Dedicated to my parents, Harivadan Patel and Jyotsana Patel
ACKNOWLEDGMENTS

I owe my gratitude to all those who have made this dissertation possible. My deepest acknowledgments go to my mentor Dr. Suming Huang, who provided me with this opportunity and guided me whole-heartedly throughout the course of this study. I am very grateful to the members of my committee Dr. Michael Kilberg, Dr. Jörg Bungert, Dr. Jianrong Lu and Dr. David Bloom, for all their useful comments, suggestions and constructive criticisms. I would like to thank all the members of the Huang lab, especially Dr. Changwang Deng for his guidance and assistance with various experimental procedures.

I am obliged to my colleague and friend Dr. Betsabeh Khoramnian, who always offered support and shared scientific ideas. I am thankful to T32 training grant in cancer research for providing me with financial assistance and to the interdisciplinary program at the University of Florida for providing an excellent platform to pursue my Ph.D. studies. I am truly indebted to all who contributed to this work either by sharing reagents or providing collaborative or administrative help.

Most Importantly, I am grateful to my family for their unconditional love and support, for never doubting my choices and for always being there. I would have not endured these years without my husband, Pritesh, who was a constant source of encouragement and patience. I am thankful to my parents-in-law for being very supportive and taking care of my daughter so that I could fulfill my career goals. Lastly, I take this opportunity to acknowledge my parents, whom this dissertation is dedicated; who made me see beyond the clouds, inspired my academic goals, and provided an excellent example of hard work, strength and perseverance.
TABLE OF CONTENTS

ACKNOWLEDGMENTS.................................................................................................................. 4

LIST OF TABLES.......................................................................................................................... 8

LIST OF FIGURES....................................................................................................................... 9

LIST OF ABBREVIATIONS.......................................................................................................... 11

ABSTRACT ................................................................................................................................... 13

CHAPTER

1 INTRODUCTION ....................................................................................................................... 15

   T-cell Acute Lymphoblastic Leukemia (T-ALL) ..................................................................... 15
   Clinical Presentation................................................................................................................. 16
   Molecular Abnormalities ......................................................................................................... 16
   Prognosis of T-ALL .................................................................................................................. 17
   Current Treatment Therapies for T-ALL .................................................................................. 17

   TAL1 (T-cell Acute Lymphoblastic Leukemia 1): A Master Regulator of
   Hematopoiesis.......................................................................................................................... 18
   TAL1 Role in Hematopoiesis ................................................................................................... 18
   TAL1 Role in T-ALL ................................................................................................................ 19
   TAL1 Gene Is Regulated By Cis-acting Factors in Erythroid Cells ....................................... 20
   TAL1 Gene Is Regulated By Trans-acting Factors In Erythroid Cells ................................. 21
   TAL1 Gene Regulation In T-ALL ............................................................................................ 22

2 MATERIALS AND METHODS ............................................................................................... 28

   Cell Culture and ShRNA-Mediated Knock-Down ................................................................. 28
   Primary Human Cord Blood Cell Culture and Erythroid Differentiation ............................ 28
   Giemsa/May-Grunwald Staining Of Erythroid Cells ............................................................ 29
   FACs Analysis of (Cluster of Differentiation) CD34+, CD36+ and (Glycophorin 
   A) GPA Surface Markers ....................................................................................................... 29
   Native and Cross-linked Chromatin Immunoprecipitation Assay ....................................... 29
   Chromosome Conformation Capture Assay (3C) ................................................................... 31
   Calculation of the Relative Cross-linking Frequency Using Image J .................................. 32
   Reverse Transcription and Quantitative Polymerase Chain Reaction ............................... 32
   CD34+ Colony Formation Assay ............................................................................................ 33
   Luciferase Reporter Assay ....................................................................................................... 33
   Soft Agar Colony Formation Assay ........................................................................................ 34

3 SET1 AND CTCF FACILITATE TAL1 PROMOTER/ENHANCER
   INTERACTION IN ERYTHROID PROGENITORS ............................................................... 40
4 TAL1 GENE REGULATION IN T-CELL LEUKEMIA

Introductory Remarks

Results

Histone Modification Profile at the TAL1 Locus in T-ALL Cells
Enhancer +51 is Epigenetically and Functionally Silent in T-ALL
SET1 Independent TAL1 Gene Regulation in T-ALL
CTCF Mediated Cell-type Specific Chromatin Loops in the TAL1 Locus
Block TAL1 Enhancer/Promoter Interaction in T-ALL
Summary

5 ROLE OF LIM DOMAIN BINDING PROTEIN 1 IN TAL1 GENE REGULATION IN ERYTHROID CELLS

Introductory Remarks

Results

Ldb1 Depletion Results in Decreased TAL1 mRNA and Protein Levels
SET1 Depletion Alters Ldb1 Levels in Erythroid Cells
Ldb1 Regulation of TAL1 Gene is Independent of H3K4me2&3
Summary

6 SUMMARY, DISCUSSION AND FUTURE DIRECTIONS

Summary and Discussion
Putative TAL1 Interacting Regions in T-ALL
Future Directions
Evaluating the Factors Regulating Cell-type Specific Role for CTCF
CTCF Depletion Studies in Erythroid and in T-ALL Cells
Blocking CTCF Activity by Utilizing Artificial Zinc-Finger DNA Binding Domains
APPENDIX

A. COHESIN’S ROLE IN TAL1 GENE REGULATION IN K562 AND JURKAT CELLS ........................................................................................................ 98

B. EFFECT OF SET1 DEPLETION ON CTCF SITE INTERACTIONS AT THE TAL1 LOCUS IN ERYTHROID CELLS ................................................................. 99

C. PRMT1-MEDIATED ASYMMETRIC DIMETHYL H4R3 CROSS-TALKS WITH H3K4 METHYLATIONS AT THE β-GLOBIN LOCUS. ..................................................... 100

LIST OF REFERENCES ............................................................................................................. 108

BIOGRAPHICAL SKETCH........................................................................................................... 125
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Primer sequences</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>1-1</td>
<td>Human TAL1 locus and protein isoforms</td>
</tr>
<tr>
<td>1-2</td>
<td>Distribution of genomic rearrangements involving TAL1 in T-cell acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>1-3</td>
<td>TAL1 is a key regulator of hematopoiesis. Shown is the brief outline of normal hematopoiesis</td>
</tr>
<tr>
<td>2-1</td>
<td>Ex-vivo differentiation of human cord blood CD34+ cells</td>
</tr>
<tr>
<td>3-1</td>
<td>Histone modification associated with TAL1 enhancer and promoter activity in erythroid cells</td>
</tr>
<tr>
<td>3-2</td>
<td>TAL1 promoter 1 and +51 enhancer interact in vivo in CD36+ erythroid progenitors</td>
</tr>
<tr>
<td>3-3</td>
<td>H3K4 methyltransferase SET1 is required for the colony formation capacity of K562 and CD34+ HSCs</td>
</tr>
<tr>
<td>3-4</td>
<td>SET1 is recruited at the TAL1 enhancer/promoter elements and is required for activation of various erythroid genes including TAL1</td>
</tr>
<tr>
<td>3-5</td>
<td>Loss of hSET1 disrupts the erythroid specific long-range interaction between +51 enhancer and the TAL1 promoter 1 at the TAL1 locus</td>
</tr>
<tr>
<td>3-6</td>
<td>Similar CTCF occupancy is observed across different cellular environments</td>
</tr>
<tr>
<td>3-7</td>
<td>CTCF organizes TAL1 locus to promote TAL1 promoter/enhancer interaction in erythroid cells</td>
</tr>
<tr>
<td>4-1</td>
<td>Histone modification profile at the TAL1 locus in T-ALL REX and HPB-ALL cells</td>
</tr>
<tr>
<td>4-2</td>
<td>Histone modification profile at the TAL1 locus in T-ALL Jurkat cells</td>
</tr>
<tr>
<td>4-3</td>
<td>TAL1 +51 enhancer is neither functionally active nor does it interact with TAL1 promoter in T-ALL cells</td>
</tr>
<tr>
<td>4-4</td>
<td>SET1 is neither recruited to the TAL1 locus nor is it required for TAL1 expression in T-ALL Jurkat cells</td>
</tr>
<tr>
<td>4-5</td>
<td>Assessment of TAL1 expression and chromosomal rearrangements in primary T-ALL patient sample</td>
</tr>
</tbody>
</table>
CTCF reorganizes TAL1 locus to block +51 enhancer mediated promoter activation in T-cell leukemia. ................................................................. 76

CTCF depletion in T-ALL Jurkat cells induces TAL1 transcription. .............. 77

Ldb1 is required for high levels of TAL1 in erythroid cells. ......................... 83

Ldb1 regulates TAL1 transcription by stabilizing long range TAL1 promoter - enhancer interaction in K562 cells. .......................................................... 84

SET1 regulates Ldb1 levels in K562 and in MEL cells. ................................. 85

Ldb1 depletion does not alter H3K4me2/3 enrichment at the TAL1 enhancer and promoter. .................................................................................. 86

Model: CTCF and epigenetic mediated chromatin looping regulated TAL1 expression in hematopoiesis and leukemogenesis .............................................. 93

No difference is observed in Cohesin occupancy at the TAL1 locus in K562 and Jurkat cells. ......................................................................................... 98

No effect of SET1 depletion on CTCF -31 and +53 site chromatin loop in erythroid K562 cells. ......................................................................................... 99

USF1 associated complexes. .......................................................................... 102

PRMT1 depletion affects global H3K4me2/3 levels. ...................................... 103

PRMT1 is required for β and α-globin transcription in EPO differentiated mES cells. ............................................................................................... 104

PRMT1 is required for active H3K4 di and tri-methylations at the mouse β-globin locus. ......................................................................................... 105

Re-introduction of PRMT1 rescues H3K4me2/3 levels ................................. 106

H4R3ame is sufficient to allow SET1 mediated H3K4me2/3 in vitro. A) Experimental strategy for in vitro methylation assay ........................................... 107
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>βHLHB1</td>
<td>Basic domain, helix-loop-helix protein, class B, 1</td>
</tr>
<tr>
<td>CD34+</td>
<td>Cluster of differentiation 34+</td>
</tr>
<tr>
<td>CD36+</td>
<td>Cluster of differentiation 36+</td>
</tr>
<tr>
<td>COMPASS</td>
<td>Complex proteins associated with Set1</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor 1</td>
</tr>
<tr>
<td>GATA1</td>
<td>Globin transcription factor 1</td>
</tr>
<tr>
<td>GPA</td>
<td>Glycophorin a</td>
</tr>
<tr>
<td>HCFC1</td>
<td>Host cell factor C1</td>
</tr>
<tr>
<td>HOX11</td>
<td>Homeobox protein 11</td>
</tr>
<tr>
<td>JAK</td>
<td>Janus Kinase</td>
</tr>
<tr>
<td>LDB1</td>
<td>LIM domain binding protein 1</td>
</tr>
<tr>
<td>LYL1</td>
<td>Lymphoblastic leukemia derived sequence 1</td>
</tr>
<tr>
<td>MAP17</td>
<td>Membrane associated protein 17</td>
</tr>
<tr>
<td>MLL</td>
<td>Mixed lineage leukemia</td>
</tr>
<tr>
<td>MYC</td>
<td>Myelocytomatosis viral oncogene</td>
</tr>
<tr>
<td>NRF</td>
<td>Nucleosome remodeling factor</td>
</tr>
<tr>
<td>OGT</td>
<td>O-linked N-acetylglucosamine (glcnac) transferase</td>
</tr>
<tr>
<td>PAF</td>
<td>RNA polymerase II associated factor</td>
</tr>
<tr>
<td>RUNX1</td>
<td>Runt-related transcription factor 1</td>
</tr>
<tr>
<td>SCL</td>
<td>Stem cell leukemia</td>
</tr>
<tr>
<td>SET1</td>
<td>Su(var), enhancer of zeste, trithorax domain protein 1</td>
</tr>
<tr>
<td>SIL</td>
<td>Scl interrupting locus</td>
</tr>
<tr>
<td>TAL1</td>
<td>T-cell acute lymphoblastic leukemia protein 1</td>
</tr>
<tr>
<td>TAL2</td>
<td>T-cell lymphoblastic leukemia protein 2</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>T-ALL</td>
<td>T-cell acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>TCL5</td>
<td>T-cell leukemia 5</td>
</tr>
<tr>
<td>TET</td>
<td>Ten-eleven translocation protein</td>
</tr>
</tbody>
</table>
Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

EPIGENETIC MECHANISMS REGULATING ACTIVATION OF THE TAL1 ONCOGENE IN NORMAL AND IN MALIGNANT HEMATOPOIESIS

By

Bhavita Patel

May 2013

Chair: Suming Huang
Major: Medical Sciences

TAL1/SCL is a critical transcription factor required for development of all hematopoietic lineages; yet, aberrant TAL1 transcription which is frequently observed in T-cell acute lymphoblastic leukemia (T-ALL), leads to leukemia manifestation. Therefore to dissect the underlying epigenetic mechanisms regulating TAL1 gene activity in normal and in malignant hematopoiesis this study was undertaken.

Here, we report that TAL1 expression is regulated by differential intra- and inter-chromosomal chromatin loops in normal and leukemia cells, respectively. These loops determine which cell-type specific enhancers interact with the TAL1 promoter. The TAL1 +51 enhancer which is specifically active in erythroid precursors and inactive in leukemic T-cells, interacts with the upstream TAL1 promoter 1 via an activating chromatin loop. hSET1 mediated H3K4 methylation facilitates this erythroid-specific long-range chromatin interaction, and regulates RNA polII recruitment at the TAL1 locus. hSET1 is further required for hematopoietic stem cells (CD34+) capacity to generate burst forming and colony forming units. Furthermore, we find that insulator protein CTCF differentially reorganizes the TAL1 locus chromatin structure keeping the
+51 enhancer in a close proximity to the TAL1 promoter in erythroid cells while blocking the same enhancer/promoter interaction in T-ALL.

In addition, we identify the role of LIM domain binding protein 1 (Ldb1) in TAL1 gene activation by regulating TAL1 +51 enhancer and promoter 1 interaction in erythroid cells. Altogether, this study provides a mechanistic insight into the epigenetic mechanisms that function to regulate long-range chromatin interactions at the TAL1 locus, to modulate TAL1 gene activity in normal and malignant hematopoiesis.
CHAPTER 1
INTRODUCTION

T-cell Acute Lymphoblastic Leukemia (T-ALL)

The term leukemia was originally coined from the ancient Greek words ‘leukos’ and ‘haima’, meaning “white blood”, resembling overabundance of immature WBCs in the patients’ blood during diagnosis. These immature leukemic cells accumulate rapidly in the bone marrow and are carried by the bloodstream to other tissues and organs including brain, liver, lymph nodes and testes, where they continue to grow and divide. Broadly, leukemia’s are classified into acute and chronic based on clinical manifestation and disease progression. Acute lymphoblastic leukemia (ALL) encompasses a group of lymphoid neoplasms that belong to either B or T cell lineage precursor cells. These neoplasms may predominantly present with extensive involvement of the bone marrow and peripheral blood, termed as lymphoblastic leukemia; or may be limited to tissue infiltration, being absent or with very limited bone marrow involvement (less than 25%), designated as lymphoblastic lymphoma. ALL can occur at any age; however, most incidences occur frequently in people under the age of 15 or over the age of 45. In children, ALLs represent 75% of all acute leukemia’s (which in turn represent 34% of all cancers in this age group), with a peak incidence at 2 to 5 years of age [1]. A variety of genetic and environmental factors have been related to ALL. Patients with Down syndrome, Bloom syndrome, neurofibromatosis type I, and ataxia-telangiectasia, show higher incidence. Furthermore, in utero exposure to ionizing radiations, pesticides and organic solvents, has also been linked to increased risk of childhood leukemia [2].

T-cell acute lymphoblastic leukemia (T-ALL), a less frequent subset of ALL, presents as a malignant disease of thymocytes and accounts for about 10-15% cases of
pediatric ALL and 20% of adult ALL. It is classified as high risk leukemia since about 30% of T-ALL patients relapse within the first two years following diagnosis [3, 4].

**Clinical Presentation**

The clinical onset of T-ALL is most often acute. Patients with T-ALL display mediastinal mass with or without pleural effusions, leading to respiratory distress. The most common symptoms of ALL include fever, fatigue and lethargy, bone and joint pain and bleeding diathesis. Laboratory analysis indicates anemia, neutropenia and leucopenia or leukocytosis. In contrast, a build of immature dysfunctional T-cells is observed that interfere with normal blood formation. Other common abnormalities include high serum uric acid and lactose dehydrogenase levels [5].

**Molecular Abnormalities**

One of the most distinguishable characteristic of T-ALL is the presence of robust chromosomal rearrangements. The most common chromosomal abnormalities include rearrangements affecting the T-cell receptor (TCR) genes, most commonly TCRα (14q11.2) and TCRβ (7q35). These TCR genes are frequently translocated to basic helix-loop-helix (bHLH) genes (MYC, TAL1, TAL2, LYL1, bHLHB1), cysteine-rich (LIM-domain) genes (LMO1, LMO2) or homeodomain genes (HOX11/TLX1, HOX11L2/TLX3, members of the HOXA cluster) [6-9]. T-ALL may harbor molecular lesion without any detectable cytogenic abnormalities. Upto 50% of T-ALL contain activating mutations of TAL1/Scl gene, correlated to late cortical stage of thymocyte maturation. About 30% of T-ALL overexpress Hox11 in absence of genetic abnormalities, related to early cortical thymocyte phenotype. Lyl1 mutations are present in 22% pediatric T-ALL, and they correlate with double-negative stage of differentiation. About 4-8% of T-ALL show MLL mutations, associated with maturation arrest at early thymocyte stage and have no
impact on prognosis [10-12]. Activating Notch mutations are present in 50% of T-ALL and inactivating mutation in CDKN2(INK4), a tumor suppressor gene are present in 80% cases on T-ALL [13].

**Prognosis of T-ALL**

The prognosis of ALL has improved dramatically over the past several decades as a result of adaptive therapy, improvements in supportive care, and optimization of the existing chemotherapy drugs. The outcome of pediatric ALL has evolved from an overall survival of less than 10% in the 1960s to approximately 75% to 80% at present [14]. However, adult patients have a less optimistic outlook. The remission rates have reached 85% to 90%, with overall survival rates of only 40% to 50% [15]. *TAL1* positive cases appear to have inferior prognosis, while Hox11 positive patients have superior prognosis. MLL positive patients have no impact on prognosis [11,12].

**Current Treatment Therapies for T-ALL**

The treatment of ALL involves short-term intensive chemotherapy (with high-dose methotrexate, cytarabine, cyclophosphamide, dexamethasone or prednisone, vincristine, L-asparaginase, and/or an anthracyclin).[14,16]. This is followed by intensification or consolidation therapy to eliminate residual leukemia, prevent or eradicate CNS leukemia, and ensure continuation of remission. Radiation may be used for patients showing evidence of CNS or testicular leukemia, although this approach is controversial at the current time, especially in children [14]. In adult patients, the use of growth factors such as granulocyte colony-stimulating factor that accelerate hematopoietic recovery has greatly improved the success rate of ALL therapy [17]. However, the overall frequency of relapse in ALL still remains a major issue. Approximately 25% children and 50% adults relapse, a rate that is highly dependent on
the immunophenotypic and genetic subtype of the leukemia [14, 16, and 17]. This has stir tremendous research interest in development of better therapies and diagnosis techniques. One such area focusses on understanding the molecular mechanisms underlying ectopic activation of various oncogenes.

**TAL1 (T-cell Acute Lymphoblastic Leukemia 1): A Master Regulator of Hematopoiesis**

T-cell acute lymphoblastic leukemia 1 (TAL1) (also known as SCL or TCL5; hereafter referred as TAL1) is a member of the basic helix-loop-helix (bHLH) family of transcription factors and plays an essential role in the development of all hematopoietic lineages. As described previously, it was cloned from a T-ALL patient with chr. (1;14) translocation [66]. It binds to DNA as a heterodimer with the product of ubiquitously expressed bHLH *E2A* or *HEB* genes, by recognizing a hexanucleotide sequence CANNTG termed E box [18]. *TAL1* is detected as early as day 8.5 in developing mice embryo and by day E9.5, it is observed in erythroid progenitors, mid brain and endothelium. It is later required for differentiation along erythroid and megakaryocyte lineages but down-regulated in B and T-cell lineages.

**TAL1 Role in Hematopoiesis**

*TAL1* is essential for specification of hematopoietic stem cells during embryonic development and in subsequent hematopoietic differentiation, continued *TAL1* expression is critical for erythroid maturation as lack of *TAL1* leads to block in erythropoiesis [19, 20]. Deletion of *TAL1* in mice leads to embryonic lethality between E9-10.5 due to complete loss of hematopoietic cells [21, 22]. Moreover *TAL1*-null embryonic stem (ES) cells are unable to generate both primitive and definitive erythropoietic cells *in vitro* and do not contribute to hematopoiesis *in vivo* in chimeric
mouse [23, 24]. These results demonstrate TAL1 as a master regulator of hematopoiesis, as loss of TAL1, leads to complete absence of erythroid, myeloid, megakaryocyte, mast, and T and B lymphoid cells (Figure 1-3). In normal hematopoiesis, where TAL1’s function has been studied extensively, it can positively as well as negatively modulate transcription of target genes by recruiting cofactors (coactivators or corepressors) proteins [25-27]. In this respect it is classified as a bifunctional regulator. The mechanistic detail of how TAL1 performs these tasks is not clearly understood and is an area of active research. An important and current question in hand is what regulates this master regulator? As TAL1 is expressed and required for erythroid cell development and oncogenic potential of T-cells [28, 29], understanding mechanisms underlining the regulation of the TAL1 gene in normal hematopoiesis and leukemogenesis is of particular interest.

**TAL1 Role in T-ALL**

Normal expression of TAL1 is restricted to the DN1-DN2 subset of immature CD4-/CD8- thymocytes with ectopic expression resulting in leukemic arrest in late cortical thymocytes [30]. Two models have been proposed for TAL1-induced leukemogenesis. In the prevailing model TAL1 acts as a transcriptional repressor by blocking the transcriptional activities of E2A, HEB, and/or E2-2 through its heterodimerization with these E-proteins or TAL1 may mediate its inhibitory effects by interfering with E2A-mediated recruitment of chromatin-remodeling complex which activate transcription [31-34]. It also been shown to associate with several corepressors including HDAC1, HDAC2, mSin3A, Brg1, LSD1, ETO-2, Mtgr1, and Gfi1-b [35]. In human T-ALL TAL1 transcriptional repression may be mediated by TAL1-E2A DNA binding and recruitment of the corepressors LSD1 and/or HP1-α [36]. In the other model
TAL1 induces leukemogenesis through inappropriate gene activation [37]. At least two genes \textit{RALDH2} and \textit{NKX3.1} are transcriptionally activated by TAL1 and GATA-3 dependent recruitment of the TAL1-LMO-Ldb1 complex [38, 39]. As a transcriptional activator TAL1 has been shown to associate with the coactivators p300 and P/CAF [40, 41]. Both of these complexes contain HAT activities. The prevalence of histone-modifying enzymes in TAL1 complexes suggests that one function of TAL1 is to regulate chromatin states of its target genes. Ectopic expression of both TAL1 and LMO1 in mice accelerated the progression to leukemogenesis. In this case thymic expression of the TAL1 and LMO1 oncogenes induced expansion of the ETP/DN1 to DN4 population and lead to T-ALL in ~120 days. The acquisition of a Notch1 gain-of-function mutation was proposed to be the rationale behind this increase in leukemia penetrance. In fact, thymic expression of all three oncogenes Notch1, TAL1 and LMO1 induced T-ALL with high penetrance in 31 days, the time necessary for clonal expansion [42]. These studies suggest that aberrant LMO proteins are key players in abnormal T-cell biology.

\textit{TAL1} Gene Is Regulated By \textit{Cis}-acting Factors in Erythroid Cells

The \textit{TAL1} gene, located on chromosome 1p32, is transcribed from two lineage specific promoters, 1a and 1b (Figure 1-1). Promoter 1a is utilized in erythroid, megakaryocytes and mast cells, while promoter 1b, is used in primitive myeloid and mast cells [43, 44]. These regulatory regions were initially identified by mapping DNase I hypersensitive sites over the mouse \textit{TAL1} locus. Later, transgenic mice studies and \textit{in vitro} reporter assays validated the function of the hypersensitive sites in regulating \textit{TAL1} expression both \textit{in vivo} as well as in cultured erythroid cells. As illustrated in figure 1-1A, \textit{TAL1} has three distinct enhancer elements, which are named based on their distance
from exon 1a; 1) -4 enhancer, which directs \textit{TAL1} expression in embryonic endothelial and hematopoietic cells [45], 2) +18/19 enhancer, in HSCs, in progenitors in fetal and adult liver and in embryonic endothelium [46], and 3) +51 enhancer, in primitive and definitive erythroid cells [47]. Besides these, various other putative regulatory sites have been identified by DNaseI hypersensitive sites mapping, and histone modification patterns [48, 49]. Furthermore, ChIP-seq and ChIP-chip assays have shown that CTCF (CCCTC binding factor) binds to -31, +40 and +57 sites at the \textit{TAL1} locus and may play an important role in the regulation of \textit{TAL1} expression in erythroid progenitor K562 cells and CD4+T-cells [49, 50]. The role of CTCF at the \textit{TAL1} locus in T-ALL cells has not been investigated, and is proposed in the current study. In addition to its regulation by these \textit{cis} elements, \textit{TAL1} primary transcript is subjected to extensive alternative splicing and translated from seven in-frame AUGs. Despite these complexities, there are two major \textit{TAL1} protein products observed, one being full length (~47kDa) and a short protein (24-28kDa) which lacks N terminus transactivation domain as a result of alternative splicing [51]. Both proteins contain DNA binding and HLH domains (figure 1-1B) and heterodimerize with E-proteins. \textit{TAL1} protein is also post translationally modified by phosphorylation and acetylation, which modulate its transcriptional activator and repressor activities [40, 41, 52, and 53].

\textbf{\textit{TAL1} Gene Is Regulated By Trans-acting Factors In Erythroid Cells}

\textit{TAL1} promoter 1a and +51 enhancer elements harbor composite E-box and two GATA motifs, which are direct binding sites for \textit{TAL1} and GATA proteins, respectively. Therefore, \textit{TAL1} may exhibit autoregulation through a positive feedback mechanism [54, 55]. In erythroid cells, \textit{TAL1} is part of a multimeric complex that includes its heterodimer interacting component E2A, as well as LIM-only domain protein LMO2, and
the LIM domain binding protein LDB1. This TAL1/E2A/LMO2/LDB1 complex interacts with GATA1 to form a “pentameric” complex [56], which is shown to be critical for the formation of long range chromatin loop in the β-globin activation [57]. Presence of composite E-box GATA motif and the assembly of the pentameric complex at both TAL1 promoter 1a and +51 enhancer elements strongly suggests role of long range chromatin interaction mechanisms in regulating TAL1 expression [49]. This idea is further strengthened by the presence of LDB1, a potential mediator of long range chromatin interactions [58]. Role of histone modifications and chromatin structure in regulating gene expression has been highlighted in various studies [59]. Histone modification patterns dictate transcriptional status of a given gene. Histone acetylation and methylation of H3K4 residues are correlated with transcriptionally active chromatin, whereas methylation of H3K9 and H3K27 residues has repressive role [60-63]. Recent studies have shown a correlation between different methylation status of H3K4 residue and regulatory elements including promoters and enhancers. H3K4 mono and dimethyl marks are enriched at enhancers, while H3K4 trimethyl is present at active gene promoters [64, 65]. This correlation can be exploited to identify regulatory elements which are active in T-cell leukemia. Furthermore, the analysis of recruitment of histone modifying enzymes at the TAL1 locus, will aid in understanding the molecular mechanism of TAL1 gene activation.

**TAL1 Gene Regulation In T-ALL**

TAL1 was first identified in T-ALL patients bearing t(1:14)(p32:q11) translocation (figure 1-2). Additional studies have shown TAL1 to be ectopically expressed in about 60% of T-ALL cases. This aberrant expression is a result of three major events; 1) 3% cases represent t(1:14)(p32:q11) translocation placing TAL1 under the control of TCRδ-
oncogene [66], 2) 30% cases show a submicroscopic 90 kb deletion, wherein the first exon of SCL interrupting locus (SIL) gene is placed upstream of the 5’noncoding part of TAL1 transcription unit [67], and 3) About 60% of T-ALL cell lines and patient samples show ectopic TAL1 expression with no detectable TAL1 gene rearrangements [68].

The question remains as to how and what mediates TAL1’s ectopic expression in T-ALL, especially with patient population that lacks TAL1 gene rearrangements. A third TAL1 promoter (promoter IV) located upstream of exon 4 (figure 1-1A), was described to be specifically used in T-cell leukemia. It directs TAL1 expression in almost all T-ALL blasts with TAL1 gene rearrangement, however in T-ALL samples lacking TAL1 genomic rearrangement, transcriptional initiation involves both promoter IV and promoter 1b, with the exception of one case, where promoter 1a was used instead [69, 70]. Apart from TAL1 activation due to its translocation to the TCR locus, 25% T-ALL patients exhibit SIL-TAL1 gene fusion, wherein 5’ noncoding region of TAL1 is fused to the SIL gene placing TAL1 under the regulation of SIL promoter/enhancer elements. SIL (also known as STIL), is a cytoplasmic protein proposed to play important role in sonic-hedgehog signaling and in the regulation of mitotic checkpoint proteins [71, 72]. It is expressed in all hematopoietic cells suggesting that the SIL promoter is in an open chromatin configuration, accessible to the transcriptional machinery in T-cells. Although the SIL promoter was suggested to control TAL1 transcription in the subset of T-ALL leukemia featuring an intact SIL-TAL1 locus chromosome structure, the prediction is largely based on the fact that the SIL promoter has the ability to activate TAL1 in the cases of interstitial deletion patients or when artificially linked to the TAL1 coding region [73, 74]. The functional link between the SIL promoter and TAL1 coding region, located
90 Kb apart, in leukemia remains unknown. While it is possible that an enhancer/promoter could mediate long range activation, it remains puzzling why the SIL promoter region only affects the \textit{TAL1} gene in T-ALL cells but not in normal erythroid cells. Understanding the mechanism of \textit{TAL1} aberrant activation in T-ALL will provide not only novel insight into TAL1 induced leukemia but also potentially provide strategies for treating the leukemia.
Figure 1-1. Human *TAL1* locus and protein isoforms. A) Cis-regulatory elements at the human *TAL1* locus, and B) Primary structure of TAL1 long and short isoforms with various functional domains.
Figure 1-2. Distribution of genomic rearrangements involving \textit{TAL1} in T-cell acute lymphoblastic leukemia. Based on published data by Bash et al., 1990; Brown et al., 1990; Ferrando and Look, 2000; Ferrando et al., 2002.
Figure 1-3. TAL1 is a key regulator of hematopoiesis. Shown is the brief outline of normal hematopoiesis. TAL1 is initially required for specification of hematopoietic stem cells (HSCs) from hemangioblasts, later for differentiation of hematopoietic progenitor cell (HPC) into common myeloid progenitors (CMP) and subsequently for maturation of erythrocytes, megakaryocytes and macrophages. Various abbreviations used are; CFU-S (colony forming unit-spleen), CFU-GEMM (colony forming unit-granulocyte erythrocyte megakaryocyte and macrophage), and CFU-GM (colony forming unit-granulocyte macrophages).
CHAPTER 2
MATERIALS AND METHODS

Cell Culture and ShRNA-Mediated Knock-Down

K562 cells were maintained in IMDM supplemented with 10% FBS (Atlanta Biologics, S11550H) at a cell density of 2 X 10^5 cells/ml. HL-60 were cultured in IMDM supplemented with 20% FBS and T-ALL cell lines were maintained in RPMI 1640 (FisherSci, MT-10-040) supplemented with 10% FBS. Constructs used for SET1 knockdown were generated by subcloning shRNA oligonucleotides into pSuper.retro.puro vector following manufacturer’s instructions (Oligoengine). Retrovirus generation and infections were performed as described previously [75]. For CTCF knockdown experiments, shCTCF lentiviral constructs were purchased from open biosystems and lentiviral generations and mammalian cell transduction was performed as described in manufacturer’s manual (Open Biosystems). All stable knockdown cells were maintained in media containing 1µg/ml puromycin (Calbiochem, 540222).

Primary Human Cord Blood Cell Culture and Erythroid Differentiation

Human cord blood derived CD34+ cells were purchased from Stem Cell Technologies and were expanded in StemSpan SFEM media (Stem cell technologies, catalog #09650) supplemented with 100ng/ml SCF (Biolegend, 579706), 20ng/ml IL-3 (Biolegend), and 20ng/ml IL-6 (Biolegend, 575706) alongwith 1% penicillin/streptomycin (FisherSci, MT-30-002). Before inducing erythroid differentiation, the cells were sorted for CD34+ using anti-CD34+-PE conjugated antibody (BD biosciences, 555822) and about one million cells were subjected to erythroid differentiation to CD36+ cells over a seven day period (refer figure 2-1). On sixth and seventh day, cells were sorted for CD36+ using anti-CD36+-FITC conjugated antibody (BD biosciences, 555454) and
used for 3C and ChIP analysis. Further, differentiation was traced by measuring β-globin and Tal1 transcript levels (figure 2-1B) and Giemsa/May-Grunwald staining of the differentiating cells (figure 2-1C).

**Giemsa/May-Grunwald Staining Of Erythroid Cells**

Differentiation of CD34+ cells was assessed by Giemsa/May-Grunwald (Sigma, 48900 and 63590) staining based on the protocol described by the manufacturer (Sigma). Briefly, 10^5 cells were pelleted on microscopic slides using cytopsin, followed by staining with May-Grunwald stain for 5 minutes at room temperature (RT). The slides were then washed with 50mMoles/L Tris pH 7.0 and stained with Giemsa (1:20 diluted) for 17 minutes at RT. Finally the slides were washed several times with dH₂O and observed under microscope and representative pictures were captured.

**FACs Analysis of (Cluster of Differentiation) CD34+, CD36+ and (Glycophorin A) GPA Surface Markers**

About >100,000 cells were washed with 0.2% FBS in 1xPBS (FACs) buffer twice. The cells were resuspended in 2mls of FACs buffer + 150μl of anti-CD34+-PE and/or anti-CD36+-FITC or anti-GPA-PE conjugated antibody (BD biosciences) in dark for 30 minutes with rotation. The cells were then washed twice with FACs buffer, and analyzed for either, CD34+, CD36+ or GPA+ on BD LSRII FACs analyzer.

**Native and Cross-linked Chromatin Immunoprecipitation Assay**

Native ChIP assays for histone modifications were performed in the absence of formaldehyde crosslinking. About 1x10^7 cells were washed with 1x PBS containing 10mM sodium butyrate (Sigma, 303410), and Protease inhibitors Pepstatin (Sigma, P5318), Leupeptin (Sigma, L2884), Aprotinin (Sigma, A1153) and 0.2mM PMSF(Sigma, P7626) at 4°C. The cells were than lysed in lysis buffer containing 10mM Tris HCL pH
7.5, 10mM NaCl, 3mM MgCl₂, 0.4% NP-40, 10mM sodium butyrate, followed by centrifugation at 2500 rpm for 5 minutes at 4°C. The nuclei pellet was resuspended in lysis buffer containing 1mM CaCl₂ (micrococcal nuclease cofactor). About 3x10⁻² U/μl of micrococcal nuclease was added and the chromatin was digested for 10 minutes at 37°C, and stopped by the addition of 10mM EDTA. The digested chromatin was then subjected to sucrose gradient centrifugation. A linear gradient of 5% to 30% sucrose was prepared in 14ml SW40 tubes (Seton, 7030). 1ml of the digested chromatin was layered on top of the gradient and nucleosomes were separated at 30000 rpm for 15 hrs. at 4°C using Beckman Ultracentrifuge. Post centrifugation, 500μl fractions were collected, and 50μl of each fraction was analyzed for the nucleosome size determination. Fractions containing mono and di nucleosomes were pooled for ChIP analysis using antibodies specific to various histone modifications: anti-H3K9/14Ac (06-599), anti-H3K4me1 (ab8895, Abcam), anti-H3K4me2 (07-030), anti-H3K4me3 (04-745), anti-H3K9me2 (07-212) and H3K27me3 (07-449) from Upstate Biotechnology. For transcription factors ChIP assay, chromatin was isolated from 1% formaldehyde (FisherSci, BP531) and EGS (Sigma, E3257) crosslinked cells, sonicated using Bioruptor TM UCD-200 (Diagenode) and subjected to IP with anti-TAL1 (sc12984, Santa Cruz Biotech), anti-GATA1 (N6, Bethyl laboratories), anti-SET1 (A300-289A, Bethyl Laboratories), anti-ASH2L (A300-107A, Bethyl Laboratories), anti-RNA polII (ab5408, Abcam), anti-CTCF (07-729, Upstate Biotechnology), anti-Rad21(ab229, Abcam), anti-TAF3 (A302-359A, Bethyl Laboratory), anti-Ldb1 (sc11198, Santa Cruz Biotechnology). Enrichment of these factors and histone modification was assessed by quantitative PCR analysis using primers (Table 1) designed across the TAL1 locus. Fold
enrichment was calculated relative to Input DNA using the formula; fold enrichment = 
\[ 2^{(Input/IP)} \]. In case of high IgG background, the value obtained was further normalized to IgG and plotted as relative fold enrichment.

**Chromosome Conformation Capture Assay (3C)**

The chromatin conformation capture (3C) assay was performed, with minor modifications [75]. In detail, 2 x 10^7 cells were cross-linked with 2% formaldehyde for 10 minutes and stopped by the addition of glycine at a final concentration of 0.125M. Cells were pelleted and washed twice with cold PBS and lysed in lysis buffer (10mM Tris, pH 8.0, 10mM NaCl, 0.2% Nonidet P-40, and Protease inhibitors Pepstatin (Sigma, P5318), Leupeptin (Sigma, L2884), Aprotinin (Sigma, A1153) and 0.2mM PMSF(Sigma, P7626) at 4°C for 90 minutes with gentle rotation. Nuclei were collected and washed with appropriate 1X restriction buffer (for BamHI: NEB buffer 3, NlaIII: NEB buffer 4, DpnII: NEB DpnII buffer) and then resuspended in restriction enzyme buffer containing 0.3% SDS at 37°C for 1 hour with shaking. TX-100 was then added to a final concentration of 1.8% to sequester SDS at 37°C for 1 hour with shaking. Chromatin was than digested with 800U of either BamHI or NlaIII or DpnII (NEB) at 37°C overnight with shaking. Next day the reaction was stopped by adding SDS to a final concentration of 1.6% at 65°C for 30 minutes. The digested chromatin was than diluted in 1 ml of T4 DNA ligation buffer (NEB, B0202) containing 1% TX-100 (Promega, H5141) and incubated at 37°C for 1 hour with shaking. 400 U of T4 DNA ligase (NEB, M0202) were added and the ligation was carried out at 16°C for 5 hours followed by 30 minutes at room temperature. The ligated chromatin was then reverse crosslinked overnight by adding 200μg of Proteinase K (Invitrogen, AM2548) at 65°C. Next day another 200μg of proteinase K were added and incubated at 42°C for 2 hours, followed by phenol chloroform extraction.
to purify the 3C DNA. Purified 3C ligated DNA was then PCR amplified using bait and test primers using PicoMaxx (Stratagene, 600650). The PCR products were cloned into pCR-TOPOII vector (Invitrogen, K4575) and sent for sanger sequencing. To control for primer efficiency, BAC template containing 182041 bp of TAL1 locus was generated and used as a standard. PCR reactions were performed in parallel for BAC standard and 3C ligated DNA. Furthermore, to control for differences in cell-type specific chromatin organization, the interactions at the TAL1 locus were normalized to ERCC3 locus. Relative crosslinking frequencies were calculated and plotted after normalization to loading control and ERCC3 control as described in [77].

**Calculation of the Relative Cross-linking Frequency Using Image J**

The relative cross-linking frequency between two loci was calculated by the ratio of cross-linked sample template versus that obtained from control templates (BAC), followed by normalization to ERCC3 control locus. The following equation was used to calculate the relative cross-linking frequency: 

\[
X(Fr1 + Fr2)= \frac{[A(Fr1+Fr2)/A(Ctrl1+Ctrl2)]_{sample}}{[A(Fr1+Fr2)/A(Ctrl1+Ctrl2)]_{control \ template}},
\]

where

- \(X\): relative cross-linking frequency between two given fragments \(Fr1\) and \(Fr2\);
- \(A(Fr1+Fr2)\): cross-linking frequency between \(Fr1\) and \(Fr2\) in the TAL1 locus;
- \(A(Ctrl1+Ctrl2)\): cross-linking frequency between two control fragments \(Ctrl1\) and \(Ctrl2\) in the ERCC3 locus [75].

**Reverse Transcription and Quantitative Polymerase Chain Reaction**

Total RNA was prepared by using the RNeasy mini isolation kit (Qiagen, 74106) according to the manufacturer’s instruction (Qiagen, MD, USA). 1 μg RNA was reverse transcribed by using the Superscript II reverse Transcriptase (Invitrogen, 18064). cDNA
was analyzed by real-time PCR (qRT-PCR) using a MyiQ Single-Color real time PCR Detection System (Bio-Rad). Primer sequences are listed in the Table 1.

**CD34+ Colony Formation Assay**

Human colony formation assay was performed as described by StemCell technology using complete Methocult™ (Stem Cell Technology, 04435). About 9.5x10^3 CD34 vector control or shSET1 cells were seeded in 4ml SFEM (Stem Cell Technology) media supplemented with 10ng/ml G-CSF, 20ng/ml SCF, 10ng/ml IL3, 10ng/ml IL6 and 6U/ml EPO (a gift from Dr. Constance Noguchi). These were then plated in triplicates on methylcellulose cell culture plates and cultured for 18 days, with fresh media supplemented with cytokines added every week. Different hematopoietic colonies; erythroid progenitors (colony-forming unit (CFU)-erythroid (CFU-E) and burst-forming unit erythroid (BFU-E)); granulocyte/macrophage progenitors (CFU-granulocyte, macrophage (CFU-GM); CFU granulocyte (CFU-G) and CFU-macrophage (CFU-M)) and multi-potential progenitors (CFU-granulocyte, erythroid, macrophage, megakaryocyte (CFU-GEMM)), were observed and counted following guidelines as described by Stem Cell Technologies. The experiment was performed with three replicates for each.

**Luciferase Reporter Assay**

*TAL1* +51 enhancer fragments corresponding to nucleotides 17618015 to 17619135 (+51 enhancer) or nucleotides 17615818-17620036 (+51 enhancer and +53 CTCF sites) based on GRCh37.p5 build of Genbank NT_032977.9 was subcloned into XhoI and MluI sites of the pGL3-SV40promoter vector. Human +51 enhancer sequence corresponds to 0.7Kb core sequence of mouse *TAL1* +40 enhancer, previously identified [47]. About 1x10^5 K562, Jurkat, Rex, HL-60, MOLT4 and HPB-ALL cells were
transfected with 0.5µg of pGL3-SV40p empty, pGL3-SV40p-Enh+51 core or pGL3-SV40p-Enh+51CTCF+53 along with 50ng of pRL-CMV vectors using Lipofactamine 2000 (Invitrogen, 11668). Luciferase activity was measured 48 hrs post transfection using Dual-Glo Luciferase Assay System according to manufacturer’s protocol (Promega, E2920). Relative luciferase activity was calculated by normalizing pGL3-SV40-Enh+51 and pGL3-SV40-Enh+51CTCF+53 to pGL3-SV40 empty vector control.

**Soft Agar Colony Formation Assay**

SETD1A-shRNA and control clones of K562 cells were seeded in soft agar for clonogenicity as described [78]. Briefly, 2.5 ml of bottom layer consisting of RPMI 1640 medium supplemented with 10% FBS plus 1% P/S and 0.65% agarose was spreaded in a 60-mm tissue culture dish. Then 5 ml of cell layer containing 1×10^4 cells with 10% FBS plus 1% P/S and 0.35% agarose was overlaid on the top. These 60-mm dishes were then placed in a bigger 10cm dish containing 1X DPBS and incubated at 37°C. About two weeks later, the colonies were counted and photographed. (BFU-E)); granulocyte/macrophage progenitors (CFU-granulocyte, macrophage (CFU-GM); CFU granulocyte (CFU-G) and CFU-macrophage (CFU-M)) and multi-potential progenitors (CFU-granulocyte, erythroid, macrophage, megakaryocyte (CFU-GEMM)), were observed and counted following guidelines as described by Stem Cell Technologies.
Figure 2-1. Ex-vivo differentiation of human cord blood CD34+ cells. Human cord blood CD34+ cells were purchased from Stem Cell Technologies and differentiated ex-vivo into erythroid precursors. (A) Ex-vivo CD34+ differentiation protocol schematic. At every time point, the cells were subjected to FACs analysis for CD34+, CD36+ and later for GPA+, the representative FACs data for Day 6 is shown. B) Assessment of TAL1 and β-globin mRNA levels during CD34+ differentiation by RT-quantitative PCR assay. C) May Grunwald-Giemsa staining of the CD34+ cells at Day 3 and 6 of erythroid differentiation.
<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer sequences (5’- 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>shRNA</strong></td>
<td></td>
</tr>
<tr>
<td>Hs SET1-1</td>
<td>Target: GAAAGAGGCTCCATCGGAAAT</td>
</tr>
<tr>
<td>Hs SET1-2</td>
<td>Target: GACAACACGATGAAATA</td>
</tr>
<tr>
<td>Hs SET1-3</td>
<td>Target: CAACGACTCAAAGTATAA</td>
</tr>
<tr>
<td>Hs c-Maf-1</td>
<td>Target: TGGTTCTCCATGACTGCAAAT</td>
</tr>
<tr>
<td>Hs c-Maf-2</td>
<td>Target: AACTTCTCGTATTTCTCCTTG</td>
</tr>
<tr>
<td>Hs c-Maf-3</td>
<td>Target: ACTTCTCGTATTTCTCCTTG</td>
</tr>
<tr>
<td>Hs c-Maf-4</td>
<td>Target: TCCAGTAGTAGCTTCCAGGT</td>
</tr>
<tr>
<td>Hs c-Maf-5</td>
<td>Target: TCATCCAGTAGTAGCTTCCCA</td>
</tr>
<tr>
<td><strong>RT-PCR</strong></td>
<td></td>
</tr>
<tr>
<td>Mm β-actin</td>
<td>For: GTGGGCCTGCTCAGGACCA</td>
</tr>
<tr>
<td></td>
<td>Rev: TGGCCCTTACGGTGCAGGG</td>
</tr>
<tr>
<td>Hs β-actin</td>
<td>For: AGAAAATCTGGCACACAC</td>
</tr>
<tr>
<td></td>
<td>Rev: AGAGGCTGACAGGGATAGCA</td>
</tr>
<tr>
<td>HoxB4 (both for Hs &amp; Mm)</td>
<td>For: TGGATCGCAGAAAGTTTCC</td>
</tr>
<tr>
<td></td>
<td>Rev: GGTTTTTTTCACCTACTGC</td>
</tr>
<tr>
<td>Hs Meis1</td>
<td>For: TCACACTGGCCTTAAGAGGA</td>
</tr>
<tr>
<td></td>
<td>Rev: CCGTAAATGGGAGTACCTG</td>
</tr>
<tr>
<td>Mm SET1</td>
<td>For: CAGACGGGCTTTTGAAGTAGTTCC</td>
</tr>
<tr>
<td></td>
<td>Rev: GAGGCTGTGGTACACTAGG</td>
</tr>
<tr>
<td>Mm TAL1</td>
<td>For: TAGCGTACGGACACGTCG</td>
</tr>
<tr>
<td></td>
<td>Rev: GCGAGGATCTCATAGCTTC</td>
</tr>
<tr>
<td>Hs TAL1</td>
<td>For: GGTGCTGACAGGAGAATT</td>
</tr>
<tr>
<td></td>
<td>Rev: AAGATACGCGCAACACTTC</td>
</tr>
<tr>
<td>Hs GATA1</td>
<td>For: GACACCTGCGCAGGAGAAAT</td>
</tr>
<tr>
<td></td>
<td>Rev: CGAGTCTGAACTCCATCTCTTC</td>
</tr>
<tr>
<td>Hs Ldb1</td>
<td>For: GCATTCAAGGACACCATTT</td>
</tr>
<tr>
<td></td>
<td>Rev: TCCGCATCATGTCGTC</td>
</tr>
<tr>
<td>Hs p4.2</td>
<td>For: CTCGGGCCTGGATCCACGT</td>
</tr>
<tr>
<td></td>
<td>Rev: ACTGACTCCACACACTCAC</td>
</tr>
<tr>
<td>Hs GPA</td>
<td>For: ATTTGACGCAATGTGAGCATCA</td>
</tr>
<tr>
<td></td>
<td>Rev: TAGATCCTGTCAGGATT</td>
</tr>
<tr>
<td>Hs NKX3.1</td>
<td>For: CCTCCCTGGCTTCGCTGTA</td>
</tr>
<tr>
<td></td>
<td>Rev: TGTACCTGAGCGCTGATA</td>
</tr>
<tr>
<td>Hs c-Myb</td>
<td>For: ACCATTTGAGAGACAGAGAG</td>
</tr>
<tr>
<td></td>
<td>Rev: TGAGGTGAGGAGTTTCTGG</td>
</tr>
<tr>
<td>Hs c-Myc</td>
<td>For: AGCTCGCTAGACGGTGG</td>
</tr>
<tr>
<td></td>
<td>Rev: GAGGTACAGTGCTGAGT</td>
</tr>
<tr>
<td>Hs MAP17</td>
<td>For: CGCACACATGTACCTGAC</td>
</tr>
<tr>
<td></td>
<td>Rev: TCTCATGCTCAGGACCTG</td>
</tr>
<tr>
<td>Hs STIL</td>
<td>For: GAAGCTTGCTCCTGATGG</td>
</tr>
<tr>
<td></td>
<td>Rev: TCAGTTCCACAAGGATGGA</td>
</tr>
<tr>
<td>Hs β-actin</td>
<td>For: AGAAAATCTGGCACACAC</td>
</tr>
<tr>
<td></td>
<td>Rev: AGAGGCTGACAGGGATAGCA</td>
</tr>
<tr>
<td>Primer</td>
<td>Primer Sequences (5’-3’)</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Hs Meis1</td>
<td>TCACACTGGCCTTAAGAGGA</td>
</tr>
<tr>
<td></td>
<td>CCGTAATGGTGATAGTCGT</td>
</tr>
<tr>
<td>Hs γ-globin</td>
<td>GGCAACCTGCTCCTGCCTC</td>
</tr>
<tr>
<td></td>
<td>GAAATGGATTGCCAAAAACGG</td>
</tr>
<tr>
<td>ChIP</td>
<td></td>
</tr>
<tr>
<td>HoxB4 (both for Hs &amp; Mm)</td>
<td>For: GCCTCTAACTTTGTTCACTTGAC</td>
</tr>
<tr>
<td></td>
<td>For: AGCCATTATTTTGGGAAAGT</td>
</tr>
<tr>
<td>Hs HoxB4 3’UTR</td>
<td>For: CCCCCGAAAAATCTATCTGC</td>
</tr>
<tr>
<td></td>
<td>Rev: CCAAAGCTGAAAACGAGGA</td>
</tr>
<tr>
<td>Hs TAL1 locus primers</td>
<td></td>
</tr>
<tr>
<td>STIL</td>
<td>For: CCGCAGTTTTCCTCAAGAGACTT</td>
</tr>
<tr>
<td></td>
<td>For: GGTGCCTGTTACGTTATGGT</td>
</tr>
<tr>
<td>Region - 16Kb</td>
<td>For: CAGTAGCAAGCCAAAGTGTAACAA</td>
</tr>
<tr>
<td>CTCF site - 31Kb</td>
<td>For: ACGTAGCTGCTGCCTGCCTG</td>
</tr>
<tr>
<td>Enhancer - 4Kb</td>
<td>For: TGGCTGCCTGACTGCAGAA</td>
</tr>
<tr>
<td>Region - 10Kb</td>
<td>For: TTTCTTTTTCTCCTCAATCTC</td>
</tr>
<tr>
<td>Upstream1a region</td>
<td>For: AGGGAGACTGCCATTGAAAT</td>
</tr>
<tr>
<td>CpG island</td>
<td>For: CCTCCCAGGGCTTTTTCTT</td>
</tr>
<tr>
<td>TAL1</td>
<td>For: CTTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>Prom1a</td>
<td>For: TGGGCCATCCCCGCACTTTAA</td>
</tr>
<tr>
<td>TAL1</td>
<td>For: CCGCCTCGAGACTCTCTTT</td>
</tr>
<tr>
<td>Prom1b</td>
<td>For: CACAGCTCGCGCATTTT</td>
</tr>
<tr>
<td>TAL1</td>
<td>For: CACAGCTCGCGCATTTT</td>
</tr>
<tr>
<td>PromIV</td>
<td>For: TGGGCCATCCCCGCACTTTAA</td>
</tr>
<tr>
<td>Enhancer +19</td>
<td>For: GGGTAGCAGCTGCTGCTGCT</td>
</tr>
<tr>
<td>CTCF site +40Kb</td>
<td>For: GGGTAGCAGCTGCTGCTGCT</td>
</tr>
<tr>
<td>Map17prom</td>
<td>For: GCCGAAAGGCACTGAA</td>
</tr>
<tr>
<td>Upstream</td>
<td>For: GCCCAGCAAGCTAAACT</td>
</tr>
<tr>
<td>Map17prom</td>
<td>For: CCGGAAAGGCACTGAA</td>
</tr>
<tr>
<td>Down region</td>
<td>For: GCTGGCATCCGAGGCTATCT</td>
</tr>
</tbody>
</table>

37
Table 2-1. Continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer Sequences (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhancer</td>
<td>For: GCCTCCTAGCTTGAGATGTC</td>
</tr>
<tr>
<td>+51Kb</td>
<td>Rev: CAGAAGTGAGACCAATGAGATCGT</td>
</tr>
<tr>
<td>CTCF site</td>
<td>For: TGGCAGTCTTCTAGTTTCG</td>
</tr>
<tr>
<td>+53Kb</td>
<td>Rev: CTCTCATCAATCTACGCTTTC</td>
</tr>
<tr>
<td>CTCF site</td>
<td>For: GTGAATTGACAGCCATCGT</td>
</tr>
<tr>
<td>+57Kb</td>
<td>Rev: GGAGGAGGAGGTGGATGTC</td>
</tr>
<tr>
<td>Region</td>
<td>For: GTG GCCCAAAGCAAGGAAT</td>
</tr>
<tr>
<td>+70Kb</td>
<td>Rev: TCTCTGGAATCTCCAAGGCAA</td>
</tr>
<tr>
<td>Nkx3.1</td>
<td>For: TTTAGCCAATGGGTGACA</td>
</tr>
<tr>
<td>promoter</td>
<td>Rev: CAGACACATGACATACCA</td>
</tr>
<tr>
<td>Nkx3.1</td>
<td>For: TTCTGTAGGTAATCTCTGAGTT</td>
</tr>
<tr>
<td>3’UTR</td>
<td>Rev: AGTGTGACAGCCAGCAAGGT</td>
</tr>
<tr>
<td>hHS2-</td>
<td>For: AAGGGGATAGAGGAGCTGAG</td>
</tr>
</tbody>
</table>

3C
A) BamHI

| Prom1aCE | Rev: CTGCTCTAGCGGTGGATTTCTAGGAGG |
| MAP17prom | For: GTTGATCACCTGGTGTCCTCCTGTTCAT |
| CE       | For: ACTCGTAGATCTGCCCTCCTC |
| Enh+51CE  | Rev: CAGCCAGCGCTTCAGT |
| Prom1a   | For: GTGATGCAACTAGTGCAAGCTGCAAGC |
| bait_L    | CAGCCAGCGCTTCAGT |
| Prom1abait | For: GACCTCTGGCTGGTAAATC |
| Enh+51    | Rev: GAGTGACCTGACTCAAG |
| MAP17prom  | For: GCCAAGCAAGGACAGACT | TTAAGGAGG |
| TAL1proIV | For: CGT TTAAACCCAGTAGGTGCT |
|          | Rev: CACGCACACTCTCTTCAGAAG |

B) NlaIII

| Enh+51   | Rev: GAGTGACCTGACTCAAG |
| Prom1a   | For: CACTCCCTCCGGTGAAAATG |
| Map17pro  | Rev: GCCCGCGCAAGCTAAACT |
| TAL1pro1a | For: CTAGCAGCGCTCAAC |
|          | Rev: TGGGCCAAATGATTACCTTT |
| TAL1proIV | For: CGTTTTAAACCCAGTAGG |
| CTCF-31_3C | For: GGCTGTGGAGGAGTAGTG |

38
### Table 2-1. Continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer Sequences (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>C) DpnII</td>
<td></td>
</tr>
<tr>
<td>CTCF-31_CE</td>
<td>Rev: CTTTCTCGCAAATGGACAAATCA</td>
</tr>
<tr>
<td>CTCF+40_3C</td>
<td>Rev: TGGAACAGGAAGGAGGAACT</td>
</tr>
<tr>
<td>CTCF+53_CE</td>
<td>Rev: CGGAGGAAGTGCTGAACC</td>
</tr>
<tr>
<td>CTCF+57_CE</td>
<td>For: TGGGAGGAACAGAGATGACA</td>
</tr>
<tr>
<td>Reg -10_3C</td>
<td>For: CAAATCAGAAAGAAAGCATCGCA</td>
</tr>
</tbody>
</table>

#### Luciferase Reporter Assay

| +51(0.7) | ATGCACGCGTCTTTATTTATATGCCATTG |
| +51(4.2) | Rev: ATGCCTCGAGAGATCCAGCCTGTA |
CHAPTER 3
SET1 AND CTCF FACILITATE TAL1 PROMOTER/ENHANCER INTERACTION IN ERYTHROID PROGENITORS

Introductory Remarks

DNA, the basic unit of life, is compacted from a linear 2nm fiber to 1400 nm metaphase chromosome to be efficiently package into the minute space of the cell nucleus. However, these several meters of our genome must remain functional and accessible to transcriptional and replication machinery. To achieve such dynamic packaging of the genomic information, DNA is found in complex with an equal mass of histone proteins to form nucleosomes, the basic unit of chromatin. Chromatin is an essential platform for all DNA-templated mechanisms including replication, repair, recombination and transcription. Each nucleosome molecule consists of nucleosome core particle containing 146bp of DNA wrapped around the histone H2A, H2B, H3 and H4 octamer, and the linker histone H1, which interacts with linker DNA between nucleosomes, thus providing further compaction [79]. Histones are lysine and arginine rich proteins synthesized during early S-phase of the cell cycle. They are highly conserved from yeast to humans and contain amino terminal tails, which are unstructured, and protrude away from the nucleosomal core. These tails are unique to each histone molecule and are subjected to various post-translational modifications including acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, and proline isomerization [80]. Histone acetylation neutralizes their positive charge and thereby aids to open the chromatin, however methylation can function to either repress or activate gene transcription. Lysine methylations; H3K4 mono, di and tri methylations are associated with active chromatin, whereas H3K9 di and H3K27 tri methylations are associated with repressive chromatin. Several histone lysine methyl
transferases have been identified which selectively modify particular histone residue to implement downstream effect on gene activity. These proteins encompass a characteristic motif of 130 amino acids called SET domain, which took its name from Drosophila Su(var) 3-9, Enhancer of zeste and Trithorax [81]. SET domain protein methyltransferases catalyze the transfer of methyl groups from the cofactor S-adenosylmethionine (AdoMet) to specific lysine residues of protein substrates, such as the N-terminal tails of histones H3 and H4. So far various SET domain containing proteins have been discovered and have been grouped into eight classes of enzymes, KMT1-8 (where K stands for lysine) [82].

**H3K4 Methyltransferase SET1 (Suvar, Enhancer of Zeste, Trithorax 1)**

The KMT2 class and its first member, Set1, in yeast Saccharomyces cerevisiae were isolated within the Set1/COMPASS functioning as histone H3K4 methylases and are well conserved in mammals. There are multiple H3K4 methyltransferases, including SET1A (here after referred as hSET1), SET1B, and MLL 1-4 in humans, comprising of the core ASH2L, RbBP5, WDR5 and HCF1 proteins. SET1 complexes contain WDR82 whereas; MLL1/2 and TRR complexes contain Menin as a unique component, respectively [82]. There is no data demonstrating the role of SET1 knockout in mammals, however deletion of SET1 in yeast leads to multiple phenotypes, due to complete absence of H3K4 mono/di/trimethylation, reinstating the importance of H3K4 methylation in gene regulation [83, 84]. However, the evidence for direct correlation between SET1, H3K4 me and the phenotype is lacking. SET1 is predominantly associated with the coding regions of highly transcribed RNA pol II genes. In particular, SET1 associates with Ser-5 phosphorylated pol II, through Paf1 complex [85]. Moreover, histone H2B ubiquitination by Rad6/Bre1, is required for H3K4 di and tri
methylation, which in turn is regulated by the Paf1 complex [86]. Thus, much remains to better understand the factors that underlie such robust phenotype observed upon Set1 deletion in yeast model system moreover; the role SET1 in mammals is poorly evaluated.

Recent studies have identified the role of TET proteins in targeting SET1 to the active chromatin regions in the genome [87]. TET2/3 interacts with OGT, which further targets SET1 complex component HCF1 to glyNacylation (addition of GluNAc moiety to the hydroxyl residue of serine or threonine residue), a modification that appears to be important for SET1/COMPASS complex integrity. Tet2/3-OGT further promotes binding of the SET1 to the chromatin, specifically to the CpG island promoters of the key hematopoietic genes including JAK1/2 and RUNX1. Work from our lab has identified USF mediated targeting of SET1 to chicken and mouse beta-globin locus. SET1 copurifies with chromatin remodeling protein NRF in a multiprotein complex of ubiquitously expressed transcription factors USF1 and USF2 in hematopoietic cells. Both SET1 and NRF activities are required for the barrier function of the chicken HS4 and USF1 aids in their recruitment to the chromatin [88]. Furthermore, unpublished data from the lab demonstrates the role of SET1 in regulation of various key hematopoietic genes including HoxB4. We have also identified SET1 to copurify with TAL1 in erthyroid cells (unpublished). Altogether, the evidence suggests a critical role for SET1 in hematopoiesis. Based on these findings, role of SET1 in TAL1 gene regulations was assessed.

**Insulators and CTCF (CCCTC-binding factor)**

Insulators are the guardians of the genome, elements first identified based on their ability to shield genes from outside influences, which might result in inappropriate
activation or repression of the gene. Insulators are divided into two classes on the basis of their activities; enhancer blocking insulators prevent distal enhancers from activating a promoter when placed between the enhancer and promoter, and barrier insulator, blocks the spread of heterochromatin and subsequent gene silencing [89]. CCCTC-binding factor CTCF, an 11 zinc-finger containing transcription factor, was the first identified mammalian enhancer blocking insulator. CTCF is also well recognized for its role in X-chromosome inactivation, gene imprinting, genome organization and promoter activation or repression. It is very intriguing as to how a single protein can regulate a multitude of cellular functions. The answer probably lays in the context specific interactions of the CTCF with diverse protein partners. The unique 11 zinc-fingers are probably utilized in a combinatorial way to recognize and bind to a variety of DNA sequences as well as interact with a plethora of cellular proteins. Furthermore, CTCF is post-transnationally modified, including poly(ADP)ribosylation, sumoylation and phosphorylation, which further influence its function. Moreover, DNA methylation blocks CTCF’s DNA binding ability, thus interfering with its gene regulation activities [90].

Recently, several genome-wide studies utilizing chromosome conformation capture assay (3C) technique have demonstrated an important role for CTCF in genome organization. CTCF’s capacity to oligomerize provides this functional advantage. Initial studies demonstrated its binding at the periphery of the nucleolus as well as to the nuclear matrix, which might occur through the nuclear phosphoprotein, nucleophosmin [91]. Later, CTCF was proposed to be associated with nuclear lamin B in the nucleus. ChIP-seq analysis for CTCF occupancy in various cell types identified extensive overlap among CTCF binding sites across different cellular environment [92,
However there is significant difference in the three dimensional interactions between these sites across cell-types. For example, significant differences were observed in CTCF association with H3K27me3 domain boundaries across cell-types. A study at the human beta-globin locus, for example, demonstrated that CTCF sites were involved in cell-type specific long-range interaction in various human cell lines despite similar occupancy [94]. Furthermore, cohesion proteins called cohesins, which play an essential role in keeping sister chromatids together during cell division, have gained tremendous interest as they colocalize with CTCF and are thought to give CTCF, its cell-type specific roles in genome organization. More advanced techniques modified from the original 3C assay like ChIA-PET and Hi-C, have further started to shed light on the CTCF’s role in governing genome plasticity and function. High-resolution CTCF-associated chromatin interactome map in mouse ES cells using ChIA-PET technique, suggests that CTCF configures the genome into distinct chromatin domains and subnuclear compartments that exhibit different epigenetic states and transcriptional activities. In contrast to its enhancer blocking roles, CTCF appears to promote interaction between promoter and enhancer elements [95]. Moreover, investigation of the three dimensional genome organization using Hi-C technique has identified CTCF to be enriched at the boundaries of the so called topological domains (large megabase-sized local chromatin interaction domains). These domains coincide with regions that constrain spread of heterochromatin and are conserved across various cell types. The boundaries are also enriched for tRNA genes, housekeeping genes, and short interspersed element (SINE) retrotransposons, indicating their role in domain formation [96].
Chromatin Looping and Transcription

Control of gene expression can occur locally and over large genomic distances. Regulatory elements are frequently positioned far upstream or downstream of the genes they control and can even influence the expression of genes that lie on separate chromosomes. A DNA regulatory element, which encompasses binding sites for various transcription factors and can activate promoter driven gene transcription independent of its location, orientation and distance, is termed an Enhancer [97, 98]. Enhancer elements are responsible for allowing spatiotemporal patterns of gene expression during development. In metazoan's enhancer's are often located long distances away from the gene they regulate, sometimes even on separate chromosomes [99]. For example, limb bud enhancer for the mouse Sonic hedgehog (Shh) gene, is more than 1 Mb from the Shh gene promoter [100, 101]; whereas olfactory receptor enhancer H is located on chromosome 14 [102]. How specific communication among distal regulatory elements is achieved has been the subject of discussion and substantial recent experimental evidence favors models for long-range control of gene expression involving chromatin looping [103, 104]. Chromatin looping represents an in vivo interaction between two DNA regulatory elements located far apart, which results into the intervening DNA to obtain a loop structure. The interaction can be mediated by transcription factors, co-activators/co-repressors as well as might involve non-coding RNA, as recently described [105]. Advances in molecular techniques specifically with the newer tools like chromosome conformation capture (3C) assay, the role of chromatin looping in gene regulation has become clearer. This approach has now been adapted to produce genome-wide map of interactions mediated by a given regulatory element (4C) or even associations among multiple sequences located throughout the
genome ("5C" and "Hi-C") [106, 107, 108]. 3C has been utilized to reveal interactions between distal sequences, at multiple gene loci. Within the mammalian β-globin locus, for example, 3C has been utilized to reveal effect of various transcription factor knockdown on gene transcription and looping interactions [109, 110, 75]. Similarly, chromatin looping has been demonstrated in the human "myb" locus, and immunoglobin "Ig" locus [111, 112]. Evidence supporting enhancer-promoter interactions is part of a large and growing body of studies that have suggested nuclear organization as a major determinant of gene expression. First, each chromosome in a nucleus is localized to its own territory, with about 20% of nuclear space dedicated to regions of interactions between neighboring chromosomes [113, 114]. In addition some regions of a chromosome loop out and can be found in chromosome territories "CT". A study of the beta-globin locus, for example found that regions extruded to CT specifically in erythroid cells, and this extrusion occurred prior to activation of high levels of beta-globin gene expression and was LCR dependent [115]. Similarly, the limb bud enhancer is required for extrusion of the Shh gene locus in CT [116]. Second, nuclear periphery in general appears to exert a repressive influence on genes that localize there. For example, tethering of reporter genes near nuclear periphery or nuclear lamina has shown to result in down regulation of the reporter and neighboring genes, although this in not true for all genes [117, 118, 119]. Third, nucleus has various self-organized bodies, including nuclear speckles, PML bodies, nucleoli, cajal bodies and others that might influence the gene expression. For example, splicing speckles represent concentration of factors, which might serve as a basis for interactions among genes that are otherwise located far apart. This case appears to hold true for erythroid-specific genes, that are found
positioned near common splicing speckles in erythoid cells [120]. Finally, genes can colocalize on the basis of shared association with specific factors. This is particularly true for RNA polymerase II. Visualization of RNA polII using antibody specific to RNA polII or labeling primary transcript have suggested that transcription is localized to RNA polII factories called “transcription factories” [121, 122]. In addition to RNA polII, other factors involved in transcription regulation appear to organize into discrete foci in the nucleus and can either directly or indirectly bring distal gene together. These include, for example, AT-rich sequence binding protein, SATB1; which is expressed in thymocytes and many other cell types; functions to bring disparate genomic regions together via long-range interactions [123, 124]. In thymocytes SATB1 is observed in a cage like distribution and binds to promoter-distal regulatory regions in multiple gene loci and loss of SATB1 results in disruption in normal gene expression patterns and locus-wide chromatin structure. Thus, understanding the mechanisms of gene expression at this scale is still an emerging field. Though 3C and its variants provide important information about enhancer function, these are inherently descriptive assays, and thus do not provide ways to distinguish between correlation and causation. Newer approaches are thus required to establish a direct link between long-range interactions and enhancer function.

Given the scope and the indispensable role of chromatin looping in gene regulation and genome organization, investigations assessing its role in oncogene activation and cancer manifestation are essential and of current interest. Based on the previous findings, TAL1-GATA1 complex are recruited to the TAL1 promoter 1 and +51 enhancer in K562 cells. Moreover, ChIP-chip studies have also demonstrated
enrichment of RNA polII at these regions in K562 cells [49], thus we hypothesized that
*TAL1* promoter 1 in erythroid cells is activated by +51 enhancer located 51 kb’s
downstream by chromatin looping mechanism.

**Results**

**Distinct Patterns of Histone Modifications are Associated with *TAL1* Enhancer/Promoter Activities**

Recent genome-wide studies predict a correlation between different types of
histone modifications such as the methylation statuses of Lys 4 residue on histone H3
tails and enhancer/promoter activities [64, 65]. Given that *TAL1* is tightly controlled by
multiple cis-regulatory elements in different stages of hematopoiesis, we examined the
H3K4me2, H3K4me3, H3K9/14ac, and H3K27me3 patterns across the 166 Kb of the
*TAL1* locus in K562, and HL-60 cells by ChIP-qPCR assay using antibodies specific to
these modifications. K562 is human erythroleukemia cell line and HL-60 is derived from
an acute myeloid leukemia patient where the *TAL1* gene is silenced. The pattern of
active and repressive histone modifications associated with gene activity in these cell
types across the entire *STIL1*-*TAL1*-MAP17 locus is shown in Figure 3-1. There are
marked enrichments of H3K4me2 and H3K9/14ac in K562 cells at the *TAL1* promoter 1,
+19, *MAP17* promoter, and the +51 enhancer. H3K4me3 is particularly enriched at the
*TAL1* promoter 1, but not at other regulatory elements (Figure 3-1A). In HL-60 cells
where *TAL1* is inactive, there are large peaks of H3K27me3 over the *TAL1* promoter IV
and promoter 1, marking the silenced *TAL1* gene (Figure 3-1B). Moreover, there is no
H3K27me3 detected in *TAL1* expressing K562 cells. Our results show that previously
defined HSC enhancer +19 and erythroid enhancer +51 are in open chromatin in K562
cells.
The Recruitment of hSET1 Correlates with TAL1 Transcriptional Activation during Hematopoiesis

hSET1 is histone methyltransferase that specifically methylates Lys 4 at histone H3 tails. Given that TAL1 colocalizes with the H3K4 methyltransferase hSET1 complex at the TAL1 target genes during erythroid differentiation [125] and both +51 enhancer and TAL1 promoter 1 contain composite E-box/GATA motifs [49], we reasoned that recruitment of hSET1 complex by TAL1 may be responsible for high levels of H3K4 methylations in the TAL1 locus in erythroid cells. To test this possibility, we examined the global interactions between TAL1 and hSET1 at the human genome comparing human primary CD34+ hematopoietic stem cells (HSCs) and CD36+ erythroid precursors using unbiased ChIP-seq technologies. Approximately 50% of intergenic bound hSET1 co-localized with TAL1 in CD36+ erythroid precursors (figure 3-2C), suggesting that TAL1 recruits the hSET1 complex to regulate its genome-wide targets in erythroid cells. As we expected, TAL1 and hSET1 complex bind to both the +51 enhancer and the promoter 1 in the TAL1 locus in primary hematopoietic cells. This binding correlates strongly with H3K4 methylations at these elements (Figure 3-2A&B). In particular, H3K4me3 was enriched around the transcription start site (TSS) of the TAL1 gene upon differentiation to CD36+ cells. Thus, our data suggest that recruitment of the hSET1 complex facilitates promoter H3K4 methylations and transcriptional activation of TAL1 expression during normal hematopoiesis.

A Long-range Chromatin Loop Mediates Enhancer/Promoter Interaction in the TAL1 Locus in Erythroid Precursors but not in T-ALL Cells

It has been previously reported that +51 enhancer is capable of driving reporter gene expression at physiological TAL1 expression sites during hematopoiesis in transgenic mice [47,126]. However, it remains unclear how the +51 enhancer activates
the TAL1 gene from 51 Kb downstream in its native chromatin location. An attractive model proposes formation/presence of a chromosome loop that brings the enhancer and the promoter in close proximity to allow enhancer mediated promoter activation. To test this possibility, we carried out chromosome conformation capture (3C) assays in CD36+ erythroid progenitor cells using TAL1 promoter 1 as bait (Figure 3-2D). We identify a long-range activating chromatin loop between TAL1 promoter 1 and +51 enhancer (Figure 3-2E). This interaction was confirmed by sequencing of the specific PCR product which revealed a fusion molecule containing sequences from both promoter 1 and +51 enhancer (Figure 3-2F). In contrast, no interactions were detected between TAL1 and MAP17 promoters. Additionally this long-range interaction was identified independently using a different restriction enzyme (data not shown). This data suggests that the +51 enhancer activates the TAL1 promoter 1 via a long-range enhancer/promoter chromatin loop.

Recent studies have highlighted that long-range chromatin interactions provide a topological basis for transcriptional regulation [127, 128]. To test whether the enhancer/promoter loop is specific for erythroid cells or is also present in TAL1 expressing T-ALL cells, 3C analysis was performed in K562, Jurkat, and HL 60 cells. For this assay, restriction enzyme NlaIII was used (Figure 3-3A) which generated on average 250-500 bp fragments across the genome facilitating detection of specific interactions between the +51 enhancer and the TAL1 promoter 1. Consistent with primary CD36+ cells, the +51 enhancer physically interacts with promoter 1 only in K562 cells (figure 3-3B). This interaction was confirmed by sequencing of the specific PCR product (not shown). This long-range interaction was neither detected between the
MAP17 promoter and the TAL1 promoter 1 nor in HL-60 cells lacking TAL1.

Interestingly, although Jurkat express the TAL1 gene, the chromatin interaction between the +51 enhancer and the TAL1 promoter 1 was not detected in these cells (Figure 3-3B) supporting the evidence that the +51 enhancer is inactive in TAL1 expressing T-ALL cells (Figure 4-3C).

**Recruitment of the hSET1 Complex is Essential for Long-range Chromatin Loop and Transcription of the TAL1 Gene**

Evidence suggests that H3K4me3 is important for establishing chromatin loops [50]. The hSET1 complex has been shown to regulate H3K4me3 methylation. Because the binding of the hSET1 complex correlates with TAL1 activation in erythroid cells (Figure 3-1 and 3-2), we further reasoned that recruitment of hSET1 mediate the physical chromatin interaction between +51 enhancer and promoter 1 at the TAL1 locus. To test this hypothesis, we generated shRNA mediated hSET1 knockdown (KD) in K562 cells (Figure 3-3C). SET1 knockdown in K562 affects its colony formation capacity on soft agar assay (Figure 3-3D and E). Furthermore, ablation of hSET1 in CD34+ HSCs resulted in a block in the ability of HSCs to differentiate into CFU-E and BFU-E colonies but not CFU-GEMM and CFU-GM colonies (Figure 3-3F and G), implying an important role for SET1 in erythroid differentiation. ChIP analysis for SET1 in K562 cells, confirmed its recruitment at the TAL1 locus, similar to the data in CD36+ cells (Figure 3-4A). Next we assessed the effect of SET1 depletion on TAL1 levels. hSET1 knockdown led to decrease in TAL1 expression in three individual hSET1 KD K562 clones (Figure 3-4B). In addition to TAL1, SET1 is also required for various other erythroid genes including TAL1 targets, which are repressed as a result of reduced TAL1 expression (Figure 3-4C and D).
Interestingly, we found disruption of the long-range TAL1 promoter 1a and +51 enhancer chromatin loop upon SET1 knockdown in K562 cells (Figure 3-4A&B). A detailed analysis revealed decrease in active histone modifications H3K4me2 and H3K4me3 at the TAL1 promoter and +51 enhancer (Figure 3-5C). In addition the recruitment of RNA PolIII to both the enhancer and the promoter of TAL1 gene is suppressed (Figure 3-5C). Consequently, active histone modifications may be associated with specific chromatin loop formation.

**CTCF Mediated Cell-type Specific Chromatin Loops in the TAL1 Locus Regulate Expression of TAL1 Gene in Erythroid and Leukemic Cells**

ChIP-chip experiments utilizing genomic tiling microarrays covering 256 kb of the human TAL1 locus identified four distinct CTCF binding sites in K562 cells, namely -31, +40, +53 and +57 elements [49,129]. CTCF site -31 is located 5’ of the TAL1 locus within the SIL gene body, while +40 is intergenic between TAL1 and MAP17, and +53/+57 are located downstream of +51 enhancer (Figure 3-6A). Given the global role of CTCF in genome organization [89, 130], it is likely that CTCF may bind differently to CTCF elements in the TAL1 locus in erythroid and leukemic cells, thereby, regulating the +51 enhancer and TAL1 promoter 1 interaction. To examine this model, we carried out CTCF ChIP analysis in K562 and Jurkat cells. Interestingly, CTCF bound to all of the CTCF elements in both cell lines (Figure 3-6B) as well as in primary CD4+ T-cells (Figure 3-6C), suggesting the binding of CTCF alone is not sufficient to modulate TAL1 gene activity. We next examined whether CTCF differentially regulates genome organization by controlling TAL1 promoter accessibility in erythroid K562 and T-ALL Jurkat cells. To address this question, we performed chromosome conformation capture (3C) assays in K562 and Jurkat cells using the -31 Kb CTCF element as a bait using
DpnII enzyme (Figure 3-7A). Figure 3-7B illustrates a specific interaction between CTCF -31 and +53/+57 sites predominantly in erythroid K562 cells but not in T-ALL Jurkat cells. As a control, there is no CTCF mediated loop formed between -10 Kb and +40 Kb CTCF elements (Figure 3-7B&C). The chromatin loop between -31 and +53 CTCF sites might bring the +51 enhancer and promoter 1 of the TAL1 gene into a close proximity, thereby facilitating the +51 enhancer mediated TAL1 promoter 1 activation. In contrast to erythroid K562 cells, a smaller chromatin loop between +40 and +53 CTCF sites is detected predominantly in T-ALL Jurkat cells (Figure 3-7B and C). Together, this data suggests a cell-type specific chromatin organization by CTCF to regulate TAL1 expression by keeping promoter/enhancer in close proximity in erythroid cells while excluding the +51 enhancer from interacting with the TAL1 promoter 1 in T-cell leukemia.

**Summary**

To understand the underlying epigenetic mechanisms of TAL1 activation in normal hematopoiesis, we performed ChIP, ChIP-seq, and chromatin conformation capture assays to investigate chromatin structure profile which correlates with transcriptional activation, at the TAL1 locus comparing erythroid progenitors cell lines and primary CD36+ cells and CD34+ HSCs. We observe changes in chromatin dynamics at the TAL1 locus from CD34+ to CD36+ cells co-relate with TAL1 promoter-enhancer usage. In CD34+ TAL1 promoter 1a, 1a and +19 enhancers are open while in CD36+, +51 erythroid enhancer acquires high enrichment for H3K4me1 mark. Further, we demonstrate that +51 enhancer interacts with the TAL1 promoter 1a via a long range chromatin loop in vivo. The recruitment of hSET1 HMT complex by TAL1 facilitates this interaction and depletion of hSET1 leads to loss of H3K4 methylation,
enhancer/promoter interaction, RNA PolII loading, and TAL1 transcription. Finally, we investigated the role of insulator protein CTCF in the regulation of TAL1 expression in normal and in malignant cells. We found that regardless of similar CTCF binding patterns at the TAL1 locus in erythroid progenitors and T-ALL cells, CTCF form differential regulatory loops to allow interaction of TAL1 promoter 1 and +51 enhancer in erythroid cells but a repressive loop in T-ALL cells. Overall, our study demonstrates, for the first time, an in vivo long range interaction between TAL1 promoter 1a and +51 enhancer, which is regulated and organized by hSET1 histone methyltransferase and insulator protein CTCF, which functions to fine tune TAL1 transcription in different hematopoietic compartments.
Figure 3-1. Histone modification associated with TAL1 enhancer and promoter activity in erythroid cells. Native chromatin immunoprecipitation assay assessing active (A) and repressive (B) histone modification profile at the TAL1 locus in K562 and HL-60 cells. The primers used for the realtime PCR are indicated on the TAL1 locus map. ChIP Fold enrichments were determined based on realtime PCR data after normalization to Input DNA. The data represents one of the three independent biological repeats.
Figure 3-2. TAL1 promoter 1 and +51 enhancer interact in vivo in CD36+ erythroid progenitors. A & B) ChIP-seq analysis for H3K4me3, H3K4me1, TAL1 and Set1 enrichment at the TAL1 locus in CD34+ and CD36+ cells. C) Genome-wide targets of SET1 and TAL1, in CD36+ cells. About 45% of SET1 targets are also bound by TAL1. D) Outline for the 3C assay using BamHI 6-cutter enzyme. E) PCR analysis of the 3C DNA indicating prom1 and +51 interaction in CD36+ cells, F) Sequencing results of TAL1 prom1a and +51 enhancer 3C interaction PCR product. ChIP-Seq analysis in CD36+ cells was performed at NHLBI, in the laboratory of Dr. Keji Zhao.
Figure 3-3. H3K4 methyltransfearse SET1 is required for the colony formation capacity of K562 and CD34+ HSCs. A) NlaIII cut sites at the TAL1 locus and PCR primers used for 3C, B) 3C analysis for TAL1 promoter1 and +51 enhancer interaction in K562, Jurkat, and HL-60 cells, C) Western blot analysis for SET1 knockdown in K562 cells, D & E) SET1 knockdown affects colony formation capacity of the K562 cells on soft agar. Representative pictures are shown in (D) and quantitation in (E), F) Western blot for SET1 knockdown in CD34+ cells, and G) Methylcellulose assay testing colony formation capacity of CD34+ cells, upon SET1 knockdown. *P-value>0.05. The colony formation assay is representative of two biological repeats, with three technical replicates in each.
Figure 3-4. SET1 is recruited at the TAL1 enhancer/promoter elements and is required for activation of various erythroid genes including TAL1. A) ChIP analysis for SET1 recruitment at the TAL1 locus. HOXB4 locus is used as a positive control. Data represented as relative fold enrichment after normalization with IgG (Background). B) RT-qPCR analyses of TAL1(B) and various erythroid genes (C) mRNA levels in the control and three KD clones harboring shRNA specific for hSET1. *P-value<0.05, **P-value<0.01. Data are shown mean ± SD. The data is cumulative of two biological repeats with three technical replicates.
Figure 3-5. Loss of hSET1 disrupts the erythroid specific long-range interaction between +51 enhancer and the TAL1 promoter 1 at the TAL1 locus. (A) The 3C analysis for the interaction between +51 enhancer and the TAL1 promoter 1 at the TAL1 locus in the vector control and shSET1 transduced clones. B) ChIP analyses of H3K4 me2 and me3 levels as well as RNA PolII recruitment at +51 enhancer and the TAL1 promoter1 upon hSET1 KD. C) Shown is real-time qPCR quantitation of the 3C products upon SET1 knockdown. 3C data is plotted after quantitation with the primers described on top and is represented as relative interaction frequency after normalization to BAC template and ERCC3 control region. Data are shown mean ± SD of three independent experiments. *P-value<0.05; **P-value<0.01.
Figure 3-6. Similar CTCF occupancy is observed across different cellular environments. A) CTCF sites at the human *TAL1* locus, B) ChIP data illustrating no differences in CTCF binding pattern in K562 and Jurkat cells. SILp, prom1a and HS2 are used as negative control regions, and C) UCSC genome browser representation of CTCF ChIP-seq data on *TAL1* locus, plotted from raw data from Cuddpah et al., 2007.
Figure 3-7. CTCF organizes TAL1 locus to promote TAL1 promoter/enhancer interaction in erythroid cells. A) Human TAL1 locus displaying four CTCF binding sites (blue fonts) and DpnII restriction sites (purple dashed lines) used for 3C analysis. B) Semi-quantitative PCR analysis of the 3C DNA product to assess interaction among CTCF sites. Specific primer set used are indicated on right along with the specific gel band size. +53 CTCF site is used as bait and -10 Kb region was used as a negative control region. C) A total of 3 independent 3C experiments were quantitated by densitometry. Shown are the mean ± SDM of 3 independent experiments. *P-value < 0.05.
CHAPTER 4
TAL1 GENE REGULATION IN T-CELL LEUKEMIA

Introductory Remarks

TAL1 gene which is normally expressed in erythroid, megakaryocyte and mast cells, is turned off early at double negative stage of normal T-cell development. However it is frequently activated in human T-cell leukemia. Unlike erythroid cells, there is scanty literature outlining mechanisms underlying ectopic activation of the TAL1 oncogene in T-ALL. As previously discussed, a third TAL1 promoter, promoter IV was described in 1998 [69]. TAL1 promoter IV appears to be inactive in Sil-TAL1 deletion cells, in normal bone marrow and in erythroid and megakaryocytic cell lines. Further studies identified a 3’ silencer that is actively bound by a cell specific factor which functions to repress promoter IV activity in these cells and this repression is released in leukemic T-cells [70]. The molecular mechanism of this repression and the identity of the trans-acting factor currently remain unknown. We are particularly interested in the 60% of T-ALL patients which overexpress TAL1 in absence of any TAL1 locus rearrangements. Jurkat and REX cell lines were established from this pool of patients and therefore provide an excellent model system. Earlier studies proposed monoallelic expression of the TAL1 gene in Jurkat cells. The CpG island 5’ of the TAL1 gene appears to be methylated in one TAL1 allele. On the other allele, the authors further identify a DNAse I hypersensitive region about 7Kb upstream of the TAL1 gene as a potential cis-regulatory element in T-ALL [131]. However no follow-up study, dissecting the role of this -7Kb element in TAL1 gene activation had been described. A recent study by Sanda et al., revealed a new cis-regulatory element located about -12Kb upstream of the TAL1 TSS, which is bound by TAL1-GATA3-Lmo1 complex [132]. The
role of this element deserves further investigation. Furthermore, since GATA1 is described to be essential for *TAL1* gene activation in normal hematopoiesis, previous studies investigating the role of various GATA proteins illustrated that, neither GATA2 nor GATA3 could activate *TAL1* promoter in T-ALL [43]. Moreover, GATA3 ChIP-seq analysis in Jurkat cells did not identify *TAL1* as a GATA3 target [133]. Distinctively, a study assessing role of HDACi in T-cell leukemia, identified that TAL1 levels are drastically reduced upon treatment with various HDACi’s including sodium butyrate, TSA, and SAHA in a dose and time dependent manner [134]. This effect of HDACi is independent of apoptosis, protein stability or alternate splicing or mRNA stability. Moreover, RNA polII recruitment is altered upon HDACi treatment at the *TAL1* promoter 1 and IV, in cell lines with intact *TAL1* gene. These studies provide interesting observations, however more in depth analysis about the molecular mechanisms need to be illustrated.

Another interesting candidate regulating *TAL1* gene in T-ALL appears to be long-noncoding RNAs. These IncRNAs range from 100nt - 9100nts, are spliced, poly-adenylated and have no protein coding potential [135]. Two IncRNAs termed A3 and A4 were described to be transcribed (A3, sense and A4, antisense to *TAL1*) from the intergenic region between MAP17 and downstream CYP4A11 genes in MCF7 breast cancer cells. Further, knockdown of A3 in MCF7 cells specifically down regulated CYP4A11 and *TAL1* transcripts. No knockdown studies for A4 were illustrated. These findings define a novel candidate for *TAL1* gene regulation in T-ALL. However, despite several attempts we failed to identify these IncRNAs (A3 and A4) in the MCF7 cells.
Moreover, we found no TAL1 expression in MCF7 cells, thus limiting our analysis to further test the role of these IncRNAs in T-ALL.

A different perspective to understand TAL1 activation in T-ALL is to understand how TAL1 is initially repressed during normal T-cell development. It is likely that these repressive mechanisms fail during leukemia manifestation, thus leading to aberrant activation of TAL1 oncogene. Unfortunately, there is no literature regarding this aspect. Overall, it is very clear that we have very limited knowledge about TAL1 gene regulation in leukemia, and since TAL1 is an essential component for T-cell acute leukemia manifestation, it is critical to dissect the mechanisms underlying its ectopic activation in T-ALL.

**Results**

**Histone Modification Profile at the TAL1 Locus in T-ALL Cells**

To assess the role of known TAL1 cis-regulatory elements in T-cell leukemia, we investigated distribution of various active and repressive histone modifications across the TAL1 neighbourhood in T-ALL. Native ChIP analysis coupled with real time PCR analysis indicated presence of high levels of H3K4me2 and H3K4me3 at the TAL1 promoter IV in Rex (figure 4-1) and Jurkat cells (figure 4-2), where this promoter is utilized. REX and Jurkat are T-ALL cell lines where TAL1 is highly activated. HPB-ALL is a TAL1 non-expressing cell line, where the TAL1 promoters and enhancers are enriched for repressive H3K27me3 mark (figure 4-1). In both REX and Jurkat cell lines, TAL1 promoter 1a and 1b appears to be enriched with active histone modifications, however, enhancer +19 and +51 are depleted of both active and repressive histone marks. Interestingly, TAL1 embryonic enhancer -4Kb appears to be open in both cell
types. This might be good candidate to look further into. *SIL* promoter is enriched for active modifications as it is expressed in all cell types studied.

**Enhancer +51 is Epigenetically and Functionally Silent in T-ALL**

Our results in erythroid precursor cells illustrate that +51 enhancer interacts with *TAL1* promoter 1 in vivo via a chromatin loop. Since we did not observe enrichment for active histone modifications at the *TAL1* +51 enhancer in T-ALL cells, we next assessed whether this *TAL1* enhancer/promoter interaction is present in various T-ALL cell lines. For this we performed similar 3C analysis using NlaIII a four cutter in various T-ALL cell lines, including K562 as a positive control. As shown in figure 4-3A, we could not detect interaction between *TAL1* promoter 1 and +51 enhancer in T-ALL cell lines in contrast to K562 cells. HL-60 and HPB-ALL are TAL1 non-expressing cell lines and therefore serve as negative controls. The specific interaction product was further sequences to confirm fusion between promoter 1 and +52 enhancer fragments (Figure 4-3B). This data is consistent with figure 3-3B.

Next, we tested whether +51 enhancer was active in T-ALL cells. Two different sized DNA fragments containing the +51 enhancer element were cloned into a SV40 minimal promoter driven luciferase reporter and introduced into K562 as well as in several T-ALL cell lines. Compared to the pGL3-SV40 vector that showed only minimal luciferase activity, the 2 Kb +51 enhancer element specifically activated transcription of the luciferase reporter in K562 cells, but not in the T-ALL cell lines, Jurkat, Rex, Molt4, and HPB-ALL (Figure 4-3C). Interestingly, the 4 Kb fragment containing +51 enhancer and downstream +53 Kb CTCF site blocked transactivation activity of the +51 enhancer in K562 cells suggesting that the +53 Kb CTCF site may block the +51 enhancer from
activating downstream neighboring genes. Together, the data revealed that the +51 enhancer is neither epigenetically nor transcriptionally active in T-ALL cells.

**SET1 Independent TAL1 Gene Regulation in T-ALL**

We observe high levels of H3K4me at the TAL1 promoters in T-ALL cell lines. Moreover, SET1 regulates TAL1 gene in erythroid cells and unpublished data from the lab has demonstrated interaction between TAL1 and SET1 in Jurkat cells. Further, GST studies have identified ASH2L, a SET1 complex component, to physically interact with TAL1 to mediated SET1 and TAL1 interaction (not shown). Therefore, we hypothesized that SET1 might play a similar role in TAL1 gene regulation in T-ALL. To test this, first we assessed SET1 recruitment at the TAL1 locus in T-ALL Jurkat cells. In contrast to erythroid cells, SET1 was absent at both the TAL1 promoter and enhancer (figure 4-4C). NKX3.1 promoter is used as a positive control. Furthermore, unlike K562 cells, we observed no enrichment of TAL1 at its enhancer and promoter in Jurkat cells (figure 4-4A & B). To confirm the effect of SET1 on TAL1 levels, in T-ALL, we stably knocked down SET1 in Jurkat cells. SET1 depletion in Jurkat cells, did not affect TAL1 mRNA and protein levels (figure 4-4D & E). These results suggest that both TAL1 and SET1 fail to bind TAL1 locus in T-cell leukemia, and therefore does not regulate TAL1 expression in T-ALL.

**CTCF Mediated Cell-type Specific Chromatin Loops in the TAL1 Locus Block TAL1 Enhancer/Promoter Interaction in T-ALL**

As described in figure 3-7, we identified a unique role for CTCF in TAL1 gene regulation, wherein interaction among CTCF sites -31 and +53 in erythroid cells might facilitate TAL1 promoter 1 and +51 enhancer interaction, while interaction between CTCF sites +40 and +53 in T-ALL Jurkat cells, might block the +51 enhancer from
activating TAL1 promoter 1 in T-cell leukemia. To further validate these findings in T-ALL patients, we obtained primary T-ALL frozen bone marrow sample from collaborators. First we assessed whether TAL1 was expressed in this patient by RT-qPCR and immunoblot analysis. As shown in figure 4-5A, T-ALL sample expresses very high levels of TAL1 when compared to positive control Jurkat cells and negative control HPB-ALL cells. Although protein extraction from the T-ALL sample was not efficient, we could still significantly observe TAL1 protein expression in T-ALL sample (figure 4-5B). Next, we assessed whether the TAL1 locus in this patient was intact. For this we designed primers as described by earlier studies evaluating the presence of various TAL1 gene rearrangements in T-ALL. As shown in figure 4-5C top panel, there are three distinct SIL-TAL1 deletion breakpoints, D1, D2 and D3, where D1 is the most prevalent (>90%). In addition to SIL-TAL deletions, t(1;14) (p32; q11) translocation is also observed in 3% of T-ALL cases (<3%). We performed semi-quantitative PCR analysis to test for the presence of these rearrangements, and we detected neither in the T-ALL patient sample. However, this evaluation is limited by the presence of only one positive control for the SIL-TAL1 d1 and therefore, we cannot rule out the possibility that the primers we used (though these primers were previously described and published) did not work efficiently. Next, to confirm our finding of CTCF’s role in Jurkat cells, we performed similar 3C analysis in T-ALL patient sample. About 1x10^7 cells were cross-linked with 1% formaldehyde and subjected to digestion with DpnII for 3C analysis. Normal human bone marrow samples were used as control for this study. Semi-quantitative PCR analysis of the 3C DNA using primer pairs, -31/+53; +40/+53, and loading control promoter IV, demonstrated specific interaction among CTCF sites +40 and +53.
predominantly in T-ALL sample (figure 4-6B). Whereas, in normal bone marrow we observed interaction between -31 and +53, consistent with the data in erythroid cells (figure 3-7). The residual interaction between -31 and +53 in T-ALL sample could be from the erythroid cells present in the T-ALL patient whole bone marrow sample. These interactions were further quantitated using Image J software (figure 4-6C).

**Summary**

In summary, unlike erythroid cells (chapter 3), in T-cell acute lymphoblastic leukemia cell lines and a primary patient sample we demonstrate that previously described *TAL1* enhancer’s +19 and +51 are epigenetically inactive, based on the histone modification profile. Furthermore, enhancer +51 fails to drive luciferase reporter in T-ALL, and is not involved in long-range interaction with the upstream *TAL1* promoter 1. Moreover, *TAL1* which is recruited to its own promoter and enhancer in CD34+, CD36+ and K562 cells (figure 3-2) does not auto-regulate itself in T-ALL. In contrast to erythroid cells, despite the enrichment for active H3K4me2/3 marks at the *TAL1* promoter 1 in T-ALL cells, SET1 depletion does not alter *TAL1* levels in T-ALL cells. This effect is due to absence of SET1 binding at the *TAL1* promoter 1 in T-ALL, therefore indicating the role of MLL HMTs in *TAL1* gene regulation in T-ALL. Finally, our studies investigating the role of insulator protein CTCF in the *TAL1* gene regulation describe that CTCF organizes the *TAL1* locus to block *TAL1* promoter activation by the downstream +51 enhancer in T-ALL, whereas in erythroid cells it may function to bring *TAL1* enhancer and promoter in close proximity. Confirmation of these findings in primary T-ALL patient samples and normal bone marrow indicates physiological functional significance for the role of CTCF and further supports the data obtained so far in T-ALL cell line model systems. Altogether, based on the histone modification profile,
TAL1 and SET1 occupancy, luciferase reporter assay and chromatin looping mediated by CTCF, we report that TAL1 +51 enhancer does not activate \textit{TAL1} gene in T-ALL, thus revealing fine tuning by the cellular mechanisms to allow proper activation of the \textit{TAL1} gene in normal, while prevent the same in malignant cells. However, rapidly dividing cancer cell, perturb the normal cellular and molecular mechanisms to their own advantage. And therefore, to dissect mechanisms underlying ectopic activation of the TAL1 oncogene in T-ALL, we investigated the role of novel T-cell leukemia specific enhancer elements genome-wide (refer to Summary and Discussion, chapter 6).
Figure 4.1. Histone modification profile at the TAL1 locus in T-ALL REX and HPB-ALL cells. A) Human TAL1 locus displaying primers designed against various cis-regulatory elements used for the ChIP analysis. B) Native ChIP analysis of active (H3-K4me2/3, K9/14Ac) and repressive (H3-K27m3) histone modifications in REX (top) and HPB-ALL (bottom) cells. Quantitative PCR analysis was performed to determine Input/Bound ratio. This analysis was performed only once with three technical replicates. Data is represented as mean ±SD from three technical repeats of single experiment.
Figure 4-2. Histone modification profile at the TAL1 locus in T-ALL Jurkat cells. A) TAL1 locus map outlining various primers used for the ChIP assay, B) Quantitative PCR analysis data for fold enrichment of various active and repressive histone modifications at various cis-regulatory regions of the TAL1 locus in Jurkat cells. The data is representative of three biological independent repeats.
Figure 4-3. TAL1 +51 enhancer is neither functionally active nor does it interact with TAL1 promoter in T-ALL cells. A) 3C analysis for interaction between TAL1 promoter 1 and +51 enhancer in various cell lines. Primers used for the PCR are indicated on right and the NlaIII cut sites are illustrated on top. B) Sanger sequencing results of the 3C interaction PCR product, and C) Luciferase reporter assay was performed in various cell lines using the constructs illustrated on left. Relative luciferase activity was calculated after normalization with renilla luciferase and to empty vector. Error bars represent mean ±SD from three biological repeats, *P-value<0.05.
Figure 4-4. SET1 is neither recruited to the TAL1 locus nor is it required for TAL1 expression in T-ALL Jurkat cells. ChIP analysis for TAL1 binding at the TAL1 locus in K562 (A) and in Jurkat (B) cells. SIL and reg -16Kb are used as negative controls, C) SET1 ChIP analysis demonstrating no SET1 binding at TAL1 enhancer/promoter in Jurkat cells. Nkx3.1 promoter is used as a positive control, and D) SET1 knockdown in T-ALL Jurkat cells did not affect TAL1 protein (D) and mRNA (E) levels in three independent clones. *P-value<0.05.
Figure 4-5. Assessment of TAL1 expression and chromosomal rearrangements in primary T-ALL patient sample. A) RT-qPCR analysis for TAL1 mRNA levels in T-ALL sample, B) Western blot analysis for TAL1 protein expression in T-ALL, and C) Semi-quantitative PCR analysis testing absence of any TAL1 locus genomic rearrangements in primary T-ALL patient sample. Various rearrangements involving TAL1 are illustrated on top. Primers were designed against the various breakpoints for semi-quantitative PCR analysis. CEM is a positive control for SIL-TAL1 D1 rearrangement.
Figure 4-6. CTCF reorganizes TAL1 locus to block +51 enhancer mediated promoter activation in T-cell leukemia. A) Human TAL1 locus displaying four CTCF binding sites (blue fonts) and DpnII restriction sites (purple dashed lines) used for 3C analysis. B) Semi-quantitative PCR analysis of the 3C DNA product to assess interaction among CTCF sites. Specific primer set used are indicated on right along with the specific gel band. -10 Kb region was used as a negative control region. C) A total of 3 independent 3C experiments were quantitated by densitometry. Shown are the mean ± SDM of 3 independent experiments. *P-value < 0.05.
Figure 4-7. CTCF depletion in T-ALL Jurkat cells induces TAL1 transcription. A) Western blot analysis demonstrating CTCF knockdown and upregulation of TAL1 protein in two independent pools of shCTCF knockdown Jurkat cells. This data is representative of two biological repeat, and B) RT-qPCR analysis measuring TAL1 mRNA levels after CTCF knockdown in Jurkat cells. Error bars represent mean ± SD from two independent biological replicates, *P-value<0.05, **P-value<0.01.
CHAPTER 5
ROLE OF LIM DOMAIN BINDING PROTEIN 1 IN TAL1 GENE REGULATION IN ERYTHROID CELLS

Introductory Remarks

Transcriptional regulation of its target genes by TAL1 is manifested in multifactorial complexes. One of most abundant and well-studied complex comprises of TAL1, E12/47, GATA1, LMO2 and Ldb1 proteins. E12 and E47 are spliced variants of E2A gene product. They belong to class I bHLH family and can homo or hetero-dimerize. E12 and E47 proteins are essential for normal T-cell differentiation. TAL1 is mostly present as a heterodimer with either of these or with HEB gene product. Lmo2 is a member of LIM (Lin 11, Isl-1, Mec-3) domain containing proteins. LIM-domains are cysteine-rich zinc-binding domains that are structurally similar to DNA-binding GATA finger domains. However, no DNA-binding activity has been demonstrated, rather they are proposed to provide protein-protein interaction interface [136]. LIM-domain binding protein 1 (Ldb1) was originally identified due to its ability to bind LIM-homeodomain (LIM-HD) or LIM-only (LMO) domain. It has no enzymatic or DNA binding activity and can homo or hetero dimerize. Deletion of Ldb1 in mice results in embryonic lethality at day E9.5-E10 due to pleiotropic effects involving defects in heart, hindbrain, and posterior axis development as well as defects in mesoderm-derived extra embryonic structure including blood islands of the yolk-sac [137]. Induced deletion of Ldb1 in hematopoietic progenitor cells results in the rapid depletion of HSCs and downregulation of many genes encoding molecules known to be required for the specification and/or maintenance of HSCs [138]. This study further revealed Ldb1 occupancy at key hematopoietic genes including TAL1, C-MYB, LYL1, and LMO2. Genome-wide analysis of the LDB1 complexes containing GATA1 and TAL1
transcription factors in mouse E13.5 fetal liver primary erythroid cells revealed that the complex functions primarily as a transcriptional activator, however, it can function as a repressor in the presence of high levels of ETO-2 and Mtgr repressor proteins [139]. These studies demonstrate a role for Ldb1 in hematopoiesis and identify its genomic targets. A study, first described by Song et al., in 2007, shed light on Ldb1’s role in mediating long-range chromatin interactions at the mouse beta-globin locus [57]. Further work demonstrated requirement for Ldb1 in the migration of the Beta-globin locus away from the nuclear periphery, which is necessary to achieve robust transcription of β-globin in nuclear transcription factories [140]. Later, the LDB1 complex was reported to mediate and activate the long-range enhancer/promoter interactions at the Myb loci upon differentiation [141]. Thus, based on these studies and given the presence of GATA-9bp-E-box motifs at the TAL1 +51 enhancer and the promoter 1a, previous studies utilizing ChIP-chip experiments demonstrated the recruitment of the Ldb1 along with TAL1, GATA, E12 and Lmo2 proteins at the TAL1 locus [49]. However the role of Ldb1 on TAL1 gene regulation is currently unknown. Moreover, it remains to be illustrated whether Ldb1 is essential for TAL1 enhancer/promoter interaction, similar to the findings at the beta-globin and Myb locus.

Results

Ldb1 Depletion Results in Decreased TAL1 mRNA and Protein Levels

As Ldb1 is an essential component of TAL1-GATA1 complex, and given that both the promoter and enhancer +51 contain E-BOX/GATA motif, we hypothesize that Ldb1 must play an essential role in TAL1 gene regulation. To assess this, we depleted Ldb1 in K562 cells using lentiviral transduction system. Five distinct Ldb1 shRNA constructs were purchased from Open Biosystems to make lentivirus. The cells were than
transduced with a mix of all to obtain maximum knockdown efficiency. Immunoblot analysis for Ldb1 and TAL1 protein levels revealed reduction in TAL1 protein level upon Ldb1 knockdown. Similarly, TAL1 mRNA was reduced significantly upon Ldb1 depletion (figure 5-1B right). This data is consistent with published data indicating role of Ldb1 in TAL1 gene expression in hematopoietic stem cells [138]. Furthermore, Ldb1 knockdown decreased expression of various other erythroid genes including Myb, p4.2, GPA, γ-globin (not shown). To test whether the decrease in TAL1 levels was subsequent to decrease in long-range TAL1 enhancer/promoter interaction, we performed 3C analysis using TAL1 promoter 1 a bait in shLdb1 K562 cells. Semi-quantitative PCR analysis of the 3C DNA with TAL1 promoter 1a and +51 enhancer primers revealed decrease in interaction upon Ldb1 kd (figure 5-2B). Realtime PCR analysis was further utilized to quantitate these differences. Various other regions including Map17 promoter and +70 regions were used as negative controls. No interaction between TAL1 promoter 1 and negative control regions was observed (figure 5-2C). The interaction was also confirmed by sanger sequencing of the PCR product (not shown).

**SET1 Depletion Alters Ldb1 Levels in Erythroid Cells**

Given that SET1 plays an important role in TAL1 gene regulation, we next wanted to test its direct versus indirect roles, for this we first assessed Ldb1 levels upon SET1 in K562 cells. Figure 5-3 illustrates SET1 depletion decreases Ldb1 mRNA and protein levels significantly not only in K562 cells (top) but also in uninduced murine erythroleukemia (MEL) cells (bottom). These results raise question about effect of SET1 on TAL1 gene regulation might be through Ldb1. Therefore, to gain insight into this, we overexpressed Flag-tagged Ldb1 in shSET1 K562 cells using pOZ-N-FH retroviral vector system. This systems offers a unique advantage to screen Ldb1 overexpressing
cells, since both exogenous gene and IL2R are made as polycistronic mRNA. Mature IL2R receptor is then localized to the cell membrane and the positive cells can then be concentrated using IL2R ab-coupled magnetic beads. Western blot analysis indicates that overexpression of Ldb1 in shSET1 cells could not rescue TAL1 levels, reinstating role of SET1 in looping and disregarding that presence of Ldb1 alone does not suffice (figure 5-4B). However, since the cells are cultured for longer time to screen for Ldb1 overexpression, SET1 kd levels weren’t maintained and therefore no conclusive analysis can be made from this data.

**Ldb1 Regulation of TAL1 Gene is Independent of H3K4me2&3**

Based on the role of SET1 on TAL1 gene regulation in erythroid cells, we propose role of H3K4me2&3 modifications in mediating long-range chromatin interactions. To test this hold true in case of Ldb1 knockdown, i.e. is the decrease in TAL1 promoter and enhancer interaction that we observe upon Ldb1 knockdown is a combination of decrease in Ldb1 as well as in H3K4 methylation levels, we performed ChIP analysis for H3K4me2 and H3K4me3 enrichment at the TAL1 locus in shLdb1 K562 cells. Surprisingly, we did not observe any reduction in H3K4 me levels at both the TAL1 promoter and enhancer upon Ldb1 depletion (figure 5-4A). Therefore, based on this data, it appears that Ldb1 acts independent of H3K4me2/3 modification on TAL1 gene, however, overexpressing Ldb1 in shSET1 cells, does not rescue TAL1 levels.

**Summary**

Assessment of Ldb1 role in TAL1 gene regulation in erythroid K562 and MEL cells, shows Ldb1 regulates TAL1 transcription by regulating long-range TAL1 promoter and enhancer interaction. Both TAL1 enhancer +51 and promoter 1, which are epigenetically active in erythroid cells are bound by Ldb1. This function of Ldb1 appears
to be independent of SET1 mediated H3K4me2/3 modifications since H3K4me2/3 modifications remain unchanged upon Ldb1 knockdown, however, since Ldb1 overexpression studies in shSET1 cells did not rescue TAL1 levels, additional experiments are needed to dissect the molecular events. Specifically because neither SET1 nor Ldb1 have DNA binding motif’s and therefore might depend on TAL1 to be recruited to the specific E-box/Gata motifs at the TAL1 +51 enhancer and promoter 1. In vivo protein-protein interaction studies have also identified interaction between Ldb1 and SET1 in K562 cell nuclear extract (not shown). Similar experiments have detected interaction between Tal1 and SET1 in K562 cells as well (not shown). Further experiments need to be performed to characterize the cross-talk between the three proteins, i.e. TAL1, SET1 and Ldb1. From SET1 depletion studies in K562 and MEL cells its appears that SET1 regulates both TAL1 and Ldb1 expression. Whereas depletion of TAL1 by siTAL1 transient transfections and Ldb1 knockdown in K562 cells does not alter SET1 levels (not shown). However, whether this effect of SET1 on Ldb1 levels is direct or indirect via TAL1 needs to be investigated. Therefore, many aspects of Ldb1 mediated TAL1 gene regulation remain unanswered and future studies aiming to address these questions are needed.
Figure 5-1. Ldb1 is required for high levels of TAL1 in erythroid cells. A) Ldb1 knockdown efficiency in K562 cells. TAL1 levels are reduced upon Ldb1 knockdown, B) RT-qPCR analysis measuring 45% reduction in Ldb1 levels, and C) RT-qPCR analysis demonstrating >50% reduction in TAL1 levels upon Ldb1 depletion. *P-value<0.05, **P-value<0.01.
Figure 5-2. Ldb1 regulates TAL1 transcription by stabilizing long range TAL1 promoter - enhancer interaction in K562 cells. A) 3C analysis illustrating decrease in TAL1 promoter/enhancer interaction upon Ldb1 knockdown, and B) Quantitation of the interaction between TAL1 enhancer and promoter upon Ldb1 depletion using real-time PCR analysis. 3C data is plotted after quantitation with the primers described on top and is represented as relative interaction frequency after normalization to BAC template and ERCC3 control region. Error bars are represented as mean ±SD from two biological repeats.
Figure 5.3. SET1 regulates Ldb1 levels in K562 and in MEL cells. A & B) SET1 knockdown in K562 cells alters Ldb1 protein (A) and mRNA (B) levels in two independent clones, and C & D) SET1 knockdown in unduced MEL cells, reduces Ldb1, TAL1 and SET1 protein (C) and mRNA (D) levels in three independent clones. The data is representative of two biological repeats.
Figure 5-4. Ldb1 depletion does not alter H3K4me2/3 enrichment at the TAL1 enhancer and promoter. A&B) ChIP experiments for enrichment of H3K4me2 (A) and H3K4me3 (B) at the TAL1 enhancer +51 and promoter 1 upon Ldb1 depletion. Reg +70 is used as a negative control, and C) Western blot analysis for TAL1 protein upon overexpression of Ldb1 in shSET1 K562 cells. Overexpression of Flag-Ldb1 in shSET1 K562 cells is shown in the western analysis. Ldb1 did not rescue TAL1 protein levels.
CHAPTER 6
SUMMARY, DISCUSSION AND FUTURE DIRECTIONS

Summary and Discussion

Hematopoiesis is a very dynamic cellular process in which multipotent HSCs give rise to diverse hematopoietic lineages. TAL1 is a critical oncogenic transcription factor required for this process [142]. Aberrant activation of TAL1 in T-lymphocytes leads to leukemic transformation in the majority of childhood T-ALL [68]. It is therefore essential to understand the molecular mechanisms that regulate TAL1 transcription activity in normal hematopoiesis and leukemic T-cells. We found that TAL1 gene is controlled in normal hematopoietic progenitor cells by a long range intrachromatin loop that brings the +51 enhancer into close proximity of the TAL1 promoter 1. The loop interaction is specific for erythroid precursor cells and is absent in other hematopoietic cells where TAL1 is silenced or even in T-ALL cells where TAL1 is expressed. Thus, an interesting question is what underlies this differential selection of the +51 enhancer usage in erythroid precursors and how the tissue specific chromatin loop is established and stabilized. Several lines of evidence support that active histone modifications such as H3 acetylation and methylations play an important role in communication between genes and distal cis regulatory elements by chromosomal loops [75, 143]. One example is the β-globin locus where the LCR can serve as a primary site to recruit transcription factors and chromatin modifying and remodeling factors and stably alter topology of the β-globin locus during transcriptional activation in erythroid cells [57, 75, and 144]. Depletion of hSET1 led to reduced H3K4 methylation, disruption of the +51 enhancer/promoter1 chromatin loop, and loss of TAL1 transcription (Figure 3-5) suggesting a plausible model that hSET1 mediated H3K4 methylation may be required.
for establishing or stabilizing the enhancer/promoter communication through a long-range chromatin loop. The loss of PolII occupancy in the hSET1 depleted cells makes it possible that hSET1 mediated H3K4 methylations remodel local chromatin and facilitate the accessibility of the basal transcription factors to the \textit{TAL1} promoter. The data provides a potential mechanism that links H3K4 methylations and long-range enhancer/promoter in erythroid cells. An explanation to the absence of TAL1 promoter/enhancer interaction in T-ALL, might stem from the observation that neither Tal1, GATA3 nor SET1 occupy \textit{TAL1} promoter and enhancer regions, moreover SET1 depletion in Jurkat cells does not alter Tal1 levels. Therefore some tissue specific activity inhibits TAL1 complex recruitment at these sites in T-ALL. Moreover, presence of active H3K4 di and tri-methyl marks at the \textit{TAL1} promoter 1, in contrast to \textit{TAL1} +51 enhancer, indicates possible role for MLL family of H3K4 methyl transferases in \textit{TAL1} gene regulation in T-ALL.

One particularly interesting finding is that CTCF mediates different chromatin loops at the \textit{TAL1} locus in erythroid cells and T-ALL cells, thereby providing another layer of regulation to ensure proper TAL1 expression (figure 3-7). CTCF has been implicated in diverse regulatory functions, including transcriptional activation and repression, insulation, imprinting, and X chromosome inactivation \cite{89, 130}. CTCF molecules are capable of interacting with each other to form a cluster and thereby creating closed looping domains \cite{90}. It has been proposed that CTCF may play a primary role in the global organization of chromatin architecture and lineage–specific gene expression. With regards to the \textit{TAL1} locus, our data revealed that CTCF differentially organized chromatin loop domains in erythroid and leukemic cells such that
the +51 enhancer was in a close proximity to the TAL1 promoter1 in erythroid cells and
the +51 enhancer was blocked from the TAL1 promoter in T-ALL cells. It is likely that
the closed chromatin loop between -31 and +53 CTCF sites also prevents the upstream
SIL promoter from activating TAL1 gene in erythroid cells. Moreover, KD of hSET1 does
not affect the CTCF mediated looping in the TAL1 locus in erythroid cells (Appendix
Figure B). This suggests that CTCF is not involved in the enhancer/promoter interaction
directly and further demonstrates a unique role for SET1 in chromatin looping. Absence
of CTCF at the TAL1 promoter 1 and at the +51 enhancer supports this idea. However,
CTCF-mediated chromatin loops probably facilitate interactions between enhancers and
promoters by bringing them into a close proximity (refer to future directions).

The observations for CTCF occupancy and differential looping at the TAL1 locus
in erythroid Vs. T-ALL cells are similar to that described at the β-globin locus in K562
and 293T cells [94]. At the human β-globin locus, though the CTCF occupancy remains
invariant across K562 and 293T cells, the specificity to CTCF’s chromatin loop is
provided by cohesion proteins called “cohesins”. Depletion of cohesion protein Rad21
alters CTCF occupancy as well as chromatin looping at the beta-globin locus.
Furthermore, co-operation between CTCF and cohesion proteins has been
demonstrated at the protocadherin α gene locus, VDJ rearrangement at the Ig locus as
well as for inter-chromatin interaction between Bcl11b and Arhgap6 loci [145-147]. To
test whether the cell specificity to the CTCF role at the TAL1 locus is provided by
cohesins, we investigated the role of cohesin proteins, rad21 and smc3 at the TAL1
locus. ChiP analysis revealed no differences in cohesin occupancy across the four
CTCF sites at the TAL1 locus between K562 and Jurkat cells (appendix A). These data
supports the idea that even CTCF interaction with its cofactor cohesin does not seem to differ in different cell types [148]. Next, whether differences in post-translationally modified CTCF protein and/or CTCF’s interaction with other factors including YY1, RNA polIII, are important for this cell specific role of CTCF will be addressed (refer to future directions).

The communication between promoters and cis regulatory enhancers by inter- or intra- chromosomal interactions has become an important mechanism that governs gene activation over long distances. Many erythroid expressed gene loci are regulated by a long-range chromatin loop that facilitates the distal enhancer accessing promoter in developmental stage specific manner. The question remains to be answered is how the tissue specific chromatin loop is established and stabilized. It was reported that the LDB1 mediated TAL1/GATA1 transcription factor complex is required for the loop formation at the β-globin locus [57]. Recent study using zinc finger mediated tethering of LDB1 to the β-globin locus in the GATA-1 null cells indicated that recruitment of LDB1 initiated long-range enhancer/promoter interaction and globin transcription [149]. In the TAL1 locus, there are four TAL1-GATA motifs present at the promoter 1a and the +51 enhancer. TAL1 transcription in hematopoiesis is tightly regulated by multiple cis regulatory elements spreading over the whole TAL1 locus. The +51 enhancer (+40 in mice) is required for TAL1 expression in primitive and definitive erythropoiesis [47]. It has been shown that the TAL1/GATA-1/LMO2/LDB1 multi-protein complex is enriched at the +51 enhancer in K562 cells [49]. Consistent with the importance of the TAL1/GATA1 complex during erythroid differentiation [150], we found that the TAL1 complex is localized to both +51 enhancer and the TAL1 promoter 1a in CD36+
erythroid precursors (Figure 3-2) and KD of LDB1 disrupts the chromatin loop between the +51 enhancer and the TAL1 promoter 1a in erythroid cells (Figure 5-2). Thus, our data support that the LDB1 mediated chromatin loop organization may initiate and facilitate TAL1 gene activation in erythroid precursors. Nevertheless, depletion of hSET1 led to reduced H3K4 methylation, disruption of the +51 enhancer/promoter 1a chromatin loop, and loss of TAL1 transcription suggesting a plausible model that hSET1 mediated H3K4 methylation may be required for stabilizing the enhancer/promoter looping initiated by LDB1 complex. It is interesting to note that active histone modifications including histone acetylation and asymmetric dimethylH4R3 are essential for maintaining active chromatin loop in the β-globin locus. However, it is possible that hSET1 mediated H3K4 methylation remodel local chromatin may also facilitate the accessibility of the TAL1/GATA1 transcription factor complex to the DNA motifs.

**Putative TAL1 Interacting Regions in T-ALL**

In T-ALL patients, TAL1 oncogene can be activated by chromosome translocation and interstitial deletion [67, 68, 73]. However, chromosomal rearrangements only account for less than 30% of cases of all with aberrant TAL1 overexpression (Figure 1-2). The question arises, how is the TAL1 oncogene aberrantly activated in the majority of T-ALL patients that lack chromosome rearrangements in the TAL1 locus and whether dysregulation of enhancer/promoter interactions leads to a disease-causing regulatory variant. Although several enhancer elements have been identified in the TAL1 locus [151], epigenetic, chromatin looping, and reporter analysis suggested that none of them is transcriptionally active in T-ALL cells (Figures 4-1,4-2 and 4-3). Thus, currently, it remains unknown, how TAL1 is activated in the majority of T-ALL patients lacking the TAL1 locus rearrangements. To understand the molecular
mechanism underlying regulation of the *TAL1* oncogene in leukemic T-cells, we employed circularized chromosome conformation capture (4C) methodology to identify new regulatory elements that activate TAL1 specifically in T-ALL leukemia. This dissertation will not cover details on the data obtained from 4C analysis in T-ALL, however the key findings are discussed. Using *TAL1* promoter 1a as the bait, we discovered that the *TAL1* promoter 1a interacts with the *TIL16* (*TAL1* interacting locus on chromosome 16) element in T-ALL cell line Jurkat, but not in erythroid progenitor K562 cells (Figure 6-1B). *TIL16* is located intergenic about ~15 Kb downstream of T-cell specific CD2BP2 gene and 5Kb upstream of the non-coding RNA gene LOC595101 on chromosome 16. The CD2BP2 protein is a cellular adapter protein that was originally identified as a binding partner of the T cell adhesion protein CD2 in the context of T cell signaling [152]. It is also called as U5-52K (U5 snRNP 52K protein), for its interaction with U5-15K protein within the splicosomal U5 snRNP [153]. The inter-chromosomal interaction between the *TIL16* and the *TAL1* promoter 1a was further confirmed by 3C assay in three *TAL1* over-expressing T-ALL cell lines, Jurkat, REX and Molt4, but not in K562 cells and by sanger sequencing of the specific interaction band. To further test this interaction in human patients, we performed 3C assay for *TAL1-TIL16* interaction in primary T-ALL patient bone-marrow sample. Consistent with our initial findings, we could detect *TAL1-TIL16* interaction in the T-ALL patient sample. This patient ectopically expresses TAL1 in absence of its gene rearrangements (not shown). These findings identify a novel *TAL1* interacting locus TIL16, with a potential enhancer activity to drive TAL1 expression in T-cell leukemia. We further identify the role of c-maf proto-oncoprotein in *TAL1* gene regulation by regulating this trans-interchromosomal
interaction in T-ALL (figure 6-1B). Evaluating the role of this novel trans-interaction in T-ALL will not only aid in answering a long standing mystery about the molecular mechanisms that underlie TAL1 activation in T-cell leukemia, but will also provide us with newer targets for better therapeutic intervention of T-cell leukemia treatment and cure.

Figure 6-1. Model: CTCF and epigenetic mediated chromatin looping regulated TAL1 expression in hematopoiesis and leukemogenesis. A) In erythroid cells, interaction between CTCF sites -31 and +53 organizes the TAL1 locus to bring TAL1 promoter 1 in close proximity to +51 enhancer, an interaction which is regulated by SET1 and LDB1, a TAL1 complex component, B) In T-cell acute leukemia, interaction between CTCF sites +40 and +53 insulates TAL1 promoter activation by +51 enhancer. However, in T-ALL cells, a novel cis-regulatory element located on chromosome 16 “TIL16” drives TAL1 expression, an interaction which is regulated by c-maf.
Future Directions

Evaluating the Factors Regulating Cell-type Specific Role for CTCF

Our findings reveal that despite no differences in CTCF occupancy at the TAL1 locus in K562, Jurkat and CD4+ T-cells, we observe cell-type specific chromatin loops involving these CTCF sites. Specifically we observe a small loop between CTCF sites +40 and +53 in T-ALL, which blocks +51 enhancer from activating TAL1 promoter 1, whereas in erythroid cells, interaction between -31 and +53 sites might function to keep TAL1 enhancer and promoter in close proximity. Following from this analysis, we next want to investigate what defines this cell-type specific role for CTCF? As discussed in the results and discussion section, we find no differences in cohesion proteins Rad21 and Smc3 occupancy at the TAL1 locus in K562 and Jurkat cells, similar to our observations for CTCF occupancy (Appendix A). Next, we’ll investigate the role of additional cohesion proteins smc1, SA1 and SA2. Recent reports highlight the role for cohesion subunit SA1, as largely responsible for cohesin accumulation at promoters and at sites bound by the insulator protein CTCF [154, 155]. Moreover, there’s some specificity attributed to each subunit. We will also investigate the role of additional CTCF interacting proteins, RNA polIII, YY1, and nucleophasmin which have been reported to influence CTCF recruitment and chromatin function [156]. For this, ChIP analysis for each of these factors will be first performed to see any differences in binding pattern across K562 and Jurkat cells. For the factors with significant differences, which might clue in driving cell-type specific regulation by CTCF, we’ll further perform depletion studies to dissect the detailed molecular mechanism.

In addition to CTCF’s function being influenced by its interacting proteins, CTCF undergoes several post-translational modifications, including acetylation,
phosphorylation, sumoylation and poly(ADP-ribo)sylation. Parylated forms of CTCF occur both in nucleus and in nucleolus, which is reflected in the change in molecular weight from 130 Kd to 180 Kd. Parylation involves addition of poly (ADP-ribose) chains (PARs) to proteins catalyzed by Poly (ADP-ribose) polymerases [157]. Parylation of CTCF is important for its role in genomic imprinting as treatment with PARP inhibitors led to loss of imprinting at 140 CTCF targets genome-wide [158]. Parylated CTCF is further illustrated to repress ribosomal gene transcription [159]. Moreover, CTCF directly interacts and activates PARP-1 protein in vivo and in vitro [160]. Thus, parylation of CTCF plays an important role in modulating CTCF’s cellular functions. Distinctly, phosphorylation of CTCF C-terminal four residues (604, 609, 610, and 612) by casein kinase II (CK II) attenuated CTCF mediated repression on c-myc gene without affecting its DNA binding [161]. The authors further demonstrate co-expression of CTCF and CK II switches function of CTCF from a repressor to activator [162]. Since phosphorylation does not alter CTCF’s DNA binding, it appears to be an interesting and a plausible factor to assess. CTCF protein can be further modified by ubiquitin like small SUMO molecules by SUMO1 and SUMO2 proteins, in vivo and in vitro. Sumoylation of CTCF contributes to its repressor function on c-myc promoter without affecting its DNA binding in vitro [163]. Given, this effect of various post-translational modifications on CTCF’s function, detail analysis for each of these modifications is proposed. First, immunoblot analysis will be utilized to detect whether CTCF is modified in erythroid and in T-ALL cells. Next ChIP analysis using modification specific antibodies will be performed to assess differences in site specific bound CTCF protein (given that both phosphorylation and sumoylation does not affect CTCF’s DNA binding but alter its function). Based on
this analysis, we'll further test the effect of modification specific inhibitors on CTCF’s activity at the *TAL1* locus.

**CTCF Depletion Studies in Erythroid and in T-ALL Cells**

Based on our findings, it is conceivable that CTCF depletion in erythroid and T-ALL cells might have distinct effects on *TAL1* expression. Interestingly, CTCF knockdown in T-ALL Jurkat cells, increases Tal1 mRNA and protein levels (figure 4-6). Whether this increase is due to *TAL1* promoter 1 and +51 enhancer interaction establishment or whether CTCF has a role in *TAL1* promoter1 and TIL16 interaction in T-ALL will be investigated. Thus, our future work will utilize 3C analysis using NlaIII restriction enzyme to assess the effect of CTCF depletion on *TAL1* locus chromatin architecture in erythroid K562 and in T-ALL Jurkat cells. To assess the role of CTCF on *TAL1* gene regulation in erythroid cells, we'll utilize similar CTCF knockdown studies using lentiviral system in K562 cells.

**Blocking CTCF Activity by Utilizing Artificial Zinc-Finger DNA Binding Domains**

The effect of CTCF knockdown may be non-specific and more robust since CTCF plays various cellular functions. Therefore, to evaluate the role of CTCF specifically in *TAL1* gene regulation, we plan to utilize artificial zinc-finger DNA binding domains (Zn-DBDs), as described earlier [164, 165]. Zinc-finger is one of the major structural motifs found in DNA-protein interaction. Various Zinc-finger proteins are key players in transcriptional regulation including Sp1, c-myc, and CTCF. Cys2-His2 zinc-finger domain represents the most common DNA-binding motif in eukaryotes and about 2% of our genome (700 genes) encodes for zinc-finger of this type. Zinc-fingers utilize
approximately 30 amino acids to specifically recognize 3bps of DNA [166]. Cys2-His2 zinc-finger domains are particularly well suited for the construction of synthetic transcription factors as they are commonly arranged as covalent tandem repeats, allowing the recognition of extended asymmetrical sequences up to 18bps. This modularity in both structure and function is of great advantage to designing artificial transcription factors [167]. Given these applicability, Zn-DBDs have been utilized to block long CAG repeats in Huntington’s disease, to repress expression of mutated huntingtin protein whereas the short wild type gene and subsequent protein product remained unaltered [168]. Similarly, various studies have utilized these artificial zinc-finger proteins for specific gene regulation [165, 169, and 170]. Thus, Zn-DBDs will provide a useful tool to specifically block a given CTCF site and thus aid in investigating the site specific effect of CTCF on TAL1 locus reorganization. For this we’ll be designing Zn-fingers against 18bp of CTCF sites -31 and +40, since +53 site appears to be non-specific. The Zn-DBDs cDNAs will be ordered from GenScript INC. These Zn-DBDs will have 2x FLAG tag at the c-terminal. The flag-tagged-Zn-DBDs will be cloned in pMSCV retroviral vectors and stably transduced into K562 and Jurkat cells. The expression, specificity and efficacy of the zinc-fingers will be validated in vivo and in vitro [165]. 3C experiments analyzing interaction among previously described CTCF sites as well as between TAL1 promoter 1/+51 enhancer and TAL1 promoter 1/TIL16 will be performed to study the effect of CTCF site inhibition on TAL1 locus organization and expression.
APPENDIX A
A. COHESIN’S ROLE IN TAL1 GENE REGULATION IN K562 AND JURKAT CELLS

As we observe differences in CTCF mediated chromatin loops in the TAL1 locus in erythroid and T-ALL, we investigated role of cohesion proteins Rad21 and Smc1 in providing the cell-type specificity the CTCF function. However, we did not observe any significant differences in cohesin occupancy at the TAL1 locus CTCF sites, indicating role of either cell-type specific CTCF interacting proteins or post-translational modifications of CTCF protein itself.

Figure A-1. No difference is observed in Cohesin occupancy at the TAL1 locus in K562 and Jurkat cells. A) TAL1 locus outlining distinct CTCF binding sites, B&C) ChIP analysis for cohesin subunit Rad21 (B) and Smc1(C) in K562 (top) and Jurkat (bottom) cells. Error bars represent mean ±SD from two independent repeats.
APPENDIX B
B. EFFECT OF SET1 DEPLETION ON CTCF SITE INTERACTIONS AT THE TAL1 LOCUS IN ERYTHROID CELLS

Our findings reveal a unique role for SET1 HMT in chromatin looping interaction between TAL1 promoter and +51 enhancer. To further test whether CTCF looping is dependent on SET1, we performed 3C analysis for CTCF site -31 and +53 in SET1 knockdown K562 cells (figure B). Semi-quantitative PCR analysis of the 3C DNA and control DNA with -31 and +53 site primers revealed no differences in interaction frequency upon SET1 knockdown. This was further confirmed by quantitation of the PCR product using realtime PCR analysis (figure B-C). This data implies a role for SET1 specifically for TAL1 promoter/enhancer interaction.

Figure B-1. No effect of SET1 depletion on CTCF -31 and +53 site chromatin loop in erythroid K562 cells. A) TAL1 locus Map displaying four CTCF sites and DpnII cut sites (red bars), B) Semi-quantitative PCR analysis of the control and 3C DNA using -31/+53 and loading control primers, and C) Realtime PCR quantitation of relative interaction frequency plotted after normalization to BAC template.
Covalent modifications of histones regulate a number of biological processes essential for normal cellular functions, including gene transcription. Whereas, each of these modifications has a specific function, how they are interrelated is an interesting and growing area of research. We previously demonstrated that asymmetric dimethylation of H4R3 residues by protein arginine methyltransferase PRMT1 potentiates histone acetylation and is essential both in vitro and in vivo for the establishment and maintenance of the active histone acetylation patterns at the chicken and mouse β-globin loci [75]. We report here a crosstalk between PRMT1 catalyzed H4R3ame2 and SET1 catalyzed H3K4me2/3 modifications. Previously we purified USF1 associated complexes in HeLa S3 cells stably transduced with a FLAG-HA-tagged chicken USF1 [88]. The purified complexes were subjected to LC-MS/MS, to identify USF-1 associated polypeptides. Interestingly, both PRMT1 and SET1 associate with USF1 in separate complexes, as determined by immunoblotting experiments of the column eluted fractions F6 to F44 (Figure C1). SET1 complex is a part of larger 1.8MDa USF-1 complex, while PRMT1 is a part of 400KDa complex. While assessing the role of PRMT1 on H3 acetylation, we found that PRMT1 is also required for global H3K4me2 and H3K4me3 levels in MEL erythroid cells (Figure C-2). This was tested in both mono-nucleosomes isolated from control and PRMT1 knockdown MEL cells, as well as crude histones prepared from these cells. We further show that the loss of PRMT1 through RNAi-mediated knock-down in murine erythroleukemia (MEL) cells prevents methylations of H3K4 at PRMT1 target beta-globin gene (Figure C-3). Therefore, we
speculate that the robust effect of PRMT1 knockdown on MEL β-globin transcription is due to effect on both histone acetylation and H3K4 methylation. Consistent with this, stable PRMT1 knockdown in mES cells, significantly reduces β-major and α-globin levels (Figure C-4). Reintroduction of rat PRMT1 into the PRMT1 knock-down cells, rescues PRMT1-mediated asymmetric dimethylH4R3, and subsequent histone H3K4 methylation patterns (Figure C-5). Rescue with rat PRMT1 also partially rescues SET1 recruitment at the β-major promoter and HS2 regions (Figure C-5b). Lastly, to dissect whether the effect on H3K4 me2/3 is due to its cross talk with H4R3ame2 or due to PRMT1 dependent SET1 recruitment, we performed in vitro methylation assay utilizing mono-nucleosomes isolated from PRMT1 kd and control cells. These nucleosomes were incubated with exogenous PRMT1 and s-adenosyl methionine (SAM), to methylate the mono-nucleosomes for 90 minutes at 30 degrees, followed by PRMT1 depletion using anti-PRMT1 antibody and protein A dynabeads. The modified nucleosomes were than incubated with PRMT1 knockdown nuclear extract and SAM. Western analysis revealed PRMT1-mediated asymmetric dimethyl H4R3 facilitates histone H3 Lys 4 methylation in vitro (Figure C-6). Thus, our results suggest an interdependent relationship between arginine and lysine methylations and their role in the establishment and maintenance of an active globin domain. Future experiments will focus on identifying whether the H3 acetylation plays a role in this H4R3ame2 and H3K4me2/3 crosstalk and to assess effect of SET1 knockdown on global H3 acetylation and H4R3ame2 levels. HDAC and p300 inhibitors will be further utilized to study the effect of H3 acetylation on H3K4 di and tri-methylation.
Figure C-1. USF1 associated complexes. Western blot analysis of column eluted USF-1 complexes fraction with PRMT1 and SET1 complex components; SET1, ASH2L and WDR5. Below is the list of polypeptides that were identified from LC-MS/MS data. SET1 and PRMT1 are eluted in different fractions with USF-1.
Figure C-2. PRMT1 depletion affects global H3K4me2/3 levels. A) Immunoblotting for various histone modifications and H3 loading control in mono-nucleosomes isolated using sucrose gradient centrifugation from control and PRMT1 knockdown cells, and B) Immunoblot analysis using crude histone isolated using acid extraction from control and PRMT1 knockdown cells.
Figure C-3. PRMT1 is required for β and α-globin transcription in EPO differentiated mES cells. A) Western analysis of whole cell extracts from control and PRMT1 knockdown in MEL cells showing PRMT1 knockdown does not alter SET1, ASH2L and USF1 levels. B) Quantitative RT-PCR analysis for β and α-globin levels in mES control and PRMT1 knockdown cells after differentiation with EPO. mES PRMT1 cells are original pooled knockdown cells.
Figure C-4. PRMT1 is required for active H3K4 di and tri-methylations at the mouse β-globin locus. A) Mouse β-globin locus displaying globin genes (red bars), hypersensitivity sites HS 1-6. B) ChIP analysis for H4R3ame2, H3K4me2 and H3K4me3 enrichment at the β-major and HS2 in control and shPRMT1 MEL cells. C) Immunoblot analysis for PRMT1 and various proteins upon PRMT1 depletion in MEL cells, and D) Quantitative RT-PCR analysis for β-major mRNA levels in control and shPRMT1 MEL cells upon differentiation with 1.5% DMSO at Day 3 and Day 5.
Figure C-5. Re-introduction of PRMT1 rescues H3K4me2/3 levels. A) Western blot displaying successful expression of Flag-tagged rat PRMT1 in shPRMT1 cells. B & C) ChIP analysis for various histone modifications and SET1 in control, shPRMT1 and PRMT1 rescue MEL cells. Re-introduction of rat PRMT1 rescues active histone modifications as well as SET1 at the mouse β-major promoter and HS2 regions. MyoD is used as a negative control region since it is not expressed in MEL cells.
Figure C-6. H4R3ame is sufficient to allow SET1 mediated H3K4me2/3 in vitro. A) Experimental strategy for in vitro methylation assay. PRMT1 is depleted using anti-PRMT1 antibody. B) Western blot analysis for H4R3ame2, H3K4me2/3; showing addition of PRMT1 and SAM catalyzes H4R3ame2, which is sufficient to allow H3K4me2/3 by SET1 enzyme present in PRMT1 knockdown nuclear extract.
LIST OF REFERENCES


<table>
<thead>
<tr>
<th></th>
<th>Authors</th>
<th>Year</th>
<th>Title</th>
<th>Journal</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>42.</td>
<td>Tremblay M, Tremblay CS, Herblot S, Aplan PD, Hebert J, Perreault C, Hoang T.</td>
<td>2010</td>
<td>Modeling T-cell acute lymphoblastic leukemia induced by the SCL and LMO1 oncogenes.</td>
<td>Genes &amp; development 24, 1093-1105</td>
<td></td>
</tr>
</tbody>
</table>


56. Wadman IA, Hirotaka O, Grutz CG, Agulnick AD, Westphal H, Forster A, and Rabbitts TH. (1997). The LIM-only protein Lmo2 is abridging molecule assembling an erythroid DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NL1 protein. EMBO J. 16, 3145-3157.


BIOGRAPHICAL SKETCH

Bhavita Patel was born in the state of Gujarat, India; first daughter of Harivadan Patel and Jyotsna Patel. She pursued her higher secondary education in Manav Kendra Gyan Mandir School, Kandari, Gujarat, and obtained senior secondary education diploma in 1998. She then joined the Biotechnology program at the Natubhai Vallabhai Patel Science College, Sardar Patel University, India and obtained Bachelors of Science with biotechnology major.

In 2002, Bhavita enrolled in the master’s program in Biochemistry, at the Maharaja Sayajirao University, Baroda, India, where she gained tremendous exposure to biological research while working with Dr. Jayashree Pohnerkar on polyketide production from Streptomyces and obtained a degree in Master of Science in biochemistry in the year of 2004.

In 2005, to pursue higher education, Bhavita got enrolled into the graduate degree program at the University of Texas at Dallas (UTD), Richardson, TX, United States of America (U.S.A). In 2006, she got married to a Floridian resident, which led her to graduate from the University of Texas with Master of Science in molecular and cellular biology, in 2007. During her master's degree at UTD, she worked with Dr. Jeff Dejong on characterizing testis specific non-coding RNAs.

To fulfill her career goals, Bhavita joined the Inter Disciplinary Program (IDP) at the University of Florida, Gainesville, FL in the year of 2008 and pursued her Doctor of Philosophy (Ph. D.) work in the laboratory of Dr. Suming Huang, identifying the role of epigenetic mechanisms in TAL1 gene activation. She has presented her work at several national and international conferences. During her Ph.D. studies, Bhavita was appointed and funded by T32 training grant in cancer biology.
In 2003, Bhavita was awarded Minaxi and Lalit award for her academic performance in Biochemistry. She has also obtained several awards including travel award from the graduate school, outstanding international student achievement award (2011), from the college of medicine at UFL and achievement award from American Society of Hematology (ASH) in the year of 2012.