatory regions, which are also bound by cohesin but not by CTCF. This observation suggests that SATB1 might interact with cohesin and recruit it onto the chromatin in the developing T-cells. To address this further, we performed the co-immunoprecipitation of SATB1 bound protein complexes which confirmed that indeed SATB1 interacts with SMC1, one of the subunits of the cohesin multi-subunit complex, in thymocytes. We then investigated the spatial localization of SATB1 and cohesin interaction by performing the immunostaining of SATB1 and cohesin in the thymocytes. Further, we show that SATB1 interacts with cohesin and co-localizes in the thymocyte nuclei. By overlaying the signal from DAPI which stains the DNA, with that of SATB1 and SMC1 immunostaining, we identified that SATB1 and SMC1 co-localizes on thymocyte DNA/chromatin indicating that SATB1 might assist cohesin in binding to the

DNA/chromatin in thymocytes. However, further biochemical characterization is required to understand the detailed mechanism of the interaction and also map their precise interaction domains.

Since these two proteins are highly co-localized in the thymocyte nucleus, we investigated whether these two proteins together maintain the T-cell genome organization. Towards this, we analyzed the publicly available ChIP-seq data of cohesin and SATB1 along with the active histone marks such as H3K4me3, H3K27ac, H3K9ac, H3K36me3, along with RNA pol II at the regulatory elements of key genes of T-cells such as *Cd3* and *Cd8*. We found that SATB1 binds to the regulatory elements of the *Cd3* locus in both DP and CDSP thymocytes, wherein both populations exhibit the expression of *Cd3* which further increases in the SP thymocytes. The majority of SATB1 bound DNA regions of the *Cd3* locus were also bound by cohesin but not by CTCF and these regulatory regions were also occupied by the active histone marks as well as RNA pol II, indicating that the *Cd3* locus is transcriptionally active in both DP and CD4SP thymocytes. We then assessed the binding profile of SATB1 on the lineage specific genes. The *Cd8* gene was expressed in the DP thymocytes and its expression was further maintained in the CD8SP T lineages and downregulated in case of CD4SP thymocytes. We found that SATB1 binds to the *Cd8* locus and a cluster of SATB1 occupancy was observed at the *Cd8* regulatory elements in the DP thymocytes. In DP thymocytes, majority of SATB1 bound regulatory DNA regions at the *Cd8* locus were also occupied by cohesin but not by CTCF. The same regulatory regions of the *Cd8* locus were also occupied by higher levels of active histone marks such as H3K4me3, H3K27ac, H3K9ac, and H3K26me3, indicating that SATB1 bound regions (here *Cd8* locus) were transcriptionally active in DP thymocytes. However, as previously reported SATB1 is not required for the expression of *Cd8* at the DP stage but is essential for the expression of *Cd8* during CD8 T lineage development from DP by directly binding to the *Cd8a* locus (Banan, 1997; Nie *et al*., 2005; Nie *et al*., 2008). Interestingly, we found that SATB1 does not bind to the*Cd8*a locus in the CD4SP thymocytes. Since *Cd8* expression is downregulated in the CD4SP thymocyes, the reduced SATB1 occupancy at the *Cd8* locus in CD4SP thymocytes is important to maintain the lineage specificity.

During T-cell development, the TCR signaling intensity plays an important role in the selection process as well as in the lineage specificity, wherein the persistent TCR signaling induces the CD4SP T-cell development, whereas cessation of TCR signaling leads to the CD8SP T-cell development (Singer, Adoro and Park, 2008). Along with the TCR signaling lineage specifying transcription factors play an essential role in the lineage specificity. The major transcription factors playing a role the lineage specificity are ThPOK in case of CD4SP and RUNX3 in case of CD8SP T-cells. ThPOK is important for the expression of CD4 T lineage specific genes including CD4 co-receptor and represses the expression of CD8 T lineage specific genes including CD8 co-receptor. On the other hand, RUNX3 allows the expression of CD8 T lineage signature genes and represses the CD4 T lineage specific genes. Along with these major transcription factors, the chromatin organizing protein such as SATB1 also plays an essential role in the thymocyte selection process and also in the expression of lineage specifying major transcription factors (Kakugawa *et al.*, 2017). Therefore in the SATB1 knockout mice, the T-cell development was blocked at the DP stage and further reduction in the number of SP thymocytes was observed (Alvarez *et al.*, 2000; Kondo *et al.*, 2016). The expression profiling of SATB1 in developing thymocytes shows that SATB1 is differentially expressed in different stages of T-cell development (Gottimukkala et al 2016; Chapters 1 and 2 in this thesis) and it is expressed abundantly in the CD4SP immature thymocytes. SATB1 regulates the expression of key genes of CD4/CD8 T lineage specificity such as ThPOK, RUNX3, and FOXP3during T-cell development. Though SATB1 was highly expressed in the CD4SP thymocytes, its occupancy was reduced at the *Cd8* locus in the CD4SP thymocytes might be to maintain the CD4 T lineage specificity in cooperation with the lineage specifying transcription factor. The dysregulation of SATB1 also results in the partial redirection of MHCI/II selected thymocytes during T-cell development (Kakugawa *et al.*, 2017). In light of these observations, it would be interesting to study whether the levels of SATB1 protein play any role towards the lineage specificity.

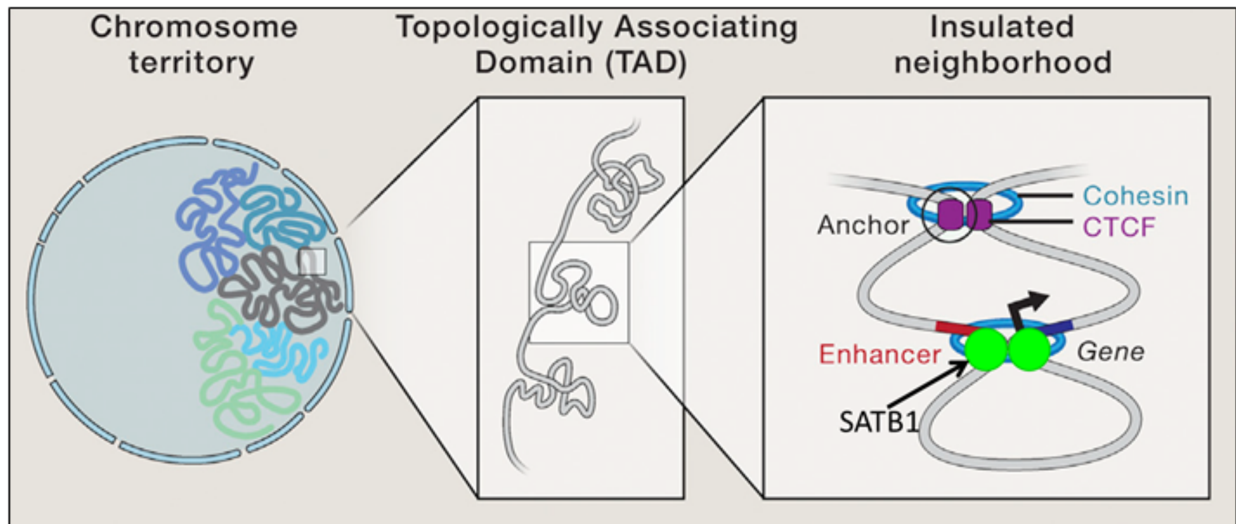


Figure 3.3. The role of SATB1 in maintaining the chromatin organization of developing T-cells. In the interphase of nucleus, the interactions between the chromatin regulatory elements folds the genome into spatially distinct domains called the topologically associated domains (TADs). Because of such higher-order chromatin interactions, the chromosomes occupy distinct territories and minimal space within the nucleus. The chromatin interactions within the TADs are called sub TADs which include the interactions between the regulatory elements such as enhancers and promoters. The insulation functions of the boundary regions and the proteins complexes such as CTCF binding to these regions maintains such domain organization. The chromatin architectural proteins such as cohesin and CTCF are the major players in maintaining the chromatin interactions, thus playing a role in the higher-order chromosome organization. The promoter enhancer interactions in the subdomains are cell type specific and mediated mostly by cohesin. Our work demonstrates that in developing T-cells, SATB1 binds to cohesin and maintains the chromatin interactions between the regulatory DNA elements, thus playing a role in T-cell genome organization. (Adapted from Hnisz, Day and Young, 2016).

In conclusion, our study shows that SATB1 and cohesin bind to the *cis* regulatory elements of key genes of developing T-cells, in the absence of CTCF binding. Therefore, our study strongly argues that the interaction of cohesin with the T lineage enriched chromatin organizer SATB1 plays an essential role in maintaining the higher-order chromatin architecture in developing T-cells.

Further studies on the effect of perturbation in any of the *Satb1* alternative promoter on the chromatin interaction at the *Satb1* locus and combinatorial expression of *Satb1* transcript variant in developing T-cells will be helpful to delineate the mechanism of *Satb1* expression in a stage specific manner during T-cell development. Additionally, characterization of the distal enhancer region will be helpful towards understanding its importance in maintaining the chromatin interactions at the *Satb1* gene locus.

3.4 Methods

3.4.1 Mice

3 week old C57BL/6 mice were used for the preparation of single cell suspension of thymocytes. All mice were bred and maintained under specific pathogen free environment and experimental procedures were performed according to the guidelines of animal house facility at IISER Pune and NCCS Pune.

3.4.2 Isolation of T-cells and Cell culture

Three week old C57/BL6 mice were used for isolation of thymus. Thymii were minced and passed through a 70 µM cell strainer to remove the debris. Single cell suspension of thymocytes were used for the immunoprecipitation and immunostaining experiments. In case of isolation of naïve $CD4^+$ T-cells, 6 week old C57BL/6 mice were used for the isolation of spleen. Spleen was used to prepare the single cell suspension and RBCs were lysed using RBC lysis buffer. Cells were passed through 70 µM cell strainer to remove debris. Single cell suspension of splenocytes was subjected to magnetic sorting by the process of negative selection using the mouse naïve $CD4^+$ T-cell isolation kit (MACs, Milteny Biotech). Now, naïve $CD4^+$ T-cells were cultured in RPMI medium supplemented with 10% fetal bovine serum, and 1%penicillin and streptomycin. Naïve $CD4^+$ T-cells were cultured in a 12 well plate which was pre-coated with 0.5 µg/ml of anti-CD3 (Clone 17A2, eBioscience) and 1.5 µg/ml of anti-CD28 (Clone 37.51, eBioscience). Cells were then polarized towards Th1 differentiation by adding 10 ng/mL of IL12, 10 ng/mL of IFNγ, and 10 ng/mL of anti-IL4 for 72 hrs. All of the above mentioned cytokines and antibodies were purchased from R&D Systems and eBiosciences. After incubation, the cells were harvested and used for quantitative qRT-PCR analysis.

3.4.3 cDNA synthesis and Quantitative PCR analysis (qRT-PCR)

Isolation of total RNA from naïve and differentiated Th1 cells was performed using Qiagen RNeasy mini kit (QIAGEN). Following DNase I (Promega) digestion, the RNA was subjected to cDNA synthesis using High capacity cDNA synthesis kit (Applied

Biosystems). Quantitative PCR analyses were performed using Sybr green qPCR master mix (Roche) at the following PCR conditions: step 1, 95°C-5 min; step 2, 95°C-45 sec, 60°C-45 sec, 72°C-1 min for 40 cycles. The following qPCR primers were used:

Satb1 (total)-F: 5'-TGATAGAGATGGCGTTGCTG-3'
Satb1 (total)-R: 5'-TTTTGAGGGTGACCACATGA-3'
P1(E1a)-F: 5'-CAAGAATCCCGGCTGCAAAG-3'
P1(E1a)-R: 5'-CCCTGAGTTGCCTCGTTCAA-3'
P2(E1b)-F: 5'-AGATTCGGAAACCAGCCTCTG-3'
P2(E1b)-R: 5'-GGACCCTTCGGATCACTCAC-3'
m18s-F: 5'-GTAACCCGTTGAACCCCATT-3'
m18s-R: 5'-CCATCCAATCGGTAGTAGCG-3'

3.4.4 ChIP-seq and HiC analysis

Publicly available data sets (GSE61428, GSE90635, GSE32311, and GSM1023418) were used for ChIP-seq analysis of H3K4me3, H3K27ac, H3K9ac, H3K36me3, SMC1, CTCF, RNA pol II and SATB1 performed in developing thymocytes at *Satb1*, *Cd3,* and *Cd8* gene loci (in DP and CD4SP) (Shih *et al.*, 2012; Zhang *et al.*, 2012; Ing-Simmons *et al.*, 2015; Kakugawa *et al.*, 2017).ChIP-seq reads were mapped by using Bowtie 2 and peak calling was performed using MACS2. MACS generated peaks were visualized by the IGV genome browser. GSE48262 data set (Nagano *et al.*, 2013) was used for the analysis of Hi-C and virtual 4C analysis at *Satb1* gene locus by using the 3D Genome Browser (Wang *et al.*, 2018).

3.4.5 Immunostaining of thymocytes

Single cell suspension of thymocytes was prepared from 3 week old C57BL/6 mice. Thymocytes were fixed with 2% of paraformaldehyde. Fixed thymocytes were subjected to the permeabilization using 0.1% Triton X-100. After permeabilization, thymocytes were incubated with mouse anti-SATB1 (BD Biosciences), and rabbit anti-SMC1 (Abcam) for 3hrs at room temperature. After intracellular staining, thymocytes were washed with 1X PBS containing 0.01% Tween 20. The following fluorochrome tagged

secondary antibodies: anti-mouse GFP, and anti-rabbit alexafluor 594 were used for secondary antibody staining at room temperature for 1 hr. After secondary antibody staining, DNA was stained using DAPI (Sigma). Cells were visualized and Z-stack images were captured using Anisotropy microscope (Carl Zeiss).

3.4.6 Co-immunoprecipitation and Immunoblotting

Single cell suspension of thymocytes was prepared as soon as 3 week old mice were dissected. Cells were lysed using NP40 based lysis buffer and the protein concentration was measured using BCA method (Thermo scientific). Nearly 500μg of protein was precleared using 1μg mouse or rabbit IgG and protein A/G dyna beads (Roche). The precleared lysate was subjected to immune precipitation by using anti-SATB1 antibody for 4hrs at 4°C. After incubation SATB1-bound protein complexes were pull down using protein A/G dyna beads (Roche) and eluted using SDS containing buffer. The eluted protein complexes were loaded on to 10% SDS-polyacrylamide gel. After electrophoresis, protein was then transferred to PVDF membrane and probed using anti-SMC1 antibody (1:1000, Abcam). The signals were visualized using ECL luminescence detection reagent (BIO-RAD) on ImageQuant LAS 4000 system (GE Healthcare Life Sciences).

3.5 References

Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, Kohwi-Shigematsu T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev *14,* 521-35.

Banan M, Rojas IC, Lee WH, King HL, Harriss JV, Kobayashi R, Webb CF, Gottlieb PD. (1997). Interaction of the nuclear matrix-associated region (MAR)-binding proteins, SATB1 and CDP/Cux, with a MAR element (L2a) in an upstream regulatory region of the mouse CD8a gene. J Biol Chem *272,* 18440-52.

Butler JE, Kadonaga JT. (2001). Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev *15,* 2515-9.

Choi OR, Engel JD. (1988). Developmental regulation of beta-globin gene switching. Cell *55*, 17-26.

Chung JH, Whiteley M, Felsenfeld G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell *74*, 505-14.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485,* 376-80.

Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schujiers J, Lee TI, Zhao K, Young RA. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159,* 374-387.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473,* 43-9.

Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337-43.

Filippova GN. (2008). Genetics and epigenetics of the multifunctional protein CTCF. Curr Top Dev Biol *80,* 337-60.

Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. (2008). A code for transcription initiation in mammalian genomes. Genome Res 18, 1-12.

Gaszner M, Felsenfeld G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet *7*, 703-13.

Gottimukkala KP, Jangid R, Patta I, Sultana DA, Sharma A, Misra-Sen J, Galande S. (2016). Regulation of SATB1 during thymocyte development by TCR signaling. Mol Immunol *77,* 34-43.

Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc Natl Acad Sci U S A *109,* 17507-12.

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-47.

Hnisz D, Day DS, Young RA. (2016). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell *167*, 1188-1200.

Ing-Simmons E, Seitan VC, Faure AJ, Flicek P, Carroll T, Dekker J, Fisher AG, Lenhard B, Merkenschlager M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. Genome Res *25*, 504-13.

Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, Weintraub AS, Hnisz D, Pegoraro G5, Lee TI, et al. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell *18*, 262-75.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467,* 430-5.

Kakugawa K, Kojo S, Tanaka H, Seo W, Endo TA, Kitagawa Y, Muroi S, Tenno M, Yasmin N, Kohwi Y, Sakaguchi S, Kowhi-Shigematsu T, Taniuchi I. (2017). Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. Cell Rep 19, 1176-1188.

Kellum R, Schedl P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell *64,* 941-50.

Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, Resch W, Baek S, Pruett N, Grøntved L, et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell *155,* 1507-20.

Kondo M, Tanaka Y, Kuwabara T, Naito T, Kohwi-Shigematsu T, Watanabe A. (2016). SATB1 Plays a Critical Role in Establishment of Immune Tolerance. J Immunol *196,* 563-72.

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84-98.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-93.

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, at al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, at al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190-5.

Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59-64.

Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017-21.

Nie H, Maika SD, Tucker PW, Gottlieb PD. (2005). A role for SATB1, a nuclear matrix association region-binding protein, in the development of CD8SP thymocytes and peripheral T lymphocytes. J Immunol *174*, 4745-52.

Nie H, Yao X, Maika SD, Tucker PW. (2008). SATB1 is required for CD8 coreceptor reversal. Mol Immunol *46*, 207-11.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-5.

Ohlsson, R., Renkawitz, R. and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet *17*, 520-527.

Ohtsuki, S., Levine, M. and Cai, H. N. (1998). Different core promoters possess distinct regulatory activities in the Drosophila embryo. Genes and Development *12*, 547-56.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422-433.

Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R,

Dekker J, Taylor J, Corces VG. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-95.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-80.

Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A *105*, 8309-14.

Sanyal A, Lajoie BR, Jain G, Dekker J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109-13.

Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, Lenhard B, Giorgetti L, Heard E, Fisher AG, Flicek P, Dekker J, Merkenschlager M. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res 23, 2066-77.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-72.

Shih HY, Verma-Gaur J, Torkamani A, Feeney AJ, Galjart N, Krangel MS. (2012). Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub. Proc Natl Acad Sci U S A *109*, E3493-502.

Shlyueva, D., Stampfel, G. and Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. Nature Reviews Genetics *15*, 272-86.

Singer, A., Adoro, S. and Park, J. H. (2008). Lineage fate and intense debate: Myths, models and mechanisms of CD4- versus CD8-lineage choice. Nature Reviews Immunology *8*, 788-801.

Smale, S. T. and Baltimore, D. (1989). The “initiator” as a transcription control element. Cell *57*, 103-13.

Smale, S. T. and Kadonaga, J. T. (2003). The RNA Polymerase II Core Promoter. Annual Review of Biochemistry. *72*, 449-79.

Sofueva S, Yaffe E, Chan WC, Georgopoulou D, Vietri Rudan M, Mira-Bontenbal H, Pollard SM, Schroth GP, Tanay A, Hadjur S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J *32*, 3119-29.

Spitz, F. and Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. Nature Reviews Genetics *13,* 613-26.

Strunnikov, A.V., Larionov, V.L., Koshland, D. (1993). SMC1: An essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family. J. Cell Biol *123*, 1635–1648.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep *10*, 1297-309.

Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, Li D, Choudhary MNK, Li Y, Hu M, et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol *19*, 151.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T. et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796-801.

Wutz G, Várnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ, et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J *36*, 3573-3599.

Yusufzai, T. M. and Felsenfeld, G. (2004). The 5-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. Proc. Natl. Acad. Sci. USA *101,* 8620-8624.

Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell *13,* 291-8.

Zhang J, Jackson AF, Naito T, Dose M, Seavitt J, Liu F, Heller EJ, Kashiwagi M, Yoshida T, Gounari F, et al. (2011). Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. Nat Immunol *13*, 86-94.

Zlatanova J, Caiafa P. (2009). CTCF and its protein partners: divide and rule. J Cell Sci. *122*, 1275-84.

Publications

A. Published:

1. Gottimukkala KP, Jangid R, **Patta I,** Sultana DA, Sharma A, Misra-Sen J, Galande S. (2016). Regulation of SATB1 during thymocyte development by TCR signaling. Mol Immunol *77,* 34-43.

2. Khare S, Shetty A, Biradar R, **Patta I**, Sathe A, Reddy PC, Chen Z, Lahesmaa R and Galande S. (2019). NF-κB signaling and IL-4 signaling regulate SATB1 expression via alternative promoter usage during Th2 differentiation. Front. Immunol. doi: 10.3389/fimmu.2019.00667. (In press).

B. Manuscript submitted

3. **Indumathi Patta**, Ayush Madhok, Satyajeet Khare, Kamalvishnu P Gottimukkala, Shilpi Giri, Dandewad V, Jyoti Misra-Sen, Vasudevan Seshadri, Girdhari Lal, and Sanjeev Galande (2019). Dynamic regulation of chromatin organizer SATB1 via TCR induced alternative promoter switch during T-cell development. (Submitted).

C. Manuscript under preparation:

4. **Indumathi Patta**, Ayush Madhok, and Sanjeev Galande (2019). Mechanism of combinatorial regulation of *Satb1* alternative promoters in T-cells. (Manuscript under preparation).