GLOBAL DNA METHYLATION AND GENE EXPRESSION ANALYSIS IN PRE-B CELL ACUTE LYMPHOBLASTIC LEUKEMIA

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

MD ALMAMUN

Dr. Kristen H. Taylor, Dissertation Supervisor

JULY 2015

The undersigned, appointed by the dean of the Graduate School,

have examined the Dissertation entitled

GLOBAL DNA METHYLATION AND GENE EXPRESSION ANALYSIS IN PRE-B CELL ACUTE LYMPHOBLASTIC LEUKEMIA

Presented by Md Almamun

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

Kristen H. Taylor

James M. Amos-Landgraf

Gerald L. Arthur

Mark A. Daniels

J. Wade Davis

DEDICATION

This work is dedicated to my parents, Abdul Aziz & Khosh Nahar Begum, my inspiration.

ACKNOWLEDGMENTS

I would first like to acknowledge my advisor, Dr. Kristen H. Taylor for her guidance and patience throughout this research project. I was able to gain knowledge, research passion and complete this project with her tremendous helps and encouragement. Thank you for being so positive and for everything you did for me. I would also like to thank my degree committee members, Dr. J. Wade Davis, Dr. Gerald L. Arthur, Dr. James M. Amos-Landgraf, and Dr. Mark A. Daniels for their advice and guidance in this endeavor.

I would like to special thank Dr. J. Wade Davis, Benjamin T. Levinson, and Nathan T. Johnson for analyzing MIRA-seq and RNA-seq data, and Annette C. van Swaay for helping me during data analysis. Thank to Dr. Stephanie D. McKay for her advice during method development. I would like to thank the members of the Taylor Lab, specifically Ambima H. Buzhyason, Alexei J. Stuckel, and Darren C. Hawkins for being so nice to me and for your suggestions in different aspects of my research works. I have truly enjoyed every day because of the great people in Dr. Taylor's Lab.

Finally, I would like to thank my parents, my wife Fauzia Huq Nur, and other family members for always supporting me. I could not have done any of this work without their support, love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALL= Acute lymphoblastic leukemia

CD= Cluster of differentiation

CLP= Common lymphoid progenitor cell

CpG= CG dinucleotide

CpGI= CpG island

DMR= Differentially methylated region

DNMT= DNA methyltransferase

FDR= False discovery rate

HCB= Human umbilical cord blood

HSC= Hematopoietic stem cell

H3K4me1= Histone H3 lysine 4 monomethylation

H3K27ac= Histone H3 lysine 27 acetylation

LincRNA= Long intergenic non-coding RNA

MBD= Methyl CpG binding domains

MeCP2= Methyl CpG binding protein 2

MIRA-seq= Methylated CpG island recovery assay (MIRA) followed by next generation sequencing

miRNA= Micro-RNA

Pro-B= Progenitor B-cell

Pre-B= Precursor B-cell

ROI= Regions of interest

TF= Transcription factor

ABSTRACT

Acute lymphoblastic leukemia (ALL) is a fast-growing cancer of lymphoblasts, and is the most common cancer diagnosed in children under the age of 15. Our complete understanding of all mechanisms responsible for ALL induction is inadequate. Moreover, the heritable and potentially stable epigenetic changes occur at a much greater rate than DNA mutations in somatic cells. DNA methylation is an epigenetic modification that plays a significant role in hematopoiesis and malignant transformation. Therefore, the identification of altered DNA methylation on key regulatory regions of the genome is critical to gaining a better understanding of ALL pathogenesis. This dissertation identified the dynamic establishment of DNA methylation during normal B-cell development, and alterations of DNA methylation in the pathogenesis of ALL. First, a protocol to isolate the subsets of precursor B-cells from human umbilical cord blood (HCB) was developed. Using this protocol, we were able to isolate sufficient numbers of pro-B, pre-BI, pre-BII and naïve B-cells from a single HCB unit. This method can be adapted for any type of cell present in HCB at any stage of differentiation. Next, genome-wide DNA methylation profiles using the methylated CpG island recovery assay followed by next generation sequencing (MIRA-seq) were generated for each of the four subsets of precursor B-cells. We report for the first time that hypermethylated loci are associated with the transition of pro-B to pre-BI cells. Differentially methylated regions were identified and the majority of the loci were present within intronic and intergenic regions that harbor putative regulatory elements. The development of methylation profiles in normal precursor B-cells will aid in elucidating the role of altered DNA methylation in the pathogenesis of precursor B-cell related disorders including ALL. In order to identify epigenetic alterations in ALL, we next determined the differentially methylated regions (DMRs) between ALL and healthy precursor B-

cells. Further, whole genome transcriptome analysis was performed to determine the regulatory potential of the DMRs. Our studies identified ALL specific epigenetically deregulated genes, novel putative enhancers, and biological pathways that showed concurrent DNA methylation and changes in gene expression during malignant transformation. These altered epigenetic marks could be used as prospective biomarkers for ALL and/or as potential targets to reset the altered DNA methylation in order to modify gene expression which may enhance the present treatment protocol for ALL and eventually improve patient outcome.

CHAPTER 1: INTRODUCTION

Cancer is defined as a large group of diseases where abnormal cells divide uncontrollably, and which then invade adjacent parts of the body and spread to other organs in a process called metastasis. Cancer is also termed as malignant tumors or neoplasms, and can affect almost any part of the human body. The International Agency for Research on Cancer (IARC) reported that cancer is one of the foremost causes of morbidity and mortality worldwide, with about 14 million new cases, and 8.2 million cancer related deaths in 2012 (World cancer Report 2014). This report also projected that the annual cancer incidence will rise to 22 million in the next two decades. Cancer arises from a single cell and many factors both inside and outside of the body contribute to the progression of cancer. Certain genetic factors and environmental factors such as physical carcinogens (ultraviolet and ionizing radiation), chemical carcinogens (asbestos, arsenic, components of tobacco smoke, aflatoxin), and biological carcinogens (certain virus, bacteria and parasites) increase the risk of cancer progression. However, the exact combination of factors accountable for a person's specific cancer, and why one person gets cancer and another does not get cancer still remains vague.

Hematological malignancies originate in the bone marrow and lymph nodes. The most common hematological malignancies include leukemia, lymphoma and myeloma. Unfortunately, the etiology of most hematological malignancies is not yet known. The type of leukemia depends on the type of blood cell that has developed cancer. For example, acute lymphoblastic leukemia (ALL) is a cancer of the lymphoblasts that represents 75% of all

acute leukemia, and also accounts for 34% of all cancers in children from 0 to 14 years of age (Onciu 2009). In the USA, the American Cancer Society projected ~ 6,250 new cases and 1,450 deaths from ALL in 2015. Primarily cancer was assumed to be merely a consequence of genetic alterations on crucial tumor-suppressor genes and oncogenes that control cell differentiation and proliferation, DNA repair, and other homeostatic functions. However, recent studies demonstrated that epigenetic modifications such as DNA methylation, histone modifications, nucleosome positioning, and micro-RNA regulate gene expression and their alterations can contribute to the neoplastic phenotype. The common identified mechanisms involved in the initiation of ALL include chromosomal translocations, hyperdiploidy, and deviant expression of proto-oncogenes. Our complete understanding of all mechanisms responsible for ALL induction is inadequate. Therefore, the identification of epigenetic mechanisms such as alteration of DNA methylation within key regulatory regions of the genome is critical to gaining a better understanding of ALL pathogenesis. DNA methylation is a reversible and heritable epigenetic modification where a methyl group is added to a cytosine base at the carbon-5 position usually within a CpG dinucleotide context (Razin and Riggs 1980). This epigenetic modification is accountable for tissue specific gene expression and plays a significant role in hematopoiesis (Hodges et al. 2011), and malignant transformation (**Figure 1.1**) (Berdasco et al. 2010; Figueroa et al. 2013). Furthermore, leukemia and pediatric tumors harbor far fewer point mutations than other solid tumors (Vogelstein et al. 2013), and heritable and potentially stable epigenetic changes can occur at a much greater rate than DNA mutations in somatic cells. Therefore, these studies were aimed to identify the dynamic establishment of DNA methylation during

normal B-cell development, and to identify the role of altered DNA methylation in the pathogenesis of ALL.

Epigenetic variation among tissue types is responsible for cellular differentiation during normal human development (Song et al. 2005; Ohgane et al. 2008). Therefore, it is extremely important to use the same type of healthy tissues as a control to identify the specific altered mechanisms responsible for the disease transformation. Since ALL is the malignancy of precursor-B cells, initially we developed a protocol to isolate the four subsets of precursor B-cells, based on the expression of cell surface antigens, from human umbilical cord blood (HCB). The HCB is readily available through cord blood banks and is enriched for immature B and T cells (Ghia et al.1996). Our developed method employs an early depletion of all non-B cells from umbilical cord blood using column purification, and requires staining with only three antibodies against CD34, CD19 and CD45 which greatly reduces the complexity and cost of performing cell sorting experiments. The cells that were recovered represent 4 stages of B-cell differentiation: 1) $CD34^{+}$; $CD19^{+}$ (late pro-B and early pre-BI); 2) $CD34^{-}$; $CD19^{+}$; $CD45^{low}$ (late pre-BI); 3) $CD34^{-}$; $CD19^{+}$; $CD45^{med}$ (pre-BII); and 4) $CD34^{-}$; $CD19^{+}$; $CD45^{high}$ (immature B-cells). Sufficient numbers of precursor B-cells for each subset were isolated and used for downstream assays requiring high quality DNA and RNA.

B-cell differentiation includes several developmental stages, beginning with pluripotent hematopoietic stem cells (HSCs), followed by differentiation into distinctive common lymphoid progenitor cells (CLP), progenitor-B cells (pro-B), precursor-B cells (pre-BI and

pre-BII), immature B-cells (naïve B-cells) and finally mature B-cells. Numerous transcription factors (TFs) and epigenetic events are required for stage-specific gene expression during normal B-cell development and differentiation (**Figure 1.2**) (Bao and Cao 2015; van Zelm et al. 2005; Pérez-Vera et al. 2011; Hystad et al. 2007; Matthias et al. 2005). In addition, regulatory enhancer elements have been shown to be crucial for tissue and developmental stage-specific gene expression (Harmston et al. 2013). Any aberration from this coordinated and precise stage-specific gene expression could lead to disease conditions including ALL. It is well-known that DNA methylation is essential for HSC self-renewal, and therefore HSCs maintain an intermediate level of methylation. The loss of methylation is observed as HSCs differentiate into myeloid cells (Bröske et al. 2009), and the gain of methylation is observed as HSCs differentiate into lymphoid cells (Hodges et al. 2011). Furthermore, alteration of DNA methylation contributes to the mis-regulation of tissue specific gene expression, which may lead to the progression of many disease states including cancer. Therefore, genome-wide assessment is critical to establish the role of DNA methylation during normal B-cell development. In this study, genome-wide DNA methylation profiles were generated using the methylated CpG island recovery assay (Rauch and Pfeifer 2005) followed by next generation sequencing (MIRA-seq) in four subsets of B-cells (pro-B, pre-BI, pre-BII, and naïve B-cell), isolated from HCB. We observed an overall loss of methylation during pro-B to pre-BI transition and the majority of the differentially methylated regions (DMRs) lie within intronic and intergenic regions. We also identified putative novel regulatory regions within the DMRs between pro-B and pre-BI. However, no differential methylation was observed in the pre-BI to pre-BII transition or in the pre-BII to

naïve B-cell transition. The methylation patterns identified in normal B-cells will aid in revealing the role of altered DNA methylation in the pathogenesis of ALL.

Aberrant DNA methylation is linked with numerous human disease conditions and contributes to the induction and progression of various cancers (Jones and Baylin 2007; Feinberg and Tycko 2004). Additionally, several research works have identified promoter hypermethylation of tumor suppressor genes (TSG) that are involved in many cellular pathways including genome maintenance, cell cycle, or apoptosis and are associated with the development of cancer (Esteller 2008; Esteller 2007; Herman and Baylin 2003). On the other hand, DNA hypomethylation is considered as hallmark of carcinogenesis which leads to genomic instability, loss of genomic imprinting, activation of silenced genes and transposable elements (Jones et al. 2007; Berdasco and Esteller 2010). Remarkably, the hypermethylation and hypomethylation patterns are cancer type- and stage-specific (Ehrlich 2009). For example, disease specific promoter and gene body methylation patterns were identified in acute myeloid leukemia (AML) and the cytogenetic subtypes of AML were distinguishable by hypomethylation marks on interspersed repeat elements (Saied et al. 2012). Additionally, recurrent mutations in DNMT3A methyltransferase were discovered in ~ 20% AML patients with a normal karyotype (Ley et al. 2010; Yamashita et al. 2010; Yan et al., 2011) and in patients with myeloproliferative neoplasms (MPN) and myelodysplastic syndromes (MDS), and are associated with an increased probability of AML progression (Fong et al. 2014). Another study found that genome-wide promoter hypermethylation was associated with better overall and event free survival in T-cell acute lymphoblastic leukemia (T-ALL) than patients with hypomethylated T-ALL (Borssén et al. 2013). Recently,

microarray studies also detected differentially methylated CpG sites in patients with ALL (Chatterton et al. 2014; Nordlund et al. 2013). However, how the certain genomic regions (specifically CpG island) become hypermethylated in some cancer types but not in others remains unclear. But it is suggested that transcriptional inactivation of a particular gene by DNA methylation may provide selective growth advantage in particular cancer type.

Despite the main focus on CpG islands promoter hypermethylation in the pathogenesis of malignancies, the importance of differential DNA methylation on other genomic regions is gaining more attention (Liang et al. 2011; Ji et al. 2010). Regulatory enhancers reside within non-coding regions of the genome and work over long distances to promote cell/tissue type specific gene expression. Active enhancer regions are frequently associated with an absence of DNA methylation (Xu et al. 2007), and alterations of enhancer methylation have been shown to be associated with the mis-regulation of nearby genes in many cancers (Aran et al. 2013). Importantly, it has been demonstrated that differential methylation of enhancers displayed a higher correlation with gene expression than differential promoter methylation in multiple cancers (Aran et al. 2013).

Our studies identified differentially methylated regions (DMRs) throughout the genome in pre-B ALL and then correlated these findings with transcriptome data. MIRA-seq was used to identify the altered DNA methylation and RNA-seq was used to identify the differentially expressed genes in ALL. We found that 9,790 regions including 1,252 gene promoters were hypermethylated in ALL and 15,492 regions including 240 gene promoters were hypomethylated in ALL compared to healthy pre-B cells isolated from HCB. Furthermore, integration of promoter methylation and differential gene expression revealed that genes involved in transcription regulation, apoptosis, cell proliferation, GTPase

activation and protein complex assembly were mis-regulated in ALL. In addition to protein-coding genes, a large number of long intergenic non-coding RNA (lincRNA), micro-RNA (miRNA), and pseudogenes were differentially methylated and/or differentially expressed in ALL. Additionally, transposable elements within the human genome are often silenced by DNA methylation (Maksakova et al. 2008), and their transcriptional activation leads to insertional mutagenesis and chromosomal rearrangements in many cancers (Lee et al. 2012). A large number of identified intergenic DMRs in ALL were associated with transposable elements and repeat sequences. Although hypomethylated DMRs were prominent within the transposable elements, centromeric alpha satellite repeats were mostly hypermethylated in ALL which may block the binding of centromere specific histone protein CENP-A, and this could result in centromere inactivation. Genome-wide differential gene expression analysis revealed that 6,433 genes including genes involved in normal B-cell development, DNA methylation and demethylation, histone acetylation and deacetylation, and genes encoding histone proteins were significantly differentially expressed in ALL. We also identified numerous protein-coding parent genes that were associated with the mis-regulation of their pseudogenes. Finally, we have identified novel putative regulatory enhancer regions that become hypermethylated in ALL. Similarly, a large number of putative enhancers were highly transcribed (eRNA) in healthy pre-B cells but their expression was decreased in ALL. Remarkably, the transcriptional inactivation of regulatory enhancers was associated with the mis-regulation of neighboring genes, which are involved in many biological processes including immune system processes, immune response, leukocyte differentiation and activation, lymphocyte activation and more.

In summary, this dissertation encompasses the development of a method for the isolation of B-cell subsets from umbilical cord blood, and optimization of the MIRA-seq protocol for small amounts of starting genomic DNA. The B-cell isolation protocol can be adapted to any type of cell present in umbilical cord blood at any stage of differentiation. Genome-wide DNA methylation profiles were generated for four subsets of B-cells at different stages of development. The developed methylation patterns in normal B-cells will aid in elucidating the role of altered DNA methylation in the pathogenesis of precursor-B cell related diseases. Our studies also identified ALL specific epigenetically deregulated genes, novel putative enhancers, and biological pathways that showed concurrent DNA methylation and changes in gene expression during malignant transformation which have the potential to serve as prospective biomarkers for ALL. Furthermore, these discoveries may lead to the development of novel therapeutic agents that can restore or remove the aberrant DNA methylation and/or reinstate normal levels of gene expression advancing the present treatment protocol for ALL and eventually improve patient outcome. Further studies are required to reveal the phenotypic consequence of restoration of altered DNA methylation and gene expression in ALL.

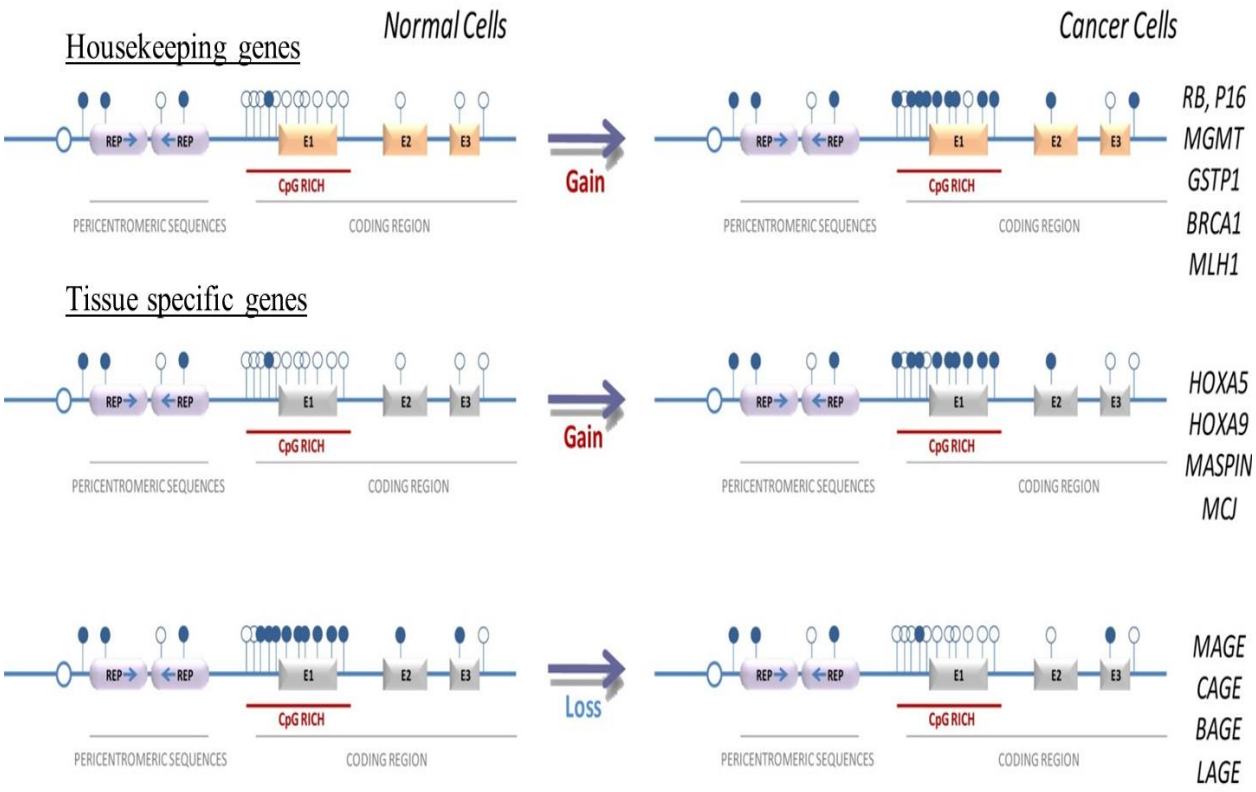


Figure 1.1: Distribution of CpG islands in promoters of housekeeping and tissue-specific genes. Promoters associated with CpG islands are found in all known housekeeping genes and half of all tissue-specific genes. Generally, housekeeping genes are unmethylated in normal cells, while tissue-specific genes may be unmethylated or methylated, depending on their requirement for lineage commitment. In a cancer cell, some of the housekeeping genes become aberrantly hypermethylated. Gains and losses of CpG hypermethylation can be observed for tissue-specific genes that are aberrantly expressed in cancer. CpG-poor regions are always found in tissue-specific genes, where expression is independent of CpG methylation status. Some examples of genes relevant to each of these circumstances are illustrated. Reproduced from Berdasco M and Esteller M. 2010; with permission, license number: 3670930270515

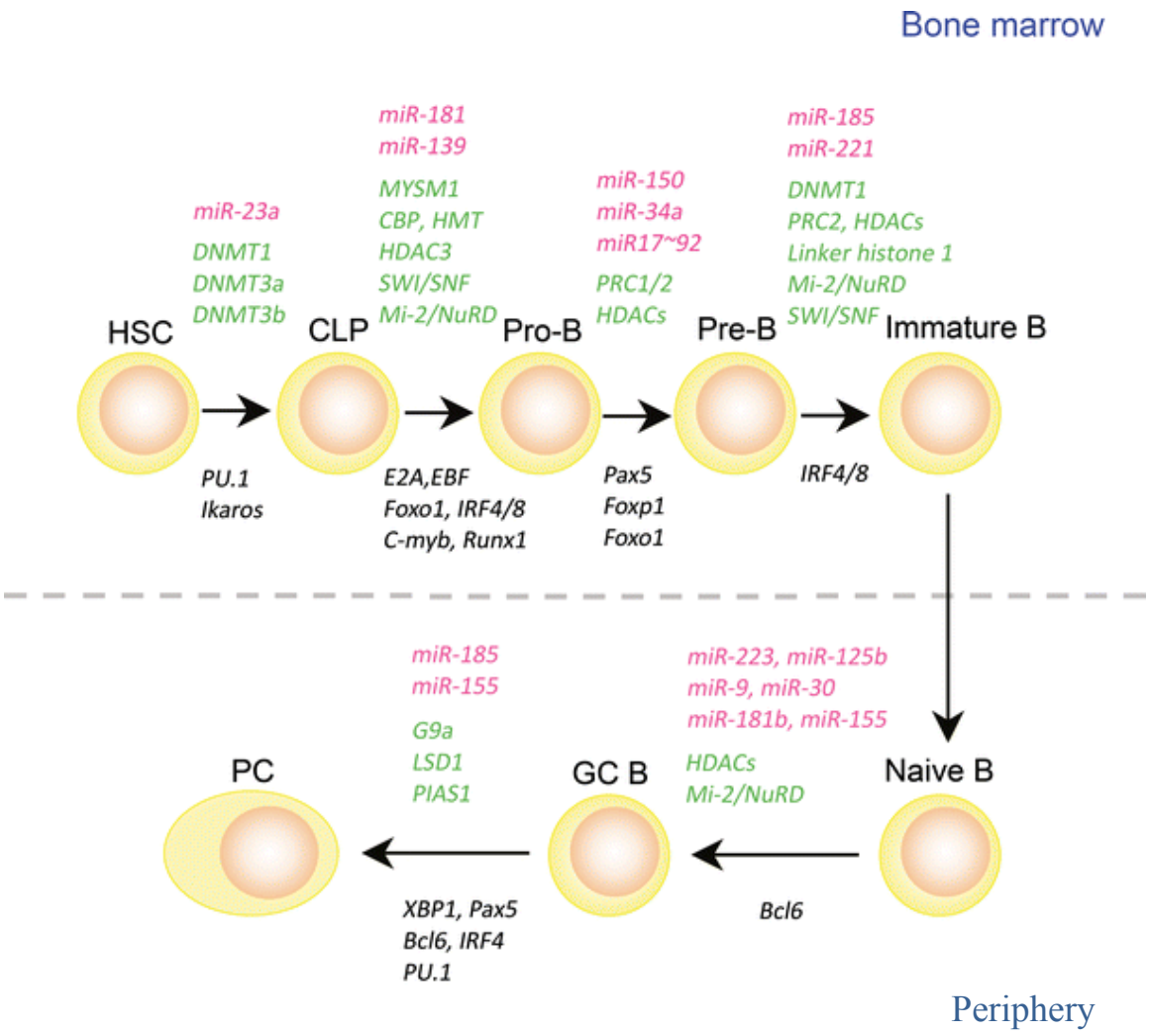


Figure 1.2: Transcriptional factors and epigenetic regulators in B cell development. Focusing on the main stages of B cell differentiation, this figure indicates the key transcriptional factors and epigenetic enzymes, chromatin remodeling complex, and miRNAs that could be involved in this process. The epigenetic enzymes and chromatin remodeling complex are in green. miRNAs are in red. Transcriptional factors are in black. Reproduced from Bao Y and Cao X. 2015; with permission, license number: 3670930535743.

CH1: REFERENCES

Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol 2013;14:R21.

Bao Y, Cao X. Epigenetic Control of B Cell Development and B-Cell-Related Immune Disorders. Clin Rev Allergy Immunol. 2015 Jun 12. [Epub ahead of print].

Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010 Nov 16;19(5):698-711.

Borssén M, Palmqvist L, Karrman K, Abrahamsson J, Behrendtz M, Heldrup J, Forestier E, Roos G, Degerman S. Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. PLoS One. 2013 Jun 6;8(6):e65373.

Bröske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, Kuhl C, Enns A, Prinz M, Jaenisch R, Nerlov C, Leutz A, Andrade-Navarro MA, Jacobsen SE, Rosenbauer F. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 2009; 41:1207-15.

Chatterton Z, Morenos L, Mechinaud F, Ashley DM, Craig JM, Sexton-Oates A, Halemba MS, Parkinson-Bates M, Ng J, Morrison D, Carroll WL, Saffery R, Wong NC. Epigenetic deregulation in pediatric acute lymphoblastic leukemia. Epigenetics. 2014 Mar;9(3):459-67.

Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009 Dec;1(2):239-59.

Esteller M. Epigenetics in cancer. N Engl J Med. 2008 Mar 13;358(11):1148-59.

Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007 Apr;8(4):286-98.

Feinberg AP, Tycko B. The history of cancer epigenetics. Nat Rev Cancer. 2004 Feb;4(2):143-53.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015 Mar 1;136(5):E359-86.

Figueroa ME, Chen SC, Andersson AK, Phillips LA, Li Y, Sotzen J, Kundu M, Downing JR, Melnick A, Mullighan CG. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. J Clin Invest. 2013 Jul 1;123(7):3099-111.

Fong CY, Morison J, Dawson MA. Epigenetics in the hematologic malignancies. Haematologica. 2014 Dec;99(12):1772-83.

Ghia P, ten Boekel E, Sanz E, de la Hera A, Rolink A, Melchers F. Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. J. Exp. Med. 1996;184:2217–2219.

Harmston N, Lenhard B. Chromatin and epigenetic features of long-range gene regulation. Nucleic Acids Res 2013;41:7185-99.

Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med. 2003 Nov 20;349(21):2042-54.

Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR, Smith AD, Hannon GJ. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell. 2011; 44:17-28.

Hystad ME, Myklebust JH, Bø TH, Sivertsen EA, Rian E, Forfang L, Munthe E, Rosenwald A, Chiorazzi M, Jonassen I, Staudt LM, Smeland EB. Characterization of early stages of human B cell development by gene expression profiling. J Immunol 2007; 179(6):3662-71.

Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 2010; 467:338-342.

Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007 Feb 23;128(4):683-92.

Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ; Cancer Genome Atlas Research Network. Landscape of somatic retrotransposition in human cancers. Science 2012; 337:967-971.

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK. DNMT3A mutations in acute myeloid leukemia. N Engl J Med. 2010 Dec 16;363(25):2424-33.

Liang P, Song F, Ghosh S, Morien E, Qin M, Mahmood S, Fujiwara K, Igarashi J, Nagase H, Held WA. Genome-wide survey reveals dynamic widespread tissue-specific changes in DNA methylation during development. BMC Genomics 2011; 12:231.

Maksakova IA, Mager DL, Reiss D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cell Mol Life Sci. 2008; 65:3329-3347.

Matthias P, Rolink AG. Transcriptional networks in developing and mature B cells. Nat Rev Immunol 2005; 5(6):497-508.

Nordlund J, Bäcklin CL, Wahlberg P, Busche S, Berglund EC, Eloranta ML, Flaegstad T, Forestier E, Frost BM, Harila-Saari A, Heyman M, Jónsson OG, Larsson R, Palle J, Rönnblom L, Schmiegelow K, Sinnett D, Söderhäll S, Pastinen T, Gustafsson MG, Lönnerholm G, Syvänen AC. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome Biol. 2013 Sep 24;14(9):r105.

Ohgane J, Yagi S, Shiota K. Epigenetics: The DNA methylation profile of tissue-dependent and differentially methylated regions in cells. Placenta. 2008;29(S):29–35.

Onciu M. Acute lymphoblastic leukemia. Hematol Oncol Clin North Am. 2009 Aug;23(4):655-74.

Pérez-Vera P, Reyes-León A, Fuentes-Pananá EM. Signaling proteins and transcription factors in normal and malignant early B cell development. Bone Marrow Res 2011; 2011:502751.

Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab Invest 2005; 85(9):1172-8.

Razin A, Riggs AD. DNA methylation and gene function. Science. 1980 Nov 7;210(4470):604-10.

Saied MH, Marzec J, Khalid S, Smith P, Down TA, Rakyan VK, Molloy G, Raghavan M, Debernardi S, Young BD. Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. PLoS One. 2012;7(3):e33213.

Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, Held WA. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A. 2005 Mar 1;102(9):3336-41.

Stewart BW, Wild CP. World Cancer Report 2014. IARC Nonserial Publication.

van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ, Lankester AC, Révész T, Staal FJ, van Dongen JJ. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. J Immunol 2005; 175(9):5912-22.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013 Mar 29;339(6127):1546-58.

World Cancer Report 2014. IARC Nonserial Publication. Stewart BW, Wild CP.

Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, Zaret KS, Weissman IL, Smale ST. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. Proc Natl Acad Sci U S A. 2007; 104:12377-12382.

Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL, Ueno T, Soda M, Hamada T, Haruta H, Takada S, Miyazaki Y, Kiyoi H, Ito E, Naoe T, Tomonaga M, Toyota M, Tajima S, Iwama A, Mano H. Array-based genomic resequencing of human leukemia. Oncogene. 2010 Jun 24;29(25):3723-31.

Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet. 2011 Mar 13;43(4):309-15.

CHAPTER 2: BACKGROUND

2.1 ACUTE LYMPHOBLASTIC LEUKEMIA (ALL)

2.1.1 ALL induction mechanisms

Acute lymphoblastic leukemia (ALL) is the most common cancer diagnosed in children under the age of 15 with a peak incidence between 2 and 5 years of age. The risk then arises gradually again after age 50. ALL is characterized by compromised early lymphoid development and can be classified as T or B cell ALL. ALLs represent 75% of all acute leukemias in children that account for 34% of all cancers (Onciu 2009). The American Cancer Society estimated ~6,250 new cases of ALL and 1,450 deaths from ALL in 2015 in the USA. It has a slender male prevalence in all age groups with an increase incidence among white children compared to black children (NCI report). ALL is a primary malignancy with only sporadic circumstances arise as a secondary condition (Shivakumar et al., 2008). A range of genetic and environmental factors are associated with the development of ALL. Patients with Down syndrome, bloom syndrome, ataxia-telangiectasia, and neurofibromatosis have increased frequency of ALL development (Spector et al., 2006) and only a few cases are connected with inheritance (<5%) (Pui et al., 2008). Additionally, in utero X-ray exposure, increased birth weight, and postnatal exposure to radiation has been associated with an increased risk of ALL (Spector et al., 2006). Leukemia studies in monozygotic twins and retrospective documentation of hyperdiploidy, leukemia specific fusion genes, rearrangement of immunoglobulin genes in archived neonatal blood spots postulated a prenatal origin of some childhood leukemia (Hong et al., 2008; Greaves 2006; Maia et al., 2003; Wiemels et al. 1999). However, putative leukemic clones with ETV6-RUNX1 fusion gene (TEL-AML1) were found in neonatal cord blood samples of 1% of newborn babies,

that is a 100 times higher frequency than the occurrence of ALL associated with *ETV6-RUNX1* fusion gene later in childhood (Hong et al. 2008; Mori et al. 2002). The variable clinical outcomes and incubation period of leukemia, and the twin concordance rate is ~ 10% with *ETV6-RUNX1* translocation suggesting that the genetic changes alone are usually not sufficient, and additional postnatal events are required for the complete leukemic transformation (Greaves and Wiemels 2003).

2.1.2 ALL classifications

The World Health Organization (WHO) classification of pre-B ALL is based on recurring cytogenetic abnormalities as mentioned below.

Classification of acute lymphoblastic leukemia according to the WHO classification (modified from Vardiman et al. 2009).

B lymphoblastic leukemia/lymphoma
B lymphoblastic leukemia/lymphoma, NOS (not otherwise specified)
B lymphoblastic leukemia/lymphoma with recurrent genetic abnormalities
B lymphoblastic leukemia/lymphoma with t(9;22)(q34;q11.2); BCR-ABL 1
B lymphoblastic leukemia/lymphoma with t(v;11q23); MLL rearranged
B lymphoblastic leukemia/lymphoma with t(12;21)(p13;q22); TEL-AML1 (ETV6-RUNX1)
B lymphoblastic leukemia/lymphoma with hyperdiploidy
B lymphoblastic leukemia/lymphoma with hypodiploidy
B lymphoblastic leukemia/lymphoma with t(5;14)(q31;q32); IL3-IGH
B lymphoblastic leukemia/lymphoma with t(1;19)(q23;p13.3); TCF3-PBX1

Hyperdiploid ALL with a modal number of 51-65 chromosomes is the most common subtypes of pediatric ALL (27-29%). Hyperdiploid ALL has distinct biological features with low number of leukemia cells at diagnosis and good response to therapy as these leukemia cells have a marked propensity to undergo apoptosis (Ito et al. 1999). Mutations in B-cell development genes are rarely present in this type of ALL (13% of patients) (Mullighan et al. 2007). *ETV6-RUNX1* (*TEL-AML1*) is another common cytogenetic subtype of pediatric ALL (22-25%) with a favorable prognosis (Loh and Rubnitz 2002; Shurtleff et al. 1995). Monoallelic deletion of the *PAX5* gene that encodes the B-cell specific transcription factor PAX5 has been found in 28% of patients (Mullighan et al. 2007). Another pre-B cell ALL with *TCF3-PBX1* fusion transcripts represents 25% of all pre-B ALL and is linked with a higher incidence of central nervous system (CNS) relapse (Jeha et al. 2009). Philadelphia chromosome-positive (*BCR-ABL1*) ALL accounts for 2-3% of all childhood ALL and usually develops in older age children (>10 years). This subtype of ALL associated with increased frequency of CNS leukemia, high leukocyte counts, L1 blast morphology, and a poor prognosis (Aricò et al. 2000; Uckun et al. 1998). The treatment outcome of patients with this abnormality is greatly improved when a tyrosine kinase inhibitor is combined with the intensive chemotherapy. Philadelphia chromosome-positive ALL are associated with an increased incidence of mutation (66%) in genes involved in B-cell development (Mullighan et al. 2007), and *IKZF1* gene that encodes the transcription factor Ikaros is deleted in 83.7% of patients (Mullighan et al. 2008*). AF4-MLL* fusion transcript is the most common MLL gene rearrangement.

Previously, ALL was classified according to the French-American-British (FAB) classification system which is based on morphological features of tumor cells, as seen on Wright-Giemsa-stained smears (Onciu 2009; Redaelli et al. 2005) and this morphological

classification of ALL had no clinical or prognostic significance. FAB classification outlined three types of ALL, designated as L1, L2, L3.

1. ALL-L1: ALL cells are small to intermediate in size with condensed nuclear chromatin, scanty cytoplasm, and absent or indistinct nucleoli. This is the most common subtype of ALL (80%).
2. ALL-L2: ALL cells are larger with modest amount of basophilic cytoplasm, finely dispersed nuclear chromatin, and noticeable nucleoli but still no vacuoles (17% cases).
3. ALL-L3: large blasts with profuse deeply basophilic and occasional vacuolated cytoplasm loutishly clumped nuclear chromatin, and inconsistently prominent nucleoli. L3 subtype contains a mature B-cell and is classified as Burkitt lymphomas (3% of cases).

2.1.3 ALL treatments and prognosis

2.1.3.1 Diagnosis:

1. *Physical examination and history:* Acute beginning of clinical symptoms is most common in ALL; although in some cases it may progress insidiously over several months. The clinical manifestations correlate with the leukemic cell load and the amount of marrow replacement that lead to cytopenias. The most common symptoms comprise fatigue, shortness of breath and lethargy due to anemia, fever due to either from cytokines released by leukemic cell or secondary infection as a result of neutropenia, bleeding diathesis related to thrombocytopenia, pain in joint and bone due to accumulation of leukemic cell in bone. Central nervous system (CNS)

complication such as headache, nausea, vomiting and cranial nerve dysfunction can also be manifest (Redaelli et al., 2005). Painless lumps may also observed in the underarm, neck, stomach, or groin due to swelling of lymphnodes

2. *Complete blood counts (CBC) with differential*: Disruptions in blood cell counts are common in ALL. Total white blood cell (WBC) counts can be very low or can be as high as >100,000/µl. As a result of excessive ALL colonization in bone marrow, patients may also have low levels of erythrocytes, neutrophils and platelets.
3. *Blood chemistry studies*: Laboratory finding include increase level of serum lactate dehydrogenase, elevated creatinine and uric acid, reduce level of circulatory immunoglobulins and disruption of calcium and phosphate metabolism.
4. *Bone marrow aspiration and biopsy*: Bone marrow aspiration is perform for the morphological, immunophenotyping, cytochemical, cytogenetic and molecular analysis of cancer cells. Differential diagnoses are performed to rule out other hematological disorders.
5. *Immunophenotype*: Immunophenotyping of leukemic lymphoblast are critical for the diagnosis of ALL. B-cell ALL expressed a variety of B-cell specific antigens including CD19, CD10, CD20, CD22, PAX-5 and cytoplasmic expression of CD79a. CD20 is a marker of mature B-cell and may weakly express in leukemic lymphoblast. Another surface antigen, CD45 show faint expression in most of the B-cell ALLs. CD34 and TdT are early progenitor B-cell markers and are often expressed by the leukemia blast. Myeloid associated antigen can be expressed in ~ 50% of acute lymphoblastic leukemia cases. Although deviant immunophenotype has no prognostic implications, they can be used to discriminate the malignantly transformed leukemic

cell from the normal progenitor cells, thus allowing detection of negligible residual leukemia.

6. *Genotyping:* Chromosomal analysis is an essential component of primary diagnosis of ALL. Fluorescence in-situ hybridization, RT-PCR, flow cytometry are using to detect specific fusion transcripts, specific chromosome with therapeutic and prognostic consequences.

Staging of ALL is performed using morphological, immunological, and genetic methods at the time of diagnosis. The subtype classification is very important for the establishment of individual patient's risk profile, to determine the treatment protocol and prognosis.

2.1.3.2 Treatment:

Chemotherapy: Patients with different subtypes of ALL may have different responses and outlooks to treatment. In order to avoid the development of resistance, newly diagnosed ALL treatment generally comprises four major components (Cooper and Brown 2015). In first step, chemotherapy is lasted for 4 to 6 weeks for the induction of remission. The aim of induction therapy is to accomplish remission by killing as many leukemic cells as possible, returning blood counts to normal and eliminating the signs of diseases for a long time. About 95% of patients achieved complete remission at the end of first step. Half of the patients those failed to induce complete remission succumb to treatment related mortality. Allogeneic bone marrow transplant is generally performed for the remaining induction failure patients. Vincristine, corticosteroids (prednisone/ dexamethasone), asparaginase, and anthracycline drugs are generally used for the remission induction. Anthracycline is often excluded for the treatment of patients from low-risk

group in order to reduce toxicity. Although dexamethasone improved central nervous system (CNS) penetration, it also increases the incidence of avascular necrosis, infection and decrease linear growth. Many asparaginase derivatives such as PEG asparaginase and Erwinia asparaginase are used to improve drug's half-life and for the reduction of immunogenicity. After complete remission, sub-microscopic residual disease is eliminated in consolidation regimens, which last ~ 6 to 9 months depending on the risk group involvement. In order to maximize synergistic effect and lessen drug resistance, different chemotherapeutic agents such as mercaptopurine, methotrexate, thioguanine, etoposide, cyclophosphamide and cytarabine are combined in consolidation phase of treatment. Maintenance chemotherapy is the longest and final stage of ALL treatment aimed to lower the risk of relapse. This phase of treatment lasts for at least 2 years and are typically comprises antimetabolite therapy with mercaptopurine and methotrexate. The last component of ALL treatment focused on clinical CNS diseases at diagnosis, and prevention for patients with subclinical disease in order to prevent CNS relapse. Several methods including direct intrathecal administration of chemotherapeutics, systemic administration of drugs that can penetrate blood - brain barrier and cranial radiation are used to eradicate the disease from CNS. Intrathecal methotrxare or triple intrathecal drugs (a combination of intrathecal methotrexate, cytarabine and hydrocortisone) are used for intrathecal chemotherapy. For systemic chemotherapy, dexamethasone, cytarabine, high-dose methotrexate and asparaginase are administrated to eliminate CNS effects. The use of cranial radiation is declining due to high risk of toxicity (intellectual disability, second malignant tumors), and only use in patients with highest risk of CNS relapse. In addition, hematopoietic stem cell transplant is considered only for the patients with utmost risk of treatment failure and/or relapse associated with hypodiploidy or induction failure.

Possible side effects of chemotherapy: Chemotherapeutic drugs attack rapidly dividing cells. In addition to rapidly dividing cancer cell, other quickly proliferating normal cells such as the lining of mouth and intestine, hair follicles, and new blood cells in bone marrow are also affected by chemotherapy. Although the side effects depends on type and dose of chemotherapeutic agents (Redaelli et al. 2005), common side effects may include mouth sore, hair loss, nausea/vomiting, loss of appetite, diarrhea, fatigue due to low number of red blood cells, increased risk of infection due to reduce number of white blood cell, and weakness in hands or feet, numbness and tingling from nerve damage. Most of the side effects are usually short-term and go away when treatment is completed and symptomatic drugs are often used to lessen or prevent these side effects.

Targeted Therapy: Recently developed drugs targeted specific parts of cancer cells and work differently than standard chemotherapeutic drugs. These targeted drugs have less side effect and some of this drugs are used in certain type of acute lymphoblastic leukemia. About 25% of ALL patients have *Philadelphia chromosome* and formed a new fusion gene termed as BCR-ABL. Tyrosine kinase inhibitors such as imatinib (Gleevec®), dasatinib (Sprycel®), nilotinib (Tasigna®), ponatinib (Iclusig®), and bosutinib (Bosulif®) are found to be useful in ALL patients with Philadelphia chromosome. These drugs are received daily as pills format, and their common side effects are generally mild and comprise nausea, diarrhea, fatique, muscle pain, and skin rashes. Others side effects include swelling around the eyes, in the hands or feet, lower red blood cell and platelets counts at the beginning of treatments. Monoclonal antibodies are using recently to target specific antigen on the surface of certain lymphocytes. For example,

monoclonal antibodies such as alemtuzumab (Campath) and rituximab (Rituxan) have been used to treat other blood diseases and now testing for use in ALL.

Researchers are identifying new genetic and epigenetic alteration in ALL those could be used as potential target for therapeutic agents with minimum side effects. Whole genome transcriptome profiles help to discover and classify different cancers by looking the pattern of several gene expressions. Many research works are in advancement to discover new chemotherapeutic agents with minimal unwanted side effects. The development of resistance to standard drugs limited the effectiveness of chemotherapy. Therefore, researchers are seeking approach to prevent or reverse drugs resistance by administrating others agents along with chemotherapy.

2.1.3.3 Prognosis

The prognosis of ALL is mainly depended on risk profile of the patients. Low risk childhood ALL (age between 1 and 10 years, being female, WBC counts <50,000/µl, no spread to other organs outside bone marrow, fast response to treatment) patients have an event-free survival (EFS) of > 85% (**Figure 2.1**). The high risk ALL is related to age younger than one year or older than 50 years, WBC counts > 50,000/µl, poor response to treatment and spread diseases outside the bone marrow has less EFS. The EFS is also varies according to the biological subtype of leukemia (**Figure 2.2**). Mostly, patients with hyperdiploidy (> 50 chromosomes), and TEL/AML1 gene fusion have a good prognosis where patients with MLL rearrangements, BCR-ABL gene fusion, and hypodioloidy are associated with bad prognosis. In general, about 80-90% of patients with ALL have complete remission at certain point during treatments. This disease

are relapse in about half of these patient, therefore the overall cure rate is ~ 40%. Younger patients tend to have higher cure rate than older patients

2.2 NORMAL B-CELL DEVELOPMENT

B-lymphocytes (B-cell) develop in the bone marrow from hematopoietic stem cell (HSC). B-cell development comprises several developmental stages beginning with pluripotent hematopoietic stem cells (HSCs). HSC asymmetrically divides into one stem cell and one differentiating cell. The differentiating cell gives rise to progenitor cells that undertake lineage commitment. Multipotent progenitors (MPP) lose their reconstituting ability and segregate chronologically into early lymphoid progenitors (ELP), followed by common lymphoid progenitors (CLP), progenitor-B cells (pro-B), precursor-B cells (pre-BI and pre-BII), immature B-cells (naïve B-cells) and finally mature B-cells. Each developmental stage has diverse biological features that are regulated by differential gene expression (Hystad et al. 2007; van Zelm et al. 2005). Alteration in stage specific gene expression may lead to the progression of many disease states including cancer. Individual stages of B-cell development can be identified by the array of intracellular and surface proteins also known as cluster of designation (CD). The pro-B cell express surface CD10, CD19, CD24, CD34, CD45, cytoplasmic CD22, CD79a, CD79b, and nuclear terminal deoxynucleotidyl transferase (TdT) (Behm 2006). The expression level of CD45 is initially weak, and then increases with cell maturation. Early B-cell complete their immunoglobulin heavy (*H-chain*) and light chain (*L-chain*) rearrangement in bone marrow before migrating to the secondary lymphoid tissues (spleen and lymph nodes) as an immature B-cell. The rearrangement of the *H-chain* ($V_{H,}$ $D_{H,}$ J_H) and *L-chain* (V_L-J_L) gene segments create a B-cell stock expressing antibodies capable of distinguishing more than 5×10^{13} different type of

antigens (Pieper et al. 2013). The activation of recombination-activating gene (RAG)-1, RAG-2 promotes the VDJ recombination in the immunoglobulin heavy (IgH) chain locus at pro-B cell stage. Following the H-chain gene rearrangement, pro-B cell differentiated into pre-B stage and undergoes 1 or 2 cell division tailed by rearrangement of gene segments encoding the κ and λ chains (Pieper et al. 2013). The expression of pre-B cell antigen receptor (pre-BCR) complex results in termination of further rearrangement by preventing the expression and activity of enzymes catalyzing the H-chain rearrangements and initiates the rearrangement of L-chain genes. Lymphoid cells that fail to produce a pre-BCR endure apoptosis. TdT, CD34, and CD10 markers disappear at the end of pre-B stage. Rearrangement and transcription of immunoglobulin light chain genes lead to the establishment of complete immunoglobulin molecules and to the development of intermediate pre-B cells into immune-competent B cell. This early immune-competent B cell (immature B cell) expressed surface IgM BCR, which is constitute of two heavy and two light chains connected by disulfide bonds. Immature B cell may co-express the pre-BCRs and IgM BCR for a short time. The multi-protein structure BCR encompassing an antigen binding membrane immunoglobulin molecule non-covalently connected with signal transducing CD79a and CD79b heterodimers. Antigenic stimulation of IgM molecules induces conformational changes in CD79a/CD79b heterodimers that function as signal transduction molecules through their cytoplasmic domains. Individual B cell fail to generate efficient IgM BCRs embark apoptosis in the bone marrow. The surface IgM expression complemented with increase expression of CD20. The immature B cell commonly referred to as a naïve B cell migrate from bone marrow to the follicular regions or B-cell zones of lymph nodes, Peyer's patches, spleen and other secondary lymphoid tissues where they differentiated into memory B-cell, marginal zone B-cell, mantle B-cell and plasma cell. Somatic hypermutation of the

immunoglobulin genes causes BCR diversification of antigen-primed B cell in the germinal center. Further rearrangement of immunoglobulin heavy chain genes lead to the production of IgD, IgA, or IgG.

Transcriptional regulation of B cell development: Numerous transcription factors (TFs) are known to be responsible for B cell differentiation (Santos and Borghesi 2011; Pérez-Vera et al. 2011; van Zelm et al. 2005; Matthias and Rolink 2005). The lymphoid versus myeloid fate choice is regulated by transcription factor PU.1, Ikaros and E2A whereas B cell specification and commitment is governed by EBF and PAX5 transcription factor (Santos and Borghesi 2011). Ikaros functions as a key overseer for the HSC progression to the lymphoid lineages by suppressing stem cell associated genes and inducing lymphoid specific genes (Ng et al. 2009). The expression of cytokine receptor FLT3, and recombination enzyme, RAG1/2 are regulated by Ikaros during lymphoid/myeloid-primed multipotent progenitor (LMPP) and early precursor B cell development. GFI-1 promotes B cell fate choice at the expense of myeloid progeny whereas the higher concentration of PU.1 (purine box factor-1) favors macrophage fate. Ikaros antagonizes PU.1 and promotes GFI-1 expression in MPPs to direct the myeloid versus lymphoid fate (Spooner et al. 2009). Transcription factor E2A is required for effective HSC self-renewal and also act as a regulator of lymphoid specification by promoting the expression of lymphoid associated genes and preventing the expression of HSC associated and non-lymphoid genes (Dias et al. 2008). Transcription factor EBF and PAX5 directed lymphoid precursors to the B cell fate. Transcriptional activation of EBF is regulated by interferon-regulatory factor 8 (IRF8) and a reduced number of pre-pro B cell and an increased number of myeloid cell were found in IRF8 knockout mice (Wang et al., 2008). PAX5 conserve B cell fate (Cobaleda et al., 2007) and

the B-lymphocyte development arrested at the transition of pre-B I to pre-B II cells in Pax5 deficient mice (Rolink et al., 2000).

2.3 EPIGENETICS

The term "epigenetics" first coined by Conrad Waddington in the 1940s as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Waddington 1942). Although the word 'epigenetics' literally means 'above the genetics', it is currently used to refer the study of mitotically or meiotically heritable changes in gene expression that do not involve a change in DNA sequence (Riggs et al. 1996). Multicellular organisms developed from a single fertilized egg, following differentiation into at least 200 differentiated cell types in mammals, and the vast majority of cells contain identical DNA sequences (with exception in lymphoid cells). Therefore, cellular differentiation in multicellular organisms may ponder as epigenetic phenomenon. Epigenetic mechanisms yield a heritable phenotype that does not severely depend on DNA sequence (Ohgane et al. 2008). Epigenetic modifications comprise any change to the both DNA and chromatin. The most studied epigenetic modification included DNA methylation, histone modifications, and ncRNA mediated gene expression. A large number of epigenetic regulators mediate those modifications. For example, DNA methyltransferases (DNMTs) add methyl-groups to the cytosine base of DNA, whereas, histone acetyltransferase (HATs), histone deacetylases (HDACs), histone methyltransferases (HMTs), and histone demethylases (HDMTs) modifies the histone tail in the nucleosome affecting the chromatin structures. Epigenetic modifications interact with each other to regulate gene expression. Both DNA methylation and histone modifications functions as mechanisms for monitoring cellular differentiation, permitting only tissue-specific and

housekeeping genes expression in somatic differentiated cells. The alterations of epigenetics modifications have the potentiality to inappropriate activation/reactivation or silencing of tissue-specific genes which could lead to human diseases.

2.4 DNA METHYLATION

DNA methylation is an epigenetic modification where a methyl group is added to a cytosine base at the carbon-5 position primarily in a CpG dinucleotide context (Razin and Riggs 1980). DNA methyltransferases (DNMTs) enzymes are required for the establishment and maintenance of DNA methylation. DNA methylation is a key epigenetic modification involved in chromatin organization, regulation of gene expression (Miranda and Jones 2007; Lande-Diner et al. 2007), genomic imprinting (Delaval and Feil 2004; Sha 2008), X chromosome inactivation (Heard 2004; Yen et al. 2007), as well as maintenance of genomic integrity by protecting against transposon and endogenous retroviruses (Howard et al. 2008; Yoder et al. 1997). DNA methylation is also involved in cellular reprogramming (Yang et al. 2007; Reik 2007; Hemberger et al. 2009), immune system development (Fitzpatrick and Wilson 2003), brain function and behavior (Sweatt 2009; McCarthy et al. 2009).

DNA methylation regulates gene expression as methylated cytosines (5Me-C) attract methyl-CpG-binding domain proteins (MBD) and promotes chromatin condensation into a transcriptionally repressive conformation (Berger 2007; Newell-Price et. al. 2000; Bird 1992). MBD protein family deacetylates histone protein by recruiting histone deacetylase enzyme (chromatin remodeling enzyme) to the methylated DNA. The histone deacetylation leads to the condensation of the chromatin causing silencing of the neighboring genes. Alternatively, DNA methylation can directly affect gene expression by interfering with the binding of transcription

factor that are delicate to methylated DNA. Several methyl-CpG-binding proteins such as MBD1, MBD2, MBD3, MBD4, MECP2, and Kaiso have been identified in mammals. Different methyl-CpG binding proteins recruit diverse transcription-regulatory complex and chromatin-remodeling proteins to the target methylated DNA. MECP2 bind methylated CpG both *in vivo* and *in vitro,* which contain a methyl-CpG-binding domain at the amino terminus and a transcription repression domain at the middle of MECP2 protein. All MBD proteins except MBD3 preferentially binds methylated DNA. MBD3 is an essential component of chromatin remodeling protein complex Mi-2-NuRD (Zhang et al. 1999; Wade et al. 1999). MBD1, MBD2 and MECP2 act as a transcription repressor whereas MBD4 is a DNA glycosylase and is required for DNA mismatch repair (Hendrich et al. 1999). Another binding protein Kaiso does not have a MBD domain but binds methylated CGCG through its zinc finger motif (Prokhortchouk et al. 2001).

DNA methylation depends upon the obtainability of methyl groups. The S-adenosylmethionine (SAM) is a major cellular methyl donor for DNA, which is derived from methionine. After losing methyl group, SAM is converted to S-adenosylhomocysteine (SAH), which is then converted to homocysteine by removal of adenosine molecule. Homocysteine can either be irreversibly removed by cystathionine-β-synthase enzyme or converted back to methionine by the addition of a methyl group. Methionine synthase is a vitamin B12 dependent enzyme that catalyzes the transfer of a methyl group from methyl-tetra-hydro-folate (MTHF) to homocysteine. In addition to MTHF, betaine can also donate methyl group to homocysteine. Therefore, folate and vitamin B12 deficiency are associated with the perturbation of DNA methylation. Folate deficient diets induce hypomethylation within the tumor suppressor gene in rat (Kim et al. 1997) and the plasma folic acid concentration is found to be associated with the

DNA methylation in human gastric cancer (Fang et al. 1997). Another mice model study demonstrated that post-weaning diet lacking folic acid, methionine, vitamin B12, and choline lead to DNA hypomethylation and loss of imprinting of Igf2 (insulin-like growth factor) (Waterland et al. 2006). DNA hypomethylation is considered an early event in cervical carcinogenesis and is strongly correlated with the cervical tissue folate and serum folate level (Fowler et al. 1998). Additionally, DNA hypomethylation was reversible in patients with colorectal adenoma by using physiological intakes of folic acid (Pufulete et al. 2005). Furthermore, genetic variant in methyl-group metabolism enzymes (methylene-tetrahydrofolate reductase, methionine synthase, and cystathionine beta-synthase) also associated with the aberrant methylation level in normal tissue and human primary solid tumors (Paz et al. 2002). Evidence also suggested that environmental exposures to chemical agents could be directly affects DNA methylation (Jirtle et al. 2007). Exposure to arsenic induces malignant transformation of human prostate epithelial cell and is associated with genomic DNA hypomethylation (Benbrahim-Tallaa et al. 2005). In mouse study, neonatal exposure to diethylstilbestrol increases the expression of proto-oncogene c-fos, and is associated with abnormal methylation pattern (Li et al. 2003). Another study demonstrates the transgenerational effect on radiation exposure. This study found that parental exposure to radiation significantly decreased global cytosine methylation along with down-regulation of DNMT1, DNMT3A, and DNMT3B enzyme in the thymus tissue of offspring (Koturbash et al. 2006).

2.4.1 DNA METHYLTRANSFERASES

Three active DNA methyltransferase (DNMT) enzymes, namely DNMT1, DNMT3A, and DNMT3B have been identified in the mammals and their catalytic domain contain highly

conserved and characteristic cytosine-methyltransferase motif (Li 2002). DNMT3L is a DNMT related protein with lack of catalytic activity. Mammalian DNMTs contain two parts: a large multidomain regulatory N-terminal part of variable size and a catalytic C terminal part (Jurkowska et al. 2011). The regulatory N-terminal part of DNMTs directed their nuclear localization and facilitates the interaction with DNA, chromatin and other proteins. All DNMTs flip their mark cytosine base out of the DNA helix and hide in a hydrophobic pocket of their active center. The catalytic cysteine residue of DNMTs causes nucleophilic attack on the carbon-6 position of cytosine leading to the formation of a covalent bond between the substrate base and enzyme. Therefore, negative charge density increased at the carbon-5 atom of the cytosine resulting addition of methyl group. After addition of methyl group, deprotonation at the carbon-5 leads to the cleavage of the covalent bond between the DNA and enzyme. Therefore, DNMTs is release from the DNA molecules.

DNMT1 is a maintenance methyltransferase and causes the restoration of DNA methylation pattern by recognizing hemi-methylated CpG sites following DNA replication. DNMT1 mostly localized to the DNA replication fork and its expression is tightly regulated during cell cycle and highest expression occurs during S phase (Chen and Li 2004). DNMT1 safeguard the inheritance of the DNA methylation pattern through cell division. Therefore, DNMT1 expression is continued after development. A full-length form of transcript variant that contain 1,616 amino acids and an oocyte-specific transcript that lacks the N-terminal 118 amino acids (DNMT1o) were identified. Both transcript variant are enzymatically active (Chen and Li 2004). DNMT1 can also catalyze *de novo* DNA methylation but its affinity for hemi-methylated DNA is much higher than unmethylated DNA. A mutation in the DNMT1 gene caused abnormal development, and embryonic death in mice (Li et. al. 1992), indicating the crucial role of

DNMT1 in early development. Furthermore, loss of DNMT1 function leads to the alteration of X chromosome inactivation (Sado et al. 2000) and also cause the losing of genomic imprinting (Howell et al., 2001). Inactivation of DNMT1 in human colorectal carcinoma cell lines lead to loss of cellular proliferation, severe mitotic defects tailed by cell death (Chen et al. 2007). In mouse fibroblast, deletion of DNMT1 causes demethylation and fibroblasts undergo a p53-dependent cell death (Jackson-Grusby et al. 2001). The reduce expression level of DNMT1 increase the susceptibility of tumor formation in mice by promoting chromosomal instability (Gaudet et al. 2003). Above mentioned gene targeting studies highlighting the importance of DNMT1 in normal development, cell proliferation and survival, as well as misregulation of DNMT1 linked with the development of malignancies.

The DNMT3 family encompasses DNMT3A, DNMT3B, and DNMT3L members. DNMT3A and DNMT3B are accountable for the establishment of *de novo* DNA methylation pattern in early mammalian development and in gametogenesis. Both DNMT3A and DNMT3B are localized within the methylated portion of chromatin including pericentromeric heterochromatin (Chen et al. 2004; Bachman et al. 2001) and mitotic chromosome. The PWWP domain of DNMT3A and DNMT3B is responsible for the tarheting of chromatin (Ge et al. 2004). The DNMT3A gene encodes at least two enzymatically active different proteins and each has distinct localization in the nucleus. The DNMT3B encodes two active and three inactive isoform (Delpu et al. 2013). DNMT3L is catalytically inactive, bind very weakly to DNA, and act as a regulatory factor. DNMT3L directly interact with the catalytic domain of DNMT3A and DNMT3B and stimulates the action of both enzymes in vitro (Kareta et al. 2006) and in vivo (Chen et al., 2005). The subnuclear and nuclear localization of DNMT3L depends on its interface with DNMT3A and DNMT3B as the study has found that DNMT3L dispersed

diffusely throughout the cytoplasm and nucleus in the absence of DNMT3A and DNMT3B (Nimura et al. 2006). Although DNMT3A and DNMT3B mainly function as a *de novo* methyltransferase, they also contribute in the maintenance of DNA methylation in heterochromatin regions (Liang et al. 2002; Chen et al. 2003). Furthermore, DNMT3A and DNMT3B do not exhibit any substantial preference between unmethylated DNA and hemi-methylated DNA. Both DNMT3A and DNMT3B are responsible for the normal mouse embryonic development (Okano et al. 1999). Homozygous deletion of DNMT3B causes several developmental defects and embryo die at the embryonic day E9.5, while DNMT3A deleted mice developed to term but die after 4 week of birth. Therefore, disturbance of any active DNMTs gene in mice is fatal, indicating the crucial role of DNA methylation in the development. Mutations of human DNMT3B gene cause hypomethylation of pericentromeric repeats, and found to be associated with a rare autosomal disorder known as ICF syndrome (Xu et al. 1999). Both DNMT3A and DNMT3B are highly expressed during embryogenesis and in undifferentiated embryonic stem cell, and decreased the expression level in differentiated tissues (Chen et al. 2004). Although DNMT3A and DNMT3B has high sequence homology and related biochemical properties, differences in knockout phenotypes reveal that this two enzymes have moderately non-overlapping biological functions. DNMT3A along with DNMT3L is essential for the adequate establishment of genomic imprinting during gametogenesis while DNMT3B is expendable in this course. DNMT3A and DNMT3B are essential for the methylation of different type of repeat sequences where DNMT3B is primarily responsible for the methylation of pericentromeric satellite repeats (Jurkowska et al. 2001).

Homozygous DNMT3L mutated mice are viable and survived to adulthood; however, male mice fail to produce mature sperms (Hata et al., 2002). DNMT3L plays a significant role

for the acquisition of DNA methylation during spermatogenesis in mice, and lacking of DNMT3L in germ cell causes reactivation of retrotransposon LTR-ERV1 element resulting aberrant chromatin packaging and meiotic failure in spermatocytes (Webster et al. 2005; Bourc'his and Bestor 2004). DNMT3L lacking female are fertile although fail to deliver live pups and embryonic death occur due to the defects in neural tubes development. Lack of DNMT3L causes loss of methylation on maternally imprinted genes leading to their biallelic expression (Hata et al. 2002). Therefore DNMT3L has a gender specific dual role: it is essential for the inactivation of repeat sequences during spermatogenesis and for establishment of maternal imprint in oocytes.

2.4.2 DNA methylation in normal development including B cell development

Each tissue/cell type has a distinctive DNA methylation profile and methylation patterns can be imitated to daughter DNA after mitosis. Therefore, DNA methylation has been proposed to act as a "cellular memory of the genome function". The global DNA methylation is established during gametogenesis and during early embryonic development (Reik et al. 2001; Jirtle RL. et al. 2007). Tissue specific DNA methylation is accountable for regulating gene expression, cellular differentiation and maintaining genomic stability during normal human development (Ohgane et. al. 2008; Song et al. 2005). The pluripotency related genes *OCT4* and *NANOG* are expressed in early embryos and in the germline, and then silenced in differentiated cells by promoter hypermethylation (Berdasco et al. 2010). Additionally, a group of non-pluripotency genes including *MAGE* are unmethylated in germline cell and then become methylated in differentiated cells (De Smet et al. 1999). DNA methylation has also been implicated in genomic imprinting, X chromosome inactivation during normal development.

Tissue specific DNA methylation mediated gene expression also found during normal development. For example, rat model study has shown that *SPHK1* gene is hypomethylated and expressed in brain tissue but hypermethylated and silence in heart tissue (Imamura et al. 2001). In another study, the expressions of *SERPINB5* in certain normal epithelial cells were associated with unmethylated promoter, and in some other normal cells like skin fibroblast, lymphocytes, bone marrow, kidney or heart do not express *SERPINB5* and are completely methylated at the promoter regions (Futscher et al. 2002). Additionally, CpG methylation profiling study demonstrated that 17% of analyze genes were differentially methylated at the 5' UTRs in 12 different somatic cells (Eckhardt et al. 2006). The housekeeping genes essential for cell metabolism, cell cycle, stress related genes or ribosomal RNA are normally unmethylated at the promoter CpG island (Caiafa et al. 2005).

Hematopoietic cells can be separated at specific stage of differentiation by flow cytometry that makes the hematopoiesis as an appropriate model system to study the role of DNA methylation during cellular differentiation. It is well-known that DNA methylation plays a critical role in the regulation of hematopoiesis, including myelopoiesis and lymphopoiesis. For example, genome-wide studies have shown that hematopoietic stem cell (HSC) has more methylation level and then precisely and consecutively lost the methylation during myeloid differentiation at common myeloid progenitor (CMP) and erythroblast stage (Hogart et al. 2012). DNA methylation is also required for the survival and self-renewal of HSCs (Bröske et al. 2009). This study has shown that deletion of DNMT1 induced cell-autonomous apoptosis of HSCs and bone marrow progenitors which ultimately lead to death of all mice. DNA methylation also regulates the fate of myeloerythroid versus lymphoid differentiation from HSCs. The low level of methylation decreased the expression of genes encoding lymphocyte progenitor markers

and transcription factors essential for B-cell development (Bröske et al. 2009; Ji et al. 2010). Similar studies in human primary cells found that HSCs maintained intermediate level of methylation whereas losing methylation associated with the myeloerythroid differentiation and gaining of DNA methylation involves in lymphoid commitment (Hodges et al. 2011). Microarray study has shown that DNA methylation predominantly decreases during pre-B cell development and the differentially methylated regions between subsequent stages of B-cell development coincided with the transcription factor binding sites (Lee et al. 2012). Collectively, all those studies determined a pivotal role of DNA methylation in hematopoiesis.

2.4.3 DNA methylation in cancer

Deviant DNA methylation pattern are associated with many human diseases and contribute to the initiation and progression of various cancers (Jones and Baylin 2007; Feinberg and Tycko 2004). DNA methylation profiles are largely altered in cancer cells and hereafter can be used to discriminate the cancer cells from normal tissue. Dnmt1-deficient mice model lacking DNA mismatch repair gene increased the susceptibility for lymphomagenesis (Trinh et al. 2002). Additionally, many studies demonstrate that the promoter hypermethylation of tumor suppressor genes (TSG) involved in different cellular pathways including cell cycle, genome maintenance, or apoptosis are associated with the development of cancer. DNA hypomethylation is a pervasive characteristic of carcinogenesis leading to genomic instability, loss of genomic imprinting, activation of silenced genes and transposable elements (Jones et al. 2007; Berdasco and Esteller 2010). Nevertheless, the pattern of hypermethylation and hypomethylation are cancer type and stage specific (Ehrlich 2009). Although the underlying mechanisms responsible for the genome-wide hypomethylation in cancer are not fully understood, there is evidence of two competing

"passive" vs. "active" demethylation process found in cancer. Passive demethylation occurs due to lack of maintenance of DNA methylation during multiple DNA replications over a series of cell divisions, whereas active demethylation occur independent of DNA replication (Wild and Flanagan 2010). Active demethylation arises much faster and depends on a class of enzymes harboring a demethylase activity. The TET family proteins (Ten Eleven Translocation proteins) enzyme assists active DNA demethylation by oxidizing methyl-cytosines to 5-hydroxyl-methylcytosines (5-hmC) followed by deamination or further oxidation of 5-hmC that leads to nucleotide mismatch, thymine-DNA glycosylase excised the mismatch and replaced by a cytosine (He et al. 2011; Song et al. 2011). Active DNA demethylation process in mammals is still difficult to define and stronger evidence about passive demethylation lead to assume that passive loss of methylation is the most likely the mechanism involves in global hypomethylation during carcinogenesis (Wild and Flanagan 2010).

Aberrant DNA hypermethylation has also been described in many cancers. DNMTs up-regulation has been reported as an early event of many cancers (Girault et al. 2003; Peng et al. 2005; Belinsky et al. 1996; Robertson et al. 1999). Hypermethylation of CpG islands in the promoter regions affects genes involve in many cellular pathways including cell cycle, cell adhesion, DNA repair, apoptosis, toxic catabolism and angiogenesis. CpG island promoter hypermethylation are reported in many genes in a wide variety of cancers, and this hypermethylation are specific to cancer types, therefore each cancer type can be allotted in a specific DNA hypermethylome (**Table 2.2**). The early report of CpG island promoter hypermethylation was found in retinoblastoma gene (RB gene) of retinoblastoma tumors (Greger et al. 1989; Sakai et al. 1991), and then hypermethylation of CpG island and resulting gene inactivation were demonstrated in tumor suppressor gene *VHL* in renal carcinoma (Herman et al.

1994), $p16^{INK4a}$ in different human cancers (Merlo et al. 1995; Herman et al. 1995), and *BRCA* in breast and ovarian cancer (Esteller et al. 2000). In addition, early stage of pancreatic cancer display aberrant CpG island hypermethylation and this methylation level gradually increases throughout the neoplastic progression (Sato et al. 2008). Promoter hypermethylation of *GSTP1* and *MGMT* genes are found in early prostatic intraepithelial neoplasia and in pre-cancerous pancreatic lesions, respectively (Brooks et al. 1998; House et al. 2003). Hypermethylation also found in the genomic region encoding miR-148a in the early stages of pancreatic cancer (Hanoun et al. 2010). Additionally tissue specific expression of *MASPIN* is silenced by aberrant hypermethylation in breast, skin, thyroid and colon cancer (Berdasco and Esteller 2010). The methylation-controlled DNAJ (MCJ) is another example of CpG island hypermethylation and gene inactivation in cancer. The methylation of *MCJ* is associated with the resistance to chemotherapy in ovarian cancer (Strathdee et al. 2004).

Alteration of DNA methylation is also a common feature in hematological malignancies. Acute myeloid leukemia (AML) specific methylation were identified within the promoters and gene bodies, and the distinct hypomethylation patterns of certain interspersed repeat elements were able to distinguished the cytogenetic subtypes of AML (Saied et al. 2012). Furthermore, recurrent DNMT3A mutation was discovered in about 20% of AML patients with a normal karyotype (Ley et al. 2010; Yamashita et al. 2010; Yan et al. 2011). AML patients with DNMT3A mutations had significantly shorter median overall survival rate and this poorer outcome were associated with intermediate-risk cytogenetic profile. DNMT3A mutants exhibited reduced enzymatic activity due to either premature truncation of this enzyme or affect single amino acid, R882 (Yan et al. 2011; Fong et al. 2014). DNMT3A mutation also found in patients with myeloproliferative neoplasms (MPN) and myelodysplastic syndromes (MDS), and are

linked to increase probability of AML progression (Fong et al. 2014). Promoter hypermethylation is found to be associated with the better overall and event free survival in T-cell acute lymphoblastic leukemia (T-ALL) compared to hypomethylated T-ALL patients (Borssén et al. 2013). Recently, microarray studies also observed the differential methylation pattern in ALL (Chatterton et al. 2014; Nordlund et al. 2013).

How the certain genomic regions (specifically CpG island) become hypermethylated in some cancer types but not in others is still remained unclear. Certain genomic regions can have a position within a specific nucleotide sequences that permit them to become hypermethylated or their location in a certain chromosomal region may make them vulnerable to epigenetic mis-regulation (Esteller 2008). Additionally, specific histone modification can mark a particular gene for hypermethylation. For example, polycomb group protein EZH2 is essential for the *de novo* DNA methylation of EZH2 target promoter where EZH2 assist as a platform for DNA methyltransferase recruitment (Viré et al. 2006; Schlesinger et al. 2007). Regardless of mechanisms, transcriptional repression of a specific gene by DNA rmethylation may provide growth advantage in particular cancer types.

2.4.4 DNA methylation as a Biomarker for diagnosis

Alteration of DNA methylation can be used as a biomarker in cancer. Sensitivity and specificity are the major criteria need to be considering for the efficacy of biomarker for cancer diagnosis. Sensitivity refers to the proportion of confirmed disease subjects who display positive detection of the marker while specificity represents the proportion of patients, negative for the disease and tested negative for the biomarker (Delpu et al. 2013). The characteristic of ideal biomarker is that no cancer patient would be negative, and no cancer free patient would be

positive for this biomarker (100% sensitivity and 100% specificity). Although ideal biomarker does not discover yet, several alteration of DNA methylation marks have emerged as potential biomarkers.

a. The tumor suppressor gene *p16INK4a* is a cyclin dependent kinase inhibitor and have an active role in Rb growth control pathway (Chang et al. 2003). *p16INK4a* promotes the formation of Rb-E2F repressive transcriptional complex and block cell cycle progression by preventing E2F dependent transcription. Hypermethylation of *p16INK4a* is found in many different type of cancer including 24% of lung cancer, 27% of colorectal cancer, and 73% of hepatocellular carcinoma patients (Delpu et al. 2013; Wong et al. 1999).

b. O-6-Methylguanine-DNA Methyltransferase (MGMT) involved in DNA repair by transferring the alkyl group at the O-6 position to a cysteine residue of MGMT molecule resulting restoration of guanine in the DNA. This reaction causes irreversible inactivation of MGMT, therefore refers as a suicide enzyme (Kaina et al. 2010). Lack of *MGMT* expression augmented DNA damage whereas higher expression increased the risk of cancer. Increase expression level of *MGMT* undercut cancer therapy that uses alkylating agents. MGMT promoter methylation was found in 46% of colorectal and 40% of brain cancer (Shen et al. 2005; Esteller et al. 2000).

c. *MLH1* (MutL Homolog 1) recruits the excision and repair machinery to the location of base mismatches. The promoter regions of this gene is hypermethylated in different cancers including sporadic colorectal cancer patient associated with microsatellite

instability (86% cases), endometrial cancer and ovarian cancer (Menigatti et al. 2001; Bischoff et al. 2012; Ozdemir et al. 2012).

d. *BRCA1* (Breast cancer type I susceptibility protein) participates in the DNA double—strand break repair by interacting with many co-factors (Venkitaraman 2002). Promoter hypermethylation of BRCA1 is found in 67% of medullary breast cancinomas, 55% of mucinous breast cancinoma, and in 31% of sporadic ovarian cancer (Esteller et al. 2001; Esteller et al. 2000; Delpu 2013).

e. *GSTP1* (glutathione S-transferase pi 1) is a detoxification enzyme that eliminates many endogenous and exogenous electrophilic compounds by the conjugation of glutathione (Townsend and Tew 2003). Promoter hypermethylation of *GSTP1* associated with the loss of expression in 73% of prostate cancer patients with a sensitivity of 73% and a specificity of 100% (Harden et al. 2003). Hypermethylation of *GSTP1* also found in large B cell lymphoma and breast cancer (Delpu et al. 2013).

f. *SEPT9* (Septin 9) involves in cytokinesis and cell cycle control. Hypermethylation of *SEPT9* is present in head and neck cancer as well as in colorectal cancer patient (Bennett et al. 2008; Lofton-Day et al. 2008). *SEPT9* hypermethylation was detected in colorectal cancer with 72% sensitivity and 90% specificity using a plasma-based methylation screening test (Grützmann et al. 2008).

The methylation marks are frequently not specific to a particular cancer, but generally preserved among different tumor types. Therefore, it is challenging to recommend a solo DNA methylation marks as a biomarker for a specific type of cancer. A combination of different methylation marks can improve the specificity of biomarkers to discriminate different type of cancer.

2.4.5 DNA methylation inhibitors for cancer treatment

Unlike genetic mutations, alterations of DNA methylation are reversible. This event has fascinated significant attention from a therapeutics perspective. DNA methylation inhibitors have the potentiality to reactivate the aberrantly methylated tumor suppressor genes. Despite a large number of inhibitors effects DNA methylation in preclinical studies, only a small number of them are clinically used in cancer treatment. There are two different types of DNA methylation inhibitors: nucleosidic and non-nucleosidic inhibitors.

Nucleosidic inhibitors form a covalent interaction with DNMTs and incorporated into replicating DNA. The main goal of this treatment is to target the replicating tumors cells whereas normal cell remain unaffected. Nucleosidic inhibitors are intent to eliminate DNA hypermethylation by interrupting the copying of aberrant methylation pattern. 5-azacytidine or azacytidine (Vidaza®, Celgene, NJ, USA) and 5-aza-2'-deoxycytidine/decitabine (Dacogen® MGI Pharma, Bloomington, MN, USA) are cytidine analogs where the carbon atom at position 5 is substituted by a nitrogen atom and connected to a ribose or a deoxyribose, respectively. After entry into the cells, cytidine analogs converted into their tri-phosphorylated active forms, and then incorporated into DNA (Gros et al. 2012). After incorporation into DNA, these analogs are recognized by DNMTs as normal cytosine but build an irreversible covalent complex due the

presence the nitrogen atom in position 5. The DNMTs enzymes then undergo proteasomal degradation leading to a cellular depletion of DNMTs (Gros et al. 2012; Delpu et al. 2013). Due to cytotoxic effect at higher dose, these compounds are used at low dose to attain only the demethylation effects with minimum cytotoxicity. Azacytidine is approved by FDA and is presently used for the treatment of myelodysplastic syndrome and acute myeloid leukemia (Lübbert et al. 2000). Transient low doses of azacytidine and decitabine treatment demonstrated decrease promoter DNA methylation, gene reexpression and antitumor effects on cultured and primary breast, colon and lung cancer cell lines (Tsai et al. 2012). The combination of azacytidine and entinostat (HDAC inhibitors) has established effectiveness in patients with non-small cell lung cancer in a phase I/II study (Juergens et al. 2011). Several phase I/II/III clinical trials are ongoing to evaluate the safety and efficiency of azacytidine and decitabine treatment in different malignancies (Delpu et al. 2013). Zebularine (Tocris Biosciences) is another nucleoside analog of cystidine. It is a cystidine deaminase inhibitor and displayed DNMTs inhibitory effects with low toxicity and anti-cancer activity (Zhou et al. 2002). Long-term oral administration of zebularine in mouse model study causes a gender-specific retraction of intestinal tumors whereas affecting a tissue-specific DNA demethylation (Yoo et al. 2008). First generation nucleosides (azacytidine, decitabine and Zebularine) demonstrated admirable result for the treatment of acute myeloid leukemia and myelodysplastic syndrome, but these inhibitors can also lead to demethylation and reexpression of pro-metastatic genes (Chik and Szyf 2011) highlighting the need for more specific DNMTs inhibitors. Non-nucleosides DNA methylation inhibitors do not incorporated into DNA molecules. Hydralazine functions as a smooth muscle relaxant and exhibited inhibitory effect on DNA methylation and reactivates methylated tumor suppressor genes in many cancer cell lines (Segura-Pacheco et al. 2003). Although the mechanisms of

demethylating effect is still not clear, hydralazine treatment of four patient with cervical cancer reestablishes the expression of methylated tumor suppressor genes without affecting genome-wide DNA methylation (Zambrano et al. 2005). Procainamide is a procaine derivative specifically inhibits DNMT1 and reactivated silenced gene expression (Lee et al. 2005). Procaine interacts with CpG rich genomic regions leading to DNA demethylation of tumor suppressor gene and displayed growth inhibitory effect on MCF7 cell line (Villar-Garea et al. 2003). Additionally, six conjugates compounds of procainamide exhibited potential inhibitory effects on murine Dnmt3A/3L complex and of human DNMT1 (Halby et al. 2012). In spite of these auspicious results, these compounds are not tested in clinical trials for their anti-cancer effects. Epigallocatechin-3-gallate (EGCG) and genistein are flavonoids compound found in green tea and soybean, respectively, displayed inhibitory effects on DNA methylation and causes reactivation of *RARbeta*, *p16INK4a,* and *MGMT* genes in KYSE 510 cells (esophageal squamous cell carcinoma cell line) (Fang et al. 2005). Despite several clinical trials are currently evaluating flavonoids as a prospective anti-cancer therapy, their accurate mechanisms of actions still remaining controversy (Diplu et al. 2013).

2.4.6 Methods for DNA methylation studies

Detection of altered DNA methylation is the complementary to cytological and genetic studies used to advance in cancer diagnosis. Development of next generation sequencing led to an outburst of techniques for studying genome-wide DNA methylation. There are many techniques available for the study of genome-wide DNA methylation profiles (Robinson et al. 2010) and these techniques mainly fall into three categories-

1. *Chemical conversion*: Bisulfite conversion of DNA is considered as a gold standard method where unmethylated cytosine is converted to uracil and methylated cytosine remains unaffected. Methylation profile of converted DNA is then recognized by different techniques such as DNA sequencing, restriction enzyme digestion or PCR. Examples are: whole genome shotgun bisulfite sequencing (WGSBS) and reduced representation bisulfite sequencing (RRBS).
2. *Enrichment of a methylated DNA*: Methylated fractions of the genome can be capture by various approaches including using an antibody against 5-methylcytosine or through methylated binding domain (MBD) proteins. Examples are: MeDIP-seq, MIRA-seq and MBD-seq.
3. *Enzymatic digestion*: Restriction enzyme that cleaves genomic DNA in a methylation specific manner can be used to detect the methylation status in various ways. HpaII enrichment by ligation-mediated PCR (HELP-seq), and methylation sensitive cut counting (MSCC) are the example of this technique.

The major criteria for selecting a particular technique consist of resolution, cost, amount of starting materials, and genome coverage. WGSBS have the best coverage and resolution among the available techniques but it also has the highest cost and required to build a more robust informatics infrastructure. MIRA-seq has the advantage of being able to distinguish the methylated DNA in all regions of the genome and the slight disadvantage that the precise location of the methylated CpG site cannot be identified. Furthermore, microarray-based technique interrogates an *a priori* selected set of CpG sites, which is not a comprehensive representation of genomic CpG methylation. New third-generation sequencing technologies are

developing that could revolutionize DNA methylation studies by distinguishing methylated cytosines from unmethylated cytosines directly on single-base resolution view of the methylome.

2.5 HISTONE MODIFICATIONS

Each nucleosome of chromatin is composed of 146/147 bp of DNA wrapped around a histone proteins octamer (two dimer of H2A-H2B, and two H3 and H4). Histone proteins (particularly H3 and H4) are posttranslationally modified at the amino terminal tails and these modifications included acetylation, methylation, phosphorylation, ubiquitylation and sumoylation. Posttranslational histone modifications changes chromatin structure and have either transcriptional activation or repression effects subjected to the type of histone modifications and the involved amino acid residues. Histone acetylation and histone methylation are the most common type of histone modifications and are enzymatically catalyzed by histone acetyltransferases (HATs) and histone methyltransferases, respectively. Histone deacetylases (HDAC) remove the acetyl groups from histone tails, and they often form complexes with methyl-DNA binding proteins (MBDs) and together they build a closed chromatin conformation which lead to transcriptional repression (Li 2002). The opposing activities of HATs and HDACs maintain a balance level of histone acetylation during normal development (Ruijter et. al. 2003). Histone acetylation utterly occurs on lysine residues of histone tails, where acetyl-CoA donates the acetyl group. Histone acetylation altered histone-DNA interaction by neutralizing the positive charge of lysine residue, and predominantly related to the transcriptional activation. The lysine residue of histone tails can also be mono (me1), di (me2), and tri-methylated (me3) and affected gene expression. Methylations of lysine residues sited at position 4 (K4me) and 36 (K36me) are

associated with transcriptional activation while residues on position 9 (K9me) and 27 (K27me) lead to transcriptional repression.

Histone modifications are also altered during tumorigenesis. The most common changes observed in cancers included deacetylation of histone H3 and H4, gain of repressive H3K9me and loss of activating H3K4me (Jones and Baylin 2007). Loss of H4K16ac (acetylation at lysine 16 of histone H4) and H4K20me3 (trimethylation at lysine 20 of histone H4) is a common trademark of human cancer (Fraga et al. 2005). Furthermore, altered global histone modifications (acetylation and methylation) patterns can be used as a predictor of prostate cancer recurrence (Seligson et al. 2005). Additionally, genetic alterations are detected in enzymes responsible for histone modifications in a tissue specific manner, and those alterations differ in solid and hematological malignancies (Berdasco and Esteller 2010).

2.6 LincRNA

A substantial portion of the human genome is transcribed without translation, generating copious short and long non-coding RNA. Non-coding RNAs can be divided into regulatory noncoding RNAs and housekeeping noncoding RNAs such as transfer RNAs, ribosomal RNA, small nuclear and nucleolar RNAs (Ponting et al. 2009). Long non-coding RNAs (lncRNAs) have a relatively longer transcript size ranging from several hundred to tens of thousands of bases and may be located within nuclear or cytoplasmic parts. LncRNAs can functions through a diverse mechanisms (**Figure 2.3**) including as a signaling molecules for transcription factors, as a decoys for transcription factors and other proteins to keep away from chromatin, as a guides molecules for chromatin-modifying enzymes to target genes or as a scaffolds to bring several proteins and form ribonucleoprotein complexes (Wang et al. 2011). Interestingly lncRNA are frequently transcribed from either strand of DNA within a protein-coding gene. LncRNA plays a

vital role in epigenetic gene regulation. For instance, lncRNA *KCNQ1ot1* and Air mediate repressive histone modifications and regulates the allele specific imprinting of *KCNQ1* and *IGF2R* genes, respectively, while *XIST* plays a critical role in X chromosome inactivation. Although protein-coding genes get more consideration in the study of tumorigenesis, recently numerous studies providing more evidence of lncRNA mediated gene regulation in normal and diseases development (Ponting et al. 2009).

Long intergenic ncRNAs (lincRNAs) are a subtype of lncRNA has been shown to affect gene expression in cancers by interacting with chromatin modification complexes. Despite the identification of more than 3,000 lincRNA in human genome, only very few of them (<1%) have been characterized (Khalil et al. 2009). LincRNA are often transcribed close (<10kb) to the protein coding genes and most probable act as a cis mechanisms (Ponting et al. 2009). LincRNAs regulate gene expression by directing chromatin-modifying complexes to precise genomic loci. About ~ 20% of expressed lincRNA are bound by polycomb repressive complex 2 (PRC2), and siRNA-mediated depletion of specific lincRNAs leads to the up-regulation of genes those normally silenced by PRC2. PRC2 complex comprises EZH2 (histone H3 lysine 27 methyltransferase), SUZ12 and EED. Nevertheless, how the lincRNA-polycomb complex identifies the target DNA is poorly understood. The expression of lincRNAs in normal human tissues occur both universally and in a tissue-specific manner suggesting that their expression occur under certain transcriptional regulation. LincRNAs are expressed in response to signaling molecules during normal development, and mis-regulated in solid and hematological malignancies (Calin et al. 2007). For example, *HOTAIR* lincRNA is overexpress in primary breast tumors and associated with bad prognosis (Gupta et al. 2010). Furthermore, enforced expression of *HOTAIR* in cancer cells targeted PRC2 complex and altered genome-wide H3K27

methylation as well as changes in gene expression leading to increase cancer invasiveness and metastasis. On the other hand, lack of *HOTAIR* can hinder cancer invasiveness suggesting that lincRNAs play direct roles in cancer epigenome modifications. Another lincRNA, *ANRIL* is an antisense transcript of *CDKN2B* and *CDKN2A* genes and interacts with a subunit of PRC1 (CBx7) leading to heterochromatin formation and transcriptional silencing (Tsai et al. 2011). The expression of lincRNA-p21 induces by DNA damage and repressed p53-dependent transcriptional response while lack of lincRNA-p21 increases the expression of many p53 mediated repressed genes (Huarte et al. 2010). Although more definitive evidence are required to confirm that the altered expression levels of lncRNA act as a etiology of disease, recent studies suggest that lncRNA added an extra layer in gene regulating network, where RNA species interact with proteins or DNA to modify the expression of protein coding genes.

2.7 MicroRNA

2.7.1 MicroRNA in B-cell development

MicroRNAs (miRNAs) are small (~22-nucleotides) noncoding RNAs that bind to their cognate mRNA leading to the translational repression or degradation of protein coding mRNA. miRNAs plays important roles in the regulation of numerous physiological process including normal development and differentiation. A single miRNA can interact with wide range of mRNAs, and the miRNAs-mediated gene regulation plays important roles in B-cell development, differentiation and activation (Li et al. 2013). Primary miRNA are transcribed by RNA polymerase II and then processed by type III RNase endonucleases Drosha and Dicer into mature miRNA. The ablation of Dicer blocked the transition from pro-B to pre-B cell differentiation and found to be associated with the up-regulation of miR-17 in Dicer-deficient

pro-B cells (Koralov et al. 2008). Additionally, miR-150, miR-181, miR-34a, miR-17~92 cluster plays a crucial role in the development of B-cell in bone marrow (Li et al. 2013). miR-150 primarily expressed in mature and resting B-cell but do not expressed in progenitors. This miR-150 target transcription factor c-Myb, which control several phases of B-cell development (Xiao et al. 2007). Ectopic expression of miR-150 in pro-B cell causes down-regulation of c-Myb leading to reduced number of mature B-cell. Another study has shown that overexpression of miR-150 block B-cell development at the pro-B to pre-B differentiation highlight the significance of miR-150 in B-cell development (Zhou et al. 2007). miR-181a increasingly expressed in bone marrow, thymus and spleen and positively regulate B cell development (Li et al. 2013). The miR-17~92 cluster (miR-17, miR-18a, miR-19a, miR-19b, miR-20a, and miR-92a) are processed from the same transcript and highly expressed in pro-B cell and then decline their expression in mature B-cells. Absence of miR-17~92 cluster expressions in pro-B cell leads to impair pro-B to pre-B differentiation (Ventura et al. 2008). Collectively, these studies indicate that miRNA plays a critical role in early B-cell development and any deviation in miRNA expression could leads to hematological malignancies.

2.7.2 MicroRNA in cancer

Mis-regulation of miRNA is a foremost event through the development of malignancies. Remarkably, human cancers can be classified based on the expression level of miRNA, where most of the miRNAs are down-regulated in cancer compared to normal tissues (Lu et al. 2005). The documentation of miR-34 as the direct target of p53 discovered for the first time that ncRNAs plays roles in this vital tumor suppressor pathways (Hermeking et al. 2007). Let-7 miRNA family negatively regulates RAS, and the lower expression of let-7 miRNA associated

with the significant higher expression of RAS in lung cancer (Johnson et al. 2005). Additionally, the expression of miR-15 and miR-16 are downregulated in chronic lymphocytic leukemia (CLL) (Weber et al. 2007). These two miRNA posttranscriptionally repressed antiapoptotic *BCL2* and the down-regulation of miR-15 and miR-16 associated with the overexpression of *BCL2* in CLL (Cimmino et al. 2005).

DNA methylation plays an important role in the regulation of miRNA expression in both normal and cancer cells (Weber et al. 2007). It was documented that the DNA hypermethylation causes transcriptional repression of miR-127 (Saito et al. 2006) and removal of methylation marks by 5-aza-2'-deoxycytidine treatment causes reactivation of miR-127 in human bladder carcinoma cell line (T24 cell line). Additionally, reactivation of miR-127 associated with the down-regulation of its potential targeted proto-oncogene *BCL6*. These finding propose that the transcriptional reactivation of miR-127 by DNA demethylation could act as tumor suppressors. Furthermore, miR-9-1, miR-124a3, miR-148, miR-152, and miR-663 were found to be hypermethylated in 34-86% of breast cancer cases (Lehmann et al. 2008). Same study shows that breast cancer cell lines treated with demethylating agent (5-aza-2'-deoxycytidine) reduces the methylation level and concomitantly up-regulated the expression of miR-9-1. The promoter regions of miR-34a and miR-148a are hypermethylated in multiple myeloma and pancreatic cancer, respectively (Delpu et al. 2013). Taken together, silencing of miRNA by DNA methylation can be an indicator of cancer development, therefore can be used as a diagnostic markers.

2.8 PSEUDOGENES

Pseudogenes are considered as genomic loci with high sequence resemblances to the corresponding functional gene. Due to accumulation of many mutations, pseudogenes lost their transcription and/or translations potency, and more than 11,000 pseudogenes are annotated in human genome (ENCODE Project Consortium. 2012). Pseudogenes play a critical role in transcriptional and post-transcriptional regulation and also have the potential to develop into new gene. Based on their origins, pseudogenes are categorized into three different groups: unitary pseudogenes, duplicated or unprocessed pseudogenes, and processed or retrotransposed pseudogenes (Xiao-Jie et al. 2015). Pseudogenes those are generated due to mutations on coding gene obliterate either their transcription or translation referred as unitary pseudogene. Therefore, unitary pseudogene does not have a fully functional complements termed as parental gene/ancestral gene or cognate gene. Pseudogenes derived from unfaithful gene duplication associated with premature stop codon or loss of promoter/enhancer functions or frame shift mutation revoking their functional activity while their parent genes remain functionally active are known as duplicated pseudogenes. Processed pseudogenes are originated through reverse-transcription of mRNA into DNA followed by integration into the genome at the new location. Therefore, introns are absent in preocessed pseudogenes whereas unitary and duplicated pseudogenes preserve their intron-exon structure. Until recently, pseudogenes have been categorized as “junk DNA” representing useless evolutionary remnants of DNA incorporated into the human genome through gene duplication or reverse transcription of mRNA. However, next generation sequencing techniques and substantial new research in non-coding RNAs has uncovered numerous function of pseudogene at the DNA, RNA or protein level.

At DNA level, pseudogenes can cause gene conversion, exonisation, homologous recombination or insertional mutations. Gene conversion is a process where a portion of parental gene sequence is replaced by its cognate pseudogene resulting inactivation of tumor suppressor gene and/or activation of oncogene. Gene conversion have been implicated in many diseases including recombination between of *JJAZ1* gene and its pseudogene in neurofibromatosis type 1 (Kehrer-Sawatzki et al. 2004), and *PMS2* gene and its pseudogene *PMS2CL* with unknown biological consequences (Ganster et al. 2010). Cytochrome P450 2A6 (*CYP2A6*) partakes in the metabolism of drugs and other pre-carcinogens. Genomic conversion from *CYP2A6* to its pseudogene *CYP2A7* created CYP2A6*1B allele that impact the rate of nicotine elimination and conceivably influence cigarette consumption and increased the risk of lung cancer (Wang et al. 2006). Pseudogene can also use the transcriptional machineries of host gene, generating a chimeric transcript or can also obtain *de novo* exon in a process called exonisation. DNA sequence can also be exchanged between parental gene and pseudogene by homologus recombination. The homologus recombination was found between *BRCA1* and its pseudogene *PsiBRCA1* in two families with breast and ovarian cancer. Recombination events took place between intron 2 of *BRCA1* and intron 2 of *PsiBRCA1* resulting a mutant alleles that lack the promoter and initiation codon of *BRCA1* gene leading to inactivation of tumorsuppressor gene (Puget et al. 2002). Additionally, insertion of pseudogene DNA into the promoter or exon of host gene leads to diminish host gene expression in cancer (Xiao-Jie et al. 2015). Collectively, pseudogene DNA mediated inactivation of protein-coding gene may offer a different layer of genetic mutations during carcinogenesis.

A group of pseudogenes are transcribed and serve significant role in post-transcriptional regulation. Pseudogenes transcripts can function in a diverse mechanisms including antisense

RNA, endogenous small-interference RNA (endo-siRNA), as well as competing for miRNA, translational machinery or RNA binding protein (RBP) (Xiao-Jie et al. 2015). Antisense transcript of pseudogene can combine with the sense mRNA of parental gene to impede its translation. Pseudogene derived endo-siRNA can be generated either from the combined sense and antisense double stranded RNAs from parental gene and pseudogene or from hairpin-shaped RNA transcribed from inverted repeat regions of pseudogene. This double stranded sense and antisense RNA hybrid as well as hairpin-shaped RNA are fragmented by ribonuclease Dicer into endo-siRNA. This endo-siRNA eventually detached into single strands and participate in RNA-induced silencing complex in a process termed RNA interference. For example, mouse model study has demonstrated that histone deacetylase-1 (Hdac1) is repressed by the siRNA produced from the sense and antisense duplex transcripts of its pseudogene (Tam et al. 2008). Pseudogene RNA that share miRNA response elements (MRE) with protein-coding gene control each other's expression by competing for the similar group of miRNA. For instance, the expression of pseudogene *OCT4-pg4* is positively correlated with the expression of its parental gene *OCT4* by competing for miR-145 in hepatocellular carcinoma (HCC). The expression of OCT4-pg4 promotes the growth of HCC cell, and associated with the poor prognosis of HCC patients (Wang et al. 2013). Another tumor suppressor gene *PTEN* (phosphatase and tensin homolog) is regulated by *PTENP1* pseudogene, where *PTEN* and *PTENP1* have only 18 mismatches throughout the coding sequence. Up-regulation of *PTENP1* in prostate cancer increases the cellular level of *PTEN* expression by competitively binding to miR-17, miR-19a, miR-20a, miR-214 families, and preventing *PTEN* mRNA from miRNA-induced suppression (Poliseno et al. 2010). Due to high sequence homology, RNA from pseudogene and parental gene can compete with each other for the transcriptional machineries and RBP. Pseudogene RNA competing for

RNA-stabilizing RBP lead to down-regulation of parental mRNA whereas pseudogene RNA competing for a RNA-degrading RBP leads to up-regulation of parental mRNA. For example, the increased expression of myosin light chain kinase pseudogene (*MYLKP1*) decrease the expression of parental gene *MYLK* and assist cancer cells proliferation by reducing the RNA stability (Han et al. 2011). Although majority of pseudogenes abrogated their protein-coding ability, a few numbers of processed pseudogenes preserve or reclaim this capacity. *NANOGP8* is one of the pseudogene for parental *NANOG* gene that encodes a protein detected by anti-*NANOG* antibody in many cancer cell lines (Zhang et al. 2006).

In summary, pseudogenes are copies of protein coding genes and retained high sequence similarities with their parental genes. Therefore pseudogene has the potentiality to regulate protein coding gene through different mechanisms as mentioned above. Recent studies piling up the evidence of pseudogene mediated gene regulation in both normal and diseases conditions. These studies offer strong indication for the functional association of pseudogenes in tumorigenesis, and also raise the likelihood that some of the human pseudogenes may be used as a diagnostic and therapeutic target.

2.9 GENOMIC REGULATORY REGIONS

The functional consequence of DNA methylation differs according to their genomic contexts. Methylation close to the TSS blocks initiation, but methylation in the gene body may have an impact on splicing and polyadenylation. Methylation in transposable elements and repeat sequences such as centromeres is important for chromosomal stability, and also suppress the expression of transposable elements. The role of DNA methylation in deregulating the activities of enhancers and other regulatory elements is only beginning to be appreciated.

2.9.1 Promoter regions

The promoters of protein coding genes harbor regulatory sequences required for the initiation of transcription. Promoter lies just upstream of the gene and polymerase tightly binds to the promoter regions and causes localized melting or separation of the two DNA strands within the promoter (Weaver 2008). Three different RNA polymerases transcribe different classes of genes. Class II promoters are recognized by RNA polymerase II and it has two parts: the core promoter and an upstream promoter element. General transcription factors bind to the core promoter and form a preinitiation complex while gene-specific transcription factor bind to the upstream promoter elements. The core promoter encompasses a TATA box centered at approximately position -33 from TSS, a TFIIB recognition element (BRE) situated just upstream of TATA box, an initiator located on the transcription start site, and a downstream promoter elements. Most of the promoters are lacking at least one of these elements. TATA box tend to be present in promoter of highly expressed specialized genes whereas housekeeping gene's promoter tend to missing them. RNA polymerase I recognized class I promoter and this type of promoter present in rRNA genes. Class III type of promoters present in genes encoding small RNAs. A large proportion of gene promoters coincided with the CpG islands and provides sites for the alteration of DNA methylation in many disease conditions including cancer. Therefore, anomalies in DNA methylation patterns of CpG islands within promoter regions of tumor suppressor genes are established as a common feature of human cancer.

2.9.2 Enhancer regions

Enhancers are regulatory sequences that can be sited upstream, downstream, or within their target gene and can control gene expression through physical interactions with gene

promoters. Enhancers can work over long distances, independently of their orientation relative to the gene promoter to stimulate cell/tissue type specific gene expression. Enhancers are considered as the key element of cell type specificity in gene expression (Bulger and Groudine 2011). Enhancer sequences are believed to contain densely clustered aggregation of transcription-factor-binding sites. Following the binding of apposite transcriptional factors, enhancer employs transcriptional co-activators and chromatin remodeling proteins. This accumulated protein complex assist DNA looping and eventually led to promoter mediated gene activation (Visel et al. 2009; Heintzman et. al. 2009). Promoter-enhancer interaction has been shown to correlate with gene expression (Sanyal et al. 2012).

Contrasting the ability to identify a gene promoter by sequencing the 5' end of its mRNA, enhancer detection relies on a number of inadequate measurements of chromatin structures and sequence functionality in a particular cell type. Distal enhancers are marked with highly stage-specific histone modifications, and the best predictor of active enhancers is the co-enrichment of H3K27ac (histone H3 lysine 27 acetylation) and H3K4me1 (histone H3 lysine 4 monomethylation) (Harmston and Lenhard 2013; Bogdanovic et al. 2012; Bulger and Groudine 2011). Enhancers may not be active in every tissue type, but still have the regulatory potential; these enhancers have been deemed to be poised enhancers. The H3K4me1 modification is present in poised enhancer, but only active enhancers exhibit the H3K27ac **(**Bogdanovic et al. 2012; Creyghton et al. 2010). Alternative criterion has previously employed for predicting enhancer is the non-coding sequence conservation. Conservation of noncoding DNA sequence through evolution indicates regulatory function. It is demonstrated that >50% of ultraconserved non-coding sequences (sequences at least 200bp in length that are 100% identical among human/mouse/rat) function as enhancers *in vivo* (Woolfe et al. 2005; Pennacchio et al. 2006).

Nevertheless enhancer cannot always be recognized using DNA sequence conservation, because biological function can be conserved without sequence conservation (Fisher et al. 2006; Hare et al. 2008). Additionally, a large percentage of functional enhancers are not subject to evolutionary restraint (Bulger and Groudine 2011).

Altered enhancer function may contribute to atypical expression of their target genes. Furthermore, genome-wide association studies (GWAS) have revealed a large number of disease susceptibility regions map to non-coding intervals (Visel et al. 2009). PU.1 is a transcription factor vital for normal hematopoietic development and its expression is directly regulated by a distal upstream enhancer element (Li et al. 2001).The presence of a SNP within the PU.1 enhancer regions lead to decrease enhancer activity and PU.1 expression in myeloid progenitors. This SNP in the PU.1 enhancer interfered with the binding of transcriptional regulator SATB1, leading to the development of acute myeloid leukemia (Steidl et al. 2007). Distal enhancers are enriched with CpG dinucleotides and their demethylation increases chromatin accessibility at enhancer region whereas control sites are not affected by complete lack of methylation. Furthermore, DNA methylation affects the interaction between cell-selective glucocorticoid receptor and enhancer DNA sequence (Wiench et al. 2011). The induction of enhancer specific transcription factor's expression as well as their subsequent recruitment to enhancers is associated with DNA demethylation (Sérandour et al. 2011). DNA methylation dictate the binding of FOXA1 transcription factor in certain types of enhancer regions. FOXA1 can bind to poised enhancers in order to trigger their transcriptional capability and the cell type specific recruitment of FOXA1depends on the differential methylation status of its binding site on enhancer regions. DNA hypomethylation allows FOXA1 binding to a specific enhancers required for those particular tissues (Sérandour et al. 2011). Altered DNA methylation pattern in

breast and prostate cancer cell lines affected the binding site of FOXA1 and other pioneer factors which lead to significant changes in tissue specific gene expression. It is quite possible that aberrant DNA methylation on transcription factor binding site could impact the expression of proto-oncogenes, tumor suppressor genes or tissue-specific gene expression patterns. Additionally, enhancer elements with low levels of H3K27ac and H3Kme1 show higher levels of DNA methylation and their activation is often complemented by their demethylation (Xu et al. 2007). DNA methylation on enhancer region is severely altered during malignant transformation where hypomethylated enhancer associated with the up-regulation and hypermethylated enhancer associated with the down regulation of cancer linked genes. The association between enhancer methylation and gene mis-regulation in cancer is much stronger than differential promoter methylation (Aran et al. 2013).

There are many documented examples of atypical enhancer function relating to cancer development. However, there is still much to be elucidated. Numerous enhancers are yet to be characterized and discovering them is no easy task. Sorting out what specific enhancers are active in unique tissues adds a further layer of difficulty in understanding enhancer behavior. Additionally, determining epigenetic abnormalities in enhancer regions is another critical avenue of research. The current research techniques have given investigators the ability to identify enhancers and now we must investigation the intergenic regions that were once thought to be devoid of anything of interest.

2.9.3 Dnase I hypersensitive site (DHS)

DNase I hypersensitive sites (DHSs) are regions of chromatin that are sensitive to cleavage by the DNase I enzyme. These DHSs describe the functional regulatory features (eg. promoters,

enhancers, insulators, locus control regions) of the complex genomes. DHSs are used to map tissue specific chromatin accessibility, and can identify the cis-regulatory modules (CRMs) in a specific tissue type (Wilken et al. 2015). The majority of DHSs identified in mouse brain tissue located in intronic (54%) and distal intergenic regions (31%) and about 50% DHSs coincided with the enhancer specific chromatin marks. The DHSs also occupied with tissue-specific transcription factors.

2.9.4 Transposable elements

Transposable elements (TE) are repeated DNA sequences that can insert themselves into the host genomes. Overall, transposable elements derived sequences make up about half of the human genome (Sinzelle et al. 2009; Criscione et al. 2014). Class I transposable elements also known as retrotransposon uses RNA intermediate and copied an RNA genome into host DNA using reverse transcriptase. Class I retrotransposons is divided into two groups; Long Terminal Repeat (LTR) and non-LTR elements (Chenais 2013). The majority of class I retrotransposons are non-LTR, long interspersed elements (LINE). Human endogenous retroviruses (HERVs) are LTRs and most of them have nonfunctional envelope genes. Class II elements also known as DNA transposons, and uses DNA mediated cut-and-paste mode of transposition by excision and reinsertion of the DNA sequence into the genome. The example of DNA transposons are hAT-Tip 100, TcMar-Tigger, TCMar-Mariner, PiggyBac and MuDR. The non-autonomus transposable elements, short interspersed nuclear elements (SINEs) are unable to insert themselves into the genome without enzymatic helps from autonomus elements. The SINEs comprises three main families in human genome; *Alu* repeat, MIR and MIR3. Although TEs are initially considered as selfish or parasitic entities, recent studies provided evidence of mutually

benefited relationship between TEs and host genome that stabilize the survival of TEs and the evolutionary importance of host genome. Insertion of transposable elements in a functional part of the genome can affect transcription of genes and disrupt the normal function of the genome. Additionally, TEs can contribute as a regulatory (promoter/enhancer) sequences to the host genes and offer alternative splice sites and cis regulatory elements. Another important contribution of TEs is the evolution of protein coding genes, and it is described that at least 4% of the protein coding genes are contributed by TEs (Nekrutenko et al. 2001). Transposable elements are generally silenced in human cell by DNA methylation (Maksakova et al. 2008; Walsh et. al. 1998) and the transcriptional activation of these elements causes transposable element mediated insertions and chromosomal rearrangements in many malignancies including breast/ovarian cancer, colon cancer, acute and chronic myeloid leukemia, Ewing sarcoma, and T-cell acute lymphoblastic leukemia (Lee et al. 2012; Chénais 2013). Losing methylation contribute the transcriptional activation of transposable elements in cancers (Ross JP et al. 2010; Watanabe Y et al. 2010; Wilson AS et al. 2007), and TEs preferentially inserted in cancer-specific hypomethyled regions (Lee et al. 2012). Furthermore, RNA polymerase II increasingly bind to retrotransposons in cancer cell line and associated with higher transcriptional activity compared to normal cell line (Criscione et al. 2014). Hypomethylation has been shown to strongly activate HERV-Fc1, an endogenous retrovirus with a large degree of Pol II enrichment (Criscione et al. 2014). Retrotransposons tend to act tumor-type specifically, acting as binding motifs for transcription factors. Binding motifs overlapping promoters, enhancers or insulator regions could confer a novel transcriptional signature (Xue et al. 2014).

2.9.5 CpG islands

A single stand of reference human genome contains over 28 million CpG dinucleotides that provide primary site for the methylation in human. CpG dinucleotide refers to a C followed by a G base in the 5' to 3' direction. Genome-wide CpG dinucleotides are underrepresented and often occur in cluster known as CpG islands. CpG island (CGI) is defined as a regions of the genome that is at least 550 bps in length, with the ratio of observed CG/expected CG > 0.65 (Takai and Jones 2004). CGIs normally lack DNA methylation and are found in the promoter regions of a large number of human genes including most tissue specific and housekeeping genes. Numerous studies provided evidence that CGIs become hypermethylated during carcinogenesis leading to inappropriate gene silencing. CGIs methylation also found in normal tissue for the tissue specific transcription regulation. Additionally, a large number of tissues specific methylated CGIs coincide to the regions far from gene promoters. CpG shore is refer to 2000 bp flanking upstream and downstream regions from CGI whereas CpG shelf refers to 2000 kb in both direction from CpG shore (**Figure 2.4**). Differential DNA methylation within CpG shore was found to have stronger correlation with gene expression than CpG islands methylation (Ji et al. 2010).

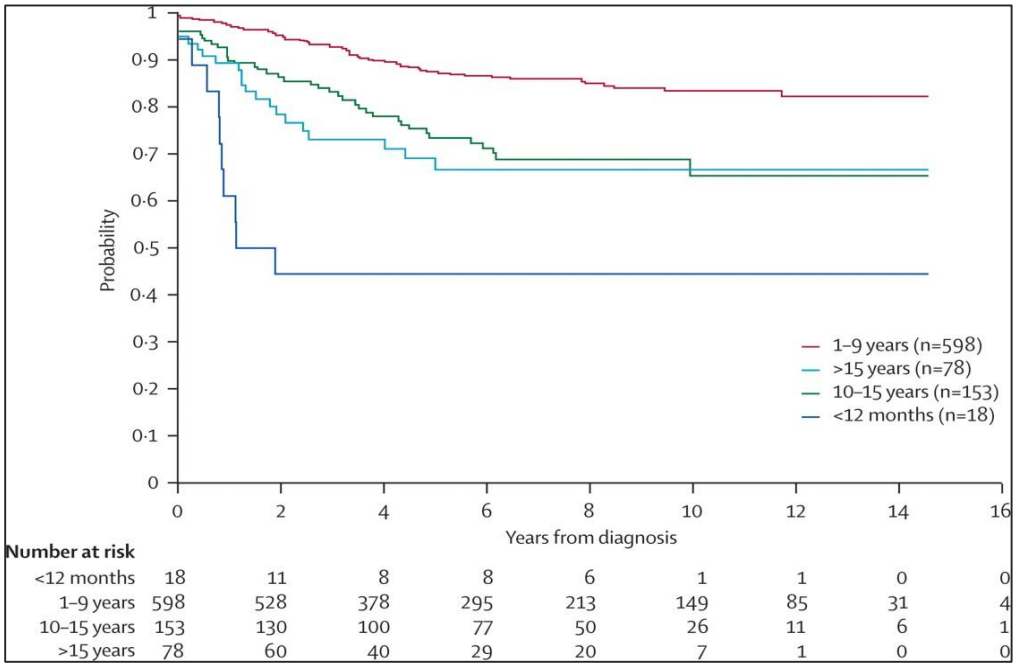


Figure 2.1: Kaplan-Meier estimates of event-free survival according to age at diagnosis of ALL. Reproduced from Pui CH. et al., 2008 with permission, license number: 3670931401057.

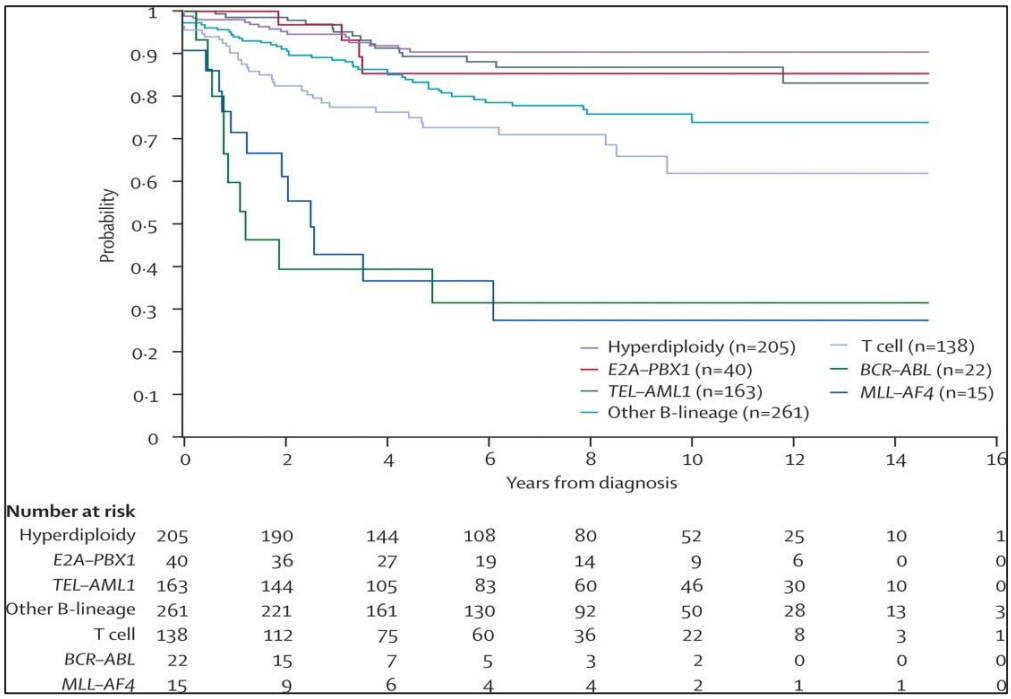


Figure 2.2: Kaplan-Meier analysis of event-free survival according to biological subtype of ALL. Reproduced from Pui CH. et al., 2008 with permission, license number: 3670931401057.

Table 2.1: Epigenetic aberrations among different tumor types. Reproduced with permission from Esteller M. N Engl J Med 2008;358:1148-1159, Copyright Massachusetts Medical Society

Type of Cancer	Epigenetic Disruption
Colon cancer	CpG-island hypermethylation (*hMLH1*, *p16^{INK4a}*, *p14^{ARF}*, *RARB2*, *SFRP1*, and *WRN*), hypermethylation of miRNAs (*miR-124a*), global genomic hypomethylation, loss of imprinting of *IGF2*, mutations of histone modifiers (*EP300* and *HDAC2*), diminished monoacetylated and trimethylated forms of histone H4
Breast cancer	CpG-island hypermethylation (*BRCA1*, E-cadherin, *TMS1*, and estrogen receptor), global genomic hypomethylation
Lung cancer	CpG-island hypermethylation (*p16^{INK4a}*, *DAPK*, and *RASSF1A*), global genomic hypomethylation, genomic deletions of *CBP* and the chromatin-remodeling factor *BRG1*
Glioma	CpG-island hypermethylation (DNA-repair enzyme *MGMT*, *EMP3*, and *THBS1*)
Leukemia	CpG-island hypermethylation (*p15^{INK4b}*, *EXT1*, and *ID4*), translocations of histone modifiers (*CBP*, *MOZ*, *MORF*, *MLL1*, *MLL3*, and *NSD1*)
Lymphoma	CpG-island hypermethylation (*p16^{INK4a}*, *p73*, and DNA-repair enzyme *MGMT*), diminished monoacetylated and trimethylated forms of histone H4
Bladder cancer	CpG-island hypermethylation (*p16^{INK4a}* and *TPEF/HPP1*), hypermethylation of miRNAs (*miR-127*), global genomic hypomethylation
Kidney cancer	CpG-island hypermethylation (*VHL*), loss of imprinting of *IGF2*, global genomic hypomethylation
Prostate cancer	CpG-island hypermethylation (*GSTP1*), gene amplification of polycomb histone methyltransferase *EZH2*, aberrant modification pattern of histones H3 and H4
Esophageal cancer	CpG-island hypermethylation (*p16^{INK4b}* and *p14^{ARF}*), gene amplification of histone demethylase *JMJD2C/GASC1*
Stomach cancer	CpG-island hypermethylation (*hMLH1* and *p14^{ARF}*)
Liver cancer	CpG-island hypermethylation (*SOCS1* and *GSTP1*), global genomic hypomethylation
Ovarian cancer	CpG-island hypermethylation (*BRCA1*)

* *BRCA1* denotes breast-cancer susceptibility gene 1, *BRG1* BRM/SWI2-related gene 1, *CBP* cyclic AMP response-element–binding protein (CREB)–binding protein, *DAPK* death-associated protein kinase, *EMP3* epithelial membrane protein 3, *EP300* E1A binding protein p300, *EXT1* exostosin 1, *EZH2* enhancer of zeste drosophila homologue 2, *GSTP1* glutathione S-transferase 1, *HDAC2* histone deacetylase 2, *hMLH1* homologue of MutL *Escherichia coli*, *ID4* inhibitor of DNA binding 4, *IGF2* insulin-like growth factor 2, *JMJD2C/GASC1* Jumonji domain-containing protein 2C, *MGMT* O^6-methylguanine–DNA methyltransferase, *MLL1* mixed-lineage leukemia 1, *MLL3* mixed-lineage leukemia 3, *MORF* monocytic leukemia zinc finger protein–related factor, *MOZ* monocytic leukemia zinc finger, *NSD1* nuclear receptor binding SET-domain protein 1, *RARβ2* retinoic acid receptor β 2, *RASSF1A* ras association domain family protein 1, *SFRP1* secreted frizzled-related protein 1, *SOCS1* suppressor of cytokine signaling 1, *THBS1* thrombospondin 1, *TMS1* target of methylation-induced silencing 1, *TPEF/HPP1* hyperplastic polyposis gene 1, *VHL* von Hippel–Lindau disease, and *WRN* Werner's syndrome.

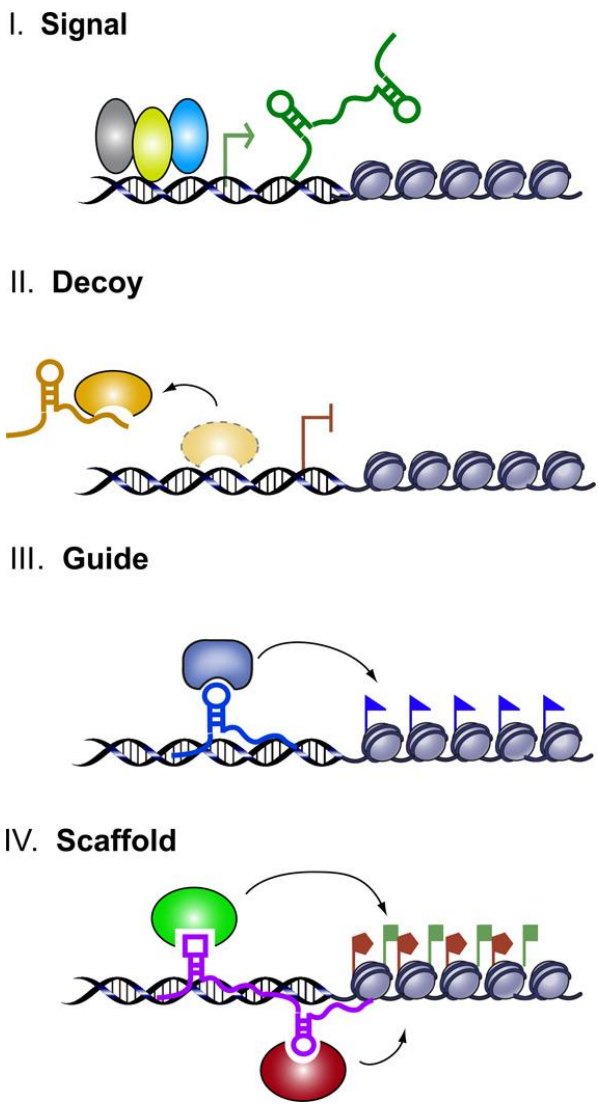


Figure 2.3: Schematic diagram of the four archetypes of lncRNA mechanism. Archetype I: As Signals, lncRNA expression can faithfully reflect the combinatorial actions of transcription factors (colored ovals) or signaling pathways to indicate gene regulation in space and time. Archetype II: As Decoys, lncRNAs can titrate away transcription factors and other proteins away from chromatin, or titrate the protein factors into nuclear subdomains. Archetype III: As Guides, lncRNAs can recruit chromatin modifying enzymes to target genes, either in cis (near the site of lncRNA production) or in trans to distant target genes. Archetype IV: As scaffolds, lncRNAs can bring together multiple proteins to form ribonucleoprotein complexes. The lncRNA-RNP may act on chromatin as illustrated to affect histone modifications. In other instances, the lncRNA scaffold is structural and stabilizes nuclear structures or signaling complexes. Figure from Wang KC and Chang HY. 2011. Reproduced with permission, license number: 3670940190396.

Figure 2.4: Relative genomic position of CpG island, CpG shore and CpG shelf.

CH2. REFERENCES

Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol 2013;14:R21.

Aricò M, Valsecchi MG, Camitta B, Schrappe M, Chessells J, Baruchel A, Gaynon P, Silverman L, Janka-Schaub G, Kamps W, Pui CH, Masera G. Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia. N Engl J Med. 2000 Apr 6;342(14):998-1006.

Bachman KE, Rountree MR, Baylin SB. Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. J Biol Chem. 2001 Aug 24;276(34):32282-7.

Behm FG. Cell biology and pathobiology, Immunophenotyping. Childhood Leukemias (Third edition). Edited by Pui C. New York, Cambridge University Press; 2006, page. 72-101.

Belinsky SA, Nikula KJ, Baylin SB, Issa JP. Increased cytosine DNA-methyltransferase activity is target-cell-specific and an early event in lung cancer. Proc Natl Acad Sci U S A. 1996 Apr 30;93(9):4045-50.

Benbrahim-Tallaa L, Waterland RA, Styblo M, Achanzar WE, Webber MM, Waalkes MP. Molecular events associated with arsenic-induced malignant transformation of human prostatic epithelial cells: aberrant genomic DNA methylation and K-ras oncogene activation. Toxicol Appl Pharmacol. 2005 Aug 15;206(3):288-98.

Bennett KL, Karpenko M, Lin MT, Claus R, Arab K, Dyckhoff G, Plinkert P, Herpel E, Smiraglia D, Plass C. Frequently methylated tumor suppressor genes in head and neck squamous cell carcinoma. Cancer Res. 2008 Jun 15;68(12):4494-9.

Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010 Nov 16;19(5):698-711.

Bischoff J, Ignatov A, Semczuk A, Schwarzenau C, Ignatov T, Krebs T, Küster D, Przadka-Rabaniuk D, Roessner A, Costa SD, Schneider-Stock R. hMLH1 promoter hypermethylation and MSI status in human endometrial carcinomas with and without metastases. Clin Exp Metastasis. 2012 Dec;29(8):889-900.

Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJ, Gómez-Skarmeta JL. Dynamics of enhancer

chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. 2012 Oct;22(10):2043-53.

Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004 Sep 2;431(7004):96-9. Epub 2004 Aug 18.

Brooks JD, Weinstein M, Lin X, Sun Y, Pin SS, Bova GS, Epstein JI, Isaacs WB, Nelson WG. CG island methylation changes near the GSTP1 gene in prostatic intraepithelial neoplasia. Cancer Epidemiol Biomarkers Prev. 1998 Jun;7(6):531-6.

Bröske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, Kuhl C, Enns A, Prinz M, Jaenisch R, Nerlov C, Leutz A, Andrade-Navarro MA, Jacobsen SE, Rosenbauer F. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 2009; 41:1207-15.

Borssén M, Palmqvist L, Karrman K, Abrahamsson J, Behrendtz M, Heldrup J, Forestier E, Roos G, Degerman S. Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. PLoS One. 2013 Jun 6;8(6):e65373.

Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell 2011;144:327-39.

Caiafa P, Zampieri M. DNA methylation and chromatin structure: the puzzling CpG islands. J Cell Biochem. 2005 Feb 1;94(2):257-65.

Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, Fabbri M, Cimmino A, Lee EJ, Wojcik SE, Shimizu M, Tili E, Rossi S, Taccioli C, Pichiorri F, Liu X, Zupo S, Herlea V, Gramantieri L, Lanza G, Alder H, Rassenti L, Volinia S, Schmittgen TD, Kipps TJ, Negrini M, Croce CM. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell. 2007 Sep;12(3):215-29.

Chang LL, Yeh WT, Yang SY, Wu WJ, Huang CH. Genetic alterations of p16INK4a and p14ARF genes in human bladder cancer. J Urol. 2003 Aug;170(2 Pt 1):595-600.

Chatterton Z, Morenos L, Mechinaud F, Ashley DM, Craig JM, Sexton-Oates A, Halemba MS, Parkinson-Bates M, Ng J, Morrison D, Carroll WL, Saffery R, Wong NC. Epigenetic deregulation in pediatric acute lymphoblastic leukemia. Epigenetics. 2014 Mar;9(3):459-67.

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet. 2007 Mar;39(3):391-6.

Chen T, Tsujimoto N, Li E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol. 2004 Oct;24(20):9048-58.

Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol. 2003 Aug;23(16):5594-605.

Chen T, Li E. Structure and function of eukaryotic DNA methyltransferases. Curr Top Dev Biol. 2004;60:55-89.

Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chédin F. Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem. 2005 Aug 1;95(5):902-17.

Chénais B. Transposable elements and human cancer: a causal relationship? Biochim Biophys Acta. 2013 Jan;1835(1):28-35.

Chik F, Szyf M. Effects of specific DNMT gene depletion on cancer cell transformation and breast cancer cell invasion; toward selective DNMT inhibitors. Carcinogenesis. 2011 Feb;32(2):224-32.

Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M, Croce CM. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A. 2005 Sep 27;102(39):13944-9.

Cooper SL, Brown PA. Treatment of pediatric acute lymphoblastic leukemia. Pediatr Clin North Am. 2015 Feb;62(1):61-73.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21931-6.

Criscione, S, Zhang Y, Thompson W, Sedivy J, Neretti N. Transcroptional landscape of repetitive elements in normal and cancer human cells. MC Genomics 2014, 15:583.

De Smet C, Lurquin C, Lethé B, Martelange V, Boon T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol Cell Biol. 1999 Nov;19(11):7327-35.

Delaval K, Feil R. Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev. 2004 Apr;14(2):188-95.

Delpu Y, Cordelier P, Cho WC, Torrisani J. DNA methylation and cancer diagnosis. Int J Mol Sci. 2013 Jul 18;14(7):15029-58.

Dias S, Månsson R, Gurbuxani S, Sigvardsson M, Kee BL. E2A proteins promote development of lymphoid-primed multipotent progenitors. Immunity. 2008 Aug 15;29(2):217-27.

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006 Dec;38(12):1378-85.

Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009 Dec;1(2):239-59.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74.

Esteller M. Epigenetics in cancer. N Engl J Med. 2008 Mar 13;358(11):1148-59.

Esteller M, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, Vanaclocha V, Baylin SB, Herman JG. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. N Engl J Med. 2000 Nov 9;343(19):1350-4.

Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, Herman JG. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst. 2000 Apr 5;92(7):564-9.

Fang JY, Xiao SD, Zhu SS, Yuan JM, Qiu DK, Jiang SJ. Relationship of plasma folic acid and status of DNA methylation in human gastric cancer. J Gastroenterol. 1997 Apr;32(2):171-5.

Fang MZ, Chen D, Sun Y, Jin Z, Christman JK, Yang CS. Reversal of hypermethylation and reactivation of p16INK4a, RARbeta, and MGMT genes by genistein and other isoflavones from soy. Clin Cancer Res. 2005 Oct 1;11(19 Pt 1):7033-41.

Feinberg AP, Tycko B. The history of cancer epigenetics. Nat Rev Cancer. 2004 Feb;4(2):143-53.

Figueroa ME, Chen SC, Andersson AK, Phillips LA, Li Y, Sotzen J, Kundu M, Downing JR, Melnick A, Mullighan CG. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. J Clin Invest. 2013 Jul 1;123(7):3099-111.

Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science. 2006;312:276–279.

Fitzpatrick DR, Wilson CB. Methylation and demethylation in the regulation of genes, cells, and responses in the immune system. Clin Immunol. 2003 Oct;109(1):37-45.

Fong CY, Morison J, Dawson MA. Epigenetics in the hematologic malignancies. Haematologica. 2014 Dec;99(12):1772-83.

Fowler BM, Giuliano AR, Piyathilake C, Nour M, Hatch K. Hypomethylation in cervical tissue: is there a correlation with folate status? Cancer Epidemiol Biomarkers Prev. 1998 Oct;7(10):901-6.

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat. Genet. 2005;37:391–400.

Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H, Domann FE. Role for DNA methylation in the control of cell type specific maspin expression. Nat Genet. 2002 Jun;31(2):175-9.

Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H, Domann FE. Ganster C, Wernstedt A, Kehrer-Sawatzki H, Messiaen L, Schmidt K, Rahner N, Heinimann K, Fonatsch C, Zschocke J, Wimmer K. Functional PMS2 hybrid alleles containing a pseudogene-specific missense variant trace back to a single ancient intrachromosomal recombination event. Hum Mutat. 2010 May;31(5):552-60.

Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R. Induction of tumors in mice by genomic hypomethylation. Science. 2003 Apr 18;300(5618):489-92.

Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem. 2004 Jun 11;279(24):25447-54.

Girault I, Tozlu S, Lidereau R, Bièche I. Expression analysis of DNA methyltransferases 1, 3A, and 3B in sporadic breast carcinomas. Clin Cancer Res. 2003 Oct 1;9(12):4415-22.

Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. Nat Rev Cancer. 2006 Mar;6(3):193-203.

Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. Nat Rev Cancer. 2003 Sep;3(9):639-49.

Greger V, Passarge E, Höpping W, Messmer E, Horsthemke B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet. 1989 Sep;83(2):155-8.

Gros C, Fahy J, Halby L, Dufau I, Erdmann A, Gregoire JM, Ausseil F, Vispé S, Arimondo PB. DNA methylation inhibitors in cancer: recent and future approaches. Biochimie. 2012 Nov;94(11):2280-96.

Grützmann R, Molnar B, Pilarsky C, Habermann JK, Schlag PM, Saeger HD, Miehlke S, Stolz T, Model F, Roblick UJ, Bruch HP, Koch R, Liebenberg V, Devos T, Song X, Day RH, Sledziewski AZ, Lofton-Day C. Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. PLoS One. 2008;3(11):e3759.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010 Apr 15;464(7291):1071-6.

Halby L, Champion C, Sénamaud-Beaufort C, Ajjan S, Drujon T, Rajavelu A, Ceccaldi A, Jurkowska R, Lequin O, Nelson WG, Guy A, Jeltsch A, Guianvarc'h D, Ferroud C, Arimondo PB. Rapid synthesis of new DNMT inhibitors derivatives of procainamide. Chembiochem. 2012 Jan 2;13(1):157-65.

Han YJ, Ma SF, Yourek G, Park YD, Garcia JG. A transcribed pseudogene of MYLK promotes cell proliferation. FASEB J. 2011 Jul;25(7):2305-12.

Harden SV, Guo Z, Epstein JI, Sidransky D. Quantitative GSTP1 methylation clearly distinguishes benign prostatic tissue and limited prostate adenocarcinoma. J Urol. 2003 Mar;169(3):1138-42.

Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS Genet. 2008;4:e1000106.

Harmston N, Lenhard B. Chromatin and epigenetic features of long-range gene regulation. Nucleic Acids Res 2013;41:7185-99.

Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 2002 Apr;129(8):1983-93.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011 Sep 2;333(6047):1303-7.

Hemberger M, Dean W, Reik W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat Rev Mol Cell Biol. 2009 Aug;10(8):526-37.

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature. 1999 Sep 16;401(6750):301-4.

Heard E. Recent advances in X-chromosome inactivation. Curr Opin Cell Biol. 2004 Jun;16(3):247-55.

Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A. 1994 Oct 11;91(21):9700-4.

Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res. 1995 Oct 15;55(20):4525-30.

Hermeking H. p53 enters the microRNA world. Cancer Cell. 2007 Nov;12(5):414-8.

Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR, Smith AD, Hannon GJ. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell. 2011; 44:17-28.

Hogart A, Lichtenberg J, Ajay SS, Anderson S; NIH Intramural Sequencing Center, Margulies EH, Bodine DM. Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. Genome Res. 2012 Aug;22(8):1407-18.

Hong D, Gupta R, Ancliff P, Atzberger A, Brown J, Soneji S, Green J, Colman S, Piacibello W, Buckle V, Tsuzuki S, Greaves M, Enver T. Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. Science. 2008 Jan 18;319(5861):336-9.

House MG, Guo M, Iacobuzio-Donahue C, Herman JG. Molecular progression of promoter methylation in intraductal papillary mucinous neoplasms (IPMN) of the pancreas. Carcinogenesis. 2003 Feb;24(2):193-8.

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene. 2008 Jan 10;27(3):404-8.

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell. 2001 Mar 23;104(6):829-38.

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010 Aug 6;142(3):409-19.

Hystad ME, Myklebust JH, Bø TH, Sivertsen EA, Rian E, Forfang L, Munthe E, Rosenwald A, Chiorazzi M, Jonassen I, Staudt LM, Smeland EB. Characterization of early stages of human B cell development by gene expression profiling. J Immunol 2007; 179(6):3662-71.

Imamura T, Ohgane J, Ito S, Ogawa T, Hattori N, Tanaka S, Shiota K. CpG island of rat sphingosine kinase-1 gene: tissue-dependent DNA methylation status and multiple alternative first exons. Genomics. 2001 Aug;76(1-3):117-25.

Ito C, Kumagai M, Manabe A, Coustan-Smith E, Raimondi SC, Behm FG, Murti KG, Rubnitz JE, Pui CH, Campana D. Hyperdiploid acute lymphoblastic leukemia with 51 to 65 chromosomes: a distinct biological entity with a marked propensity to undergo apoptosis. Blood. 1999 Jan 1;93(1):315-20.

Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E, Jaenisch R. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet. 2001 Jan;27(1):31-9.

Jeha S, Pei D, Raimondi SC, Onciu M, Campana D, Cheng C, Sandlund JT, Ribeiro RC, Rubnitz JE, Howard SC, Downing JR, Evans WE, Relling MV, Pui CH. Increased risk for CNS relapse in pre-B cell leukemia with the t(1;19)/TCF3-PBX1. Leukemia. 2009 Aug;23(8):1406-9.

Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 2010; 467:338-342

Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. Nat Rev Genet. 2007 Apr;8(4):253-62.

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ. RAS is regulated by the let-7 microRNA family. Cell. 2005 Mar 11;120(5):635-47.

Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007 Feb 23;128(4):683-92.

Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B, Sebree R, Rodgers K, Hooker CM, Franco N, Lee B, Tsai S, Delgado IE, Rudek MA, Belinsky SA, Herman JG, Baylin SB, Brock MV, Rudin CM. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. Cancer Discov. 2011 Dec;1(7):598-607.

Jurkowska RZ, Jurkowski TP, Jeltsch A. Structure and function of mammalian DNA methyltransferases. Chembiochem. 2011 Jan 24;12(2):206-22.

Kaina B, Margison GP, Christmann M. Targeting O^6-methylguanine-DNA methyltransferase with specific inhibitors as a strategy in cancer therapy. Cell Mol Life Sci. 2010 Nov;67(21):3663-81.

Kareta MS, Botello ZM, Ennis JJ, Chou C, Chédin F. Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. J Biol Chem. 2006 Sep 8;281(36):25893-902.

Kehrer-Sawatzki H, Kluwe L, Sandig C, Kohn M, Wimmer K, Krammer U, Peyrl A, Jenne DE, Hansmann I, Mautner V. High frequency of mosaicism among patients with neurofibromatosis type 1 (NF1) with microdeletions caused by somatic recombination of the JJAZ1 gene. Am J Hum Genet. 2004 Sep;75(3):410-23.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009 Jul 14;106(28):11667-72.

Kim YI, Pogribny IP, Basnakian AG, Miller JW, Selhub J, James SJ, Mason JB. Folate deficiency in rats induces DNA strand breaks and hypomethylation within the p53 tumor suppressor gene. Am J Clin Nutr. 1997 Jan;65(1):46-52.

Koralov SB, Muljo SA, Galler GR, Krek A, Chakraborty T, Kanellopoulou C, Jensen K, Cobb BS, Merkenschlager M, Rajewsky N, Rajewsky K. Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. Cell. 2008 Mar 7;132(5):860-74.

Koturbash I, Baker M, Loree J, Kutanzi K, Hudson D, Pogribny I, Sedelnikova O, Bonner W, Kovalchuk O. Epigenetic dysregulation underlies radiation-induced transgenerational genome instability in vivo. Int J Radiat Oncol Biol Phys. 2006 Oct 1;66(2):327-30.

Laird PW. Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics 11, 191-203.

Lande-Diner L, Zhang J, Ben-Porath I, Amariglio N, Keshet I, Hecht M, Azuara V, Fisher AG, Rechavi G, Cedar H. Role of DNA methylation in stable gene repression. J Biol Chem. 2007 Apr 20;282(16):12194-200.

Lee BH, Yegnasubramanian S, Lin X, Nelson WG. Procainamide is a specific inhibitor of DNA methyltransferase 1. J Biol Chem. 2005 Dec 9;280(49):40749-56.

Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ; Cancer Genome Atlas Research Network. Landscape of somatic retrotransposition in human cancers. Science. 2012 Aug 24;337(6097):967-71.

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK.. DNMT3A mutations in acute myeloid leukemia. N Engl J Med. 2010 Dec 16;363(25):2424-33.

Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet. 2002 Sep;3(9):662-73.

Li J, Wan Y, Ji Q, Fang Y, Wu Y. The role of microRNAs in B-cell development and function. Cell Mol Immunol. 2013 Mar;10(2):107-12. doi: 10.1038/cmi.2012.62. Epub 2013 Jan 14.

Li Y, Okuno Y, Zhang P, Radomska HS, Chen H, Iwasaki H, Akashi K, Klemsz MJ, McKercher SR, Maki RA, Tenen DG. Regulation of the PU.1 gene by distal elements. Blood. 2001 Nov 15;98(10):2958-65.

Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol. 2002 Jan;22(2):480-91.

Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M, Song X, Lesche R, Liebenberg V, Ebert M, Molnar B, Grützmann R, Pilarsky C, Sledziewski A. DNA methylation biomarkers for blood-based colorectal cancer screening. Clin Chem. 2008 Feb;54(2):414-23.

Loh ML, Rubnitz JE. TEL/AML1-positive pediatric leukemia: prognostic significance and therapeutic approaches. Curr Opin Hematol. 2002 Jul;9(4):345-52.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. Nature. 2005 Jun 9;435(7043):834-8.

Lübbert M. DNA methylation inhibitors in the treatment of leukemias, myelodysplastic syndromes and hemoglobinopathies: clinical results and possible mechanisms of action. Curr Top Microbiol Immunol. 2000;249:135-64.

Maia AT, van der Velden VH, Harrison CJ, Szczepanski T, Williams MD, Griffiths MJ, van Dongen JJ, Greaves MF. Prenatal origin of hyperdiploid acute lymphoblastic leukemia in identical twins. Leukemia. 2003 Nov;17(11):2202-6.

Maksakova IA, Mager DL, Reiss D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cell Mol Life Sci. 2008;65:3329-47.

Matthias P, Rolink AG. Transcriptional networks in developing and mature B cells. Nat Rev Immunol 2005; 5(6):497-508.

McCarthy MM, Auger AP, Bale TL, De Vries GJ, Dunn GA, Forger NG, Murray EK, Nugent BM, Schwarz JM, Wilson ME. The epigenetics of sex differences in the brain. J Neurosci. 2009 Oct 14;29(41):12815-23.

Menigatti M, Di Gregorio C, Borghi F, Sala E, Scarselli A, Pedroni M, Foroni M, Benatti P, Roncucci L, Ponz de Leon M, Percesepe A. Methylation pattern of different regions of the MLH1 promoter and silencing of gene expression in hereditary and sporadic colorectal cancer. Genes Chromosomes Cancer. 2001 Aug;31(4):357-61.

Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat Med. 1995 Jul;1(7):686-92.

Miranda TB1, Jones PA. DNA methylation: the nuts and bolts of repression. J Cell Physiol. 2007 Nov;213(2):384-90.

Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, Hows JM, Navarrete C, Greaves M. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proc Natl Acad Sci U S A. 2002 Jun 11;99(12):8242-7.

Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui CH, Relling MV, Evans WE, Shurtleff SA, Downing JR. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007 Apr 12;446(7137):758-64.

Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, White D, Hughes TP, Le Beau MM, Pui CH, Relling MV, Shurtleff SA, Downing JR. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008 May 1;453(7191):110-4.

National Cancer Institute (NCI). SEER Cancer Statistics Review, 1975-2006. http://seer.cancer.gov/archive/csr/1975_2006/results_merged/sect_13_leukemia.pdf

Nekrutenko A, Li WH. Transposable elements are found in a large number of human protein-coding genes. Trends Genet. 2001 Nov;17(11):619-21.

Ng SY, Yoshida T, Zhang J, Georgopoulos K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity. 2009 Apr 17;30(4):493-507.

Nimura K, Ishida C, Koriyama H, Hata K, Yamanaka S, Li E, Ura K, Kaneda Y. Dnmt3a2 targets endogenous Dnmt3L to ES cell chromatin and induces regional DNA methylation. Genes Cells. 2006 Oct;11(10):1225-37.

Nordlund J, Bäcklin CL, Wahlberg P, Busche S, Berglund EC, Eloranta ML, Flaegstad T, Forestier E, Frost BM, Harila-Saari A, Heyman M, Jónsson OG, Larsson R, Palle J, Rönnblom L, Schmiegelow K, Sinnett D, Söderhäll S, Pastinen T, Gustafsson MG, Lönnerholm G, Syvänen AC. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome Biol. 2013 Sep 24;14(9):r105.

Ohgane J, Yagi S, Shiota K. Epigenetics: the DNA methylation profile of tissue-dependent and differentially methylated regions in cells. Placenta. 2008 Mar;29 Suppl A:S29-35.

Okano M1, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999 Oct 29;99(3):247-57.

Onciu M. Acute lymphoblastic leukemia. Hematol Oncol Clin N Am 2009, 23(4):655-674.

Ozdemir F, Altinisik J, Karateke A, Coksuer H, Buyru N. Methylation of tumor suppressor genes in ovarian cancer. Exp Ther Med. 2012 Dec;4(6):1092-1096.

Paz MF, Avila S, Fraga MF, Pollan M, Capella G, Peinado MA, Sanchez-Cespedes M, Herman JG, Esteller M. Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors. Cancer Res. 2002 Aug 1;62(15):4519-24.

Peng DF, Kanai Y, Sawada M, Ushijima S, Hiraoka N, Kosuge T, Hirohashi S. Increased DNA methyltransferase 1 (DNMT1) protein expression in precancerous conditions and ductal carcinomas of the pancreas. Cancer Sci. 2005 Jul;96(7):403-8.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006;444:499–502.

Pérez-Vera P, Reyes-León A, Fuentes-Pananá EM. Signaling proteins and transcription factors in normal and malignant early B cell development. Bone Marrow Res 2011; 2011:502751.

Pieper K, Grimbacher B, Eibel H. B-cell biology and development. Journal of Allergy and Clinical Immunology. April 2013, 131(4): 959–971.

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010 Jun 24;465(7301):1033-8.

Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009 Feb 20;136(4):629-41.

Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. 2001 Jul 1;15(13):1613-8.

Pufulete M, Al-Ghnaniem R, Khushal A, Appleby P, Harris N, Gout S, Emery PW, Sanders TA. Effect of folic acid supplementation on genomic DNA methylation in patients with colorectal adenoma. Gut. 2005 May;54(5):648-53.

Puget N, Gad S, Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S. Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot. Am J Hum Genet. 2002 Apr;70(4):858-65.

Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. Lancet. 2008 Mar 22;371(9617):1030-43.

Redaelli A, Laskin BL, Stephens JM, Botteman MF, Pashos CL. A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL). Eur J Cancer Care (Engl). 2005 Mar;14(1):53-62.

Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature. 2007 May 24;447(7143):425-32.

Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science. 2001 Aug 10;293(5532):1089-93.

Robertson KD1, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, Jones PA. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res. 1999 Jun 1;27(11):2291-8.

Robinson MD, Statham AL, Speed TP, Clark SJ. Protocol matters: which methylome are you actually studying? Epigenomics. 2010 Aug;2(4):587-98.

Rolink AG, Schaniel C, Busslinger M, Nutt SL, Melchers F. Fidelity and infidelity in commitment to B-lymphocyte lineage development. Immunol Rev. 2000 Jun;175:104-11.

Ross JP, Rand KN, Molloy PL. Hypomethylation of repeated DNA sequences in cancer. Epigenomics. 2010 Apr;2(2):245-69.

Ruijter AJMD, Gennip AHV, Carson HN, Kemp S, Kuilenburg ABPV. Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem. J. 2003; 370:737-749.

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol. 2000 Sep 15;225(2):294-303.

Santos PM, Borghesi L. Molecular resolution of the B cell landscape. Curr Opin Immunol. 2011 Apr;23(2):163-70.

Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012 Sep 6;489(7414):109-13.

Saied MH, Marzec J, Khalid S, Smith P, Down TA, Rakyan VK, Molloy G, Raghavan M, Debernardi S, Young BD. Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. PLoS One. 2012;7(3):e33213.

Saito Y, Liang G, Egger G, Friedman JM, Chuang JC, Coetzee GA, Jones PA. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell. 2006 Jun;9(6):435-43.

Sakai T, Toguchida J, Ohtani N, Yandell DW, Rapaport JM, Dryja TP. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. Am J Hum Genet. 1991 May;48(5):880-8.

Sato N, Fukushima N, Hruban RH, Goggins M. CpG island methylation profile of pancreatic intraepithelial neoplasia. Mod Pathol. 2008 Mar;21(3):238-44.

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet. 2007 Feb;39(2):232-6.

Sérandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Métivier R, Salbert G, Eeckhoute J. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res. 2011 Apr;21(4):555-65.

Segura-Pacheco B, Trejo-Becerril C, Perez-Cardenas E, Taja-Chayeb L, Mariscal I, Chavez A, Acuña C, Salazar AM, Lizano M, Dueñas-Gonzalez A. Reactivation of tumor suppressor genes by the cardiovascular drugs hydralazine and procainamide and their potential use in cancer therapy. Clin Cancer Res. 2003 May;9(5):1596-603.

Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK. Global histone modification patterns predict risk of prostate cancer recurrence. Nature. 2005;435:1262–1266.

Sha K. A mechanistic view of genomic imprinting. Annu Rev Genomics Hum Genet. 2008;9:197-216.

Shen L, Kondo Y, Rosner GL, Xiao L, Hernandez NS, Vilaythong J, Houlihan PS, Krouse RS, Prasad AR, Einspahr JG, Buckmeier J, Alberts DS, Hamilton SR, Issa JP. MGMT promoter methylation and field defect in sporadic colorectal cancer. J Natl Cancer Inst. 2005 Sep 21;97(18):1330-8.

Shivakumar R, Tan W, Wilding GE, Wang ES, Wetzler M. Biologic features and treatment outcome of secondary acute lymphoblastic leukemia--a review of 101 cases. Ann Oncol. 2008 Sep;19(9):1634-8.

Shurtleff SA, Buijs A, Behm FG, Rubnitz JE, Raimondi SC, Hancock ML, Chan GC, Pui CH, Grosveld G, Downing JR. TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. Leukemia. 1995 Dec;9(12):1985-9.

Sinzelle L, Izsvák Z, Ivics Z. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci. 2009 Mar;66(6):1073-93.

Song CX, Clark TA, Lu XY, Kislyuk A, Dai Q, Turner SW, He C, Korlach J. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat Methods. 2011 Nov 20;9(1):75-7.

Spector LG, Charbonneau B, Robison LL. Epidemiology and etiology. Childhood Leukemias (Third edition). Edited by Pui C. New York, Cambridge University Press; 2006, page. 48-66.

Spooner CJ, Cheng JX, Pujadas E, Laslo P, Singh H. A recurrent network involving the transcription factors PU.1 and Gfi1 orchestrates innate and adaptive immune cell fates. Immunity. 2009 Oct 16;31(4):576-86.

Strathdee G, Davies BR, Vass JK, Siddiqui N, Brown R. Cell type-specific methylation of an intronic CpG island controls expression of the MCJ gene. Carcinogenesis. 2004 May;25(5):693-701.

Sweatt JD. Experience-dependent epigenetic modifications in the central nervous system. Biol Psychiatry. 2009 Feb 1;65(3):191-7.

Takai D, Jones PA. Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. Mol Biol Evol. 2004 Mar;21(3):463-7.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. 2008 May 22;453(7194):534-8.

Townsend DM, Tew KD. The role of glutathione-S-transferase in anti-cancer drug resistance. Oncogene. 2003 Oct 20;22(47):7369-75.

Trinh BN, Long TI, Nickel AE, Shibata D, Laird PW. DNA methyltransferase deficiency modifies cancer susceptibility in mice lacking DNA mismatch repair. Mol Cell Biol. 2002 May;22(9):2906-17.

Tsai HC, Li H, Van Neste L, Cai Y, Robert C, Rassool FV, Shin JJ, Harbom KM, Beaty R, Pappou E, Harris J, Yen RW, Ahuja N, Brock MV, Stearns V, Feller-Kopman D, Yarmus LB, Lin YC, Welm AL, Issa JP, Minn I, Matsui W, Jang YY, Sharkis SJ, Baylin SB, Zahnow CA.

Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. Cancer Cell. 2012 Mar 20;21(3):430-46.

Uckun FM, Nachman JB, Sather HN, Sensel MG, Kraft P, Steinherz PG, Lange B, Hutchinson R, Reaman GH, Gaynon PS, Heerema NA. Clinical significance of Philadelphia chromosome positive pediatric acute lymphoblastic leukemia in the context of contemporary intensive therapies: a report from the Children's Cancer Group. Cancer. 1998 Nov 1;83(9):2030-9.

Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell. 2002 Jan 25;108(2):171-82. Cancer Res. 2003 Aug 15;63(16):4984-9.

Villar-Garea A, Fraga MF, Espada J, Esteller M. Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. Cancer Res. 2003 Aug 15;63(16):4984-9.

Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F. The Polycomb group protein EZH2 directly controls DNA methylation. Nature. 2006 Feb 16;439(7078):871-4.

Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature 2009;461:199-205.

van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ, Lankester AC, Révész T, Staal FJ, van Dongen JJ. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. J Immunol 2005; 175(9):5912-22.

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet. 1999 Sep;23(1):62-6.

Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature Genet 1998;20:116-117.

Wang J, Pitarque M, Ingelman-Sundberg M. 3'-UTR polymorphism in the human CYP2A6 gene affects mRNA stability and enzyme expression. Biochem Biophys Res Commun. 2006 Feb 10;340(2):491-7.

Wang L, Guo ZY, Zhang R, Xin B, Chen R, Zhao J, Wang T, Wen WH, Jia LT, Yao LB, Yang AG. Waterland RA, Lin JR, Smith CA, Jirtle RL. Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. Hum Mol Genet. 2006 Mar 1;15(5):705-16.

Watanabe Y, Maekawa M. Methylation of DNA in cancer. Adv Clin Chem. 2010;52:145-67

Weber B, Stresemann C, Brueckner B, Lyko F. Methylation of human microRNA genes in normal and neoplastic cells. Cell Cycle. 2007 May 2;6(9):1001-5.

Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ, Thurman RE, Stamatoyannopoulos JA, Hager GL. EMBO J. 2011 Jun 24;30(15):3028-39.

Wild L, Flanagan JM. Genome-wide hypomethylation in cancer may be a passive consequence of transformation. Biochim Biophys Acta. 2010 Aug;1806(1):50-7.

Wilken MS, Brzezinski JA, La Torre A, Siebenthall K, Thurman R, Sabo P, Sandstrom RS, Vierstra J, Canfield TK, Hansen RS, Bender MA, Stamatoyannopoulos J, Reh TA. DNase I hypersensitivity analysis of the mouse brain and retina identifies region-specific regulatory elements. Epigenetics Chromatin. 2015 Feb 28;8:8.

Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. Biochim Biophys Acta. 2007 Jan;1775(1):138-62.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 2005;3:e7.

Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellström-Lindberg E, Tefferi A, Bloomfield CD. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009 Jul 30;114(5):937-51.

Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J, Bronson RT, Crowley D, Stone JR, Jaenisch R, Sharp PA, Jacks T. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell. 2008 Mar 7;132(5):875-86.

Wang H, Lee CH, Qi C, Tailor P, Feng J, Abbasi S, Atsumi T, Morse HC 3rd. IRF8 regulates B-cell lineage specification, commitment, and differentiation. Blood. 2008 Nov 15;112(10):4028-38.

Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011 Sep 16;43(6):904-14.

Weaver RF. Molecular Biology, Fourth edition. McGraw-Hill Publication.

Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, Craig J, Harrison DK, Aung H, Phutikanit N, Lyle R, Meachem SJ, Antonarakis SE, de Kretser DM, Hedger MP, Peterson P, Carroll BJ, Scott HS. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proc Natl Acad Sci U S A. 2005 Mar 15;102(11):4068-73.

Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, Saha V, Biondi A, Greaves MF. Prenatal origin of acute lymphoblastic leukaemia in children. Lancet. 1999 Oct 30;354(9189):1499-503.

Wong IH, Lo YM, Zhang J, Liew CT, Ng MH, Wong N, Lai PB, Lau WY, Hjelm NM, Johnson PJ. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. Cancer Res. 1999 Jan 1;59(1):71-3.

Xiao C, Calado DP, Galler G, Thai TH, Patterson HC, Wang J, Rajewsky N, Bender TP, Rajewsky K. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell. 2007 Oct 5;131(1):146-59.

Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X. Pseudogene in cancer: real functions and promising signature. J Med Genet. 2015 Jan;52(1):17-24.

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Péquignot E. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature. 1999 Nov 11;402(6758):187-91.

Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, Zaret KS, Weissman IL, Smale ST. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. Proc Natl Acad Sci U S A. 2007 Jul 24;104(30):12377-82.

Xue B, Lin H. An expanding universe in the non-coding genome in cancer biology. Carcinogenesis. 2014 Jun;35(6):1209-16.

Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL et al. Array-based genomic resequencing of human leukemia. Oncogene. 2010 Jun 24;29(25):3723-31.

Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet. 2011 Mar 13;43(4):309-15.

Yang X, Smith SL, Tian XC, Lewin HA, Renard JP, Wakayama T. Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. Nat Genet. 2007 Mar;39(3):295-302.

Yen ZC, Meyer IM, Karalic S, Brown CJ. A cross-species comparison of X-chromosome inactivation in Eutheria. Genomics. 2007 Oct;90(4):453-63.

Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol. 1997 Jul 18;270(3):385-95.

Yoo CB, Chuang JC, Byun HM, Egger G, Yang AS, Dubeau L, Long T, Laird PW, Marquez VE, Jones PA. Long-term epigenetic therapy with oral zebularine has minimal side effects and prevents intestinal tumors in mice. Cancer Prev Res (Phila). 2008 Sep;1(4):233-40.

Zambrano P, Segura-Pacheco B, Perez-Cardenas E, Cetina L, Revilla-Vazquez A, Taja-Chayeb L, Chavez-Blanco A, Angeles E, Cabrera G, Sandoval K, Trejo-Becerril C, Chanona-Vilchis J, Duenas-González A. A phase I study of hydralazine to demethylate and reactivate the expression of tumor suppressor genes. BMC Cancer. 2005 Apr 29;5:44.

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999 Aug 1;13(15):1924-35.

Zhou B, Wang S, Mayr C, Bartel DP, Lodish HF. miR-150, a microRNA expressed in mature B and T cells, blocks early B cell development when expressed prematurely. Proc Natl Acad Sci U S A. 2007 Apr 24;104(17):7080-5.

Zhou L, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. J Mol Biol. 2002 Aug 23;321(4):591-9.

CHAPTER 3: ISOLATION OF PRECURSOR-B CELL SUBSETS FROM HUMAN UMBILICAL CORD BLOOD

3.1 ABSTRACT

Umbilical cord blood is highly enriched for hematopoietic progenitor cells at different lineage commitment stages. We have developed a protocol for isolating precursor B-cells at four different stages of differentiation. Because genes are expressed and epigenetic modifications occur in a tissue specific manner, it is vital to discriminate between tissues and cell types in order to be able to identify alterations in the genome and the epigenome that may lead to the development of disease. This method can be adapted to any type of cell present in umbilical cord blood at any stage of differentiation.

This method comprises 4 main steps. First, mononuclear cells are separated by density centrifugation. Second, B-cells are enriched using biotin conjugated antibodies that recognize and remove non B-cells from the mononuclear cells. Third the B-cells are fluorescently labeled with cell surface protein antibodies specific to individual stages of B-cell development. Finally, the fluorescently labeled cells are sorted and individual populations are recovered. The recovered cells are of sufficient quantity and quality to be utilized in downstream nucleic acid assays.

3.2 INTRODUCTION

In order to identify aberrations that are present in disease, it is vitally important that we use healthy tissues or cells that correspond to the tissue or cell type affected by the disease. One reason for this is that epigenetic variation among tissue types is responsible for regulating gene expression and is critical for cellular differentiation during normal human development (Song et al. 2005; Ohgane et al. 2008). A second reason is that aberrant tissue specific gene regulation may have dire consequences on normal development and is known to contribute to a multitude of disease states including cancer. Therefore, a better understanding of a disease that involves hematopoietic cells requires knowledge of healthy hematopoietic cells.

The development of hematopoietic cells in the bone marrow proceeds through a systematic order of events characterized by changes in the expression of cell surface markers (Brown et al. 2007). Studies involving adult participants have shown that bone marrow usually contains a low number of precursor B-cells (Clark et al. 1987, Caldwell et al. 1991); whereas studies involving pediatric participants have shown that the percentage of precursor B-cells is relatively high in individuals less than 5 years of age (Tucci et al. 1991). Umbilical cord blood is used as a source of hematopoietic stem cells in the treatment of blood related disorders and malignancies, is readily available via cord blood banks and is enriched for immature B and T cells (Ghia et al.1996) which are the target cells of multiple disorders including leukemia and lymphomas.

Precursor B-cells in the bone marrow have been extensively phenotyped (Noordzij et al. 2002) and can be defined by the presence of specific cell surface markers that can be used to sort these cells into distinct subsets. Normal B-cell differentiation proceeds through a series of stages

in the bone marrow beginning with the earliest pro-B cells and culminating in immature or naïve B-cells. According to van Zelm and colleagues (van Zelm et al. 2005), pro-B cells are characterized by the presence of CD34 and in the transition to stage 2 (Pre-BI) CD19 is acquired. Stage 3 (Pre-BII) cells no longer express CD34 and begin to express cytoplasmic IgM. Finally, a defining characteristic of stage 4 (immature B-cells) is the expression of surface IgM. The sorting strategy described in this protocol was first described by Caldwell and colleagues6 and includes the use of only 3 cell surface markers which greatly reduces the complexity and the cost of performing cell sorting experiments. In their work, a relationship between CD45 and the stages of B-cell differentiation was established. They observed that B-cells in the bone marrow display variable levels of expression of CD45. Specifically, cells that expressed high levels of CD45 corresponded to cells that expressed surface IgM (immature B-cells), those that expressed an intermediate level of CD45 corresponded to cells that expressed cytoplasmic IgM (pre-BII cells), and those that expressed low levels of CD45 corresponded to cells that did not express cytoplasmic IgM (pre-BI cells). This protocol uses the strategy developed by Caldwell and colleagues (Caldwell et al. 1991) to isolate subsets of precursor B-cells from umbilical cord blood (**Figure 3.1**) which can be used in downstream assays requiring high quality nucleic acids such as the methylated-CpG island recovery assay (MIRA) (Rauch et al. 2009) and quantitative real time PCR assays. The method employs an initial separation using magnetic beads to deplete all non-B cells from umbilical cord blood and requires staining with only 3 antibodies (CD34, CD19 and CD45). The cells that are recovered represent 4 stages of B-cell differentiation: 1) CD34+; CD19+ (late pro-B and early pre-BI); 2) CD34-; CD19+; CD45low (late pre-BI); 3) CD34-; CD19+; CD45med (pre-BII); and 4) CD34-; CD19+; CD45high (immature B-cells).

3.3 PROTOCOL

3.3.1 Isolation of mononuclear cells from umbilical cord blood

1. Prepare EDTA-PBS buffer- add 5 ml bovine serum albumin (BSA) stock solution to 95 ml rinsing buffer (1:20 dilution). Degas the buffer and keep the buffer on ice. IMPORTANT: Failure to degas the buffer may result in less than optimal results when isolating CD19+ B-cells because bubbles may block the isolation column.
2. Prepare 50 ml conical tubes for density gradient centrifugation. Determine the number of tubes required for processing the cord blood (1 tube can process 8 ml blood) and add 15 ml of Ficoll-Paque PLUS to each tube.
3. Dilute 8 ml of cord blood with 24 ml DPBS (1X) and carefully layer the diluted cord blood mixture on top of the Ficoll-Paque PLUS in each of the 50 ml conical tubes. Do not mix the blood and Ficoll-Paque PLUS. IMPORTANT: To avoid mixing of the cord blood and Ficoll-Paque PLUS, hold the tube at a 45 degree angle and layer the blood mixture slowly.
4. Centrifuge at 400 x g for 40 min at 20 °C. Mononuclear cells (MNC) will remain at the plasma-Ficoll-Paque PLUS interface whereas granulocytes and erythrocytes sediment due to higher density at the osmotic pressure of Ficoll-Paque PLUS. Label seven 5 ml round-bottom tubes to be used in Part 3 with the sample ID, date and the following:

No.	Sample ID	Labeled	Purpose
1	HCB001	Unstained	To keep the unstained cells for normalization during Flow Cytometry
2	HCB001	7AAD	To keep 7-AAD stained cells
3	HCB001	+++	To keep CD19, CD34, CD45 labeled B-cells
4	HCB001	34+	To collect CD19+/CD34+ cells during cell sorting
5	HCB001	45low	To collect CD19+/CD34+/CD45low cells during cell sorting
6	HCB001	45med	To collect CD19+/CD34+/CD45med cells during cell sorting
7	HCB001	45high	To collect CD19+/CD34+/CD45high cells during cell sorting

Coat tubes No. 4, 5, 6, and 7 with 2% FBS and place all tube on ice.

5. Aspirate the upper plasma layer carefully and avoid contact with the mononuclear cell layer. Using a 10 ml glass pipette, carefully transfer the mononuclear cell layer to a new 50 ml conical tube. Combine the mononuclear cells from three tubes together into a single 50 ml tube.
6. Fill the tube with PBS, mix gently and centrifuge at 300 x g for 10 min at 20 °C. Carefully aspirate the supernatant without disturbing the cell pellet. Repeat 1x. After the first wash, resuspend the pellets and transfer to a single 50 ml conical tube.
7. Gently resuspend the cell pellet in 200 µl of PBS. Remove 1 µl of the cell suspension, add it to 1 ml of PBS in a 1.5 ml microcentrifuge tube and set aside for counting (count cells after placing the 50 ml cell suspension in the centrifuge). Fill the tube with PBS and centrifuge at 200 x g for 15 min to remove platelets. Remove the supernatant completely without disturbing the cell pellet.

3.3.2 Modified B-cell isolation procedure from mononuclear cells using MACS separation

1. Resuspend the pellet from step 1.7 with 160 µl cold (4 °C) EDTA-PBS buffer.
2. Add 40 µl B-CLL biotin antibody cocktail. Pipette to mix well and incubate for 10 min at 4 °C.

3. Wash cells - add 1 ml cold (4 °C) EDTA-PBS buffer per 10 million cells (cell number determined in step 1.7) and centrifuge at 300 x g for 10 min. Aspirate the supernatant completely and then resuspend the cell pellet in 320 µl cold (4 °C) EDTA-PBS buffer.
4. Add 80 µl anti-biotin MicroBeads, mix well and incubate for 15 min at 4 °C.
5. Wash cells - add 1 ml cold (4 °C) EDTA-PBS buffer per 10 million cells and centrifuge at 300 x g for 10 min. Resuspend up to 100 million cells in 500 µl of cold (4 °C) EDTA-PBS buffer. Scale buffer volume according to cell number.
6. Prepare MACS Columns and MACS Separators during centrifugation (Step 2.5). Place an LS column in the magnetic field of the MACS separator, add 3 ml of cold (4 °C) EDTA-PBS buffer onto the column and allow the buffer to drip through the column. Use a receptacle to catch the flow through. Discard the flow through.
7. Place a 50 ml conical tube below the column and pipette the cell suspension onto the column. Allow the unlabeled cells to drip through the column and into the 50 ml conical tube. IMPORTANT: These are the cells that will be used for antibody labeling and cell sorting. Do not discard the flow through. To recover all of the cells from step 2.5, add an additional 1 ml of cold (4 °C) EDTA-PBS buffer to the walls of the conical tube. Mix gently and pipette the remaining cell suspension onto the column. Repeat 1x. Proceed to step 2.8 using the unlabeled cells collected in the conical tube after passing through the column.
8. Fill the conical tube containing unlabeled cells with PBS and centrifuge at 300 x g for 10 min. Carefully remove the supernatant without disturbing the cell pellet. Add 200 µl of EDTA-PBS buffer and gently resuspend the cells.

3.3.3 Antibody labeling and preparation for cell sorting

1. Aspirate 1 µl of the cell suspension and place into tube labeled "unstained" (Step 1.4). Add 500 µl EDTA-PBS buffer and store the tube on ice.
2. Add 20 µl of each antibody (CD19, CD34 and CD45) per million cells into the remaining cell suspension. Mix well and place the tube in the dark for 30 min at RT. CD antibodies are light sensitive, complete steps 2-4 with the lights turned off.
3. After 30 min of incubation with antibodies, aspirate 40 µl of the cell suspension and place it into the tube labeled "7AAD". Add 1 ml PBS into the tube and centrifuge at 500 x g for 2 min. Remove the supernatant and resuspend the cells in 100 µl of binding buffer. Add 7 µl of 7AAD (7-Aminoactinomycin D) and incubate in the dark for 10 min at RT. During the 10 min incubation complete step 3.4.
4. Add PBS to the top of the tube containing the remaining cells labeled with CD antibodies. Centrifuge the tube at 500 x g for 3 min. Remove the supernatant, add 500 µl EDTA-PBS buffer and resuspend the cell pellet. Transfer the entire volume of the cell suspension into the tube labeled "+++". To recover all of the cell suspension add an additional 1 ml EDTA-PBS buffer to the 50 ml conical tube and transfer into the tube labeled "+++".
5. After incubation (Step 3.3), add 300 µl binding buffer into 7AAD tube. The "unstained", "7AAD" and "+++" samples and the "34 $^{+}$", "45^{low}", "45^{med}" and "45^{high}" tubes are ready for flow cytometry.

3.3.4 Cell sorting using the MoFlo XDP flow cytometer

1. Set-up MoFlo for sorting: Align lasers; stabilize droplet stream; determine drop delay.

2. Prepare single color compensation controls using Invitrogen AbC Mouse bead kit (or similar) according to manufacturer's instructions. Run single color controls, adjusting the voltages of fluorescence channels for optimum separation of positive and negative populations. Set compensation coefficients and apply compensated parameter to collection protocol. Re-establish compensation coefficients each time a new lot of antibody is obtained.
3. Create protocol including plots as shown in **Figure 3.2.**
4. Run unstained sample, set voltage and gain for forward and side scatter, and identify negative fluorescence populations. Because DNA recovery is the goal of the sort, the threshold should be set low (≤1%) so that DNA-containing debris does not contaminate the recovered samples. *Ensure that the Aerosol Evacuation System is running at all times that live human cells are being run on the MoFlo.
5. Run a small aliquot of the +++ sample (~ 50,000 events) to set the gating strategy (**Figure 3.2**). Because B-cell progenitors do not scatter identically to mature B-cells, backgate the ungated CD19+/CD34+ population onto the SSC vs FSC plot to ensure the lymphocyte gate includes potential Pro-B cells. From this sample also set the negative fluorescence population for 7AAD.
6. Run the 7AAD sample to determine sample viability. Only sort cells from a sample with high viability at the outset (≥95%) as dead cells will stain indiscriminately and may contaminate sort populations.
7. Set up sort decisions to collect four populations: $CD19^{+}/CD34^{+}$; $CD19^{+}/CD34^{-}/CD45^{low}$; $CD19^{+}/CD34^{-}/CD45^{med}$; and $CD19^{+}/CD34^{-}/CD45^{high}$. Include the lymphocyte and doublet discrimination gates in the sort decisions of all populations.

8. To prevent clogs in the sorting tip, filter +++ sample through a 40 μm cell strainer immediately before sorting and re-filter if any aggregation occurs in sample during sort.
9. Sort the cells into collection tubes (Step 1.4) coated with 2% FBS in PBS in a cooled or ice-packed tube holder.
10. Immediately after sort is completed, extract DNA.

3.4 REPRESENTATIVE RESULTS

Between samples variation plays a role in the success of the cell sort (Table 3.1). The samples with good success rates have low levels of contaminating debris (Figure 3.3A) and the samples with poor success rates have high levels of contaminating debris (Figure 3.3B). Between sample variation can be somewhat controlled if the cord blood sample was collected within 24 hr of shipment (O/N first priority).

The flow sorted cells from each precursor B-cell subset are of sufficient quantity and quality to perform nucleic acid isolations (Table 3.2). The DNA isolated is of high quality and can be used in downstream analyses. We routinely utilize this DNA in MIRA (Rauch et al. 2009) to enrich for methylated DNA (Figure 3.4).

3.5 DISCUSSION

The factor with the greatest impact on the success of the protocol is the presence of contaminating debris. If requesting blood from a cord blood bank it is important to have the blood shipped as soon after collection as possible. In addition, samples that are classified as

lymphocytosis contain higher numbers of lymphocytes; however, these samples do not have adequate numbers of precursor B-cells and should not be used. To increase the probability of obtaining sufficient numbers of cells for each of the precursor subsets we recommend beginning with at least 85 ml of cord blood.

It is important to note that unlike adult peripheral blood separations, when fractionating umbilical cord blood the mononuclear layer is often contaminated with red blood cells (Kanof EM, et al. 1996). This protocol includes an extra wash step to remove contaminating platelets. Protocols describing mononuclear separations from cord blood suggest including a lysis step to remove unwanted red cells. We do not recommend this step because it produces contaminating debris and has a negative impact on the success of the cell sort.

In order to reduce the number of cells that have to be distinguished by flow cytometry it is necessary to perform a B-cell enrichment prior to cell sorting. The protocol provided by Miltenyi Biotec, that accompanies the B-Cell Isolation Kit (B-CLL), has been optimized for peripheral blood and recommends using 10 µl B-CLL biotin antibody cocktail per 10 million cells. This step uses antibodies against CD2 (T cells; NK cells), CD4 (T cells), CD11b (granulocytes; monocytes; macrophages), CD16 (NK cells; macrophages; mast cells), CD36 (platelets), CD235a (erythroid cells) to deplete non-B cells from the umbilical cord blood. It is important to use the kit described and not the B-cell isolation kit II because the former kit contains an antibody against CD43 which is present on pro-B cells. During our optimization of this protocol, we found that a total of 40 µl of B-CLL biotin antibody cocktail is sufficient to produce a positive cell sort result. In addition we found that using as little as 5 µl more of the B-CLL biotin antibody cocktail had adverse effects on the cell sort. Therefore, we highly

recommend using 40 µl of B-CLL biotin antibody cocktail for total mononuclear cell numbers between 175-250 million. If starting with lower or higher numbers of cells it may be necessary to scale the reagents accordingly.

There are numerous conflicting publications describing the markers that may be used to distinguish sub-populations of B-cells. Most of the discrepancy can be explained by the fact that differentiation is an on-going process and that the presence or absence of cell surface markers occurs in gradual manner instead of in an all or nothing manner. In this protocol CD34-/CD19+and the intensity level of CD45+ (low, medium, high) expression was used to distinguish pre-BI, pre-BII and immature B-cells with an increase in the level of CD45 expression corresponding to the progression of differentiation6. Pro-B cells have been described by some to be CD19+/CD34+ (LeBien 2002) while others have shown that pro-B cells are CD19-/CD34+ and pre-BI cells are CD19+/CD34+ (Noordzij et al. 2002; van Zelm et al. 2005). Based on these discrepancies we have designated the CD19+/CD34+cells as late pro-B cells/early pre-BI cells. It is important to note that this strategy was developed to isolate subsets of precursor B-cells that correspond to the cells affected by the disease acute lymphoblastic leukemia. However, there are multiple sorting strategies that may be employed depending on the cell subtype of interest.

Antibody-fluorochrome combinations should be selected based on the capabilities of the available cell sorter. As a general rule, the brightest fluorochrome in your panel should be used to label the least populous antigen and vice versa. In detecting double-stained populations, especially rare events such as these, doublet discrimination (**Figure 3.2B**) is a vitally important part of the gating strategy. This ensures that double-positive events are truly single cells stained with both antibodies and not merely two single-stained cells adhered to one another.

High speed, 4-way cell sorting necessarily creates a controlled aerosol environment. Therefore, live human cell sorting should be performed with great care. Published safety and decontamination guidelines are available and should be reviewed and implemented before the sort (Schmid et al. 2007). Approval from Institutional Biosafety Committees or their equivalents may be required prior to sorting. For proper decontamination after sorting, bleach should be added to the waste container to a final concentration of 10%. Similarly, all samples tubing as well as all surfaces in the immediate area should be thoroughly decontaminated with a freshly made 10% bleach solution.

In summary, this protocol provides a strategy for obtaining rare populations of precursor B-cells and may be modified to isolate any rare population present in cord blood including hematopoietic stem cells and immature T-cells. Recently, immature B cells have been identified in the peripheral blood of individuals with advanced HIV (Malaspina et al. 2006). Therefore, the utility of this method extends beyond the study of blood cancers. Finally, we have not yet performed RNA isolations on the flow sorted cells; however, this method should be adaptable with the caveat that during cell sorting the cells should be sorted directly into trizol or an equivalent RNA compatible solution such as RLT from the RNeasy kit available through Qiagen.

Table 3.1: Representative cell sort statistics. Rows 1-5 are counts provided by the cord blood facility. The remaining rows are data collected in our laboratory. The poor sort contained high levels of debris and a low percentage of cells in the lymphocyte gate.

Cord Blood Facility Data	**Poor Sort**	**Good Sort**
Working Blood Volume (mL)	110	104
White Blood Cells [10^3/ µl]	8.68	11.19
Lymphocytes [10^3/ µl]	3.88	3.76
Lymphocyte (%)	44.70	33.6
Red Blood Cells [10^6/ µl]	3.65	3.05
In-house Data		
Cell Number after Ficoll-Paque	206 M	309 M
Cell Number after MACS® Separation	22 M	16.5 M
Total Events Count (MoFlo XDP)	30.57 M	19.97 M
Viability (%)	98.00	98.00
Cells in Lymphocyte Gate (%)	17.00	50.00
$CD19^+/CD34^+$		
Total Cell Number	**10959**	**48316**
% of Total Events	0.04	0.24
Efficiency	88%	88%
$CD19^+/CD34^-/CD45^{low}$		
Total Cell Number	**16619**	**26941**
% of Total Events	0.05	0.13
Efficiency	89%	87%
$CD19^+/CD34^-/CD45^{med}$		
Total Cell Number	**469745**	**507540**
% of Total Events	1.54	2.54
Efficiency	89%	89
$CD19^+/CD34^-/CD45^{high}$		
Total Cell Number	**1896062**	**2047142**
% of Total Events	6.2	10.25
Efficiency	91%	90%

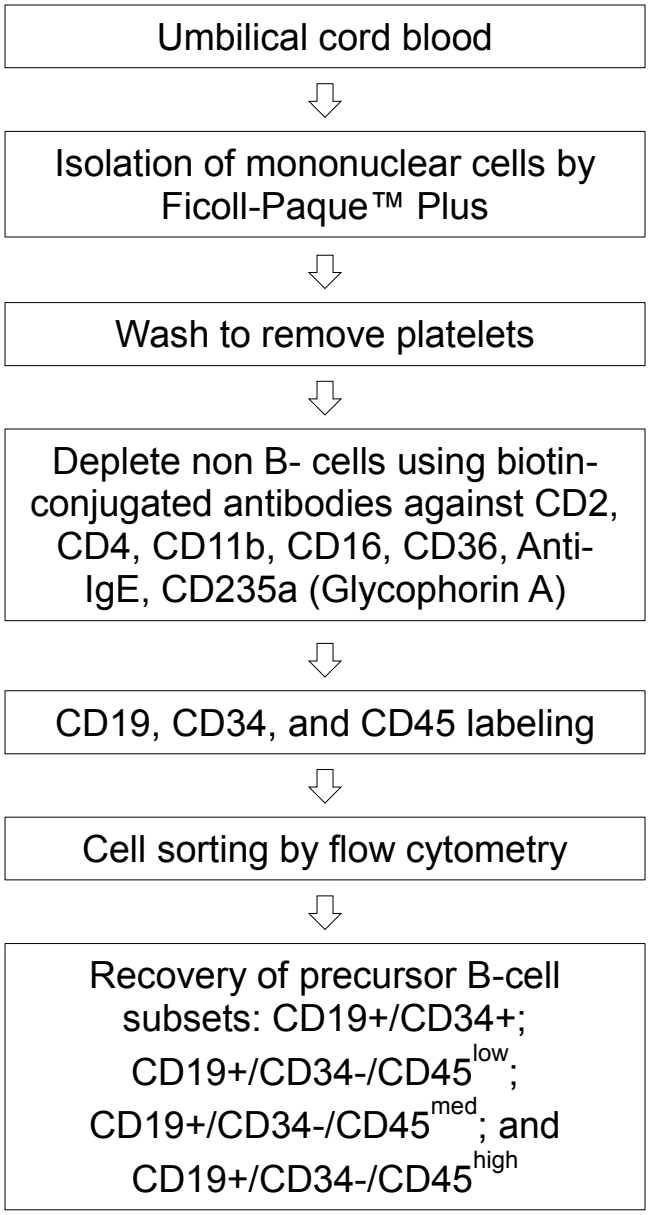


Figure 3.1: **Flow chart of cell isolation procedure**. Umbilical cord blood is processed and mononuclear cells are isolated using Ficoll-paque™ plus. An additional wash step is included to remove contaminating platelets. B-cells are isolated from mononuclear cells using the MACS human B-cell isolation kit (B-CLL). This step uses biotin-conjugated monoclonal antibodies (CD2, CD4, CD11b, CD36, Anti-IgE and CD235a) to remove all non-B cells. The recovered B-cells are labeled with CD19-APC, CD34-PE and CD45-FITC antibodies then sorted using the MoFlo XDP flow cytometer to recover precursor B-cell subsets: $CD19^{+}/CD34^{+}$; $CD19^{+}/CD34^{-}/CD45^{low}$; $CD19^{+}/CD34^{-}/CD45^{med}$; and $CD19^{+}/CD34^{-}/CD45^{high}$

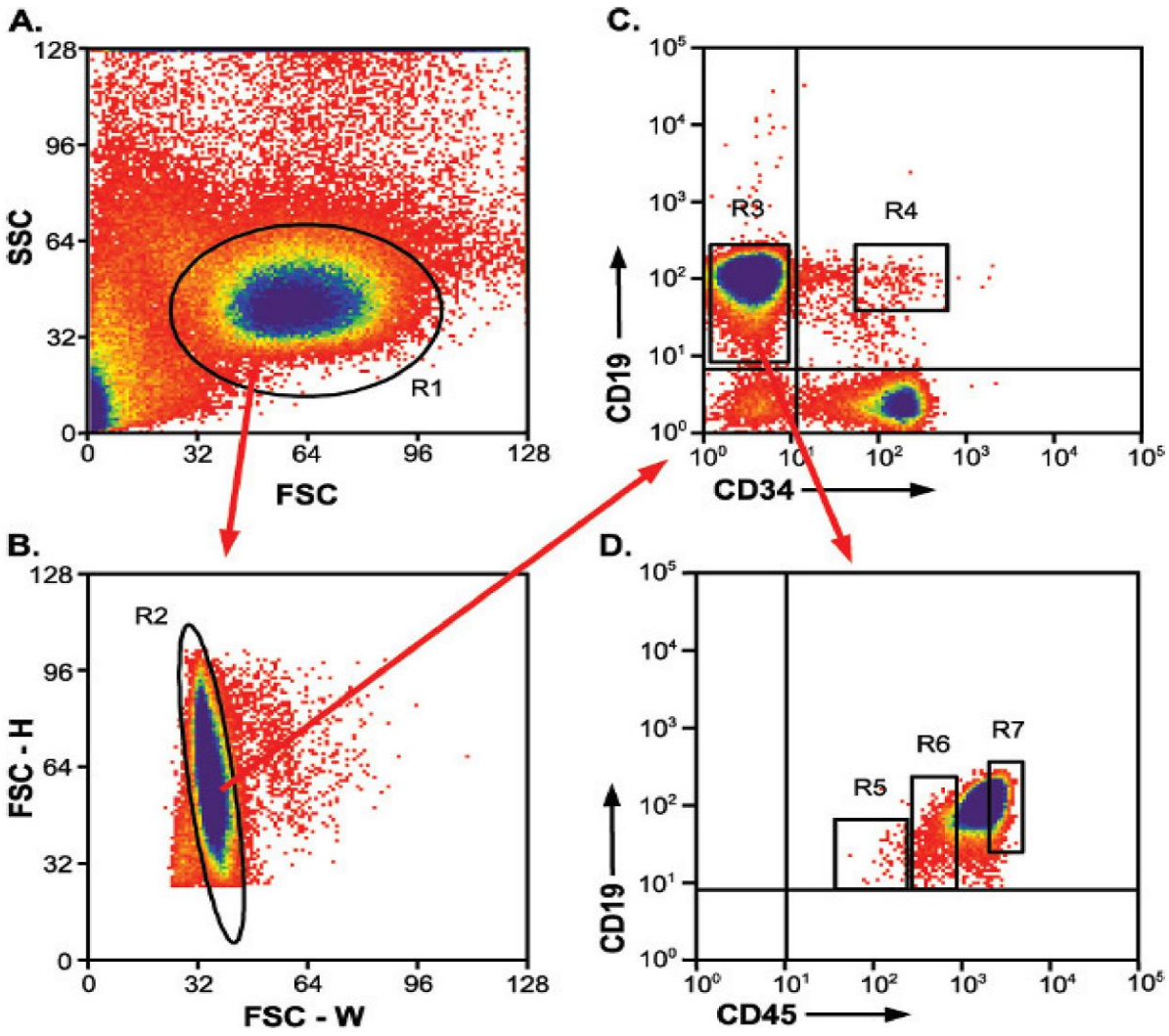


Figure 3.2: **Gating strategy for identifying and sorting populations**. Red arrows indicate the gate that is applied to subsequent plot. R4, R5, R6 and R7 gates indicate sorted populations.

A) Forward vs. side scatter plot showing enriched lymphocyte population. R1, lymphocyte gate.
B) Height vs. width of forward scatter plot for identifying single cells versus doublets. R2, single cell gate.
C) CD19-APC positive cells fall into CD34-PE negative and positive populations. R3, $CD19^{+}/CD34^{-}$ gate; R4, $CD19^{+}/CD34^{+}$ gate indicating desired sort population.
D) $CD19^{+}/CD34^{-}$ cells fall into three somewhat distinct CD45-FITC populations. R5 and R6, $CD19^{+}/CD34^{-}/CD45^{low}$ and $CD19^{+}/CD34^{-}/CD45^{med}$ populations, respectively. R7, because of high cell numbers in the $CD19^{+}/CD34^{-}/CD45^{high}$ population, this sort gate encompasses only a portion of the population. All cells of this population do not need to be collected for downstream analyses.

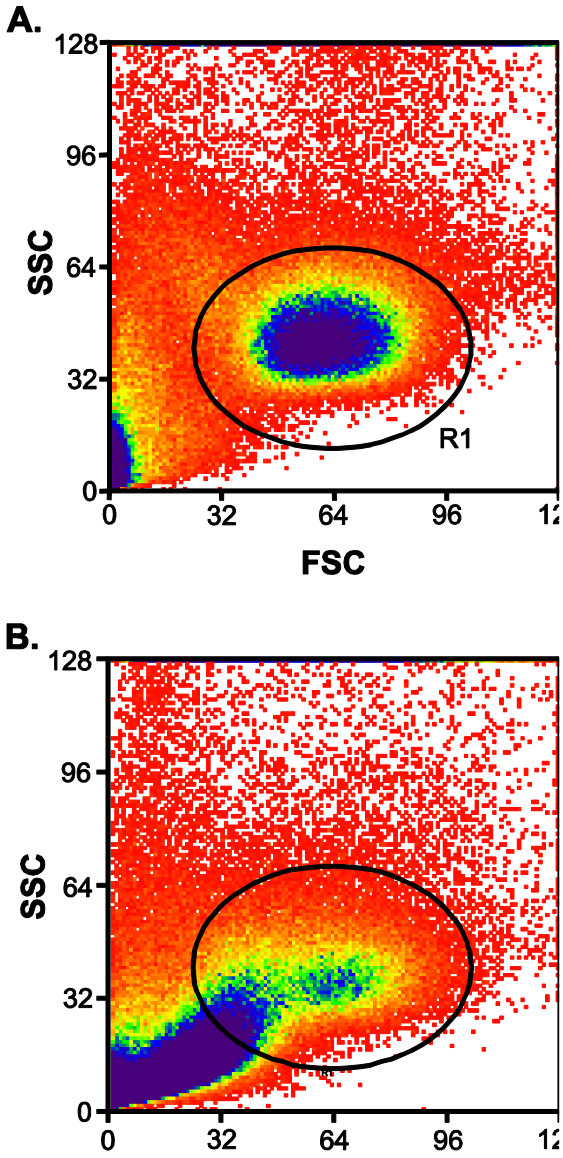


Figure 3.3: Contaminating debris after column enrichment.

a) Low levels of contaminating debris after column enrichment. R1, lymphocyte gate.
b) High levels of contaminating debris after column enrichment

Table 3.2: Sorted subsets of B-cell from human umbilical cord blood.

Sample ID	**Pro-B**	**Pre-BI**	**Pre-BII**	**Naïve B-cell**
	Cell Numbers	Cell Numbers	Cell Numbers	Cell Numbers
HCB1	34,764	49,122	866,140	3,997,888
HCB2	36,753	75,776	1,232,526	7,328,168
HCB3	40,201	27,005	563,853	5,338,513
HCB4	20,800	65,314	1,621,849	2,290,070
HCB5	21,305	98,983	1,157,008	1,846,331
HCB6	24,168	63,075	1,817,684	3,172,421
HCB7	33,912	43,644	785,269	1,837,981
HCB8	48,316	26,941	507,540	2,047,142
HCB9	28,844	35,757	408,855	3,381,926
HCB10	44,489	35,263	483,934	3,665,502

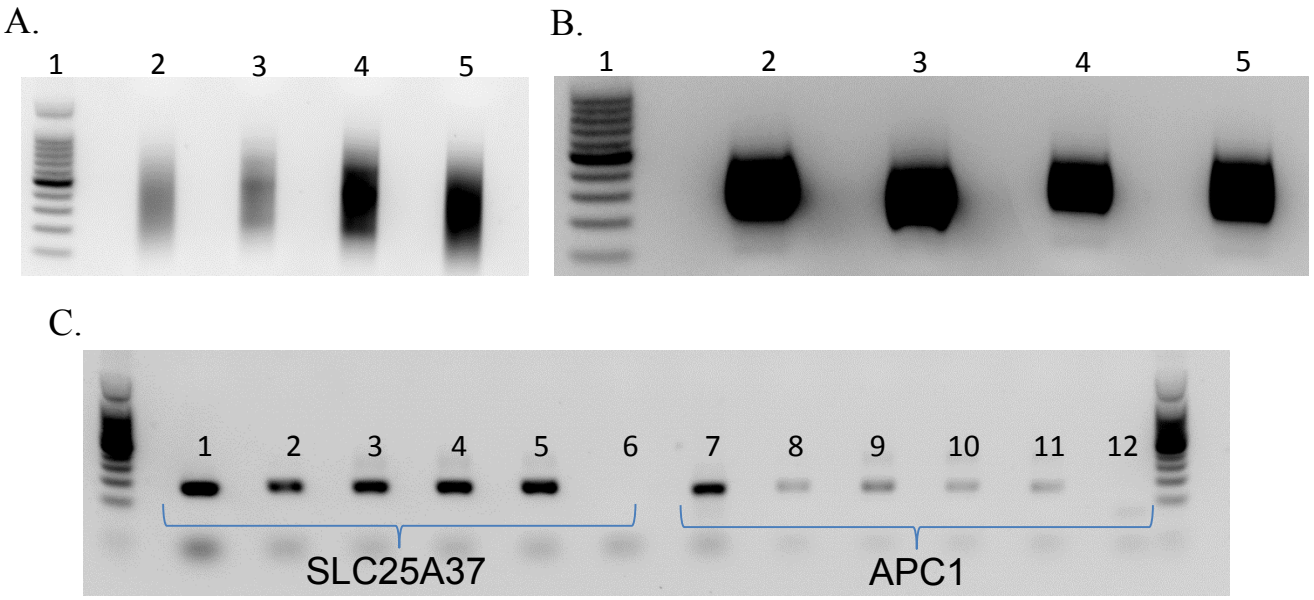


Figure 3.4: Enrichment of methylated DNA from subsets of precursor B-cells.

A) Sonicated DNA isolated from precursor B-cell subsets is of high quality. For library construction DNA is sonicated to an average of 200-600 bp. Lane 1: Promega 100 bp ladder (catalog #G2101); Lane 2: Total DNA isolated from CD19+/CD34+ cells; Lane 3: Total DNA isolated from $CD19^{+}/CD34^{-}/CD45^{low}$ cells; Lane 4: 100 ng DNA isolated from $CD19^{+}/CD34^{-}/CD45^{med}$ cells; Lane 5: 100 ng DNA isolated from $CD19^{+}/CD34^{-}/CD45^{high}$ cells. DNA was sonicated using the Diagenode Bioruptor on High for a total of 9 minutes (30 sec ON; 30 sec OFF). DNA was column purified and visualized on a 1% agarose gel with SYBR green nucleic acid gel stain.

B) MIRA libraries run on 1% agarose gel to check the size of DNA fragment.

C) After MIRA using the ActivMotif MethylCollector ™ Ultra kit, PCR with SLC25A37 (1-6) and APC1 (7-12) is performed to confirm enrichment of methylated DNA. 1&7: Sonicated genomic DNA (no MIRA); 2&8: MIRA-CD19+/CD34+ DNA; 3&9: MIRA-$CD19^{+}/CD34^{-}/CD45^{low}$ DNA; 4&10: MIRA-$CD19^{+}/CD34^{-}/CD45^{med}$ DNA; 5&11: MIRA-$CD19^{+}/CD34^{-}/CD45^{high}$ DNA; 6&12: Water control. Higher amplification of SLC25A37 confirms the enrichment of methylated DNA in the subsets of precursor B-cells.

CH3: REFERENCES

Brown G, Hughes PJ, Michell RH, Rolink AG, Ceredig R. The sequential determination model of hematopoiesis. Trends Immunol. 2007;28(10):442–448.

Caldwell CW, Poje E, Helikson MA. B-cell precursors in normal pediatric bone marrow. American Journal of Clinical Pathology. 1991;95(6):816–823.

Clark P, Normansell DE, Innes DJ, Hess CE. Lymphocyte subsets in normal bone marrow. Blood. 1986;67(6):1600–1606.

Ghia P, ten Boekel E, Sanz E, de la Hera A, Rolink A, Melchers F. Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. J. Exp. Med. 1996;184:2217–2219.

Kanof ME, Smith PD, Zola H. Isolation of whole mononuclear cells from peripheral blood and cord blood. Curr Protoc Immunol. 2001 May;Chapter 7:Unit 7.1.

LeBien TW. Fates of human B-cell precursors. Blood. 2000;96(1):9–23.

Loken MR, Shah VO, Dattilio KL, Civin CI. Flow cytometric analysis of human bone marrow. II. Normal B lymphocyte development. Blood. 1987;70(5):1316–1324.

Malaspina A, Moir S, Ho J, Wang W, Howell ML, O'Shea MA, Roby GA, Rehm CA, Mican JM, Chun TW, Fauci AS. Appearance of immature/transitional B cells in HIV-infected individuals with advanced disease: correlation with increased IL-7. Proc Natl Acad Sci U S A. 2006 Feb 14;103(7):2262-7.

Noordzij JG, de Bruin-Versteeg S, Comans-Bitter WM, Hartwig NG, Hendriks RW, de Groot R, van Dongen JJ.. Composition of precursor B-cell compartment in bone marrow from patients with X-linked agammaglobulinemia compared with healthy children. Pediatric Research. 2002;51(2):159–168.

Ohgane J, Yagi S, Shiota K. Epigenetics: The DNA methylation profile of tissue-dependent and differentially methylated regions in cells. Placenta. 2008;29(S):29–35.

Rauch TA, Pfeifer GP. The MIRA method for DNA methylation analysis. Methods Mol. Biol. 2009;507:65–75.

Schmid I, Lambert C, Ambrozak D, Marti GE, Moss DM, Perfetto SP; International Society of Analytical Cytology.. International society for analytical cytology biosafety standard for sorting of unfixed cells. Cytometry A. 2007;71(6):414–437.

Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, Held WA.. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A. 2005 Mar 1;102(9):3336-41.

Tucci A, Mouzaki A, James H, Bonnefoy JY, Zubler RH.. Are cord blood B cells functionally mature? Clin. Exp. Immunol. 1991;84(3):389–394.

van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ, Lankester AC, Révész T, Staal FJ, van Dongen JJ. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. J. Immunol. 2005;175(9):5912–5922.

CHAPTER 4: GENOME-WIDE DNA METHYLATION ANALYSIS IN PRECURSOR B-CELLS

4.1 ABSTRACT

DNA methylation is responsible for regulating gene expression, cellular differentiation and maintaining genomic stability during normal human development. Furthermore, it plays a significant role in the regulation of hematopoiesis. In order to elucidate the influence of DNA methylation during B cell development, the genome-wide DNA methylation status of pro-B, pre-BI, pre-BII and naïve-B cells isolated from human umbilical cord blood was determined using the methylated CpG island recovery assay (MIRA) followed by next generation sequencing. On average, 182 – 200 million sequences were generated for each precursor B-cell subset in 10 biological replicates. An overall decrease in methylation was observed during the transition from pro-B to pre-BI; whereas no differential methylation was observed in the pre-BI to pre-BII transition or in the pre-BII to naïve B-cell transition. Most of the methylated regions were located within intergenic and intronic regions not present in a CpG island context. Putative novel enhancers were identified in these regions and found to be differentially methylated between pro-B and pre-BI cells. The genome-wide methylation profiles are publically available and may be used to gain a better understanding of the involvement of atypical DNA methylation in the pathogenesis of malignancies associated with precursor B-cells.

4.2 INTRODUCTION

B-cell development comprises several developmental stages beginning with pluripotent hematopoietic stem cells (HSCs), followed by common lymphoid progenitor cells (CLP), progenitor-B cells (pro-B), precursor-B cells (pre-BI and pre-BII), immature B-cells (naïve B-cells) and finally mature B-cells. Each developmental stage has diverse biological features that are regulated by differential gene expression (Hystad et al. 2007; van Zelm et al. 2005) Numerous transcription factors (TFs) are known to be responsible for lineage-specific gene regulation (van Zelm et al. 2005; Pérez-Vera et al. 2011; Matthias et al. 2005). In addition, regulatory elements, such as enhancers, have been shown to be critical for tissue and developmental stage-specific gene expression (Harmston et al. 2013); however, the identification of enhancers and how they are regulated has not been described in precursor B-cell development.

DNA methylation is an epigenetic modification by which a methyl group is added to a cytosine base at the carbon-5 position in a CpG dinucleotide (Razin and Riggs 1980). DNA methylation regulates gene expression by attracting methyl-CpG-binding domain proteins (e.g., MeCP2 and MBDs) which promote chromatin condensation into a transcriptionally repressive conformation (Berger 2007; Newell-Price et al. 2000; Bird 1992). Tissue specific DNA methylation is responsible for regulating gene expression, cellular differentiation and maintaining genomic stability during normal human development (Ohgane et al. 2008; Song et al. 2005). It is well established that DNA methylation plays a role in the regulation of hematopoiesis, including myelopoiesis and lymphopoiesis. For example, DNA methylation is requisite for HSC self-renewal and deficiency of methylation leads to differentiation into myeloerythroid cells (Bröske et al. 2009); whereas an increase in DNA methylation is associated with lymphoid commitment (Hodges et al. 2011). Furthermore, the alteration of tissue specific

gene expression by DNA methylation may lead to the progression of many disease states including cancer.

The major focus of methylation studies has historically been centered around the gains or losses of methylation within the promoter regions of genes and the impact of these modifications on the regulation of gene expression. More recently, studies have highlighted the importance of alternative regulatory regions such as transcriptional enhancers in the regulation of gene expression (Zhang et al. 2013; Aran et al. 2013). Unlike the ability to identify a promoter based on its proximal location adjacent to a gene, enhancer detection relies on a number of imperfect measures of chromatin regulators such as histone modifiers in a particular cell type. These important regulatory elements are often found within noncoding regions of the genome, recruit transcriptional coactivators and like tissue-specific promoters are responsible for cell type specific gene regulation.

Genome-wide assessment of DNA methylation in both healthy and diseased tissue is critical to understanding the functional consequence of altered DNA methylation. In this study, genome-wide DNA methylation profiles were generated using the methylated CpG island recovery assay (Rauch and Pfeifer 2005) followed by next generation sequencing (MIRA-seq) for four subsets of B-cells at different stages in development which were isolated from human umbilical cord blood (HCB). The development of methylation profiles in normal cells will aid in elucidating the role of altered DNA methylation in the pathogenesis of acute lymphoblastic leukemia (ALL), a malignancy associated with stage specific precursor B-cells.

4.3 MATERIALS AND METHODS

4.3.1 Isolation of precursor B-cell subsets

The cell sorting strategy developed by Caldwell and colleagues (Caldwell et al. 1991), was employed to isolate 4 subsets of precursor B-cells from 10 HCB samples from healthy individuals provided by the St. Louis Cord Blood Bank as previously described (Almamun et al. 2013). Briefly, mononuclear cells were separated by density centrifugation using Ficoll-Paque PLUS (GE Healthcare Bio-Sciences AB; cat. no. 17-1440-03). Next, non B-cells were depleted from the mononuclear cells using a cocktail of biotin conjugated monoclonal antibodies and anti-biotin monoclonal antibodies conjugated to magnetic beads with the human B cell Isolation Kit (MACS Miltenyi Biotec; order no. 130-093-660). The unlabeled B-cells that passed through the column were then fluorescently labeled with antibodies against cell surface antigens (CD19, CD34, and CD45) specific to individual stages of B-cell development. Finally, four subsets of fluorescently labeled cells were sorted and recovered in individual tubes as pro-B ($CD34^{+}/CD19^{+}$); pre-BI ($CD34^{-}/CD19^{+}/CD45^{low}$); pre-BII ($CD34^{-}/CD19^{+}/CD45^{med}$); and naïve ($CD34^{-}/CD19^{+}/CD45^{high}$) B-cells (**Figure 4.1**).

4.3.2 DNA isolation and MIRA-seq library preparation

Genomic DNA was isolated from each precursor B-cell subset immediately after flow sorting with the QIAamp® DNA Micro Kit (Qiagen; cat.no. 56304) according to manufacturer's instructions and eluted in 30 µl of nuclease free distilled water. Due to the low number of pro-B cells isolated from HCB samples, approximately 100 ng of DNA was recovered. Therefore, library construction for each of the 4 subsets was performed using a starting amount of 100 ng of DNA. The DNA was sonicated to generate 200- to 600-bp fragments using the Bioruptor®

standard (Diagenode; cat. no. B01010002) at 4ºC. Sonication was performed using the high power setting with alternating 30 s on/off intervals for a total of 9 minutes. After 5 min, ice was added to the water bath to keep the samples cool. Sonicated DNA fragments were purified and concentrated using the MinElute PCR purification kit (Qiagen; cat. no. 28004). The purified fragments were electrophoresed on a 1% agarose gel at 96 volts for 45 min for size selection (**Figure 4.2A**). DNA bands between 200 bp and 600 bp were excised and then extracted using the MinElute Gel Extraction kit (Qiagen; cat no. 28604) according to manufacturer's instructions.

The NEBNext® DNA Library Prep Master Mix Set for Illumina kit (New England BioLabs; cat. no. E6040S) was utilized for library construction according to manufacturer's instructions with modifications (**Figure 4.2B**). Briefly, end repair and dA tailing was performed with purification after each step using Agencourt® AMPure® XP Beads (Beckman Coulter; cat. no. A63881). NEBNext Adaptors (New England BioLabs; cat. no. E7335S) were ligated to the repaired DNA and purified using 1.0X AMPure XP beads. After adaptor ligation, MIRA was performed for the enrichment of methylated DNA using the Methyl Collector™ Ultra kit (Active Motif; cat. no. 55005) which utilizes a His-MBD2b/MBD3L1 protein complex to bind methylated DNA fragments (**Figure 4.2C**). PCR enrichment of recovered methylated DNA fragments was then performed using reagents from NEBNext® DNA Library Prep Master Mix Set for Illumina with multiplex oligos for 16 cycles. PCR products were purified using Agencourt® AMPure® XP Beads (Beckman Coulter; cat. no. A63881) and run on 1% agarose to check the size of library products (**Figure 4.2D**). To validate the enrichment of methylated DNA, end point PCR amplification was performed for SLC25A37, a gene that should be methylated and therefore present in the methylation enriched sample, and APC1, a gene that should be

unmethylated and therefore depleted in the methylation enriched sample. The forward primer 5'-CCCCCTGGACGTCTGTAAG-3' and reverse primer 5'-GGCATCTGGTAGATGACACG-3' were used for SLC25A37 (annealing temperature 58°C; amplicon size 158 bp), and the forward primer 5'-ACTGCCATCAACTTCCTTGC-3' and reverse primer 5'-GCGGATTACACAGCTGCTTC C-3' were used for APC1 (annealing temperature 56°C; amplicon size 162 bp). Fragment analysis for each validated library was performed prior to sequencing (**Figure 4.2E**). Four quality-tested MIRA-seq libraries were multiplexed in either 2nM or 10nM concentrations and sequenced on the HiSeq 2000 (1x100 bp reads) at the University of Missouri DNA Core Facility.

4.3.3 Data processing, alignment, and peak identification

Adaptor sequences were trimmed from reads and subsequent short reads were discarded using cutadapt 1.2.1 (Martin 2011). Reads were then aligned to the human reference sequence (GRCh37 with SNP135 masked) with bowtie2 (version 2.1.0) (Langmead and Salzberg 2012) using default parameters. Aligned files were then converted to BAM files and sorted by coordinate using SAMtools (version 0.1.19-44428cd) (Li et al. 2009) "view" and "sort", respectively. Duplicate reads were removed from the BAM files using Picard-tools (version 1.92) "MarkDuplicates" with default parameters. The resulting BAM files were indexed with SAMtools "index". Peaks were identified using MACS2 (version 2.0.10.20130712) (Zhang et al. 2008) "callpeak" with default parameters. In order to standardize peak calls between samples, one file with all of the peak calls from each sample was created, sorted, and then merged with bedtools (version 2.17.0) (Quinlan et al., 2010) "mergeBed" to create a file of unified peaks. Samples were assigned a peak if their own peak overlapped with the unified peaks. Peak results are available as "Custom annotation tracks" for download and viewing in the UCSC genome

browser (http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=taylorkh&hgS_otherUserSessionName=MIRAseq).

4.3.4 Annotation and enhancer prediction

Methylated peaks were annotated with HOMER (Hypergeometric Optimization of Motif EnRichment), version 4.3, using the Perl script "annotatePeaks.pl" with default parameters to identify their genomic location (Heinz et al. 2010). The X and Y chromosomes were excluded from the analysis because the genders of the individual cord blood samples were unknown. Genome-wide CpG island positional information from the UCSC table browser was used to identify the position of methylation peaks within a CpG island context. In order to predict enhancer-like regions, publically available ChIP-seq data for histone modifications (H3K4me1 and H3K27ac) in the lymphoblastoid cell line (GM12878) were obtained from the ENCODE Data Coordination Center at UCSC. The GM12878 cell line is derived from peripheral blood B-cells and is the most closely related cell type to precursor B-cells with publically available histone data.

4.3.5 Differentially methylated regions of interest

Differentially methylated regions (DMRs) between the various stages of B-cell development were identified using the same computational methodology proposed by Chavez and colleagues (Chavez et al. 2010). A saturation analysis was first performed on each preprocessed and aligned sample (BAM file) to ensure sufficient coverage depth (and therefore reproducibility): any sample with insufficient depth (based on saturation correlation < 0.90) was excluded from further analyses. Following this additional quality control check, the data were normalized using a CpG

coupling factor–based method. This approach is built on the theory of coupling factors (Down et al. 2008) and is implemented using a mixture model that takes into consideration the distribution of CpGs in the reference genome. Identification of DMRs was based on the normalized data, and the aforementioned approach does not necessarily require an input ('control') sample to be sequenced. Read counts were binned using 100 bp windows with 300 bp overlaps (based on an expected fragment size of 400 bp). This resulted in approximately 3 million windows, which we term 'regions of interest' (ROIs). ROIs with modest signal representation were excluded (i.e., mean across all samples being compared < 20 counts) which is akin to traditional non-specific filtering. This filtering approach is independent of the test statistic so as not to introduce bias (Bourgon et al. 2010), while increasing power and reducing false discoveries. In the remaining ROIs, testing for differential methylation was performed using the edgeR package called via the MEDIPS package in R/Bioconductor. The ROIs with a false discovery rate (FDR; based on Benjamini-Hochberg approach) <5% were extracted and these regions were merged if immediately adjacent (distance between bp=1) to other ROIs meeting this criteria. Hyper- and hypomethylated ROIs were merged separately so that only putatively consistent ROIs were combined. The reported log fold change for merged ROIs is the maximum (hypermethylated) or minimum (hypomethylated) log fold change for any of its constituent ROIs. It should be noted that the DMR approach is not a peak-based approached, so the aforementioned peak identification and annotation work serves a complimentary role and does not involve identification of DMRs.

4.4 RESULTS

4.4.1 Genomic distribution of DNA methylation during B-cell development

To better understand the role of methylation in early B-cell development, 10 MIRA-seq libraries were constructed each for pro-B, pre-BI, pre-BII and naïve B-cell subsets isolated from the same individual. A sorting strategy that utilized the level of CD45 expression as a discriminator between pre-BI, pre-BII and naïve B-cells (Caldwell et al. 1991) and a stringent gating strategy was employed to ensure that the 4 distinct populations of cells were collected with no overlap (Almamun et al. 2013). On average, 180 million reads were generated for each sample, of which 92% of the reads were aligned to the human reference genome. Within each B-cell subset, the coverage was approximately 60x when taken across all biological replicates, resulting in a high-resolution map. After removing duplicates to eliminate the preferential amplification of certain fragments that may be introduced as a result of low amounts of starting material, 40 – 79 million unique reads remained in each subset (**Figure 4.3 and Table 4.1**).

Methylation peaks present in at least 8 of the 10 samples were used to investigate the genome-wide distribution of DNA methylation in each precursor B-cell subset. More than 201,000 peaks were shared across subsets and less than 5% of the peaks were unique to any individual subset. Similar genomic distributions were observed for each subset with an overwhelming majority of the methylation peaks present in intergenic and intronic regions (**Figure 4.4**). Historically, genome-wide methylation assays have focused on the promoters of genes, CpG islands (CGIs), CpG shores and more recently CpG shelves (Jones 2012; Ji et al. 2010). Therefore, we also investigated the distribution of methylation peaks within a CGI context. Strikingly, the vast majority of methylation peaks in healthy precursor B-cell subsets do

not occur within a CGI context. A total of 3-4% overlapped with a CGI, 5-6% were found within a CpG shore, and 3-4% were found within a CpG shelf (**Figure 4.5**).

4.4.2 Differentially methylated regions in precursor B-cell differentiation

A total of 14,294 hypomethylated and 4,210 hypermethylated DMRs were identified and plotted using Circos (Krzywinski et al. 2009) at an FDR of 5% during the transition from pro-B to pre-BI cells (**Figure 4.6**). The genomic distribution of DMRs differs between the hypomethylated and hypermethylated loci (**Figure 4.7**). More than 95% of the hypomethylated loci lie within intronic and intergenic regions; whereas approximately 80% of the hypermethylated loci are present in intronic or intergenic regions. Quite interestingly, no DMRs were observed between the pre-BI to pre-BII cell transition or between the pre-BII to naïve B-cell transition. This is not surprising considering that these cells are more similar than pro-B cells which still possess the stem-like cell surface marker CD34.

In the pre-BI cells, ~1% of the hypomethylated DMRs and ~3% of the hypermethylated DMRs were found within a gene promoter (**Figure 4.7**). To gain insight into the potential biological consequence of the loss or gain of promoter methylation in pre-BI cells we performed functional annotation of promoter DMRs using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (Huang et al. 2009). No gene sets or functional groups were significantly enriched in promoter DMRs. Therefore, we sought to identify potential functional DMRs present in intergenic and intronic regions. We first identified enhancer related histone marks in the lymphoblastoid cell line GM12878, a surrogate for precursor B-cells. The location of the histone marks representing potential poised and active enhancers (H3K4me1 and H3K27ac) were then superimposed on the intergenic and intronic DMRs (Harmston et al. 2013; Bulger et al. 2011; Xu et al. 2012; Bogdanovic et al. 2012). A total of 467 potential enhancer-

like DMRs were identified that possessed H3K4me1 and H3K27ac; 197 were hypomethylated and 270 were hypermethylated in pre-BI cells compared to pro-B cells. To examine the possible impact of differentially methylated enhancer-like regions, a list of target genes for each DMR was generated from the nearest upstream and downstream gene. More than 200 putative target genes were identified in the hypomethylated DMRs and 396 were identified in the hypermethylated DMRs. Consistent with the differentially methylated promoter loci, no gene sets or functional groups were enriched in the enhancer associated DMRs. However, the enhancer-like DMRs contained genes involved in the regulation of leukocyte activation and the B-cell receptor (BCR) signaling pathway. Specifically, the enhancer regions of genes implicated in lymphocyte differentiation (*EGR1, FOXP1* and *KLF6*) and in the BCR signaling pathway (*CARD11, LILRB3* and *CD81*) gained methylation during the transition from pro-B to pre-BI cells. In contrast, the BCR signaling pathway genes *VAV3, GRB2, PPP3CA*, and *NFATC1* lost methylation during the transition from pro-B to pre-B cells (**Table 4.2**). Therefore, it is reasonable to infer that the regulation of these genes during B-cell differentiation is affected by the presence or absence of methylation in cell-type specific enhancers.

It has previously been shown that DNA methylation decreases during cellular differentiation of myeloid cells (Hogart et al. 2012; Lee et al., 2012), but increases when comparing early progenitor cells to fully differentiated lymphoid cells (Ji et al. 2010; Hodges et al. 2011). In order to visualize the changes in methylation as precursor B-cells differentiate, the average level of methylation at DMR loci was examined across each subset. Two prominent patterns arose, each beginning with a decrease in methylation between the pro-B and pre-BI stage followed by: 1) a steady state of methylation being observed in subsequent stages; or 2) a slight increase in

methylation in subsequent stages. The second pattern observed in our data, helps to elucidate when the increase in methylation begins to occur in lymphoid differentiation.

4.5 DISCUSSION

Epigenetic modifications are well known to play important roles in disease. In order to better understand and identify aberrant disease associated epigenetic events, it is imperative to create epigenomic data sets for all normal tissue types. The ENCODE consortia set out to define the functional elements in the human genome including epigenetic marks such as histone modifications and DNA methylation. This effort has produced genome-scale epigenetic data sets for multiple cell types but does not include every cell type. Precursor B-cell ALL is the most common malignancy seen in children and also affects adults. Precursor B-cells are present at a low frequency in pediatric bone marrow and at even lower frequencies in adult bone marrow. In order to generate DNA methylation profiles in precursor B-cells we obtained umbilical cord blood from healthy donors which is rich in precursor B-cells and more easily attainable than bone marrow from healthy children. For the first time, we provide whole-genome DNA methylation profiles for 4 subsets of B-cells derived from umbilical cord blood.

In this study, MIRA-seq followed by next generation sequencing was used to examine genome-wide DNA methylation during the development of B cells. Caldwell and colleagues have shown that $CD19^+$ cells that express high levels of CD45 correspond to naïve B-cells which express surface IgM, $CD19^+$ cells that express intermediate levels of CD45 correspond to pre-BII cells which express cytoplasmic IgM, and $CD19^+$ cells that express low levels of CD45 correspond to pre-BI cells which do not express cytoplasmic IgM (Caldwell et al. 1991).

Therefore, we isolated 4 subsets of precursor B-cells: 1) pro-B ($CD34^{+}/CD19^{+}$); 2) pre-BI ($CD34^{-}/CD19^{+}/CD45^{low}$); 3) pre-BII ($CD34^{-}/CD19^{+}/CD45^{med}$); and 4) naïve B-cell ($CD34^{-}/CD19^{+}/CD45^{high}$) (Almamun et al. 2013). DMRs were identified to determine the dynamic range of DNA methylation as B-cells differentiate. Many DMRs lost methylation during the transition from pro-B to pre-BI cells consistent with a previous study (Lee et al. 2012). In addition, more than 4,000 regions were identified that gained methylation during the transition from pro-B to pre-BI cells. Contradictory to Lee and colleagues (Lee et al. 2012), there were no DMRs at a 5% FDR during the pre-BI to pre-BII transition nor during the pre-BII to naïve B cell transition indicating that methylation levels remain unchanged during this period. This can be explained by the fact that MIRA-seq identifies genomic regions that are methylated; whereas the Infinium array utilized by Lee and colleagues surveys individual CpG sites. Our results suggest that any of these subsets (pre-BI, pre-BII, naïve-B) could serve as an appropriate "control" cell population in studies aimed at identifying differential methylation in diseases of precursor B-cells such as ALL. It is interesting to note that many precursor B-ALL patients retain (or gain) the stem-like cell surface marker CD34 calling into question whether the cells are more related to pre-B or pro-B cells.

MIRA-seq has the advantage of being able to distinguish the methylated DNA in all regions of the genome and the slight disadvantage that the precise location of the methylation cannot be distinguished. The vast majority of methylated regions in normal precursor B-cells were present in intergenic and intronic loci. Perhaps these loci are important in maintaining genomic stability (Jones et al. 2012); however, it is equally plausible that these regions harbor important regulatory sequences, such as enhancers, that are affected by methylation (Bulger et al. 2011; Shore et al. 2010). In fact, more than 450 putative enhancer-like DMRs were identified in

pre-BI cells compared to pro-B cells indicating cell-type specific gene regulation. Putative targets of the enhancer-like DMRs play fundamental roles in lymphopoiesis and BCR signaling and included the genes *FOXP1, KLF6, CARD11, LILRB3, CD81, VAV3, GRB2, PPP3CA*, and *NFATC1*. When other regulatory DMRs were considered, none of the promoters associated with genes involved in B-cell differentiation were differentially methylated. Therefore, intergenic and intronic regulatory regions may be the driving force in B-cell differentiation.

In summary, we provide genome-wide DNA methylation profile at four stages of B-cell development. We report for the first time hypermethylated loci associated with the transition of pro-B to pre-BI cells. The majority of the differentially methylated regions identified lie within intronic and intergenic regions and some of these regions overlap with putative regulatory regions. This study supports the use of umbilical cord blood as a source for precursor B-cells and establishes a baseline methylation profile for precursor B-cell subsets. It is likely that the DMRs identified in regulatory regions will be important in identifying the mechanisms responsible for the dysregulation of gene expression associated with precursor B-cell disorders such as ALL.

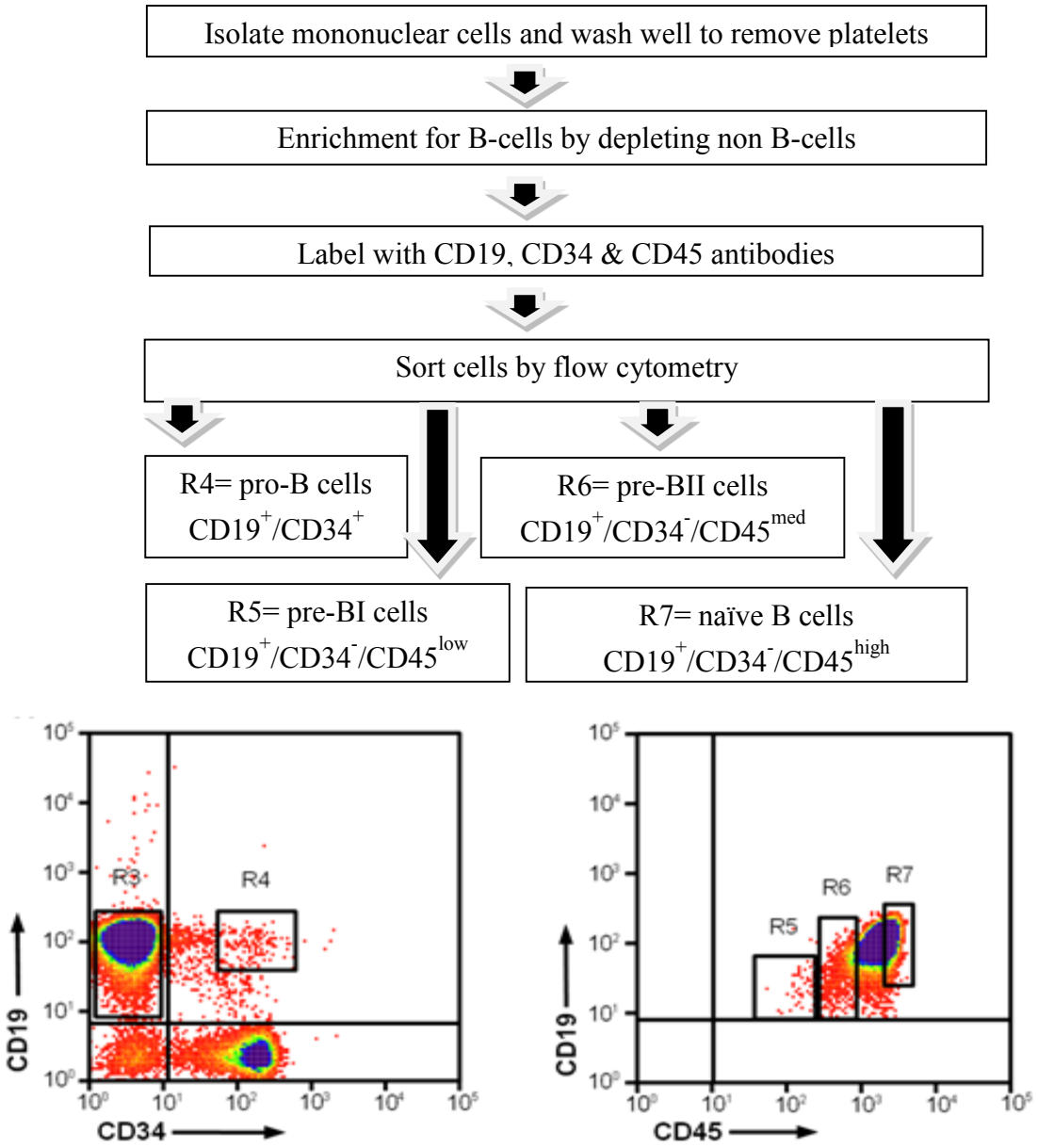


Figure 4.1. Isolation of precursor B-cell subsets from human umbilical cord blood. Mononuclear cells were isolated using density gradient centrifugation to remove all non B-cells. B-cells were labeled with cell surface antibodies and sorted into four separate tubes. R4: pro-B cells; R5: pre-BI cells; R6: pre-BII cells; R7: naïve B cells.

Table 4.1: Read and alignment statistics for individual samples

Subset	Sample	Total Reads	Fraction Mapped	Unique Reads
Pro-B	HCB1	159,375,936	0.946	34,890,917
	HCB2	275,775,450	0.902	37,054,047
	HCB3	143,136,524	0.933	20,938,997
	HCB4	150,983,572	0.933	34,713,018
	HCB5	221,920,825	0.920	64,801,910
	HCB6	256,016,969	0.913	36,113,336
	HCB7	204,552,858	0.942	56,437,859
	HCB8	186,406,308	0.929	34,739,657
	HCB9	197,700,606	0.939	45,116,675
	HCB10	171,626,945	0.886	32,579,069
Pre-BI	HCB1	168,552,888	0.952	85,329,586
	HCB2	265,314,820	0.901	67,984,512
	HCB3	84,521,014	0.938	39,905,728
	HCB4	148,457,990	0.933	47,064,491
	HCB5	240,194,057	0.919	83,332,904
	HCB6	188,877,538	0.944	131,520,066
	HCB7	147,961,433	0.942	57,073,380
	HCB8	148,151,509	0.933	65,601,840
	HCB9	175,217,423	0.941	62,713,505
	HCB10	258,120,121	0.935	155,091,010
Pre-BII	HCB1	128,568,614	0.939	41,948,710
	HCB2	240,907,679	0.898	78,846,617
	HCB3	150,168,486	0.944	62,616,516
	HCB4	195,937,594	0.938	63,730,137
	HCB5	237,968,679	0.918	51,527,842
	HCB6	182,755,070	0.946	124,485,754
	HCB7	228,160,559	0.937	98,691,194
	HCB8	281,061,395	0.926	67,410,961
	HCB9	152,443,649	0.941	64,389,853
	HCB10	137,008,710	0.928	88,156,469
Naïve B	HCB1	156,540,399	0.943	44,593,245
	HCB2	268,175,606	0.898	71,107,061
	HCB3	140,510,823	0.939	65,425,259
	HCB4	253,253,459	0.933	78,647,622
	HCB5	231,601,747	0.909	43,301,897
	HCB6	171,904,801	0.941	99,929,288
	HCB7	219,665,529	0.941	65,994,827
	HCB8	216,408,028	0.927	35,548,885
	HCB9	121,564,595	0.943	44,267,598
	HCB10	226,597,896	0.918	136,066,654

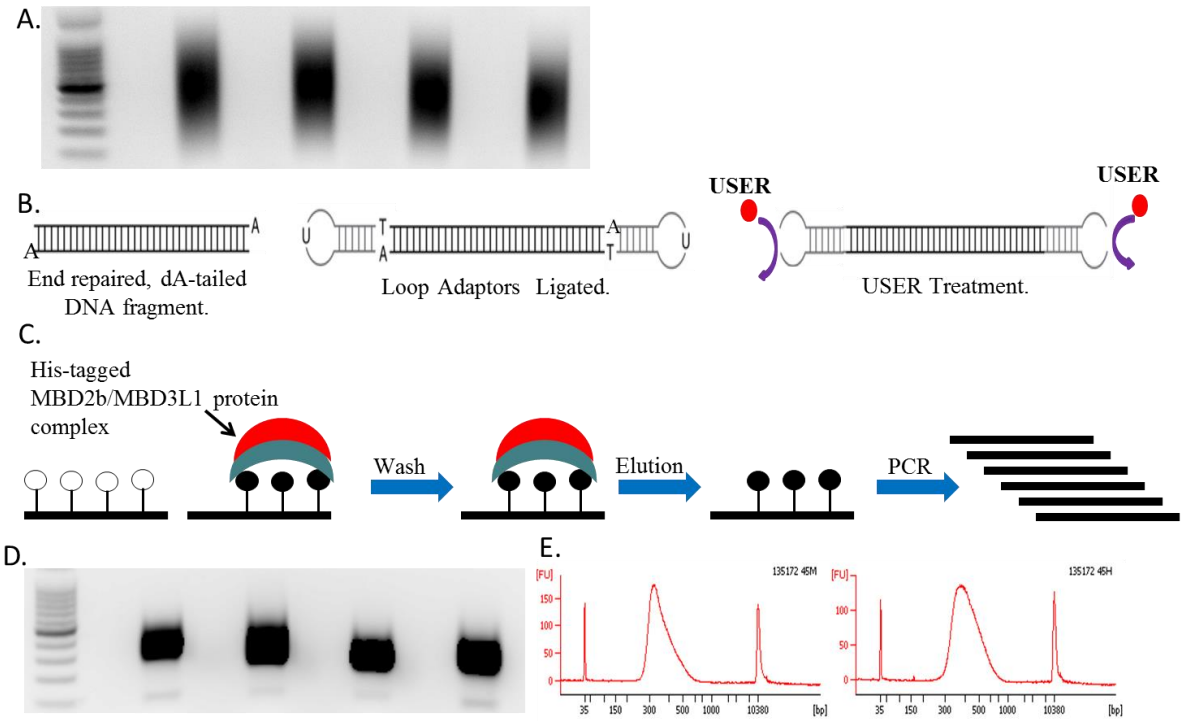


Figure 4.2: MIRA-seq library preparation. (A) Sonicated DNA run on 1% agarose gel. (B) Adaptor ligation at both ends of fragmented DNA. (C) Methylated CpG island recovery assay. (D) MIRA-seq library run on 1% agarose gel after PCR amplification. (E) Fragment analysis of MIRA-seq library.

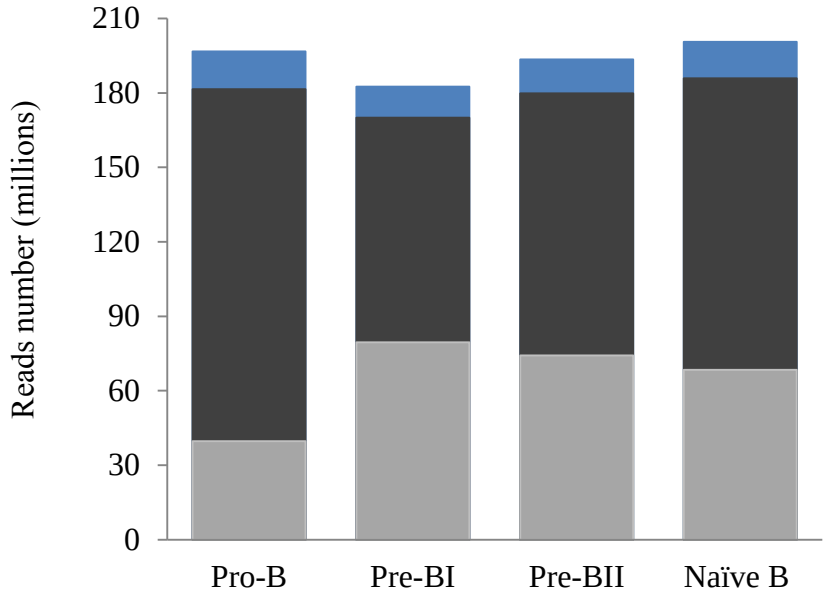


Figure 4.3. Average read and alignment statistics. Read and alignment statistics were averaged across all individuals for each precursor B-cell subset. The top of each bar represents the total number of sequencing reads (blue), the total number of mapped reads (dark gray), and the total number of unique reads (light gray).

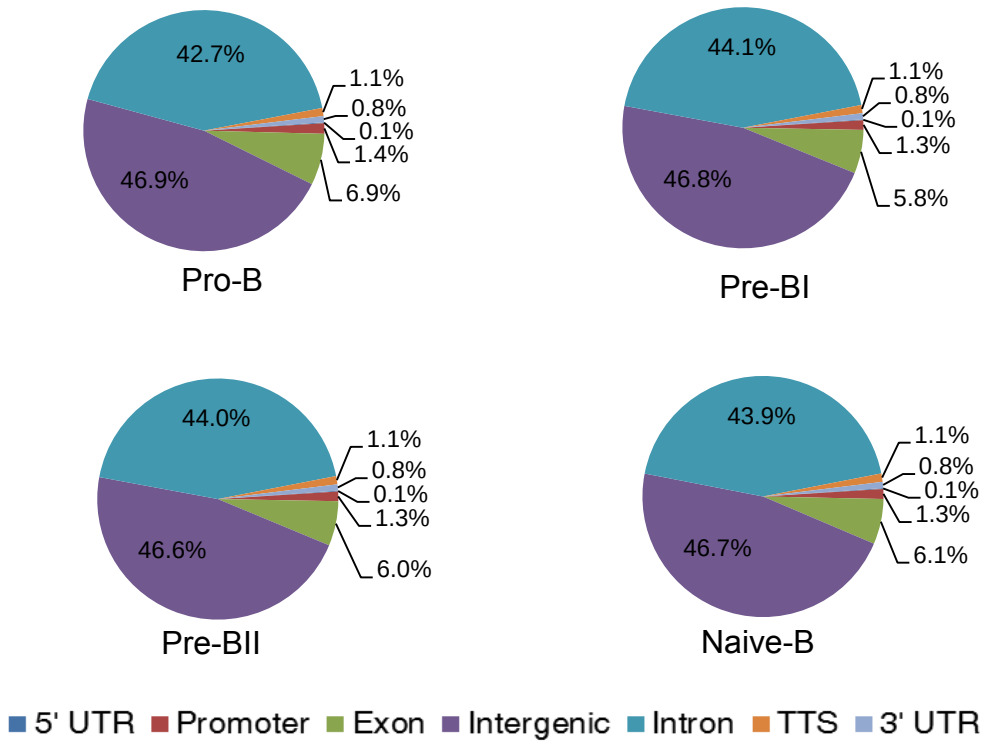


Figure 4.4: Genomic distribution of methylation peaks in pro-B, pre-BI, pre-BII and naïve-B cells.

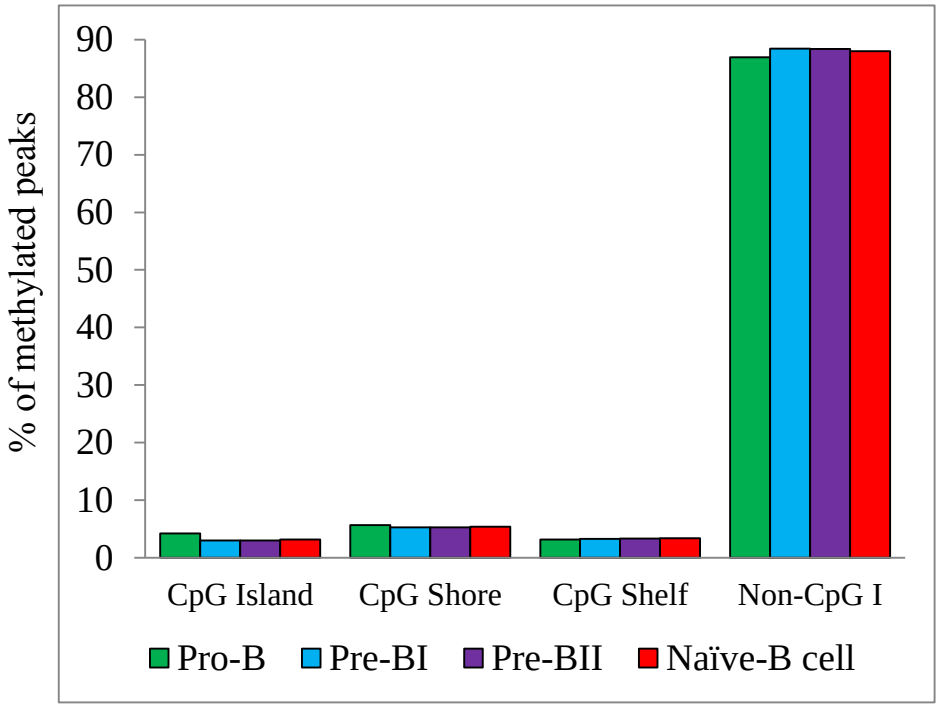


Figure 4.5: Percentages of methylation peaks associated with CpG islands, CpG shores and CpG shelves.

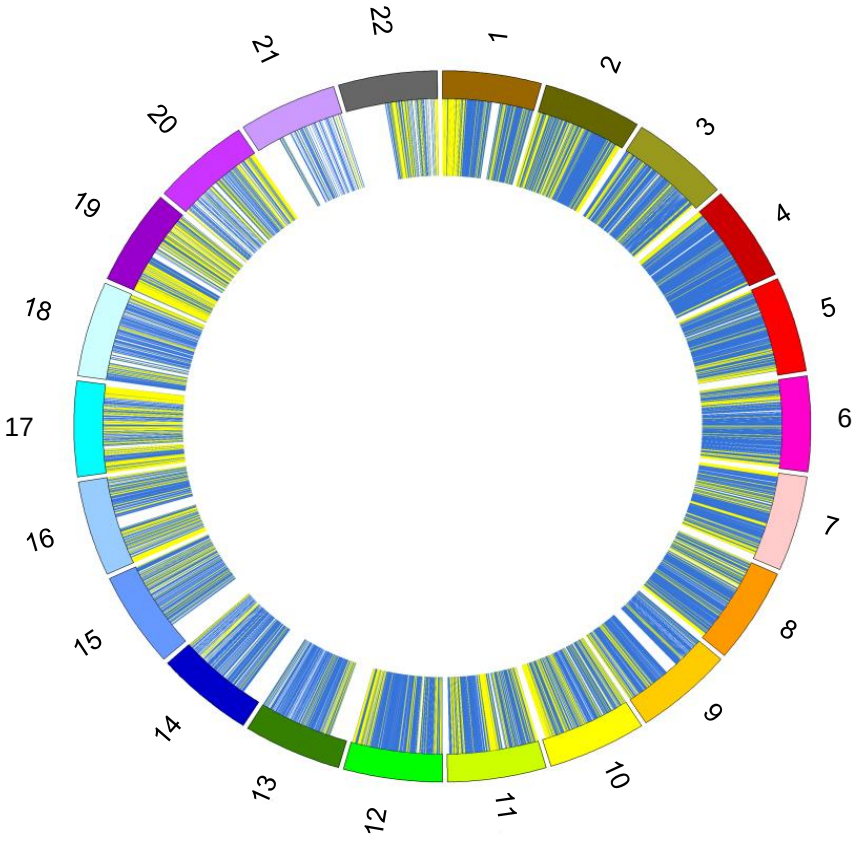


Figure 4.6:. **Differentially methylated regions in the pro-B to pre-BI transition**. Circos plots representing chromosomes 1-22. The X and Y chromosomes were excluded from analysis. A total of 4,210 loci gained methylation (hypermethylated in pre-BI) and 14,294 loci lost methylation (hypomethylated in pre-BI). Blue represents hypomethylated regions in pre-BI cells and yellow represents hypermethylated regions in pre-BI cells.

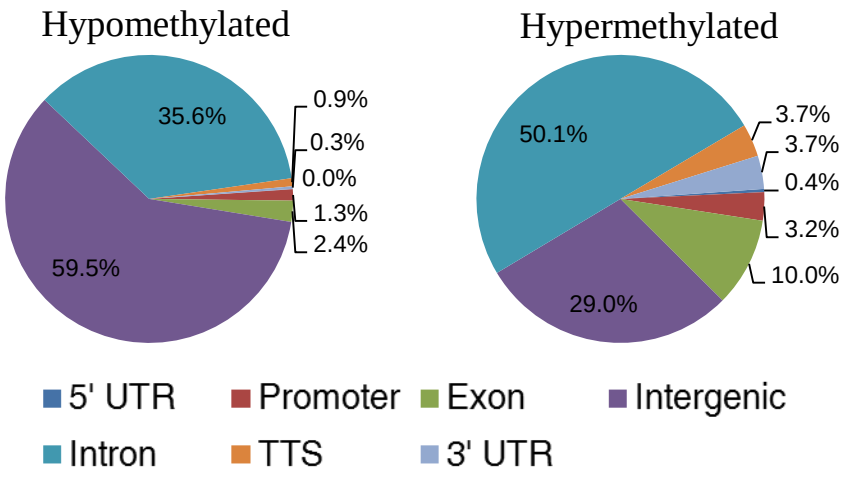


Figure 4.7: Genomic distribution of hypomethylated and hypermethylated loci during the pro-B to pre-BI transition.

Table 4.2: Differentially methylated enhancer target genes involved in leukocyte activation and BCR signaling pathway.

	Gene	Chr	DMR_start	DMR_end	Location	Distance from TSS	Fold change
Hypomethylated enhancer	VAV3	1	108,508,801	108,508,900	Intronic	-1,306	1.318
	PPP3CA	4	102,216,101	102,216,500	Intronic	52,327	1.550
	GRB2	17	73,397,201	73,397,600	Intronic	4,388	1.604
	NFATC1	18	77,284,801	77,285,500	Intronic	12,4825	1.536
Hypermethylated enhancer	LILRB3	19	54,715,301	5,4715,400	Intergenic	31,202	1.420
	CD81	11	2,411,401	2,411,500	Intronic	13,994	1.300
	KLF6	10	3,822,601	3,822,800	Intronic	4,772	1.412
	EGR1	5	137,799,101	137,799,300	Intergenic	-1,980	1.387
	FOXP1	3	71,276,301	71,276,400	Intronic	17,965	1.432

CH4: REFERENCES

Almamun M, Schnabel JL, Gater ST, Ning J, Taylor KH. Isolation of precursor B-cell subsets from umbilical cord blood. J. Vis. Exp 2013; 74.

Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013; 14(3):R21.

Berger LS. The complex language of chromatin regulation during transcription. Nature 2007; 447:407-412.

Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJ, Gómez-Skarmeta JL. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res 2012; 22(10):2043-53.

Bird A. The Essentials of DNA Methylation. Cell 1992; 70:5-8.

Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci U S A 2010; 107:9546-51.

Bröske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, Kuhl C, Enns A, Prinz M, Jaenisch R, Nerlov C, Leutz A, Andrade-Navarro MA, Jacobsen SE, Rosenbauer F. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 2009; 41(11):1207-15.

Bulger M and Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell 2011; 144(3):327-39.

Caldwell CW, Poje E, Helikson MA. B-cell precursors in normal pediatric bone marrow. Am J Clin Pathol 1991; 95(6):816-823.

Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H, Herwig R, Adjaye J. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. Genome Res. 2010; 20(10):1441-50.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Gräf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Bäckdahl L, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJ, Durbin R, Tavaré S, Beck S. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 2008; 26:779-85.

Harmston N, Lenhard B. Chromatin and epigenetic features of long-range gene regulation. Nucleic Acids Res. 2013; 41(15):7185-99.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010; 38(4):576-589.

Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR, Smith AD, Hannon GJ. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell. 2011; 44(1):17-28.

Hogart A, Lichtenberg J, Ajay SS, Anderson S. NIH Intramural Sequencing Center, Margulies EH, Bodine DM. Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. Genome Res 2012; 22(8):1407-18.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc 2009; 4(1):44-57.

Hystad ME, Myklebust JH, Bø TH, Sivertsen EA, Rian E, Forfang L, Munthe E, Rosenwald A, Chiorazzi M, Jonassen I, Staudt LM, Smeland EB. Characterization of early stages of human B cell development by gene expression profiling. J Immunol 2007; 179(6):3662-71.

Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 2010; 467(7313):338-42.

Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 2012; 13:484-492.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. Genome Res 2009; 19:1639–1645.
Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012; 9:357-359.

Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK, Zheng S, Dou X, de Smith A, Chokkalingam A, Buffler P, Ma X, Wiemels JL. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. Nucleic Acids Res 2012; 40(22):11339-51.

Li H.*, Handsaker B.*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 2009; 25:2078-9.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 2011; 17:10-12.

Matthias P, Rolink AG. Transcriptional networks in developing and mature B cells. Nat Rev Immunol 2005; 5(6):497-508.

Newell-Price J, Clark AJL, King P. DNA Methylation and Silencing of Gene Expression. Trends Endocrinol Metab 2000; 11(4):142-148.

Ohgane J, Yagi S, Shiota K. Epigenetics: the DNA methylation profile of tissue-dependent and differentially methylated regions in cells. Placenta 2008; 29 Suppl A:S29-35.

Pérez-Vera P, Reyes-León A, Fuentes-Pananá EM. Signaling proteins and transcription factors in normal and malignant early B cell development. Bone Marrow Res 2011; 2011:502751.

Quinlan A, Hall I. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26(6): 841–842.

Razin A and Riggs AD. DNA methylation and gene function. Science 1980; 210(4470):604-610.
Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab Invest 2005; 85(9):1172-8.

Shore A, Karamitri A, Kemp P, Speakman JR, Lomax MA. Role of Ucp1 enhancer methylation and chromatin remodelling in the control of Ucp1 expression in murine adipose tissue. Diabetologia 2010; 53(6):1164-73.

Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, Held WA. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A 2005; 102(9):3336-41.

van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ, Lankester AC, Révész T, Staal FJ, van Dongen JJ. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. J Immunol 2005; 175(9):5912-22.

Xu J, Shao Z, Glass K, Bauer DE, Pinello L, Van Handel B, Hou S, Stamatoyannopoulos JA, Mikkola HK, Yuan GC, Orkin SH. Combinatorial assembly of developmental stage specific enhancers control gene expression programs during human erythropoiesis. Dev Cell 2012; 23:796-811.

Zhang B, Zhou Y, Lin N, Lowdon RF, Hong C, Nagarajan RP, et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res. 2013; 23(9):1522-40.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9(9):R137.

CHAPTER 5: INTEGRATED METHYLOME AND TRANSCRIPTOME ANALYSIS REVEALS NOVEL REGULATORY ELEMENTS IN PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA

5.1 ABSTRACT

Acute lymphoblastic leukemia is the most common cancer diagnosed in children under the age of 15. In addition to genetic aberrations, epigenetic modifications such as DNA methylation are altered in cancer and impact gene expression. To identify epigenetic alterations in ALL, genome-wide methylation profiles were generated using the methylated CpG island recovery assay followed by next-generation sequencing. More than 25 000 differentially methylated regions (DMR) were observed in ALL patients with ~90% present within intronic or intergenic regions. To determine the regulatory potential of the DMR, whole-transcriptome analysis was performed and integrated with methylation data. Aberrant promoter methylation was associated with the altered expression of genes involved in transcriptional regulation, apoptosis and proliferation. Novel enhancer-like sequences were identified within intronic and intergenic DMR. Aberrant methylation in these regions was associated with the altered expression of neighboring genes involved in cell cycle processes, lymphocyte activation and apoptosis. These genes include potential epi-driver genes such as *SYNE1, PTPRS, PAWR, HDAC9, RGCC, MCOLN2, LYN, TRAF3, FLT1* and *MELK* which may provide a selective advantage to leukemic cells. In addition, the differential expression of epigenetic modifier genes, pseudogenes, and noncoding RNAs was also observed accentuating the role of erroneous epigenetic gene regulation in ALL.

5.2 INTRODUCTION

Acute lymphoblastic leukemia (ALL) is a hematological malignancy associated with precursor B-cells. ALL is the most common type of cancer in children with an annual occurrence rate of 35 to 40 cases per 1 million people in the United States (NCI report). The development and differentiation of B-cells comprises numerous stages and is a highly synchronized and controlled process governed by stage-specific gene expression (Hystad et al. 2007; van Zelm et al. 2005). Any deviation from normal stage-specific gene expression could lead to disease conditions including ALL. The general known mechanisms underlying the induction of ALL include chromosomal translocation, hyperdiploidy, and aberrant expression of proto-oncogenes. Advancement in deciphering additional mechanisms that may be responsible for the induction of all ALL is lacking. Therefore, the identification of key regulatory regions in the genome that may impact the development of ALL is critical to gaining a better understanding of ALL pathogenesis.

DNA methylation is responsible for tissue specific gene expression and plays a significant role in hematopoiesis (Almamun et al. 2014; Hodges et al. 2011), and malignant transformation (Berdasco et al. 2010; Figueroa et al. 2013). A reduced level of CpG methylation was one of the first epigenetic alterations to be found in human cancer when compared with normal-tissue counterparts (Feinberg et al. 1983). Although hypermethylation of CpG islands within gene promoters has been the main focus of studies on malignant cells, the role of differential DNA methylation in other regions is gaining favor (Liang et al. 2011; Ji et al. 2010). One such region harbors transcriptional enhancers which reside within non-coding regions of the genome and are known to work over long distances to promote cell/tissue type specific gene expression. Active enhancers are often accompanied by DNA demethylation (Xu et al. 2007),

and alterations in enhancer methylation are seen in malignant transformation. Recently, it has been shown that differential methylation of these regions exhibit a higher correlation with gene expression than differential promoter methylation (Aran et al. 2013).

As a step towards better understanding the consequence of altered DNA methylation on gene expression in pre-B ALL, MIRA-seq was used to identify altered DNA methylation throughout the genome and then correlated with transcriptome data. We show that differential comparisons of DNA methylation between normal and diseased tissue can identify potential regulatory regions of the genome and that when paired with gene expression data the functionality of the regulatory regions can be determined.

5.3 MATERIALS AND METHODS

5.3.1 Patient samples

De-identified patient samples were obtained under full ethical approval of the institutional review board at the University of Missouri. A total of 20 pre-B ALL patient samples (**Table 5.1**) and pre-BI and pre-BII cells from 10 healthy individuals were used for this study. ALL patient samples contain at least 88% blasts. Normal control pre-BI and pre-BII cells were isolated from 10 human umbilical cord blood (HCB) samples as previously described (Almamun et al. 2013). Briefly, mononuclear cells were isolated by density gradient centrifugation using Ficoll-Paque PLUS (GE Healthcare Bio-Sciences AB, Björkgatan, Uppsala, Sweden) followed by depletion of all non B-cells with biotin conjugated antibodies cocktail and anti-biotin monoclonal antibodies conjugated to magnetic beads using human B cell Isolation Kit (MACS Miltenyi Biotec, Bergisch Gladbach, Germany). For the methylation studies, purified B-cells were fluorescently labeled with antibodies against cell surface antigen (CD19, CD34, CD45; BD

Biosciences, San Jose, CA, USA) specific to individual stages of B-cell development. Finally, the fluorescently labeled cells were sorted as pre-BI ($CD19^{+}/CD34^{-}/CD45^{low}$) and pre-BII ($CD19^{+}/CD34^{-}/CD45^{med}$). Because no regions of differential methylation were observed in pre-BI versus pre-BII cells, transcriptomes were generated for precursor B-cells which include both pre-BI and pre-BII subsets. To obtain this population of cells, purified B-cells were fluorescently labeled with antibodies against CD19 and IgM and precursor B-cells ($CD19^{+}/IgM^{-}$) were isolated by flow cytometry (**Figure 5.1**).

5.3.2 Antibodies

The following antibodies were used for flow cytometry and non B-cell isolation through column purification: BD Pharmingen™ PE Mouse Anti-Human CD34 (cat. no. 560941, BD Biosciences, San Jose, CA, USA); BD Pharmingen™ APC Mouse Anti-Human CD19 (cat. no. 555415, BD Biosciences, San Jose, CA, USA); CD45 FITC (cat. no. 347463, BD Biosciences, San Jose, CA, USA); BD Pharmingen™ PE Mouse Anti-Human IgM (cat. no. 555783, BD Biosciences, San Jose, CA, USA); B cell Isolation kit (order no. 130-093-660, MACS Miltenyi Biotec, Bergisch Gladbach, Germany).

5.3.2 MIRA-seq library preparation

Genomic DNA from ALL patient samples was isolated using DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. MIRA-seq libraries for normal precursor B-cells were prepared as previously described (Almamun et al. 2014). For ALL patient samples, 1.0 µg of DNA from each ALL patient was sonicated with alternating 30 seconds on/off intervals for a total of 9 minutes to generate 200- to 600-bp fragments. A small

portion of sonicated DNA was run on 1% agarose gel to ensure the sonication accuracy. The remaining sonicated DNA fragments were concentrated and purified using the MinElute PCR purification kit (Qiagen, Hilden, Germany). Adaptor ligation to fragmented DNA followed by MIRA using MethylCollector^{TM} Ultra kit (Active Motif, Carlsbad, CA, USA) was performed according to manufacturer's protocols and as previously described (Almamun et al., 2014). After size selection of enriched methylated DNA on 1% agarose gel, PCR amplification of recovered methylated DNA fragments was performed for 11 cycles and then purified with the MinElute gel extraction kit (Qiagen, Hilden, Germany). In order to validate the enrichment of methylated DNA, end point PCR amplification of methylated SLC25A37- and unmethylated APC1- regions was performed with the following primer pairs: 5'-CCCCCTGGACGTCTGTAAG-3' (forward) and 5'-GGCATCTGGTAGATGACACG-3' (reverse) for SLC25A37, and 5'-ACTGCCATCAACTTCCTTGC-3' (forward) and 5'-GCGGATTACACAGCTGCTTC C-3'(reverse) for APC1. Quantity and fragment analysis was performed using Qubit and Bioanalyzer before sequencing. Four high quality MIRA-seq libraries were multiplexed in 10nM concentrations and sequenced on the Illumina HiSeq 2000 (1x100 bp reads) at the DNA Core Facility, University of Missouri-Columbia.

5.3.3 Identification of methylated peaks and differentially methylated regions in ALL

MIRA-seq data processing and methylated peaks for individual samples were identified using MACS2 pipeline as previously described (Almamun et al. 2014). Briefly, following adaptor trimming, sequences were aligned to the human reference sequence (GRCh37 with SNP135 masked) with bowtie2 (version 2.1.0). Patient sample A32 had an insufficient numbers of reads and was excluded from subsequent analyses. Picard-tools (version 1.92) were used to remove duplicate reads from the BAM files. The resulting BAM files were indexed with

SAMtools “index”. Methylated peaks were identified using MACS2 (version 2.0.10.20130712) (Zhang et al. 2008) with default parameters. Unified peak locations across the samples were created using bedtools (version 2.17.0). Individual sample was assigned a peak when their own peak overlapped with the unified peaks. ALL and HCB peaks were included if the peak was present in at least 17 biological replicates. Differentially methylated regions (DMRs) between the ALL and control precursor B-cells isolated from HCB were identified as described previously (Almamun et al., 2014). The coverage depth for each sample was analyzed and any sample with insufficient depth (saturation correlation < 0.90) was omitted from further analyses. Following normalization of data using a CpG coupling factor–based method (Down et al. 2008), DMRs were identified. Initially regions of interest (ROIs) were determined based on read counts within 100 bp windows with a 300 bp overlaps (expected fragment size of 400 bp). Non-specific filtering was performed by discarding the ROIs with modest signal representation (< 20 mean counts across all samples). Differentially methylated regions were identified from the remaining ROIs using the edgeR package called via the MEDIPS package in R/Bioconductor. The ROIs with <5% false discovery rate (FDR; Benjamini-Hochberg) and at least a 2-fold change were identified as a DMRs. ROIs immediately adjacent to one another were combined into a single DMR. Hyper- and hypomethylated ROIs were merged separately so that only putatively consistent ROIs were combined. The reported log fold change for merged DMRs is the maximum $\log_2$ fold change for any of its constituent ROIs. All MIRA-seq data were deposited in NCBI Sequence Read Archive (Accession SRP058314).

5.3.4 Annotation and enhancer prediction

Methylated peaks and differentially methylated regions were annotated with HOMER (Hypergeometric Optimization of Motif EnRichment), version 4.3, using the default setting to

identify genomic locations (Heinz et al. 2010). The X and Y chromosomes were excluded from the analysis as the genders of individual normal samples were unknown. CpG island positional information from the University of California Santa Cruz (UCSC) table browser was used to determine the position of methylation peaks within a CpG island context. The genomic locations of enhancers were identified based on the enrichment of histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac) modifications in the GM12878 cell line (lymphoblastoid) available from ENCODE.

5.3.5 RNA-seq library preparation and data analysis

RNA samples were also obtained from the pre-B ALL patients (20 samples) utilized in the MIRA-seq assays and from eight normal precursor B-cells isolated from HCB. RNA sequencing libraries were constructed with the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) and sequenced on the Illumina HiSeq 2000 (1x100 bp reads) at the University of Missouri DNA Core Facility. The reads were preprocessed to remove poor quality reads of <20 using FastX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were aligned to hg19 using Tophat (v2.0.13) with default settings. On average 23.6 million and 30.3 million sequence reads were aligned to the genome for ALL and healthy pre-B samples, respectively. Differential gene expression between ALL and healthy precursor B-cells were determined using Cufflinks with default parameters (version 2.2.1) (Trapnell et al. 2013). The read counts along with FPKM values and their variances were calculated by cuffdiff 2 and the log fold change and p-value was calculated for each gene. Multiple testing corrections using Benjamini-Hochberg was also performed (q value). The same cutoffs for FDR and fold change used in the analysis of methylated ROIs were used to determine differential expression. All functional annotations were performed using the

Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (Huang et al. 2009). All RNA-seq data were deposited in NCBI Sequence Read Archive (Accession SRP058414).

5.4 RESULTS

5.4.1 Genome-wide DNA methylation profiles

MIRA-seq was utilized to generate genome-wide DNA methylation profiles for 19 pre-B ALL patient samples from diagnostic bone marrow. Normal precursor B-cell populations (pre-BI and pre-BII) were isolated from 10 human umbilical cord blood (HCB) samples to generate methylation profiles for healthy tissue to be used as a comparator (Almamun et al. 2014). On average, 188 million reads were generated for HCB samples and 176 million reads were generated for ALL patient samples (**Figure 5.2A, Table 5.2**). Methylation peaks were more abundant in HCB samples (305,736) than in ALL samples (162,832) and across all chromosomes revealing an overall genome-wide reduction of methylation in ALL (**Figure 5.2B**). Genomic distribution analysis showed that ~90% of the methylated peaks were located within intronic and intergenic regions (**Figure 5.3A**). The distribution of methylation peaks relative to CpG islands (CGIs) revealed that 9,814 CGIs were methylated in HCB samples and 11,015 CGIs were methylated in ALL samples but the overwhelming majority of methylated peaks were present in regions of the genome not associated with CGIs (**Figure 5.3B**).

5.4.2 Differentially methylated regions in ALL

To determine methylation patterns distinct to ALL, differentially methylated regions (DMRs) between ALL and HCB samples with at least a 2-fold change and an FDR of $\leq$ 5% were

identified (**Figure 5.4**). A total of 15,492 regions lost methylation and 9,790 regions gained methylation in ALL compared to the normal HCB samples and the genomic distribution of loci harboring DMRs differed in the hypomethylated versus hypermethylated DMRs (**Figure 5.5A and B**). Hypermethylation was more prevalent in the 5' regulatory regions of genes than hypomethylation. The majority of the DMRs coincided with intergenic and intronic genomic regions. DMRs have applicability as disease specific biomarkers and may also play regulatory roles in the expression of genes that are involved in the pathogenesis of ALL. To further elucidate the importance of DMRs, we sought to identify the DMRs with regulatory potential.

DMRs are associated with regulatory sequences: The promoters of protein coding genes harbor regulatory sequences required for the initiation of transcription. A total of 1,568 differentially methylated gene promoters were identified (corresponding to 1,252 hypermethylated genes and 240 hypomethylated genes) in ALL. To explore the association of DNA methylation and gene expression, MIRA-seq data and RNA-seq data were correlated. Sixty-two promoter DMRs were hypermethylated and down-regulated in ALL (**Table 5.3**) and were significantly enriched for genes involved in the regulation of transcription, and apoptosis; whereas 37 promoter DMRs were hypomethylated and up-regulated (**Table 5.4**), and were significantly enriched for genes involved in GTPase activation, the regulation of cell proliferation, and those that play a role in protein complex assembly. Additionally, hypermethylated DMRs were identified in the promoters of 3 tumor suppressor genes, *MTSS1*, *PAWR*, and *EXT1*, and corresponded with a significant decrease in gene expression.

In addition to protein coding gene promoters, differential methylation was observed within 1,000 bp upstream or downstream of the TSS in non-coding RNAs and pseudogenes (**Figure 5.6A**). MicroRNAs (miRNA) are non-coding RNAs that regulate expression through imperfect

base-pairing with the 3'UTR of multiple target genes. A total of 69 miRNAs were differentially methylated in ALL including miR-375, miR-196a, miR-3545, miR-9-1/2/3, miR-124-1/3, and miR-34b, which have been implicated in human malignancies (Chim et al. 2011; Wang et al. 2013; Chatterton et al. 2010). RNA-seq libraries were prepared from polyA RNA and excluded the capture of miRNA; therefore, correlation studies between methylation and gene expression were not performed for miRNA. The regulatory potential of DMRs associated with miRNAs warrants further attention. Long intergenic non-coding RNAs (lincRNAs) are emerging as key regulators of numerous cellular processes and regulate the expression of multiple target genes. Differential methylation occurred in 65 lincRNAs. Of these, hypomethylation and up-regulation was observed in AC002398.5, DIO3OS and LINC00642. Lastly, 55 pseudogenes were differentially methylated in ALL. No correlations between expression and promoter methylation was observed in the pseudogenes; however, pseudogenes, much like lincRNAs, have the potential to epigenetically regulate their parental genes and were further investigated.

It is well known that transposable element activities are often silenced by DNA methylation (Maksakova et al. 2008), and that transcriptional activation of these elements results in transposable element mediated insertions and chromosomal rearrangements in many cancers (Lee et al. 2012). Many of the intergenic DMRs were associated with transposable elements and repeat sequences (**Figure 5.6B**). Non-autonomous short interspersed nuclear elements (SINE) were the most abundantly present transposable element within the differentially methylated intergenic regions followed by long terminal repeat (LTR), autonomous long interspersed nuclear elements (LINE) and satellite repeats**.** Centromeric alpha satellite repeats were often hypermethylated in ALL which may block CENP-A and result in centromere inactivation.

DMRs are associated with predicted regulatory sequence: Differential methylation predominately occurred in intergenic and intronic regions in ALL. One third of the intronic DMRs (3,341) were located within 150 base pairs of the 5' or 3' splice sites and could potentially alter appropriate splicing in ALL. To investigate whether the intergenic and intronic DMRs coincided with the location of regulatory enhancer elements, the sites for intergenic and intronic DMRs were overlaid with ENCODE Chip-seq data for enhancer related histone marks (H3K4me1 and H3K27ac) in the GM12878 lymphoblastoid cell line. Overall, 765 intergenic and intronic DMRs overlapped with potential enhancer like regions (eDMR). Of these, 453 were hypomethylated and 312 were hypermethylated. Enhancer methylation has been shown to have a stronger association with gene deregulation than promoter methylation in cancer (Aran et al. 2013). To investigate the association between enhancer methylation and gene expression in our data, lists were constructed of the nearest upstream and downstream gene to identify the potential target genes for each eDMR. A total of 81 genes exhibited significantly decreased expression in ALL that corresponded with hypermethylation of potential neighboring eDMRs, and 111 genes showed significantly increased expression that corresponded with eDMR hypomethylation. Functional annotation clustering revealed that down-regulated genes with eDMR hypermethylation included those involved in cell cycle processes, cell division, regulation of gene expression, cytoskeleton, and a large number of zinc finger proteins; whereas up-regulated genes with eDMR hypomethylation included those involved in lymphocyte activation, cell migration, apoptosis, DNA replication and DNA metabolic processes.

Gene body DMRs are associated with gene expression: Associations between gene body methylation and gene expression were also observed. Increasing gene body methylation along with promoter methylation has been shown to have a stronger repressive effect on gene

expression during normal B-cell development than promoter methylation alone (Lee et al. 2012). However, the effect of gene body methylation in the absence of promoter methylation is less clear. Both inverse and positive correlations between gene body methylation and gene expression were observed. Gene body hypermethylation and a significant decrease in expression was observed in 261 genes and included protein kinases (*CDK5R1, NRBP1, LYN, NUAK2, PHKB, BLK, PRKAG2, MKNK2, SMG1, TRIO, GAK, PRKD2, ULK1, RIOK3, WNK4, MAP3K9, PDGFRA, NEK8, DCLK2, TLK2, LRRK1, CDC42BPB, CAMK1D*) , cell morphogenesis genes (*CDK5R1, GDF7, ULK1, LAMA5, NR4A2, MAPK8IP3, SEMA3B, MYCBP2, NFATC1*), lymphocyte differentiation genes *(CHD7, IL7, CEBPG, HDAC9, FOXP1),* chromatin modifiers (*RSF1, CREBBP, BANP, ARID1B, UIMC1, CHD8, CHD7, WHSC1L1, PHF21A, TLK2, IRF4, HDAC9, RERE*), and regulators of MAPK, JNK, JUN kinase activity. Conversely, gene body hypomethylation and a significant increase in expression was observed in 815 genes and included the DNA methyltransferases (*DNMT3A* and *DNMT1),* anti-apoptotic genes (*IL2RB, PRDX2, BCL2L1, TCF7L2, DAPK1, AKT1, ATF5, BAX, TGM2, NOS3, THBS1* and *MYO18A),* and telomere organization genes *(TERT* and *TNKS1BP1).* Additionally, many genes showed positive correlations between methylation and expression. For example, several members of the protein tyrosine phosphatase family that regulate many cellular processes such as cell growth, mitotic cycle, cellular differentiation, and malignant transformation were up-regulated and hypermethylated in ALL. Alternately, genes that play roles in B-cell activation were down-regulated and hypomethylated in ALL.

5.4.3 B-cell development genes and epigenetic modifiers are aberrantly expressed in ALL

To investigate the deregulation of gene expression in ALL, genome-wide gene expression profiling of ALL patients and healthy precursor B-cells was performed using RNA-seq. A total

of 3,699 genes were significantly up-regulated in ALL versus healthy samples and 2,734 genes were significantly down-regulated. Forty-three genes known to play roles in B-cell differentiation and activation were differentially expressed and may contribute to the pathogenesis of ALL (**Table 5.5**). The aberrant expression of epigenetic modifiers was also observed (**Figure 5.7**). The DNA methylation catalyzing enzymes *DNMT1*, *DNMT3A*, and *DNMT3B* were significantly up-regulated in ALL. Conversely, 2 genes known to actively demethylate DNA (Pastor et al. 2013), *TET2* and *TET3*, were significantly down-regulated in ALL. In addition, 22 genes encoding histone proteins were significantly up-regulated in ALL. Finally, the chromatin activating histone lysine acetyltransferases (*MGEA5, CDYL, CREBBP, EP300,* and *NCOA3),* were down-regulated and the chromatin inactivating histone deacetylases (*HDAC11* and *SIRT2)* were up-regulated in ALL.

5.4.4 Differential expression of transcripts with epigenetic regulatory functions

LincRNAs epigenetically regulate gene expression by a number of diverse mechanisms including recruitment of histone methyltransferases through polycomb repressor complex 2 to modify chromatin states (Nagano and Fraser 2011; Wang and Chang 2011), and the differential expression of lincRNA has been shown to play critical roles in many diseases (Gupta et al. 2010; Trimarchi et al. 2014). Differential expression analysis of lincRNAs in ALL patients compared to healthy controls revealed 197 lincRNAs were differentially expressed Among them, 104 lincRNAs were upregulated (**Table 5.6)** and 93 were down-regulated in ALL **(Table 5.7**).

Pseudogene transcripts play a significant role in cancer pathogenesis and are differentially expressed in different types of cancer (Han et al 2014; Xiao-Jie et al. 2015). The relationship between differentially expressed pseudogene transcripts on the expression of parent

gene targets was diverse in our data (**Figure 5.8**). In some instances, the up-regulation of a pseudogene was associated with the down-regulation of its parent gene. For example, the pseudogenes *GRK6P1* were up-regulated and associated with down-regulation of their parent genes *GRK6.* Interestingly, *GRK6* phosphorylates the activated forms of G protein-coupled receptors (GPCRs) thereby instigating their deactivation. Further, the overexpression of GPCRs is known to contribute to cancer cell proliferation (Dorsam et al. 2007). Thus, the up-regulation of *GRK6P1* may lead to the constitutive activation of GPCRs and contribute to the proliferation of cancer cells. Conversely, in other instances the down-regulation of a pseudogene was associated with the up-regulation of its parent gene. For example, the down-regulation of *AC007041.2, RP11-368P15.1* and *KRT18P4* was associated with the up-regulation of *DRG1* and *NDUFB3,* and *KRT18* respectively. *KRT18* (cytokeratin 18) is involved in multiple cellular processes including apoptosis, mitosis, cell cycle progression and cell signaling and is hypothesized to be involved in carcinogenesis through multiple signaling pathways (Weng et al. 2012). Therefore, the pseudogene mediated up-regulation of *KRT18* may lead to the aberrant regulation of signaling pathways in ALL.

A positive correlation was also observed in which up-regulated pseudogene transcripts were associated with up-regulated parent gene transcripts and down-regulated pseudogene transcripts were associated with down-regulated parent gene transcripts. In these cases the pseudogene transcripts may up-regulate their parent gene by competing with endogenous RNAs that share miRNA response elements (Poliseno et al. 2010), or by competing for RBP that degrade their parent gene and vice versa. In ALL, the up-regulation of pseudogenes *RP11-423H2.1, FAM86C2P*, and *HMGB1P41* was associated with the up-regulation of their parent genes *THOC3, FAM86A,* and *HMGB2*. Previous studies have shown that *HMGB2* is

overexpressed in a variety of cancers and that there is a decline in the proliferation of cancer cells when siRNA is used to knockdown expression of *HMGB2* (Sharma et al. 2008; Kwon et al. 2010), suggesting a putative role in the pathogenesis of ALL. Likewise, some pseudogenes were down-regulated and their parent genes were also down-regulated. Specifically, the down-regulation of *PABPC1P3* was associated with the down-regulation of its parent gene, *PABPC1* which encodes a poly(A) binding protein involved in stabilizing the 5' cap of mRNA. The down-regulation of *PABPC1* has also been reported in esophageal cancer (Takashima et al. 2006). It is possible that the pseudogene mediated down-regulation of *PABPC1* results in unstable mRNA transcripts and contributes to the pathogenesis of ALL.

5.5 DISCUSSION

On average more than 50 million unique mapped MIRA-seq reads were generated providing genome-wide coverage of the methylome in 19 pediatric ALL patients. Importantly, these profiles were compared to healthy precursor B-cells isolated from umbilical cord blood, the normal counterparts of malignant precursor B-cells to identify DMRs. To determine the regulatory potential of DMRs, transcriptomes were also generated and differential expression was determined between ALL patients and normal controls. Previous studies in ALL have identified inverse correlations between gene expression and DNA methylation in CGIs, CGI shores and gene promoters (Busche et al. 2013). In this study 99% of DMRs associated with a CGI were hypermethylated in ALL; however, these only accounted for a small number of the total DMRs. In fact, more than 80% of DMRs were identified in intronic or intergenic regions and not within a CGI context. Since DMRs can be used as biomarkers and as targets for novel therapeutics, we sought to identify the most likely candidates with regulatory potential.

Strikingly 70% of the intergenic DMRs were concomitant with functional regulatory elements including transposable elements, enhancers, transcription factor binding sites, ncRNA and pseudogenes. Inverse and positive correlations between DNA methylation in regulatory regions and gene expression were observed. In addition, inverse and positive correlations were observed between gene body methylation and expression. The cause and effect of DNA methylation within gene bodies is not fully understood; however, mechanisms leading to faulty gene expression have been postulated including the regulation of transcriptional elongation (Lorincz et al., 2004), cell-type specific selection of alternative promoters (Maunakea et al. 2010), modulating alternative RNA splicing (Maunakea et al. 2013), or defining alternative polyadenylation sites (Wood et al. 2008).

Genes that are regulated by DNA methylation and provide a selective growth advantage to cancer cells have been referred to as epi-driver genes (Vogelstein et al. 2013). The ability to weed out driver epi-mutations from passenger epi-mutations is crucial in the quest to delineate potential therapeutic targets from a multitude of passenger events. Integrated DNA methylation and gene expression analysis identified potential epi-driver genes including *SYNE1* (cytokinesis), *PTPRS* (signaling molecule), *PAWR* (pro-apoptotic gene), *HDAC9* (downstream target of *KRAS*), *RGCC* (cell-cycle regulator) and *MCOLN2* (unknown function), which were hypermethylated in the 5'regulatory region and down-regulated in ALL. These genes have also been shown to be hypermethylated and/or down-regulated in other malignancies (Nagai et al. 2010; Okudela et al. 2014; Vlaicu et al. 2008), indicating the potential for tumor suppressor activity and supporting the role of DNA methylation as a regulator of gene expression. Although the function of *MCOLN2* is unclear, the B-cell lineage specific activator PAX5 regulates its expression, strongly implicating its involvement in early B-cell development (Valadez and

Cuajungco 2015). Taken together, the down-regulation of these genes due to DNA methylation may play important roles in the development of ALL.

Perhaps the most paramount finding of this study was the identification of potential regulatory enhancers (eDMR). In relation to this, potential epi-drivers regulated by DNA methylation of an eDMR were also identified. Three of the genes with hypermethylated promoter DMRs (*SYNE1*, *PTPRS* and *MCOLN2*) also possessed a hypermethylated eDMR. In addition, *LYN* and *TRAF3* were down-regulated in ALL patients and associated with a hypermethylated eDMR. *LYN* plays an important role in the regulation of B-cell differentiation, proliferation, survival and apoptosis, and *TRAF3* negatively regulates the activation of the NF-κB2 pathway in B-cells. Conversely, *FLT1* and *MELK* were up-regulated and associated with a hypomethylated eDMR. Both genes have previously been shown to be upregulated in cancer (Van Limbergen et al. 2014; Alachkar et al. 2014). Furthermore, *FLT1* has been shown play a role in the proliferation of tumor cells (Lichtenberger et al. 2010), and suppression of *MELK* expression by siRNA has been shown to inhibit the growth of cancer cells. Therefore, the aberrant expression of these genes due to DNA methylation may provide a survival advantage to malignant cells and play a role in pediatric ALL.

In summary, novel differentially methylated regulatory regions and differentially expressed genes were identified that may contribute to the pathogenesis of ALL. As expected, genes associated with B-cell development and epigenetic modifier genes were differentially expressed. The *de novo* DNA methyltransferases (*DNMT3A, DNMT3B*) responsible for the establishment of DNA methylation patterns and chromatin inactivating deacetylase genes were up-regulated; whereas chromatin activating acetyltransferase genes were down-regulated in ALL. The result of the aberrant expression of the epigenetic modifier genes observed in this study may effectively

be the inactivation of key genes that contribute to ALL. In addition, pseudogenes and lincRNAs genes were also aberrantly expressed in ALL and have functional roles in epigenetic regulation through diverse mechanisms including behaving as antisense RNA, endo-siRNA, competing endogenous RNA or competing for RNA-binding proteins to regulate their target genes. Finally, for the first time, putative transcriptional enhancers were identified that were differentially methylated and associated with the expression of a neighboring gene. Importantly, these may be used as prospective biomarkers for ALL and/or as targets for novel therapeutic agents that can restore altered DNA methylation and gene expression back to the normal state with the ultimate goal of improving treatment therapies and patient outcomes.

Table 5.1. Pre-B ALL patients characteristics

Patient ID	Blast rate (%)	Age (months)	WBC, $10^3/\mu l$	Sex	Immunophenotype	Cytogentics
A4	88	4	7.8	M	19;10	hyperdiploidy
A15	94	36	7.8	M	19;10	hyperdiploidy
A18	97	17	4.3	F	19;10	46, XX-15der(1) t(1;?),del(6)(q21),t mar
A19	88	36	3.7	M	19;10	hyperdiploidy
A20	92	120	3.6	M	19;10	46, XY
A21	91	36	6.6	M	19;10	46, XY t(3;19)(p25;p13)
A22	94	60	2.5	F	19;10	47, XX +21; 48, XX
A23	96	180	2.3	M	19;10	46, XY del(6)(q21;q27)
A24	94	108	3.7	M	19;10	45, -7 -9 +der(9) t(8;9)(q112;p11)
A25	96	48	13.7	M	19;10	46, XY
A26	91	48	4.3	M	19;10	47, XY
A28	96	36	1.5	M	19;10	none available
A29	93	24	10.2	F	19;10;20	46, XX
A30	94	24	3.7	F	19;10;20wk	46, XX
A31	94	132	18.8	M	19;10;20	45, XY -7
A32	92	36	3.4	M	19;10;20	none available
A33	88	180	4.5	M	19;10;20	46, XY
A35	97	22	25.9	M	19;10;20	46, XY
A36	91	72	2.7	F	19;10;20	46, XX
A37	93	20	2.5	M	19;10	hyperdiploidy

Table 5.2: Illumina sequencing data for MIRA-seq.

Sample ID	Total reads	% mapped	Unique mapped reads
A4	158,898,810	92	94 ,95,865
A15	360,924,065	87	102,021,923
A18	143,609,332	60	65,103,827
A19	125,564,694	92	62,912,100
A20	262,984,833	87	22,641,915
A21	156,947,890	52	66,061,458
A22	228,014,591	88	93,055,414
A23	184,851,358	85	109,086,484
A24	188,016,523	93	87 ,94,239
A25	134,869,456	78	65,509,709
A26	126,328,156	84	66,745,501
A28	135,392,530	89	81,825,617
A29	111,707,016	84	60,404,578
A30	152,276,253	92	72,383,507
A31	196,300,860	92	88,702,037
A33	81,958,128	89	48,596,416
A35	295,098,948	88	114,910,735
A36	218,117,998	88	97,881,476
A37	87,262,577	91	51,645,043
HCB1_pre-BI	168,552,888	95	85,329,586
HCB2_pre-BI	265,314,820	90	67,984,512
HCB3_pre-BI	84,521,014	94	39,905,728
HCB4_pre-BI	148,457,990	93	47,064,491
HCB5_pre-BII	240,194,057	92	83,332,904
HCB6_pre-BI	188,877,538	94	131,520,066
HCB7_pre-BI	147,961,433	94	57,073,380
HCB8_pre-BI	148,151,509	93	65,601,840
HCB9_pre-BI	175,217,423	94	62,713,505
HCB10_pre-BI	258,120,121	93	155,091,010
HCB1_pre-BII	128,568,614	93	41,948,710
HCB2_pre-BII	240,907,679	90	78,846,617
HCB3_pre-BII	150,168,486	94	62,616,516
HCB4_pre-BII	195,937,594	94	63,730,137
HCB5_pre-BII	237,968,679	92	51,527,842
HCB6_pre-BII	182,755,070	95	124,485,754
HCB7_pre-BII	228,160,559	94	98,691,194
HCB8_pre-BII	281,061,395	93	67,410,961
HCB9_pre-BII	152,443,649	94	64,389,853
HCB10_pre-BII	137,008,710	93	88,156,469

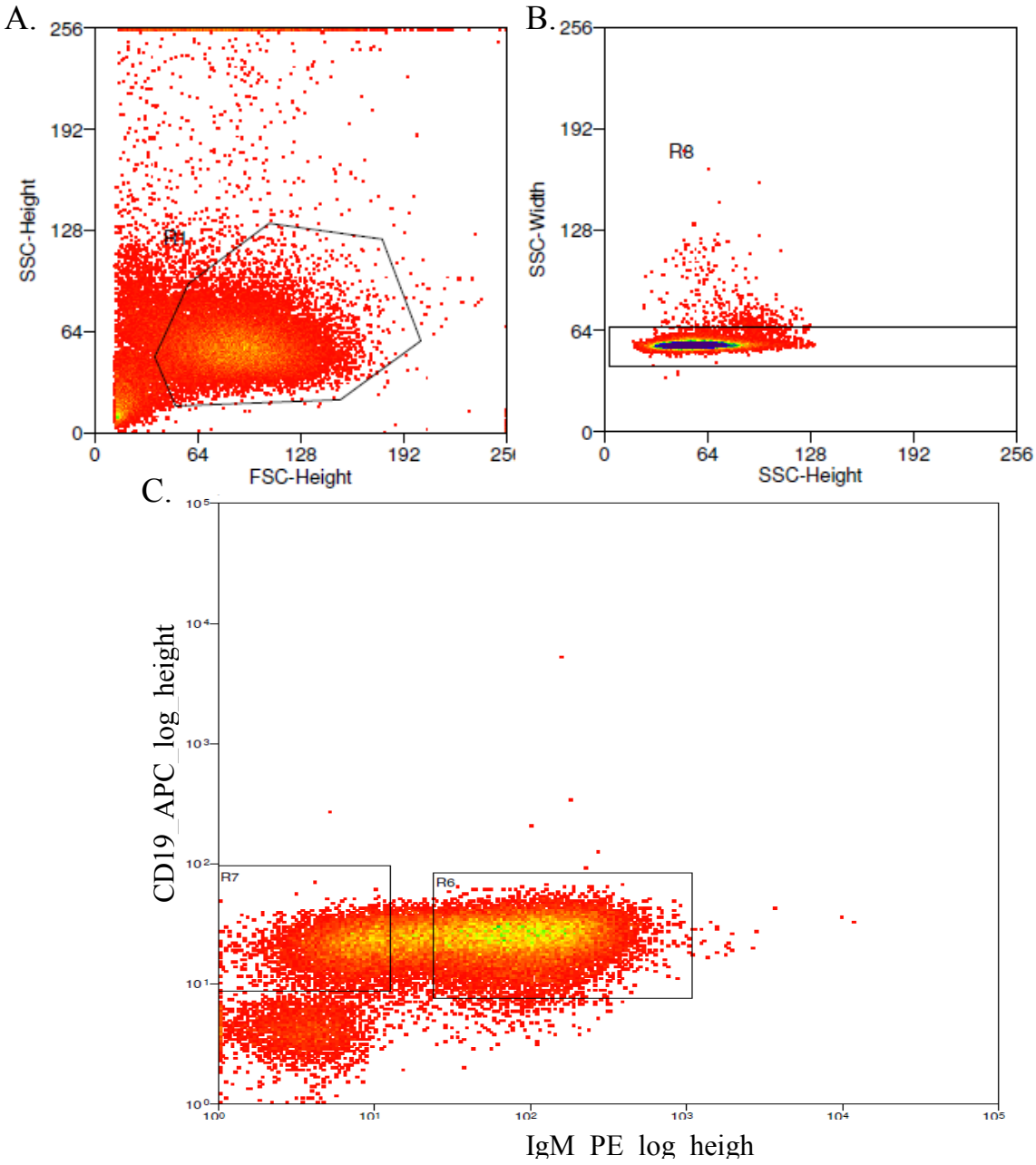


Figure 5.1: Gating strategy for isolation of pre-B cell ($CD19^+$, IgM^-) from HCB. (A) Lymphocyte gate. (B) Single cell gate. (C) Pre-B cell ($Cd19^+$,IgM^-)

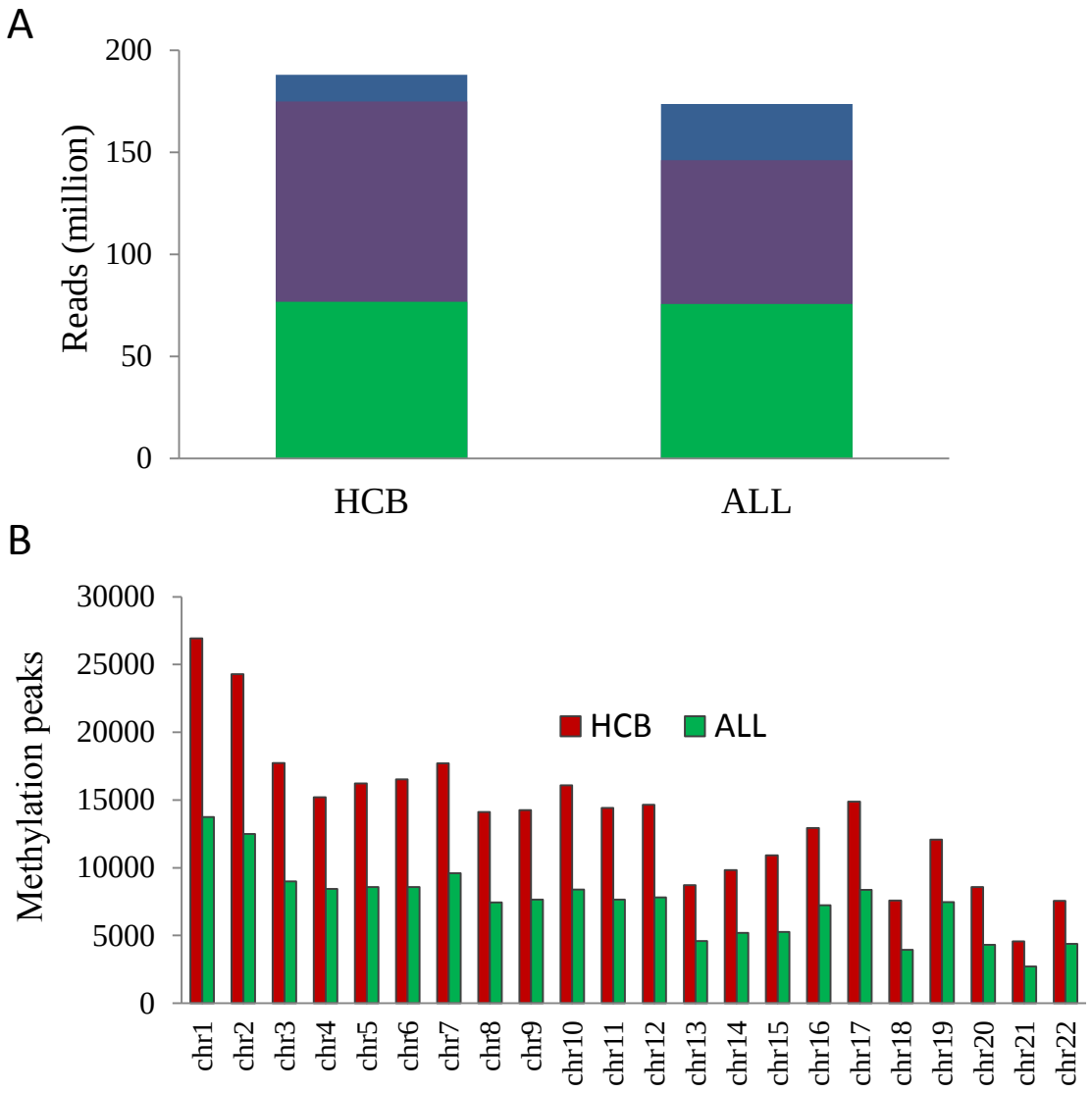


Figure 5.2. **Genome-wide DNA methylation profiles in HCB and ALL.** (A) Average read and alignment statistics. Reads were averaged across all individuals for HCB and ALL samples. The top of each bar represents the total number of reads for each category. Blue bars: total reads; Purple bars: reads mapped; Green bars: unique reads. (B) Chromosome-wise methylation peaks. The X and Y chromosomes were excluded from analysis.

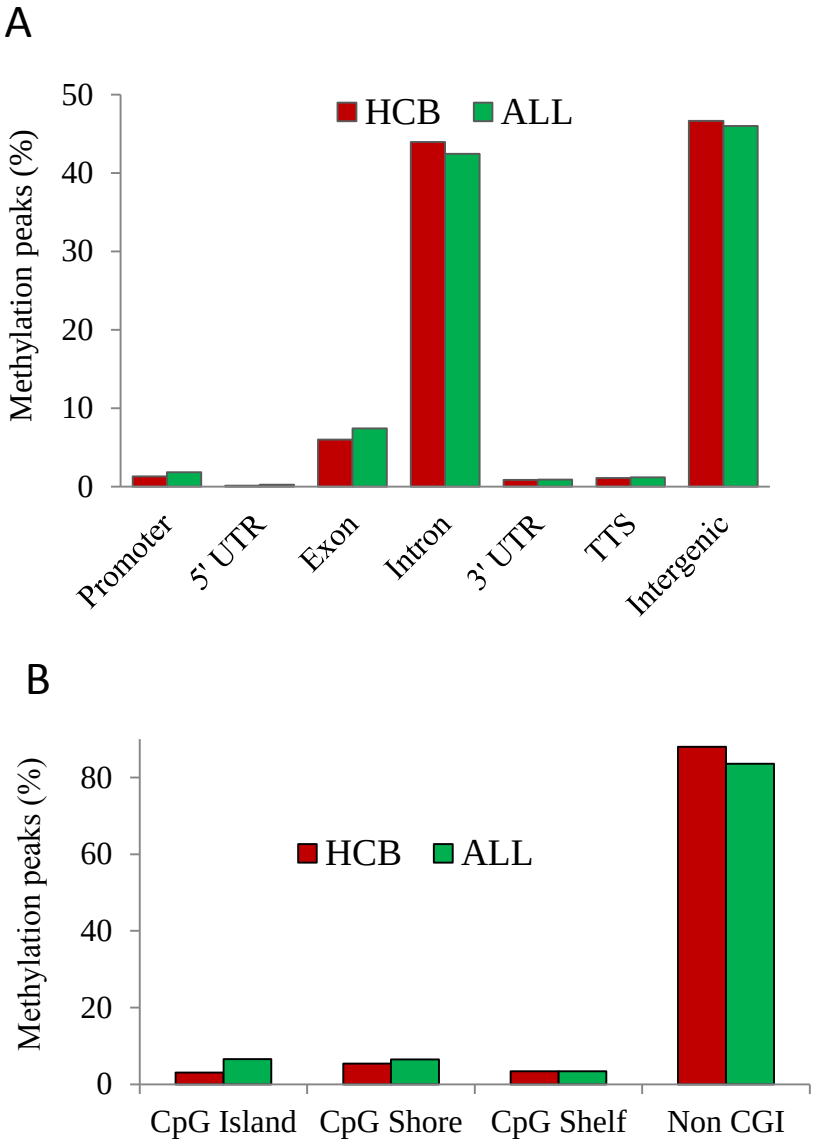


Figure 5.3: Distribution of methylation peaks. (A) Genomic distribution of methylation peaks. (B) Distribution of methylation peaks in CGI context.

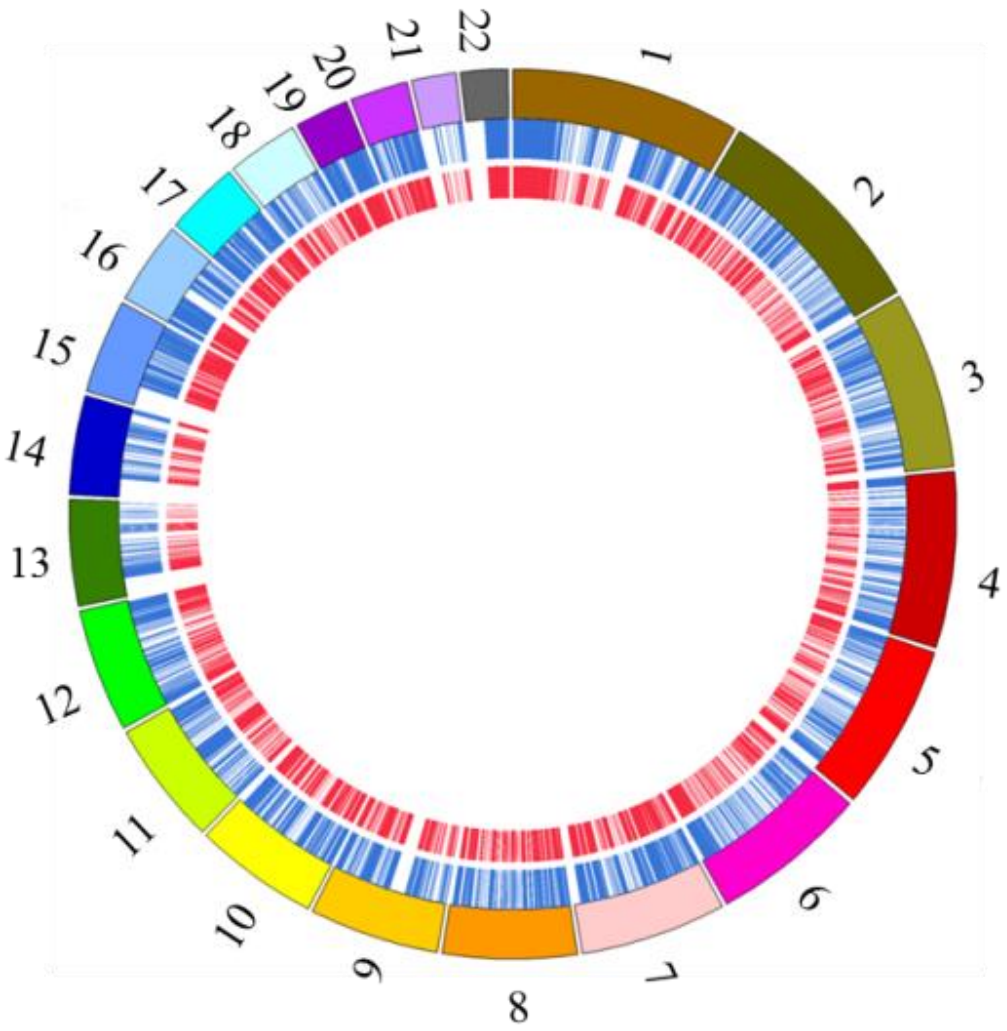


Figure 5.4: Differentially methylated regions in ALL. Blue represents hypo- and red represents hyper-methylated regions in ALL**.** Circos plots representing chromosomes 1-22. The X and Y chromosomes were excluded from analysis.

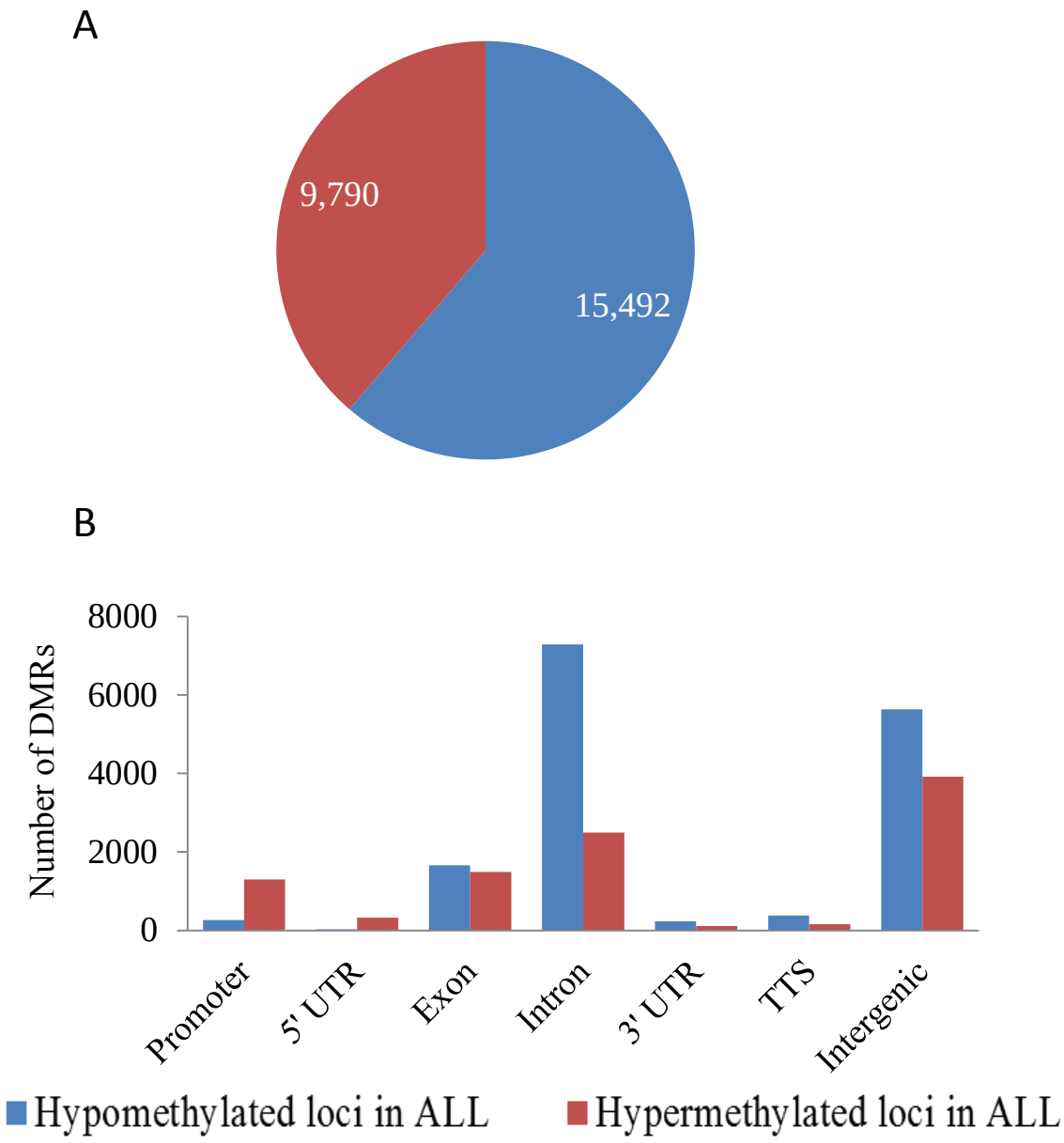


Figure 5.5: Hypomethylated (blue) and hypermethylated (red) regions in ALL. (A) Total number of DMR. (B) Genomic distribution of hypo- and hypermethylated DMR.

Table 5.3: Gene promoter hypermethylation associated with significant decreased expression in ALL. *Log_2 fold change difference in expression (FPKM).

Gene	Fold change*	q_value	Gene	Fold change*	q_value
ABCB4	-3.17	0.00045	MCOLN2	-4.06	0.00045
BAIAP3	-2.52	0.00082	SRSF12	-1.92	0.01748
HDAC9	-2.42	0.00045	BTG3	-1.66	0.00361
SLC6A16	-3.00	0.00045	CBS	-1.78	0.00045
ASNS	-1.34	0.03243	SNED1	-3.06	0.00045
RHOBTB1	-2.75	0.00115	WDR43	-0.94	0.03366
TMEM131	-1.71	0.00045	GPR98	-2.40	0.04144
PKD2L2	-2.66	0.00286	KIAA0355	-2.13	0.00045
GNAS	-1.13	0.00045	TMEM88	-3.00	0.01560
RGCC	-1.97	0.00045	DCLK2	-3.27	0.00361
SYT17	-3.63	0.00231	MTSS1	-2.97	0.00045
PTPRS	-2.59	0.00045	FRMD5	-1.96	0.00503
PDE4C	-3.82	0.00045	CCDC36	-2.01	0.02025
PRUNE2	-1.47	0.02624	GOLGA8A	-1.45	0.03322
SETX	-1.43	0.00045	ZDHHC21	-2.03	0.00175
SLC1A2	-2.36	0.01682	DOK7	-2.27	0.00458
KHDRBS2	-4.83	0.01077	BASP1	-3.79	0.00045
CNR1	-4.89	0.00045	PAWR	-2.77	0.00677
ARL4A	-3.72	0.00045	MLF1	-3.58	0.00045
AHNAK	-2.17	0.00115	GALNT11	-1.45	0.00458
RFPL3	-6.17	0.00082	HKR1	-1.14	0.00045
SYNE1	-1.30	0.00045	EXT1	-1.65	0.02961
CEPT1	-2.46	0.00045	SDR42E1	-1.89	0.04046
IFT172	-0.82	0.03298	AHNAK2	-1.29	0.02217
NUPL1	-1.79	0.00045	KCNQ5	-3.31	0.00115
DEF8	-2.27	0.00045	CD47	-1.29	0.02272
SKAP1	-2.72	0.00082	CDC42BPB	-1.98	0.00887
SLC13A5	-1.71	0.00744	PKHD1L1	-5.38	0.00336
THNSL2	-2.29	0.01364	RFPL3S	-4.93	0.00045
OSMR	-2.38	0.00045	ADRA2B	-2.90	0.03193
AP1S3	-3.53	0.00045	ZNF571-AS1	-1.81	0.00045

Table 5.4: Gene promoter hypomethylation associated with significant increased expression in ALL. *Log_2 fold change difference in expression (FPKM).

Gene	Fold change*	q_value	Gene	Fold change*	q_value
DPEP1	8.21	0.00045	IFITM3	5.64	0.00045
RNH1	0.86	0.00925	EVI5L	1.58	0.01381
NGFR	6.57	0.00175	SELENBP1	3.25	0.00045
NDST1	2.48	0.00045	SLC39A4	2.21	0.00634
KIF22	1.78	0.00045	MARVELD1	6.20	0.00045
PEBP1	2.12	0.00115	XKR8	3.65	0.01930
CHN2	4.61	0.00045	EVPL	4.23	0.00204
ENG	2.98	0.00045	CDH4	5.48	0.01020
CD81	1.26	0.01112	RGMA	7.47	0.00082
NDUFS8	1.70	0.00045	GAS6	2.85	0.00045
CDK2AP1	3.42	0.00045	BRF1	1.93	0.00045
HPCAL1	1.53	0.01615	BCR	1.39	0.00045
SOCS2	8.75	0.00045	LINC00963	1.30	0.01257
CALML4	1.56	0.01095	NEU4	3.15	0.00045
TMEM204	2.78	0.04802	AC002398.5	3.23	0.04233
DAB2IP	4.62	0.00045	LINC00642	4.70	0.03035
SLCO2B1	1.73	0.02105	BRK1	1.54	0.01364
RAB40B	3.04	0.00045	DIO3OS	1.65	0.01846
SECTM1	2.96	0.00045			

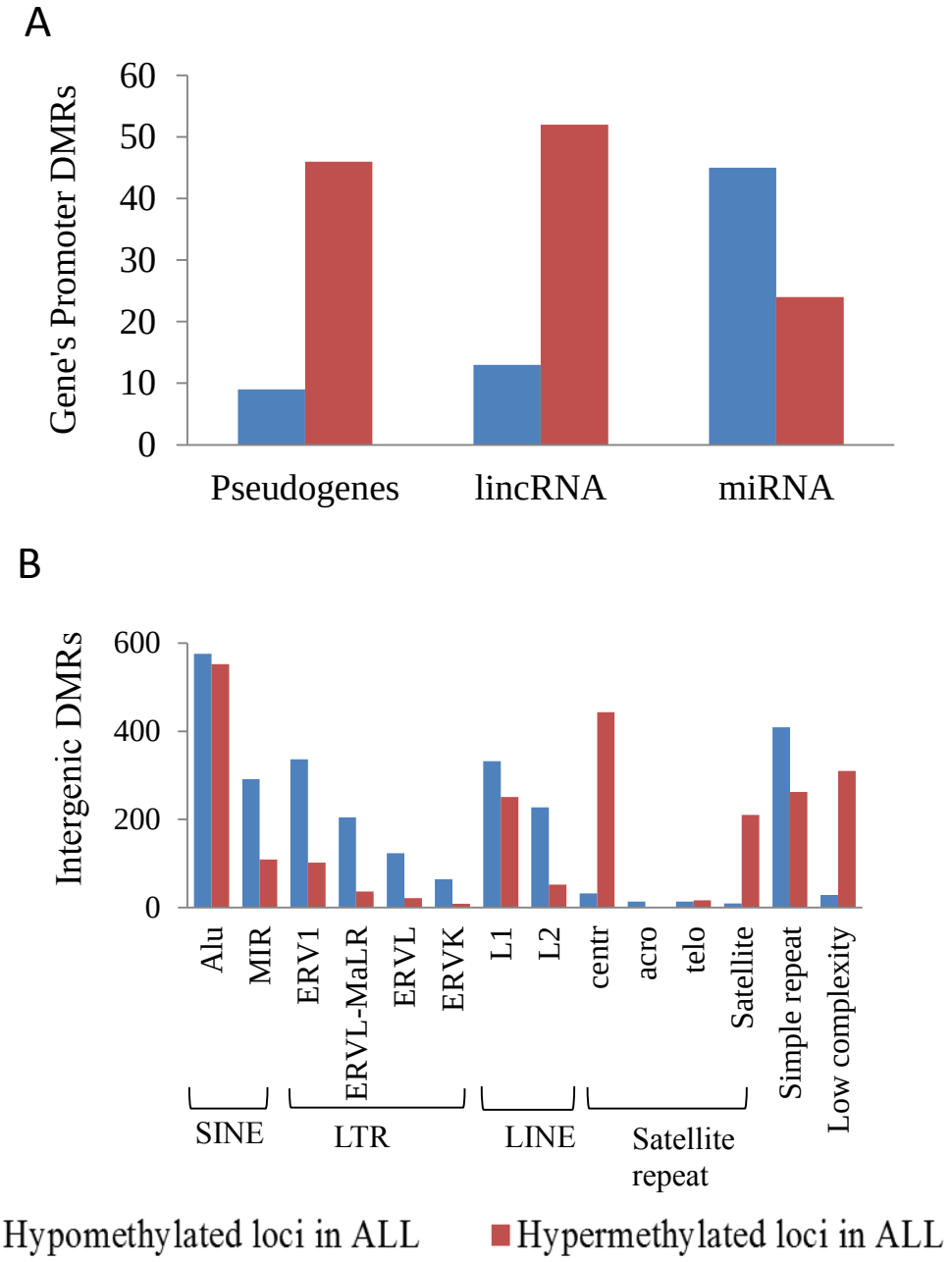
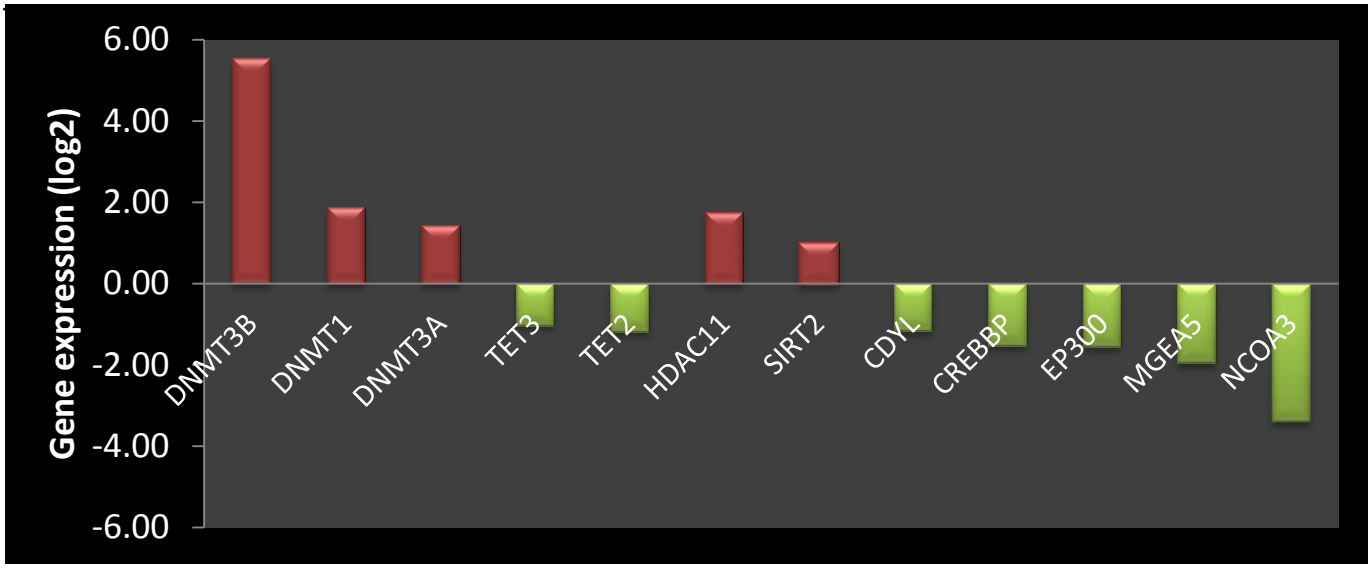


Figure 5.6: DMRs associated with ncRNA and transposable elements. (A) DMRs associated with the 5'regulatory region of pseudogenes and noncoding RNA. (B) Intergenic DMRs associated with transposable elements and repeat sequences.

Table 5.5: Differentially expressed genes in ALL involved in B-cell development. Positive log fold change= up-regulated in ALL, negative fold change= down-regulated in ALL.* Log2 fold change.

Genes	Fold change*	Genes	Fold change
DNTT	10.25	CEACAM1	1.21
VPREB1	9.08	PAXBP1	1.12
RAG1	8.81	MALT1	-1.05
RAG2	8.52	IGHM	-1.17
IGLL1	7.37	ETS1	-1.39
FCER1G	6.41	BCL2	-1.40
LEF1	5.49	HLA-DMB	-1.58
TNFSF4	4.91	RFX1	-1.63
HMGB2	3.49	BCL10	-2.08
LCP2	3.30	IL24	-2.16
OAS3	3.21	BTG1	-2.18
VPREB3	3.15	HLA-DQB1	-2.39
IL18R1	2.94	IRF4	-2.46
BST1	2.87	FCGR2B	-2.56
CD59	2.63	ADAM8	-2.60
CTSC	2.31	CARD11	-3.24
SOX4	2.15	ADAM19	-3.59
ADA	1.91	MS4A1	-3.87
IGJ	1.89	IL4R	-4.00
LRRC8D	1.84	LYN	-4.081
NOTCH1	1.62	IRF8	-4.298
TCF3	1.22		

A



B

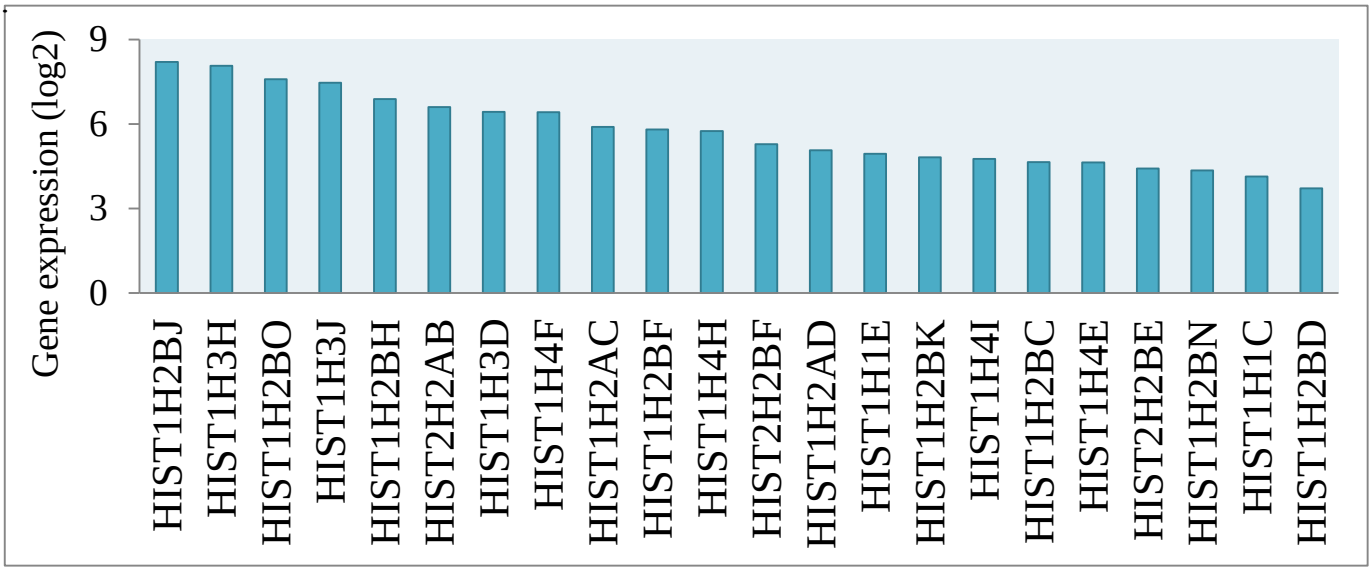


Figure 5.7: **Aberrant expression of epigenetic modifiers in ALL (log_2 fold change).** (A). DNA methyltransferases DNMT1, DNMT3A, and DNMT3B were significantly up-regulated in ALL. DNA demethylases TET2 and TET3 were significantly down-regulated in ALL. Histone deacetylases (HDAC11 and SIRT2) were up-regulated in ALL and histone lysine acetyltransferases (MGEA5, CDYL, CREBBP, EP300, and NCOA3)were down-regulated in ALL. (B). Genes encoding histone proteins were significantly up-regulated in ALL.

Table 5.6: Up-regulated lincRNA in ALL. * fold change was not calculated due to the expression values in healthy pre-B was zero.

LincRNA	Fold change (Log2)	LincRNA	Fold change (Log2)	LincRNA	Fold change (Log2)
RP11-301G19.1	11.0216	AC016735.2	3.77333	RP11-161M6.2	2.03259
SNORD3A	8.93891	RP5-1024G6.8	3.76655	CTD-2587H19.1	2.00297
RP11-76E17.3	8.53287	CRNDE	3.74897	MIR4453	1.97689
RP11-301H24.4	7.37013	RP1-60O19.1	3.70083	PART1	1.92267
RP11-322E11.5	7.19327	SNHG15	3.69053	MIR4435-1HG	1.87498
XXbac-B33L19.6	7.13743	LINC00925	3.68746	TPTEP1	1.7972
LINC00861	6.97658	RP11-1149M10.2	3.62006	CTC-503J8.6	1.78794
LINC00958	6.50334	LINC00461	3.61917	RP11-620J15.3	1.76433
LINC00487	6.34398	CTD-3220F14.1	3.60779	LINC00085	1.75603
AL589743.1	6.29196	RP11-676J15.1	3.59706	RSBN1L-AS1	1.70091
SNORA71B	6.06094	AC019118.3	3.54765	DIO3OS	1.6467
TERC	5.92382	KIF25	3.51118	RP11-435B5.5	1.61682
RP11-35J10.5	5.80256	CTD-3049M7.1	3.42807	HCG11	1.55098
CTD-2291D10.4	5.65644	TEX41	3.3701	NEAT1	1.53412
SCARNA2	5.6246	RP11-212I21.4	3.35001	RP11-782C8.5	1.31346
LINC00922	5.45859	RP11-439C15.4	3.30158	RP11-1246C19.1	1.2426
RP11-134O21.1	5.28229	CTD-2037K23.2	3.28292	AC017002.2	*
RMRP	5.23731	CTB-175P5.4	3.25416	AF064858.8	*
RP11-410D17.2	5.18407	AC002398.5	3.22895	AL132796.1	*
RP11-439L8.4	5.16466	RP11-290F5.2	3.21663	RP11-260A9.6	*
LINC00664	5.04427	BAIAP2-AS1	3.20252	RP11-400K9.4	*
RP11-114G22.1	5.00214	AP000476.1	3.17655	RP11-403N16.2	*
LINC01013	4.9257	LINC00426	3.05976	RP11-403N16.3	*
CTA-373H7.7	4.77356	AF127936.7	3.00507	RP11-462G22.1	*
CTD-2008L17.1	4.75478	AC006129.1	2.83324	RP11-470M17.2	*
KIF25-AS1	4.72123	LINC00539	2.79902	RP11-473E2.3	*
LINC00642	4.70373	AF127936.3	2.79756	RP11-542A14.2	*
LINC01016	4.1583	CTD-3162L10.1	2.78842	RP1-16A9.1	*
U1	4.15398	AC093323.3	2.49259	RP11-768F21.1	*
MIR4458HG	4.14194	RP11-222A11.1	2.4111	RP11-923I11.5	*
RP5-943J3.2	4.05018	MIAT	2.38481	RP1-223B1.1	*
RNU12	4.03313	RP5-1043L13.1	2.35784	RP1-35C21.2	*
CASC15	4.02091	MIR181A1HG	2.29911	RP3-380B8.4	*
RP11-231E4.2	3.96416	LINC00152	2.21827	RP3-477M7.5	*
PRKCQ-AS1	3.89534	RP11-427H3.3	2.05548		

Table 5.7: Down-regulated lincRNA in ALL.

LincRNA	Fold change (Log2)	LincRNA	Fold change (Log2)	LincRNA	Fold change (Log2)
RP11-359N11.2	-7.7081	RP11-57A19.5	-4.06137	RP11-38J22.6	-2.46354
RP3-448I9.1	-7.70383	RP11-255C15.3	-4.02432	RP13-516M14.1	-2.45035
RP11-359N11.1	-6.96853	RP11-159D12.2	-3.90345	RP11-1100L3.8	-2.38801
AC092620.3	-6.82846	CTC-308K20.2	-3.86445	RP11-506N2.1	-2.36851
XXbac-BPG170G13.32	-6.74266	RP11-255C15.4	-3.84445	RP11-215G15.5	-2.34521
RP11-446E9.2	-6.61287	RP11-803P9.1	-3.8444	RP11-449H11.1	-2.2574
RP11-2H8.2	-6.02993	RP1-224A6.9	-3.83327	RP11-381O7.3	-2.25124
AC079767.4	-5.95251	BACH1-IT2	-3.79543	RP11-444D3.1	-2.2412
RP1-149A16.17	-5.73057	CTD-2184D3.6	-3.73482	RP11-584P21.2	-2.24101
LINC00926	-5.60209	CTC-301O7.4	-3.42549	RP11-220I1.1	-2.1667
RP11-277L2.4	-5.60045	RP11-58E21.3	-3.34217	RP11-430B1.2	-2.1069
AC007386.2	-5.51412	CTC-490E21.10	-3.28387	RP5-1159O4.1	-2.08079
RP5-887A10.1	-5.27281	RP5-1139I1.1	-3.2399	RP11-960L18.1	-2.0125
RP11-309L24.9	-4.99748	RP4-798A10.2	-3.16312	RP11-861A13.4	-2.00773
RP1-149A16.3	-4.85137	CTD-2105E13.13	-3.11287	RP11-727A23.11	-1.99819
RP11-58E21.5	-4.84268	RP11-94L15.2	-3.10704	RP11-16E12.2	-1.94671
RP11-309L24.4	-4.78704	RP11-23P13.7	-3.07923	CASC7	-1.86443
AC104024.1	-4.65636	RP11-420G6.4	-3.01464	RP11-391M1.4	-1.84355
CTD-2540L5.6	-4.57979	RP1-257A7.5	-2.92551	LINC00304	-1.70882
RP11-253I19.3	-4.5339	7SK	-2.90883	CTC-297N7.8	-1.66828
RP11-693J15.5	-4.53358	LINC00494	-2.8041	LINC00663	-1.63642
RP11-227G15.8	-4.5324	RP11-731F5.1	-2.78373	AC018464.3	-1.5734
RP11-67L2.2	-4.51188	RP11-817O13.6	-2.77691	CTC-444N24.11	-1.5555
RP11-13P5.1	-4.50307	AC007880.1	-2.73331	RP11-37B2.1	-1.5532
RP11-231E19.1	-4.46962	SNX29P2	-2.70062	RP11-553L6.5	-1.51511
RP11-424G14.1	-4.41193	CTA-250D10.23	-2.67063	LINC00174	-1.33374
RP11-307C12.13	-4.35791	RP11-4F5.2	-2.64443	MIR24-2	-1.31403
AC007879.2	-4.34379	RP11-562A8.5	-2.64233	RP11-284N8.3	-1.31296
PWRN1	-4.28807	AC137934.1	-2.60019	RP11-480A16.1	-1.21416
CYP2A6	-4.15048	CTD-2270L9.4	-2.56409	RP11-206L10.9	-0.930285
RP11-258C19.7	-4.10157	RP5-1085F17.3	-2.50301	LINC00657	-0.827787

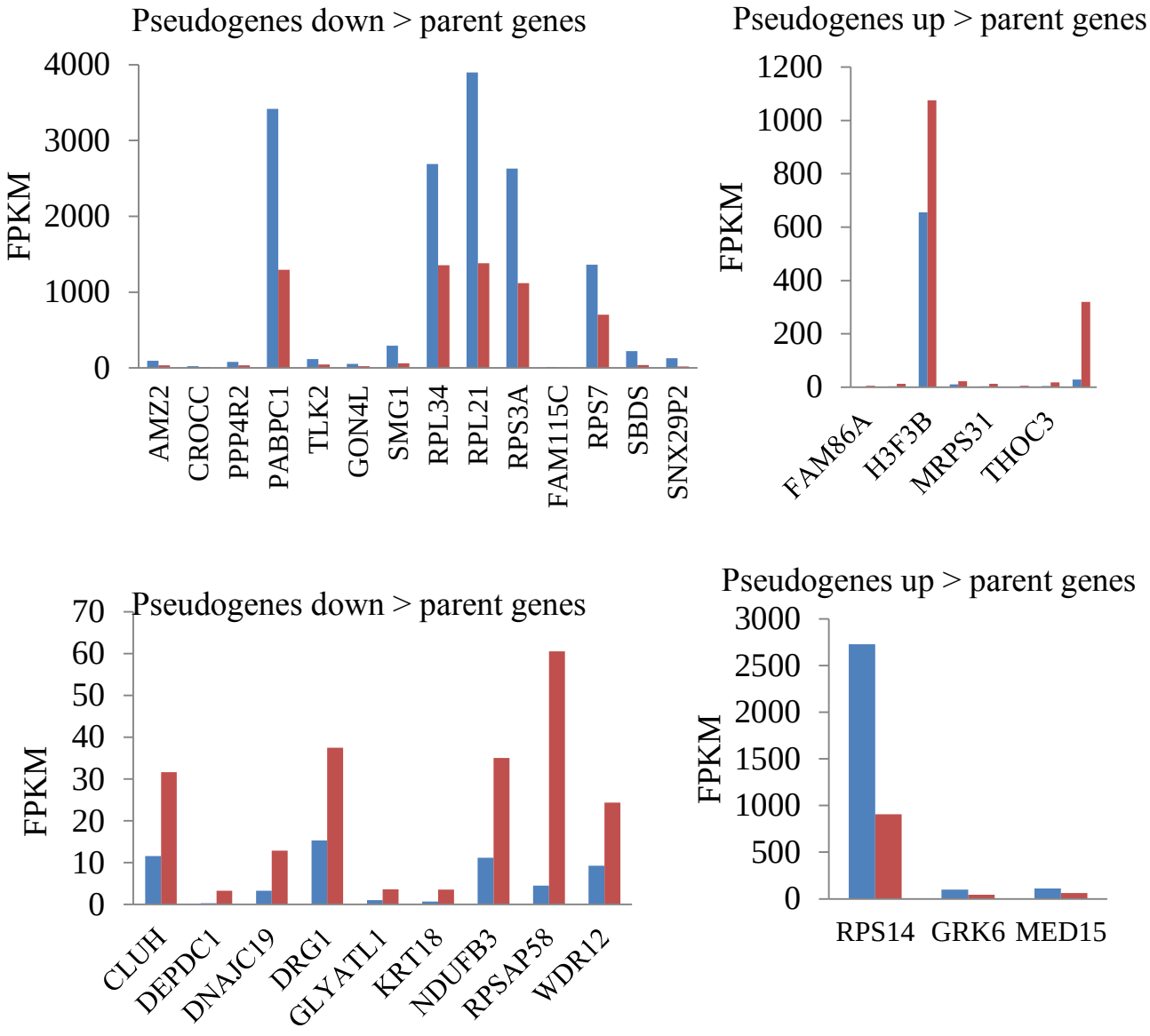


Figure 5.8: The differential expression of parent gene associated with the mis-regulation of pseudogenes in ALL. Blue bar= Expression (FPKM) value in HCB; red bar= expression value in ALL.

CH5: REFERENCES

Alachkar H, Mutonga MB, Metzeler KH, Fulton N, Malnassy G, Herold T, et. al. Preclinical efficacy of maternal embryonic leucine-zipper kinase (MELK) inhibition in acute myeloid leukemia. Oncotarget 2014; 5:12371-12382.

Almamun M, Levinson BT, Gater ST, Schnabel RD, Arthur GL, Davis JW, et al. Genome-wide DNA methylation analysis in precursor B-cells. Epigenetics 2014; 9:1588-1595.

Almamun M, Schnabel JL, Gater ST, Ning J, Taylor KH. Isolation of precursor B-cell subsets from umbilical cord blood. J. Vis. Exp 2013; 74:1-9.

Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013; 14:R21.

Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010; 19:698-711.

Busche S, Ge B, Vidal R, Spinella JF, Saillour V, Richer C et al. Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia. Cancer Res 2013; 73:4323-4336.

Chatterton Z, Morenos L, Saffrey R, Craig J, Ashley D, Wong N. DNA methylation and miRNA expression profiling in childhood B-cell acute lymphoblastic leukemia. Epigenomics 2010; 2:697-708.

Chim CS, Wan TS, Wong KY, Fung TK, Drexler HG, Wong KF. Methylation of miR-34a, miR-34b/c, miR-124-1 and miR-203 in Ph-negative myeloproliferative neoplasms. J Transl Med 2011; 9:197.

Dorsam RT, Gutkind JS. G-protein-coupled receptors and cancer. Nat Rev Cancer 2007; 7:79-94.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 2008; 26:779-85.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983; 301:89-92.

Figueroa ME, Chen SC, Andersson AK, Phillips LA, Li Y, Sotzen J, et al. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. J Clin Invest 2013; 123:3099-3111.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010 ; 464:1071-1076.

Han L, Yuan Y, Zheng S, Yang Y, Li J, Edgerton ME, et al. The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. Nat Commun. 2014; 5:3963.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010; 38:576-589.

Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell. 2011; 44:17-28.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc 2009; 4:44-57.

Hystad ME, Myklebust JH, Bø TH, Sivertsen EA, Rian E, Forfang L, et al. Characterization of early stages of human B cell development by gene expression profiling. J Immunol 2007; 179:3662-3671.

Ji H, Ehrlich L, Seita J, Murakami P, Doi A, Lindau P, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 2010; 467:338-342.

Kwon JH, Kim J, Park JY, Hong SM, Park CW, Hong SJ, et al. Overexpression of high-mobility group box 2 is associated with tumor aggressiveness and prognosis of hepatocellular carcinoma. Clin Cancer Res 2010; 16:5511-5521.

Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, et al. Landscape of somatic retrotransposition in human cancers. Science 2012; 337:967-971.

Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK et al. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. Nucleic Acids Res 2012; 40:11339-11351.

Liang P, Song F, Ghosh S, Morien E, Qin M, Mahmood S, et al. Genome-wide survey reveals dynamic widespread tissue-specific changes in DNA methylation during development. BMC Genomics 2011; 12:231.

Lichtenberger BM, Tan PK, Niederleithner H, Ferrara N, Petzelbauer P, Sibilia M. Autocrine VEGF signaling synergizes with EGFR in tumor cells to promote epithelial cancer development. Cell 2010; 140:268-279.

Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 2004; 11:1068-1075.

Maksakova IA, Mager DL, Reiss D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cell Mol Life Sci. 2008; 65:3329-3347.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 2010; 466:253-257.

Maunakea AK, Chepelev I, Cui K, Zhao K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 2013; 23:1256-1269.

Nagano T, Fraser P. No-nonsense functions for long noncoding RNAs. Cell 2011; 145:178-181.

Nagai MA, Gerhard R, Salaorni S, Fregnani JH, Nonogaki S, Netto MM, et al. Down-regulation of the candidate tumor suppressor gene PAR-4 is associated with poor prognosis in breast cancer. Int J Oncol 2010; 37:41-49.

National cancer institute report. (http://www.cancer.gov/cancertopics/pdq/treatment/childALL/HealthProfessional/page1#Reference1.3).

Okudela K, Mitsui H, Suzuki T, Woo T, Tateishi Y, Umeda S, et al. Expression of HDAC9 in lung cancer--potential role in lung carcinogenesis. Int J Clin Exp Pathol 2014; 7:213-220.

Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 2013; 14:341-356.

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010; 465:1033-3108.

Sharma A, Ray R, Rajeswari MR. Overexpression of high mobility group (HMG) B1 and B2 proteins directly correlates with the progression of squamous cell carcinoma in skin. Cancer Invest 2008; 26:843-851.

Takashima N, Ishiguro H, Kuwabara Y, Kimura M, Haruki N, Ando T, et al. Expression and prognostic roles of PABPC1 in esophageal cancer: correlation with tumor progression and postoperative survival. Oncol Rep 2006; 15:667-671.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 2013;31:46-53.

Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell 2014; 158:593-606.
Valadez JA, Cuajungco MP. PAX5 is the transcriptional activator of mucolipin-2 (MCOLN2) gene. Gene 2015; 555:194-202.

Van Limbergen EJ, Zabrocki P, Porcu M, Hauben E, Cools J, Nuyts S. FLT1 kinase is a mediator of radioresistance and survival in head and neck squamous cell carcinoma. Acta Oncol 2014; 53:637-645.

van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ et al. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. J Immunol 2005; 175:5912-5922.

Vlaicu SI, Cudrici C, Ito T, Fosbrink M, Tegla CA, Rus V, et al. Role of response gene to complement 32 in diseases. Arch Immunol Ther Exp (Warsz) 2008; 56:115-122.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013; 339:1546-1558.

Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011 Sep 16; 43:904-914.

Wang LQ, Kwong YL, Kho CSB, Wong KF, Wong KY, Ferracin M, et al. Epigenetic inactivation of miR-9 family microRNAs in chronic lymphocytic leukemia - implications on constitutive activation of NFκB pathway. Mol Cancer 2013;12:173

Weng YR, Cui Y, Fang JY. Biological functions of cytokeratin 18 in cancer. Mol Cancer Res 2012; 10:485-493.

Wood AJ, Schulz R, Woodfine K, Koltowska K, Beechey CV, Peters J, et al. Regulation of alternative polyadenylation by genomic imprinting. Genes Dev 2008; 22:1141-1146.

Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X. Pseudogene in cancer: real functions and promising signature. J Med Genet 2015; 52:17-24.

Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, et al. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. Proc Natl Acad Sci U S A. 2007; 104:12377-12382.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008; 9:R137.

CHAPTER 6: TRANSCRIPTIONAL INACTIVATION OF REGULATORY ENHANCERS ASSOCIATED WITH THE MIS-REGULATION OF NEIGHBORING GENES.

6.1 ABSTRACT

Enhancers are capable of regulating gene expression in a distance and orientation independent manner, and considered as key regulators of tissue-specific gene expression. Recent evidence shows that enhancers are transcribed to generate long non-coding RNA (eRNA) and contribute to enhancer function. Altered DNA methylation within the enhancer regions may contribute to atypical eRNA expression and ultimately lead to mis-regulation of their nearby protein-coding genes. Based on enhancer specific histone marks (H3K4me1 and H3K27ac), this study identified active enhancers in the GM12878 lymphoblastoid cell line, a surrogate of normal B-lymphocytes. We find that a large number of active enhancers become hypermethylated (FDR $\leq$ 5% and fold change $\geq$ 2) in ALL and are associated with the expression of nearby potential target genes. Additionally, many enhancers' transcripts were significantly down-regulated in ALL compared to healthy pre-B cells. This study also investigates the association between enhancer methylation and differential eRNA expression (q-value $\leq$0.05) in ALL. A group of enhancers were transcriptionally inactivated due to gaining methylation in ALL. Importantly, enhancer hypermethylation and down-regulation of eRNAs were associated with the significant down-regulation of nearby *GRB2, ISG20, ELOVL5, IRF4, IFNGR2, ADRA2B, ASTL, MS4A1, GCM1* genes in ALL. Conversely, enhancer hypomethylation and up-regulation of eRNAs were associated with the up-regulation of nearby *COL5A1, RXRA, ALG12* genes. Collectively, we

have demonstrated the DNA methylation mediated transcriptional inactivation of regulatory enhancers in ALL, which leads to the mis-regulation of their target genes.

6.2 INTRODUCTION

Enhancers are regulatory sequences that can be located far away from the gene they regulate, and can control target gene expression through physical interactions with gene promoters. Enhancers are the key element of cell type specificity in gene expression. Altered enhancer activity may contribute to the mis-regulation of target genes and ultimately lead to the pathogenesis of many diseases. Indeed, genome-wide association studies (GWAS) have discovered a large number of disease susceptibility regions that do not overlap with protein–coding genes but rather map to non-coding intervals (Visel et al. 2009). Furthermore, mutations, insertions and deletions in enhancer regions has been shown to change the expression of respective target genes by altering the recruitment of transcription factors and other cofactors in many disease conditions (Herz et al. 2014). DNA methylation regulates tissue specific gene expression and plays a significant role in hematopoiesis. Aberrant DNA methylation in vital regulatory regions of the genome, like enhancers, may cause initiation of many hematopoietic disease conditions including precursor-B cell acute lymphoblastic leukemia (ALL). Altered DNA methylation in enhancers has been shown to affect the binding site of FOXA1 and other pioneer factors in breast and prostate cancer cell line which leads to significant changes in tissue specific gene expression (Sérandour et al. 2011). In our previous studies (chapter #5), we have identified a large number of differentially methylated regions located within the intergenic and intronic regions in ALL. These differentially methylated regions could coincide with enhancer

regions. Further, RNA-seq data revealed that a vast number of genes are differentially expressed in ALL while their promoter regions were not differentially methylated. These results highlight the importance of studying the relationship between non-promoter methylation and gene expression in ALL.

About 50% of gene promoters exhibit one or more long-range interaction, and sometimes they may interact with as many as 20 distal elements. Transcriptionally active genes interact with more distal elements than do silenced genes (Sanyal et al. 2012). Similarly, the same distal element can also interact with more than one transcriptional start site (Sanyal et al. 2012). In addition, to facilitate the transcription of cognate target genes, enhancers can act as alternative TSS and produce both short non-polyadenylated transcripts and long polyadenylated multi-exonic transcripts that imitate the structure of the host gene with the exception of an alternative first exon (Natoli et al. 2012). Recently, several studies have demonstrated that enhancers are transcribed to produce long non-coding RNA (Lam et al. 2014; Lai et al. 2014; Melgar et al. 2011; Wang et al. 2011), and this enhancer-derived RNA (eRNA) plays a crucial role in regulating gene expression, normal development and pathogenesis of many disease conditions. Activation of enhancers initiate eRNA transcription which then recruits other looping factors such as the cohesin complex which assists/stabilizes chromosomal looping between the enhancer and the promoter of their target gene (Li et al. 2013). eRNA also facilitates the loading of RNA Pol II, and transcription initiation complex to the promoter of the target gene (Lam et al. 2014). Further experiments have demonstrated that eRNA expression is strongly correlated with the expression of nearby genes (Kim et al. 2010; De Santa et al. 2010), emphasizing a functional consequence of eRNA mediated transcription. The transcriptional augmentation of human growth hormone (hGH) is dependent on the intensity of enhancer transcription but not on the

structure of its ncRNA (Yoo et al. 2012). Furthermore, ncRNAs are related to enhancer function in Kcnq1 imprinted domain (Korostowski et al. 2011). The biallelic expression of imprinted Kcnq1 during mid-gestation is regulated by tissue and stage specific chromatin loop formation between regulatory elements and promoter of Kcnq1 and this biallelic expression is regulated by the non-coding Kcnq1ot1 transcript. eRNA produced from p53-bound enhancer region (p53BERs) is required for the efficient transcription enhancement of the tumor suppressor p53 gene and the depletion of eRNA had a significant inhibitory consequence on p53-induced cell cycle arrest (Melo et al. 2013; Lai et al. 2014). Another study has shown that nuclear receptor Rev-Erb-α regulated gene expression by inhibiting the expression of eRNAs in mouse macrophages, and targeted degradation of eRNAs reduces the mRNA expression level of nearby genes (Lam et al. 2013). These studies demonstrated the functionality of eRNA during development, and provided evidence that targeting eRNA to diminish the expression of protein coding mRNA could be a practical approach to control tissue-specific gene expression.

This study has identified potential active regulatory enhancer regions in normal lymphoblastoid cell line (GM12878). The majority of active enhancers were occupied by transcription factors and a large number of them gained methylation in ALL compared to healthy pre-B cells. In addition, we have identified differential enhancer transcripts (eRNA) in ALL. The expression level of eRNAs were associated with the methylation status of enhancers. Comparable to long non-coding RNA, eRNAs are postulated to be tissue specific and therefore targeting of eRNAs provides a promising therapeutic potentiality. Further investigation of enhancer transcription and the functional roles of eRNAs will upsurge our understanding of complex mechanisms of gene expression and may also progress the field and provide novel therapeutic interventions for human diseases including cancers.

6.3 MATERIALS AND METHODS

6.3.1 ChIP-seq data

Chromatin signatures can be used to identify the enhancer like region. Active enhancers are greatly enriched for H3K4me1 and H3K27ac marks. H3K4me1 represents a generalized mark for enhancer region while H3K27ac is linked with the activation of tissue specific enhancer. Due to lack of both histone modifications data for precursor B-cells, ChIP-seq H3K4me1 and H3K27ac data of GM12878 cell line (B-Lymphocytes cell type) were obtained from publically available ENCODE Data Coordination Center at UCSC (https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeBroadHistone). Lymphoblastoid cells line, GM12878 was established by Epstein - Barr virus transformation of peripheral blood mononuclear cells using phytohemagluttinin as a mitogen (Coriell Cell Repositories; Coriell Institute for Medical Research). The overlapping regions between two histone marks were determined by intersect-BEDTools (Quinlan and Hall 2010). Regions that are co-enriched for H3K4me1 and H3K27ac were considered as potential active enhancer regions in normal B-cell. Enhancer regions were annotated with HOMER (Hypergeometric Optimization of Motif EnRichment), version 4.3, using default parameters to identify their genomic location (Heinz et. al. 2010).

6.3.2 RNA-seq data

RNA-seq data were generated from 20 ALL and 8 healthy pre-B as previously described in chapter# 5. Briefly, RNA-seq libraries were prepared using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) and then sequenced on the Illumina HiSeq 2000 (1x100 bp reads) at the University of Missouri DNA Core Facility. FastX toolkit was used to remove poor quality reads of <20. After quality check, reads were aligned to

reference human genome (hg19) using Tophat (v2.0.13) with default settings. Differential gene expressions were determined using Cufflinks with default parameters (version 2.2.1). The read counts along with FPKM values and their variances were calculated by cuffdiff 2 and the log fold change and p-value was calculated for each gene and isoforms. Multiple testing corrections was also performed (q-value) using Benjamini-Hochberg.

6.3.3 Differentially methylated regions between ALL and healthy pre-B cell

We used the differentially methylated regions (DMRs) identified in chapter #5. DMRs were identified from MIRA-seq data of 20 ALL and 10 healthy pre-B cell samples. The genomic regions with <5% false discovery rate (FDR; Benjamini-Hochberg) and at least 2 fold change were considered as DMRs.

6.3 RESULTS

6.3.1 Enhancer identification in normal B lymphocytes (GM12878)

Based on co-enrichment of histone marks (H3K4me1 and H3K27ac), a total of 16,212 intergenic and 44,370 intronic regions were identified as potential active enhancer regions in B lymphocytes (**Figure 6.1A**). The GREAT tool (http://bejerano.stanford.edu/great/public/html/; McLean et al. 2010) was used to identify the TSS of two nearest genes within 1Mb in each direction. We found 15,083 intergenic enhancers have two genes while 1,064 have one gene and 65 enhancers do not have any potential target genes within 1Mb distance in each direction. More than 50% of possible target genes' TSS are located within 50-500kb and about 36% located inside 5-50kb whereas only ~14% are located within either 0-5kb or >500 kb from their regulatory enhancers (**Figure 6.2**). Gene Ontology (GO) analysis revealed that intergenic and

intronic enhancers' target genes are associated with immune response, B cell activation, leukocyte differentiation and other relevant biological processes providing further evidence that the identified regions have the capacity to function as regulatory enhancers in normal lymphocytes (**Figure 6.3**). Enhancers contain binding sites for many transcription factors (TF) that function during development and differentiation (Serandour et al. 2011; Natoli and Andrau 2012). To elucidate the occupancy of TF within the identified enhancers, ChIP-Seq data from the GM12878 cell line for 77 transcription factor targeting antibodies from the ENCODE project was used. The majority of identified intergenic (14,497) and intronic (35,260) enhancers were occupied with TFs that play roles in normal development including B-cell development and differentiation (**Figure 6.4**).

The poised enhancer is marked by H3K4me1 alone. Overall, 56,748 poised enhancer regions were identified where 27,116 enhancers were located within intergenic regions and 30,639 enhancers were located within intronic regions of the genome **(Figure 6.1B**). GO analysis revealed that intergenic enhancers' target genes were associated with leukocyte activation, lymphocyte differentiation, and negative regulation of MAP kinase activity emphasizing the importance of studying poised enhancer regions **(Figure 6.5**). A total of 1726 potential target genes of cognate poised intergenic enhancers were up-regulated and 1123 potential targets genes were down-regulated in ALL when compared to healthy pre-B cell.

6.3.2 Alteration of enhancer methylation in ALL

Distal regulatory elements including enhancers are associated with lower levels of DNA methylation (Stadler et al. 2011). Gaining DNA methylation within the enhancer regions may prevent the binding of transcriptional machinery and ultimately lead to transcription inactivation (Serandour et al. 2011). To investigate whether the active enhancers coincide with

hypermethylated regions in ALL, the enhancer regions were overlaid with the previously identified differentially methylated regions (DMRs) in ALL (chapter # 5) using a custom python script (Python 3.3.0). Overall 245 intergenic and 545 intronic enhancer regions became hypermethylated in ALL. The target genes (2 nearest genes within 1Mb in each direction) of hypermethylated enhancers were found to be associated with several biological processes including lymphocyte activation and differentiation. Integration of gene expression data revealed that enhancer hypermethylation in ALL was associated with the down-regulation of 204 potential target genes. Transcriptional silencing of the target genes by enhancer hypermethylation may provide survival and growth advantages for leukemia cells in ALL. For example, the down-regulated genes involved in B cell proliferation and activation (*IL7, IL13, PAWR, PLCG2, MS4A1, HDAC9, ICOSLG, FOXP1*), BCR signaling pathways (*LYN, GRB2, RELA, PLCG2, NFATC1*), transcriptional regulation, Jak-STAT signaling pathway (*IL7, GRB2, IL4R, IL21R, IL13, IFNGR2*) and tumor suppression (*SIK1, ERF, EXT1, TES, FOXP1, PAWR, PLCE1, BANP, PTPN2, SIAH1*) were associated with the hypermethylation of their regulatory enhancer regions, emphasizing the significance of enhancer methylation in ALL.

The loss of DNA methylation may allow poised enhancers to become activated in ALL. Hence, we have determined the differential methylation status of poised enhancer in healthy pre-B cells. Overall, 424 intergenic poised enhancers lost methylation in ALL compare to healthy pre-B cell. We further studied these 424 hypomethylated enhancers in ALL those were poised in normal lymphoblastoid cell line but could become active in ALL. A total of 568 target genes were associated with hypomethylated enhancer in ALL. The results also showed that enhancer hypomethylation was associated with significant up-regulation of 121 target genes (**Table 6.1**). These aberrantly up-regulated genes were involved in several biological processes including

lymphocyte activation, cell cycle processes, protein complex assembly, chromosomal organization, negative regulation of apoptosis, GTPase activation, and cytoskeleton. Furthermore, hypomethylated poised enhancers were associated with the down-regulation of 71 target genes involved in multiple functional groups including B cell homeostasis and protein kinase activity (**Table 6.2**).

6.3.3 Enhancer transcript (eRNA) associated with the expression of neighboring genes

Active enhancers are transcribed to generate eRNA and play roles in regulating gene expression during normal development (Natoli and Andrau 2012; Lam et al, 2014). Pol II associated eRNAs are either non-polyadenylated or polyadenylated transcripts and the length varies from less a kilo-base pair to several kilo-base pairs. From RNA-seq data, a considerable number of unannotated transcripts were identified within the intergenic regions in both healthy pre-B and ALL samples. Therefore, intergenic enhancers and intergenic transcripts with a length > 200 bp were comprehensively analyzed in ALL and healthy pre-B cells. All transcripts associated with isoforms of protein coding genes, pseudogenes, or lincRNA were excluded from the analysis. A total of 1,903 intergenic enhancers were transcribed in healthy pre-B cells and ALL samples where 319 eRNA were significantly differentially expressed in ALL (q-value <0.05; **Figure 6.6**). Enhancers were active in the normal cell line; therefore, more transcripts were expected in normal B-cells. Indeed, a majority of differentially expressed eRNA had significantly higher expression in healthy pre-B cells compared to ALL samples (284- vs 35 eRNAs). This result suggests that a large number of regulatory enhancers become transcriptional inactivated during ALL pathogenesis. The two nearest potential target genes of differentially expressed enhancers were identified within 1 Mb in each direction. As sometimes multiple enhancer regions are located in between the same target neighboring genes, a total of 230 genes

target genes were identified for the down-regulated enhancers and 32 target genes were identified for the up-regulated enhancers in ALL. GO analysis revealed that potential target genes associated with differentially expressed enhancers were associated with the immune system and leukocyte differentiation and activation (**Figure 6.7**). Disease Ontology analysis showed that several of the target genes were also involved in different hematological malignancies (**Figure 6.8**).

To determine the effect of enhancer transcripts on neighboring target gene expression, RNA-seq data was used to compare the expression level of target genes in ALL and HCB. Significant down –regulation of 89 target genes were associated with significantly down-regulated regulatory enhancers in ALL (**Table 6.3**). On the other hand, up-regulation of seven targets genes was associated with the significantly up-regulated regulatory enhancers in ALL (**Figure 6.9**). Several target genes are associated with normal B cell differentiation and activation; therefore enhancer mediated mis-regulation may contribute to the pathogenesis of leukemia. For example, *ICOSLG* induces B-cell proliferation and differentiation, and *SWAP70* is required for normal B-cell migration into secondary lymphoid organs (Pearce et al., 2006). Both genes were down-regulated in ALL, and were associated with transcriptional inactivation of regulatory enhancer. Furthermore, *IRF4, IRF8* genes participate in B-cell development including pre-B cell differentiation (Shukla et al. 2014), and both genes were down-regulated in ALL in association with decreased level of eRNA. Additionally, *IRF8* regulates the expression of apoptosis-related genes and acts as a tumor suppressor gene in chronic myeloid leukemia (CML) (Gabriele et al. 1999). The germline mutation of *IRF8* results in CML-like disease syndromes in mice (Holtschke et al. 1996, Tamura et al. 2003). Therefore, the significant down-regulation of *IRF8* in ALL patients could be negatively regulating the apoptotic genes, leading to persistence

and proliferation of lymphoid cells in ALL. The epigenetic modifier *JARID2* regulates histone methyltransferase, and its down-regulation in ALL was associated with the transcriptional inactivation of enhancer region. Decreased expression of *JARID2* is also found in acute monocytic leukemia (AMOL) and B-chronic lymphocytic leukemia (B-CLL), and is responsible for stimulating the proliferation of leukemia cells through acceleration of the G1/S transition (Su et al. 2015). Furthermore, the down-regulation of proapoptotic *HES1* (Hairy/Enhancer of Split1) is mediated by decreased eRNA expression in ALL. On the other hand, up-regulation of several important genes were associated with the up-regulation of regulatory eRNAs in ALL. The up-regulation of serine/threonine kinase *NEK2* in ALL was associated with the significantly higher expression of eRNA. The higher expression of *NEK2* obstructs B-cell development by increasing the amount of immature B cells in the bone marrow (Gu et al. 2014). Taken together, a group of genes have been identified that are mis-regulated by differential expression of regulatory enhancers in ALL. These results highlight the noteworthy role of enhancer transcripts in the regulation of target gene expression in normal development and in ALL pathogenesis.

As DNA methylation is associated with transcriptional repression, the effect of differential DNA methylation on the differential expression of eRNA was determined in ALL. A total of 77 enhancers with eRNA transcription were differentially methylated (FDR < 5%; Fold change ≥ 2) in ALL, and negatively correlated with the differential expression of eRNA (R^2= -0.50). Hypermethylation of enhancer regions was associated with the significant down-regulation (q-value ≤ 0.05) of thirteen eRNAs; whereas hypomethylation was associated with the up-regulation of five eRNAs in ALL (**Figure 6.10**).

Differential enhancer methylation and eRNA expression affected the expression of nearby target genes (**Figure 6.11**). Interestingly, enhancers' hypermethylation and down-

regulation of eRNAs were associated with the significant down-regulation of nearby genes *GRB2, ISG20, ELOVL5, IRF4, IFNGR2, ADRA2B, ASTL, MS4A1, GCM1* in ALL. On the other hand, enhancers' hypomethylation and up-regulation of eRNAs were associated with the up-regulation of *COL5A1, RXRA*, and *ALG12*. The adapter protein, growth factor receptor-bound 2 (GRB2) recruits Bruton's tyrosine kinase (Btk) and plays important roles in B-cell proliferation (Engels et al. 2009; Ackermann et al. 2011; Seda and Mraz 2015). Contrary to solid tumors (Li et al. 2014; Gui et al. 2012), *GRB2* is down-regulated in ALL, and is associated with enhancer hypermethylation and decreased eRNA expression. Additionally, *ELOVL5* gene encodes a plasma membrane protein and negatively regulates the expression of genes involved in gluconeogenesis by phosphorylation of *FOXO1* (Tripathy and Jump 2013) and was found to be down-regulated in ALL. The membrane protein *ELOVL5* has reduced expression in prostate cancer (Romanuik et al. 2009) but the functional role of reduced *ELOVL5* expression in prostate cancer remains unknown. Additionally, we found that the decreased expression of *IRF4*was associated with enhancer hypermethylation and significant down-regulation of eRNA in ALL. The transcription factor IRF4 along with IRF8 is essential for differentiation of pre B-cells to immature B cells. B-cell differentiation in mice lacking IRF4 was blocked at the pre-B cell stage and the mice exhibited a hyper-proliferative phenotype (Shukla and Lu 2014). Adrenoceptor Alpha 2B (*ADRA2B)* is a G protein-coupled receptor and its expression level was associated with the mis-regulation of enhancer methylation. Furthermore, Interferon Gamma Receptor 2 (*IFNGR2*) is required to initiate the IFNγ signal transduction and IFNγ is essential for the initiation of multiple cellular responses, including inhibition of cell growth and variation of cell differentiation. IFNγ is also considered as antitumor molecule (Wang et al. 2014). The present study elucidated the decreased expression of *IFNGR2* in ALL associated with mis-regulation of

an enhancer region. Conversely, the up-regulation of *COL5A1*, a type V collagen, is associated with the enhancer hypomethylation and significantly increased expression of eRNA in ALL. The continuous increase expression of *COL5A1* is also found during ovarian cancer progression in human, and the depletion of *COL5A1* in mice reduces cell migration, invasion and progression of tumor (Cheon et al. 2014). Collectively, this study identified DNA methylation mediated mis-regulation of eRNA transcription in ALL, and this differential eRNAs expression were associated with the differential expression of potential target genes.

6.4 DISCUSSION

The majority of the non-coding part of the human genome is occupied by many different functional elements, and a higher number of genomic regions contain enhancer-like features than contain promoter-like features (ENCODE Project Consortium, 2012). However, the detection of enhancers has always been a daunting task. With the advent of whole genome sequencing data from many diverse species, efforts were made to recognize regulatory enhancers based on sequence conservation. Nevertheless, it has become clear that only a certain percentage of enhancers are conserved at the sequence level and not all non-coding conserved sequence functions as enhancers (Royo et al. 2011). Genome-wide ChIP-seq experiments found that enhancers are enriched for precise chromatin modifications, particularly high levels of H3K4me1 and low level of H3K4me3 (Heintzman et al. 2007). Next, it was found that H3K27ac specifically marks active enhancers (Creyghton et al. 2010). The co-enrichment of histone marks, H3K4me1 and H3K27ac is considered to be the best predictor of active enhancers where H3K4me1 is a poised enhancer mark, and H3K27ac is associated with tissue specific enhancer activation (Bogdanovic et al. 2012; Bulger et al. 2011). This study identified active and poised

distal regulatory enhancer regions in the lymphoblastoid GM12878 cell line based on histone marks, which could mimic the regulatory enhancer regions in healthy B-cell. Distinctive sets of transcription factors bind to promoters and at distal regulatory enhancers and this binding is required for the establishment/maintenance of long-range promoter–enhancer interactions (Lan et al. 2012). About 90% of identified intergenic enhancers and ~ 80% of intronic enhancers were occupied by multiple transcription factors highlighting the functional importance of identified enhancers**.** A large number of identified enhancer regions occupied transcription factor, EP300 (histone acetyltransferase), that is responsible for the acetylation of histone tail. We also observed an enrichment of bound POLR2A within the identified enhancers. POLR2A is a subunit of RNA polymerase II, which synthesizes mRNA precursors and numerous functional non-coding RNAs, and transcriptional repression of this gene arrested cellular proliferation (Yamada et al., 2013). The binding of POLR2A within the enhancer regions provided evidence that the identified enhancers were transcriptionally active. Transcription factor PML functions as a tumor suppressor by regulating p53 response to oncogenic signals, and PML was found to be bound with a large number of enhancers. Other transcription factors including PAX5, IRF4, IKZF1, EBF1, BCL3 and EZH2, known to play important roles in normal B-cell development, were bound to multiple enhancers. Since the active enhancers were predicted in the lymphoblastoid cell line, the functional groups of cognate target genes were associated with lymphocyte activation and differentiation. This result demonstrated that tissue specific activation of regulatory enhancers occurred in adjacent to the tissue specific genes. Additionally, poised enhancers at specific stage of development can become activated during normal development, and atypical activation of regulatory enhancers could lead to many disease conditions including malignancies. This study also identified a group of poised enhancers, which has potential target

genes involve in MAP kinase activity and lymphocyte differentiation. Interestingly 1,726 potential target genes of cognate poised enhancers were up-regulated and 1,123 genes were down-regulated in ALL. The mis-regulation of stage specific gene expression mediated by abnormal enhancer activation may contributed to the development of disease.

Although distal regulatory elements typically display low levels of methylation, enhancers are often enriched with CpG dinucleotides. The methylation status is cell type specific and is associated with the permissiveness of the enhancers (Wiench et al. 2011). Enhancer methylation is also remarkably transformed in cancers, and genes that are associated with hypermethylated enhancers have lower gene expression while genes that are associated with hypomethylated enhancers are up-regulated in cancers (Aran et al. 2013). We demonstrated that a large number of active enhancers gained methylation during ALL transformation, and that the down-regulation of a subset of the target genes, including genes involved in B cell proliferation and activation (*IL7, IL13, PAWR, ICOSLG, FOXP1*), and BCR signaling pathway (*GRB2, LYN, RELA, NFATC1, PLCG2*) were associated with hypermethylated enhancers in ALL. In addition, a substantial number of poised enhancers lost methylation in ALL and were associated with the up-regulation of their target genes. Therefore, enhancer methylation mediated transcriptional mis-regulation of genes involved in normal B-cell development and differentiation could play a vital role for the selective growth advantage and survival of leukemic cell.

The recent discovery that enhancer regions are transcribed to RNA augments an additional layer to the complexity of gene regulation. Enhancer transcription is a well conserved phenomenon as evidenced by the presence of transcripts across multiple cell types in diverse species (Lam et al. 2014). We have shown that a large number of enhancers are transcribed and significantly differentially expressed in ALL. Since eRNA production is considered as the final

step of enhancer activation, the down-regulation of eRNA in ALL suggesting that a vast number of regulatory enhancers were transcriptionally inactive in ALL. eRNA is correlated with the expression of neighboring genes, and the targeted degradation of eRNA by RNA interference (siRNA) reduces the expression of nearby protein-coding genes (Lam et al. 2014). Remarkably, this study identified neighboring genes whose expression levels were associated with the differential expression of eRNA in ALL. Genes involved in several critical biological processes including normal B-cell migration (*SWAP70*), proliferation (*ICOSLG*, *IRF8*), and apoptosis (*HES1*) were mis-regulated in ALL concomitant with the differential expression of eRNA. Furthermore, we have identified the enhancer regions that were silenced by DNA methylation in ALL. Differential enhancer methylation was negatively correlated with the expression of eRNA, where hypermethylated enhancers were associated with the down-regulation of eRNA and hypomethylated enhancers were associated with the up-regulation of eRNA expression. Furthermore, DNA methylation mediated transcriptional inactivation of eRNAs were associated with the mis-regulation of nearby potential target genes *GRB2, ISG20, ELOVL5, IFNGR2, ADRA2B, ASTL, MS4A1, GCM1,* and *COL5A1* in ALL.

In summary, we have identified the potential regulatory enhancer regions in a normal lymphoblastoid cell line and the majority of the enhancers were occupied by transcription factors required for normal B-cell development. Additionally, the genes in proximity to the potential enhancers were involved in immune response and B-cell activation. We have shown a large number of enhancers were hypermethylated in ALL and were associated with the mis-regulation of neighboring genes' expression. A large number of enhancers were highly expressed in normal pre-B cell and then down-regulated during ALL pathogenesis. Furthermore, the down-regulation of eRNAs were associated with the methylation status of enhancer regions and affected the

expression of their nearby protein-coding genes. eRNAs are postulated to be tissue specific, hereafter targeting of eRNAs offers an encouraging therapeutic potentiality.

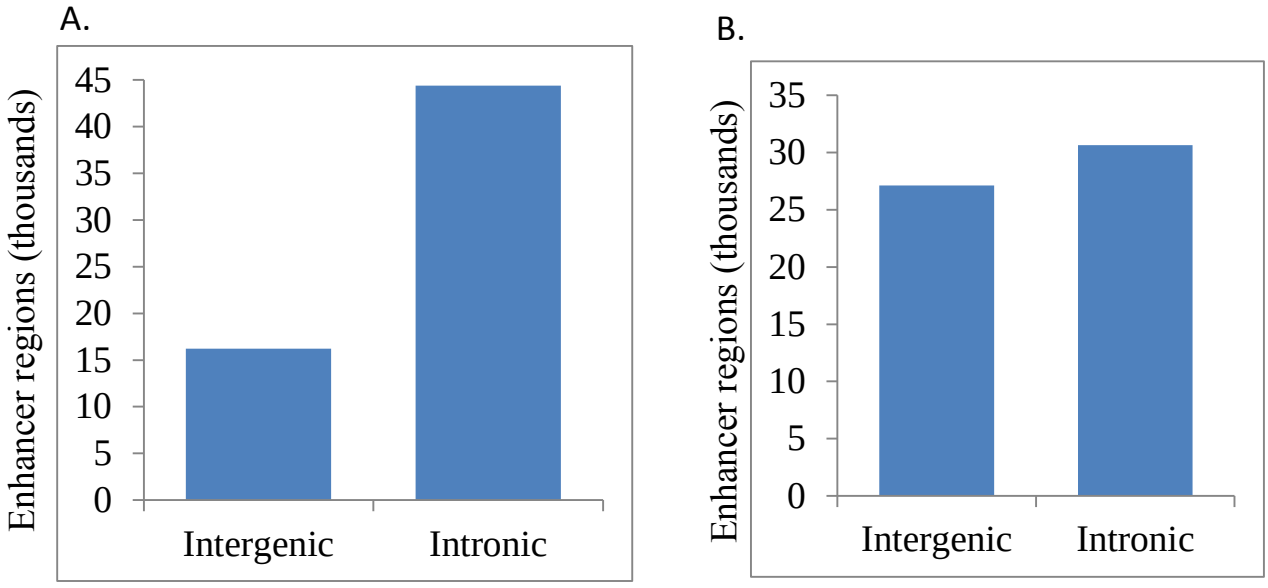


Figure 6.1: Potential intergenic and intronic enhancer regions in normal B-lymphocytes (GM12878 cell line). (A) Active enhancers. (B) Poised enhancers.

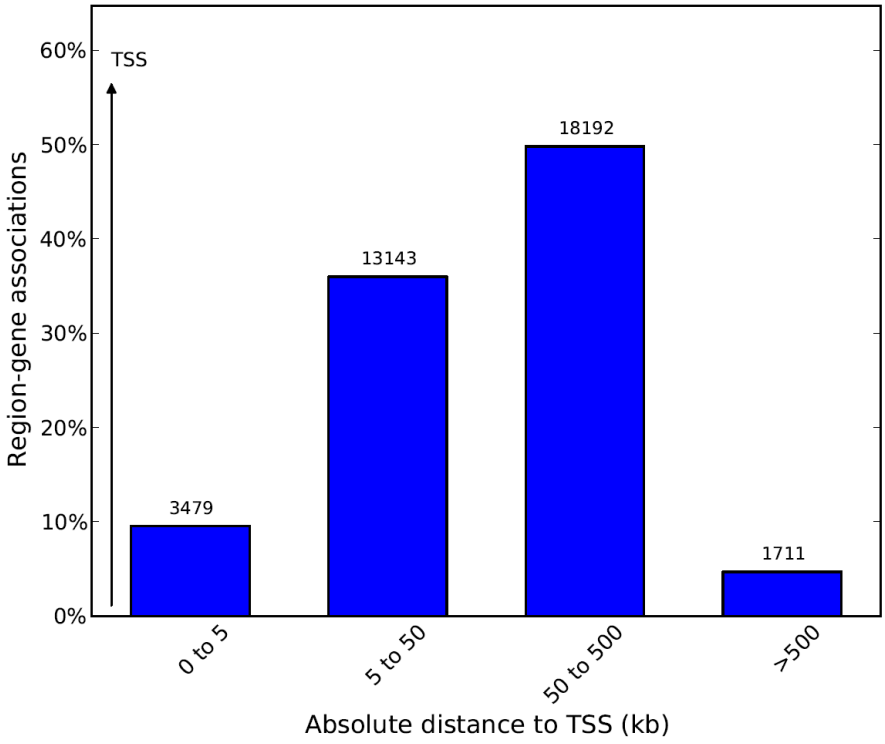


Figure 6.2: Absolute distance between intergenic enhancer regions and 2 nearest regulatory gene's TSS.

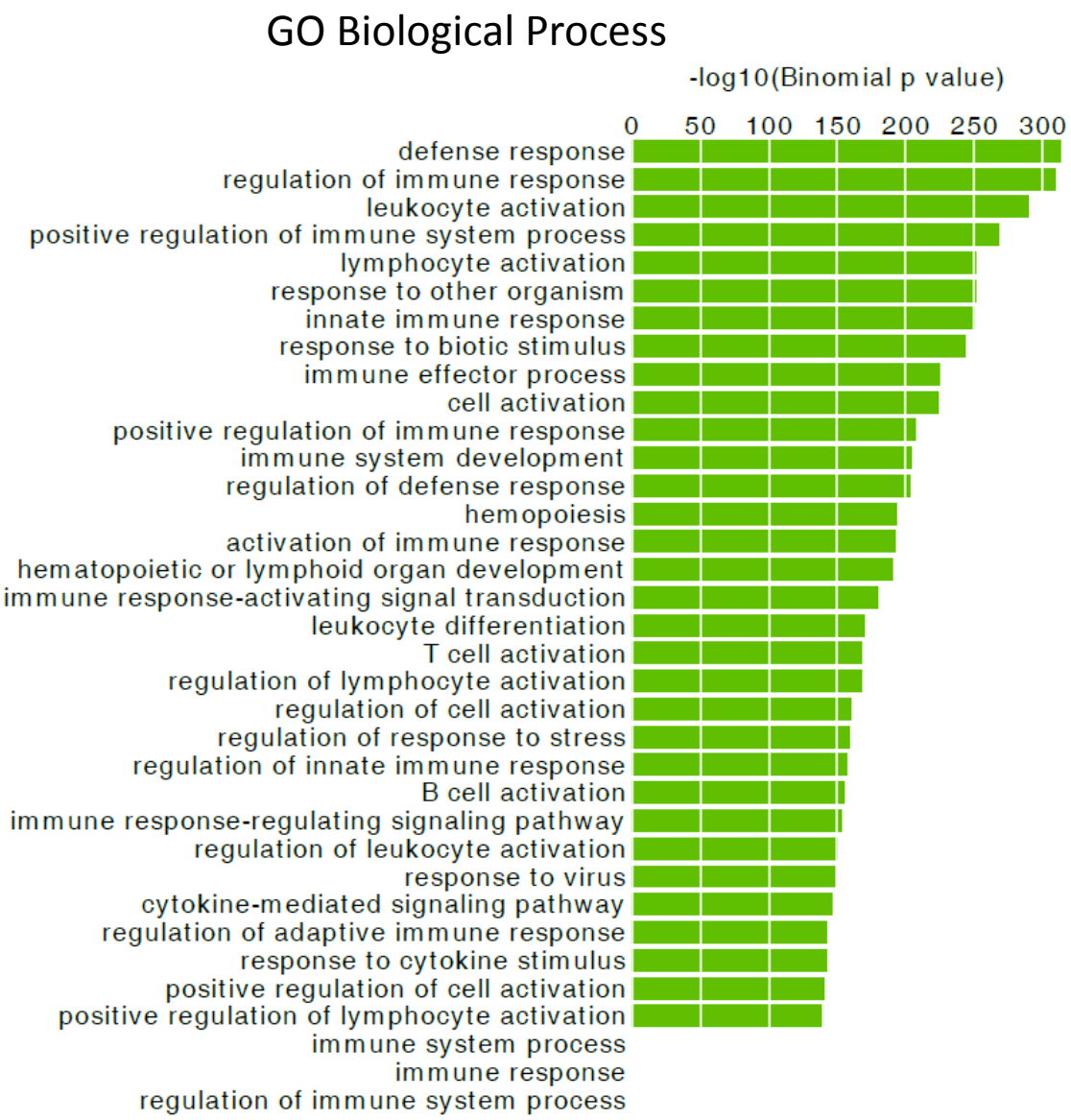


Figure 6.3: GO Biological process of possible target genes of intergenic enhancers**.**

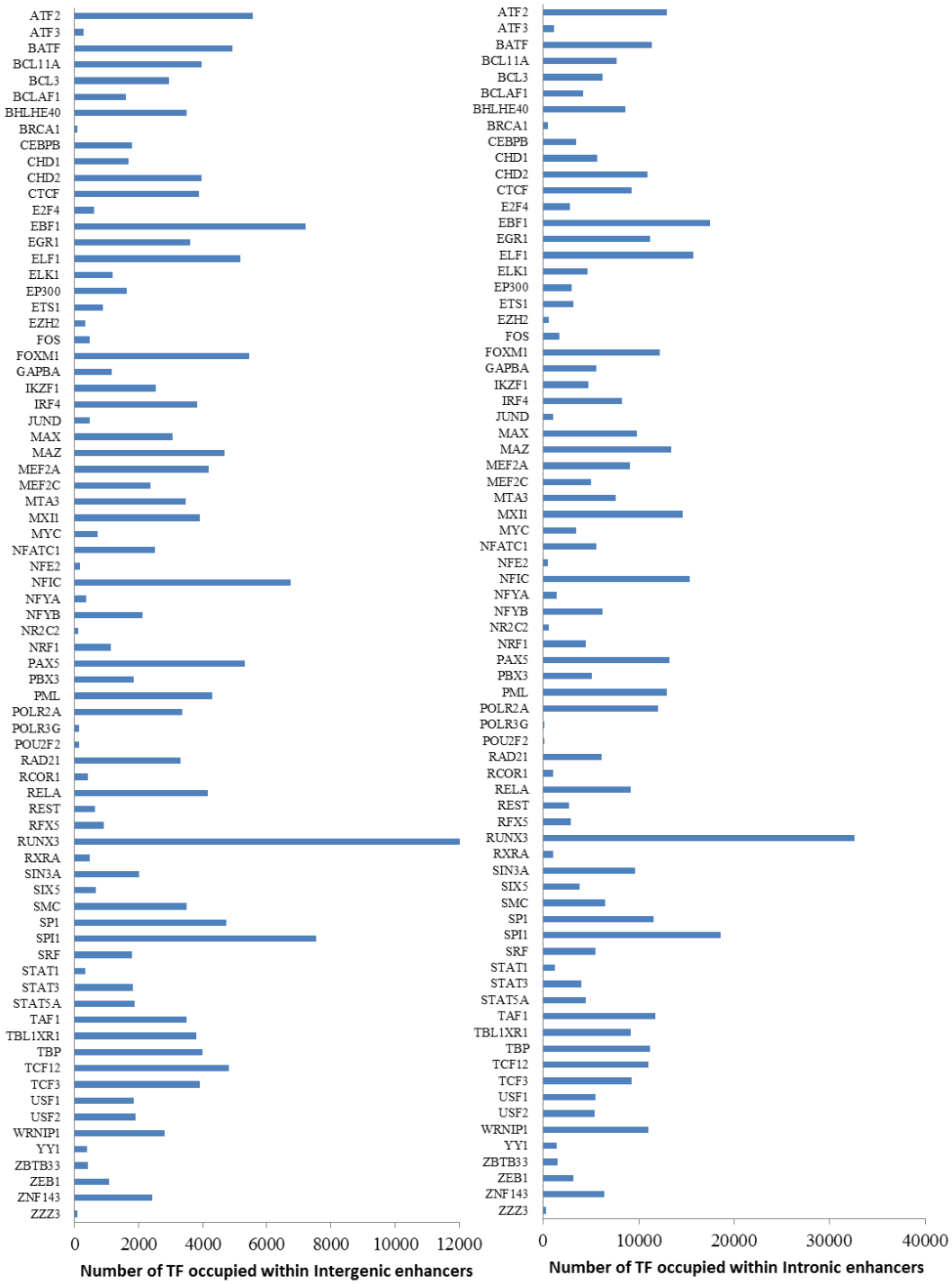


Figure 6.4: Intergenic and intronic enhancers are occupied with TFs involved in normal development including B-cell developments and differentiation.

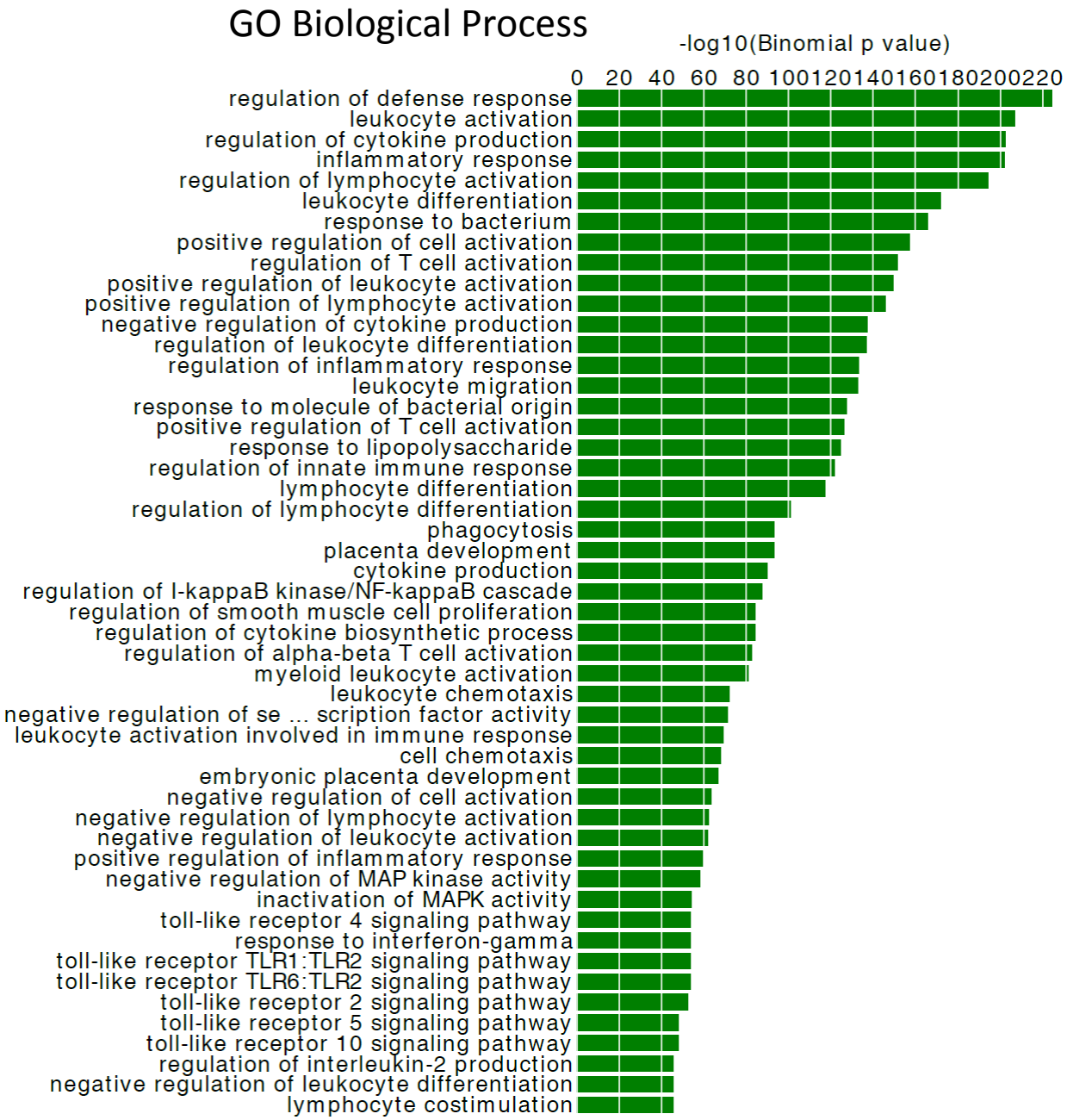


Figure 6.5: GO biological process associated with the poised intergenic enhancer's target genes.

Table 6.1: Significantly up-regulated genes related with the hypomethylation of poised enhancers in ALL

Gene name	fold change (log2)	Gene name	fold change (log2)	Gene name	fold change (log2)
LINGO1	9.22868	SCARB1	3.57723	SPIN1	2.06352
LILRA2	7.60091	RMI2	3.5091	STK32C	1.95926
BCL11B	7.58517	CPNE7	3.43258	OAT	1.9245
RGMA	7.46903	FAM132B	3.42906	APCDD1	1.90354
EGFL7	6.91028	PTGDS	3.40332	CD6	1.9018
RRM2	6.69486	PCBP3	3.16051	PLCD1	1.86844
NEURL1B	6.67143	COL6A1	3.1338	SEPN1	1.84394
HTRA3	6.58754	GIMAP2	3.08106	UMODL1	1.8341
CLIC5	6.49555	C10orf128	3.07786	PRR5	1.81845
ETS2	6.37726	EPAS1	2.98679	SRGN	1.80312
MARVELD1	6.20498	SECTM1	2.96059	SLC22A23	1.78273
KCNH2	6.16954	ABCG2	2.86714	SLC27A1	1.77181
FAM20C	6.01389	ELN	2.84787	HAAO	1.71623
TMEM255B	5.66309	TCF7L2	2.82986	TBC1D14	1.70607
IFITM3	5.63674	MGAT4A	2.81732	PQLC3	1.68345
MKI67	5.62574	TPRA1	2.7352	ZNF414	1.65035
ACKR3	5.60261	CHST15	2.6411	GLUL	1.62675
SIRPG	5.60056	IL15RA	2.63746	NOTCH1	1.61978
APBA2	5.42897	FADS2	2.59988	ITM2C	1.58206
SYNGR1	5.29295	NQO2	2.51979	HPCAL1	1.52733
CDCA8	5.02594	FBLN1	2.49152	ARHGEF10	1.52648
CD7	4.98177	DIP2C	2.47851	KIAA0513	1.5043
SPN	4.92697	LRRC4	2.47205	LRPAP1	1.48683
ARHGEF17	4.8952	MBOAT2	2.44192	CEP72	1.46653
CDT1	4.82467	MGMT	2.42902	GPR133	1.45225
SPC24	4.80203	CTNNA1	2.42521	TMPO	1.44712
ASAP2	4.8003	PLIN3	2.37846	TOB1	1.44558
PALM	4.79404	CDK20	2.34645	ATXN10	1.43927
ID2	4.65611	A2M	2.31512	TSC22D1	1.43614
GNAQ	4.64171	MAN1C1	2.29405	42256	1.36815
PLXND1	4.53693	KLRG1	2.29128	UBAC1	1.35359
CPT1A	4.4797	SUPT3H	2.26947	RABL5	1.3305
HRH2	4.47051	CTSD	2.26393	CHST12	1.29562
B4GALNT4	4.2531	LILRA5	2.26303	LDLR	1.28377
NOS1AP	4.00251	NKD2	2.1853	ZNF354B	1.25194
EPHB3	3.94459	SRC	2.12971	UBL3	1.2401
POC1A	3.93676	DBF4B	2.12782	PPP1R13B	1.22936
AQP3	3.70141	PSMG1	2.12319	C12orf49	1.19376
LAT	3.68098	NPHP4	2.12039	TXNL4A	1.1341
RGS12	3.67257	PDIA4	2.11805		
QPRT	3.64937	MRPL21	2.08209		

Table 6.2: Significantly down-regulated genes related with the hypomethylation of poised enhancers in ALL.

Gene name	fold change (log2)	Gene name	fold change (log2)
IKBKB	-0.735	PPM1K	-1.679
PIK3R5	-0.832	CTSZ	-1.752
MED15	-0.835	SPAG9	-1.754
MKLN1	-0.925	PDXK	-1.760
INSIG1	-0.935	FBRS	-1.760
ABR	-0.966	JUND	-1.851
APLP2	-1.019	CCR7	-1.885
DENND3	-1.034	COX4I1	-1.897
CELF2	-1.065	ETF1	-1.901
USP6NL	-1.072	WIPI2	-1.910
HSPA9	-1.083	KIAA1683	-1.942
ARID1B	-1.084	USP22	-1.967
ARHGAP12	-1.105	NOC3L	-2.028
RPL13	-1.135	CYLD	-2.104
TSGA10	-1.215	FBXW11	-2.151
GAPVD1	-1.219	CSNK1D	-2.175
CRLF3	-1.266	SENP2	-2.227
VPS41	-1.275	RBM38	-2.263
USP10	-1.275	NPIPB11	-2.310
HTT	-1.324	CELSR1	-2.378
DYRK1A	-1.387	KLF6	-2.467
ASB7	-1.403	SLCO4A1	-2.592
KDM4B	-1.412	PTPN1	-2.641
DUSP1	-1.425	GLIS2	-2.730
PRKAG2	-1.426	ADRA2B	-2.901
PRDM10	-1.434	TNFRSF13B	-3.035
LPCAT1	-1.441	ZNF516	-3.127
ABCG1	-1.455	C9orf72	-3.152
ULK1	-1.477	IFNGR1	-3.358
ABCA5	-1.497	PPP4R1L	-3.377
GRHL1	-1.592	SIK1	-3.593
ZEB1	-1.594	TNS4	-3.850
NSMF	-1.601	BCL2L11	-4.261
COL4A3BP	-1.606	IRF8	-4.298
OSGIN1	-1.633	PMEPA1	-5.014
ARHGEF1	-1.657		

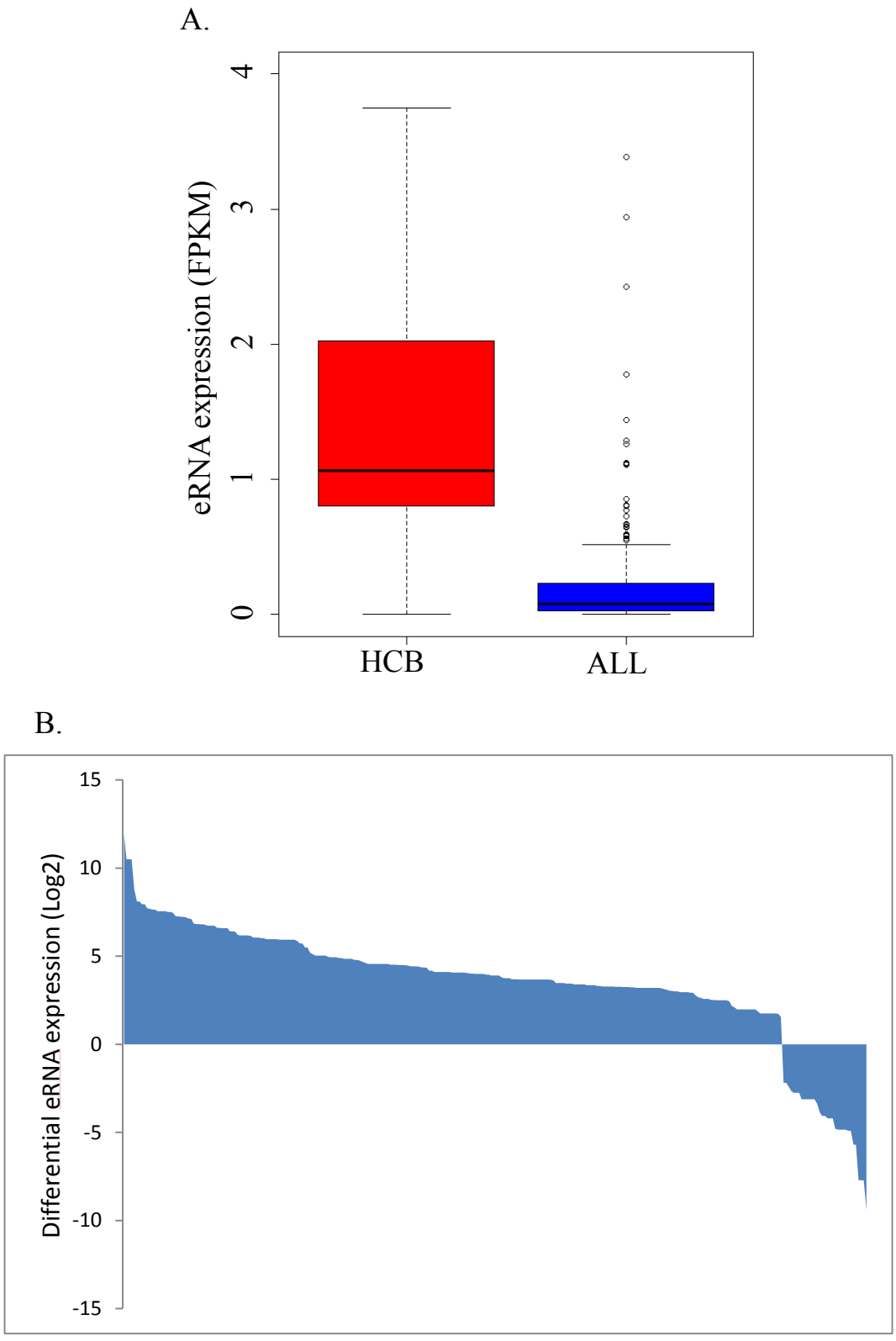


Figure 6.6: Enhancer expression. (A) eRNA expression level in HCB and ALL. (B) Significantly differentially expressed eRNA between HCB and ALL. Positive fold change= overexpression in HCB, negative fold change= overexpression in ALL.

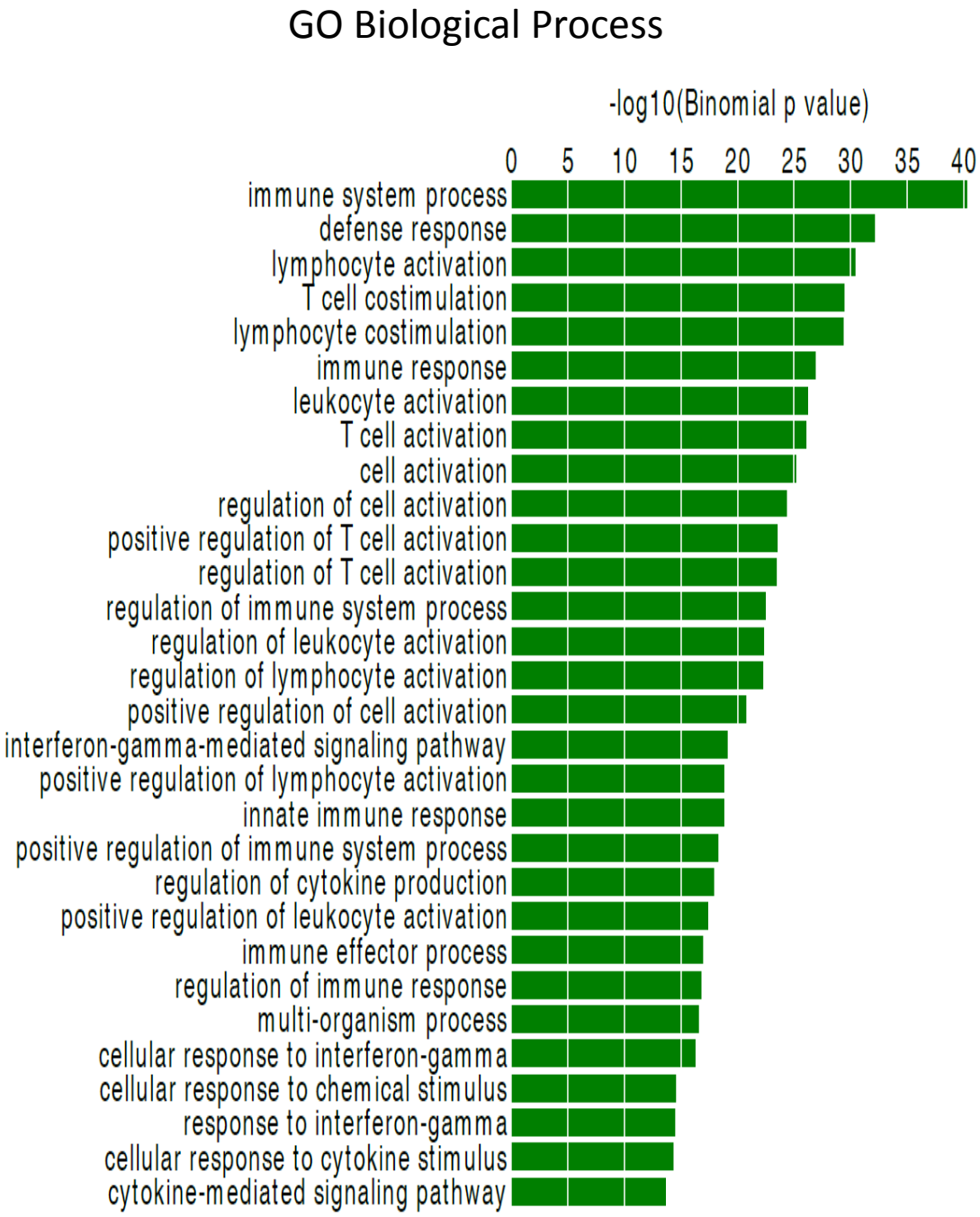


Figure 6.7: Biological process affected by the target genes of significantly differentially expressed enhancers.

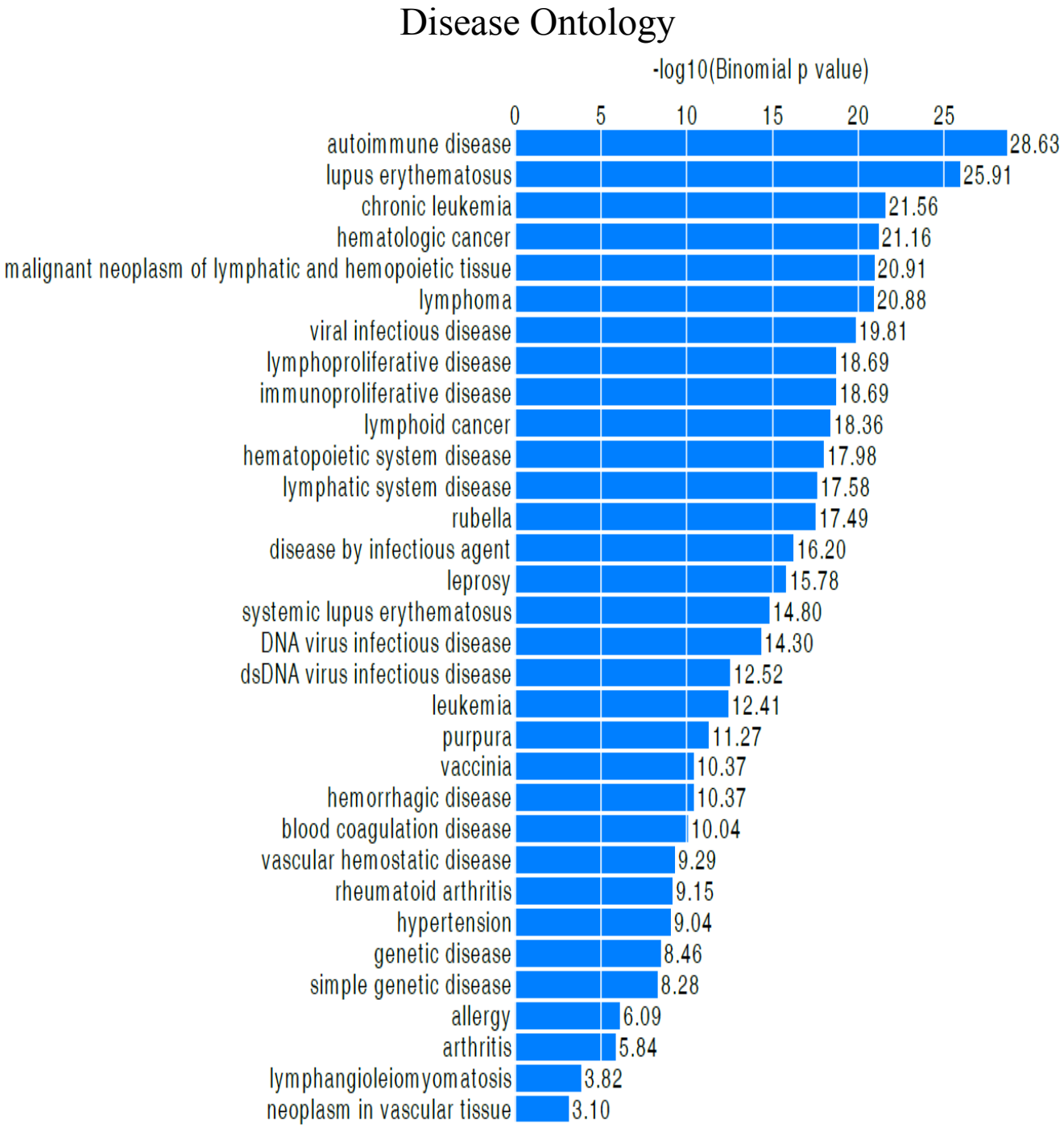


Figure 6.8: Disease ontology analysis of target genes associated with differentially expressed eRNA.

Table 6.3: Significant down-regulation of target genes associated with decreased eRNA expression in ALL.

Target genes	**Fold change (log2)**	**Target genes**	**Fold change (log2)**	**Target genes**	**Fold change (log2)**
ADAM28	-4.42	HLA-DOB	-3.36	RPL21	-1.50
ADRA2B	-2.90	HNRNPC	-1.00	RPS11	-1.44
ARHGEF1	-1.66	ICOSLG	-1.45	RPS14	-1.59
ARL5A	-0.88	IFNGR2	-2.48	RUNX3	-2.93
ASTL	-3.02	IL17C	-3.27	SCAF11	-1.92
B4GALT1	-2.88	IL21R	-4.16	SEC62	-1.32
BTLA	-4.24	IL4R	-4.00	SGMS1	-1.49
C11orf24	-1.97	IRF4	-2.46	SHOC2	-1.04
C17orf99	-2.27	IRF8	-4.30	SIK1	-3.59
CD48	-2.36	ISG20	-1.55	SLAMF1	-2.26
CDC42SE2	-1.74	ITPKB	-2.62	SLC38A1	-2.11
CDH22	-3.49	ITSN2	-1.79	SNX9	-5.00
CHD7	-0.72	JARID2	-2.18	SWAP70	-2.66
CLIC4	-2.18	LAPTM5	-1.77	SYPL1	-0.88
CTNNB1	-1.33	LPGAT1	-0.87	TAGAP	-2.37
CTSH	-1.82	LY9	-3.67	TMC8	-1.17
CTSK	-1.85	MALT1	-1.05	TMPRSS13	-2.77
CXXC5	-1.66	MAP3K8	-2.11	TNRC6B	-1.86
CYTH1	-2.65	MKNK2	-2.28	TOM1	-1.83
CYTH4	-0.83	MOB3A	-2.35	TPCN2	-2.18
ELOVL5	-1.89	MS4A1	-3.87	TRIM69	-2.34
EML4	-2.80	NEB	-1.46	TSC22D2	-1.67
FAM159A	-2.89	OPA1	-1.29	TSSK3	-4.12
GCM1	-5.08	PER1	-3.13	USP12	-1.88
GPBP1	-2.11	PLEKHG1	-1.75	USP7	-2.32
GPR160	-1.22	PPP1R2	-2.06	WBP11	-1.44
GRB2	-1.31	PRDM2	-2.40	WIPF1	-1.62
HCRTR1	-2.80	RALGPS2	-2.89	ZDHHC14	-2.49
HERPUD2	-0.99	RASL11A	-2.77	ZNF34	-2.66
HES1	-4.29	RGS2	-2.48		

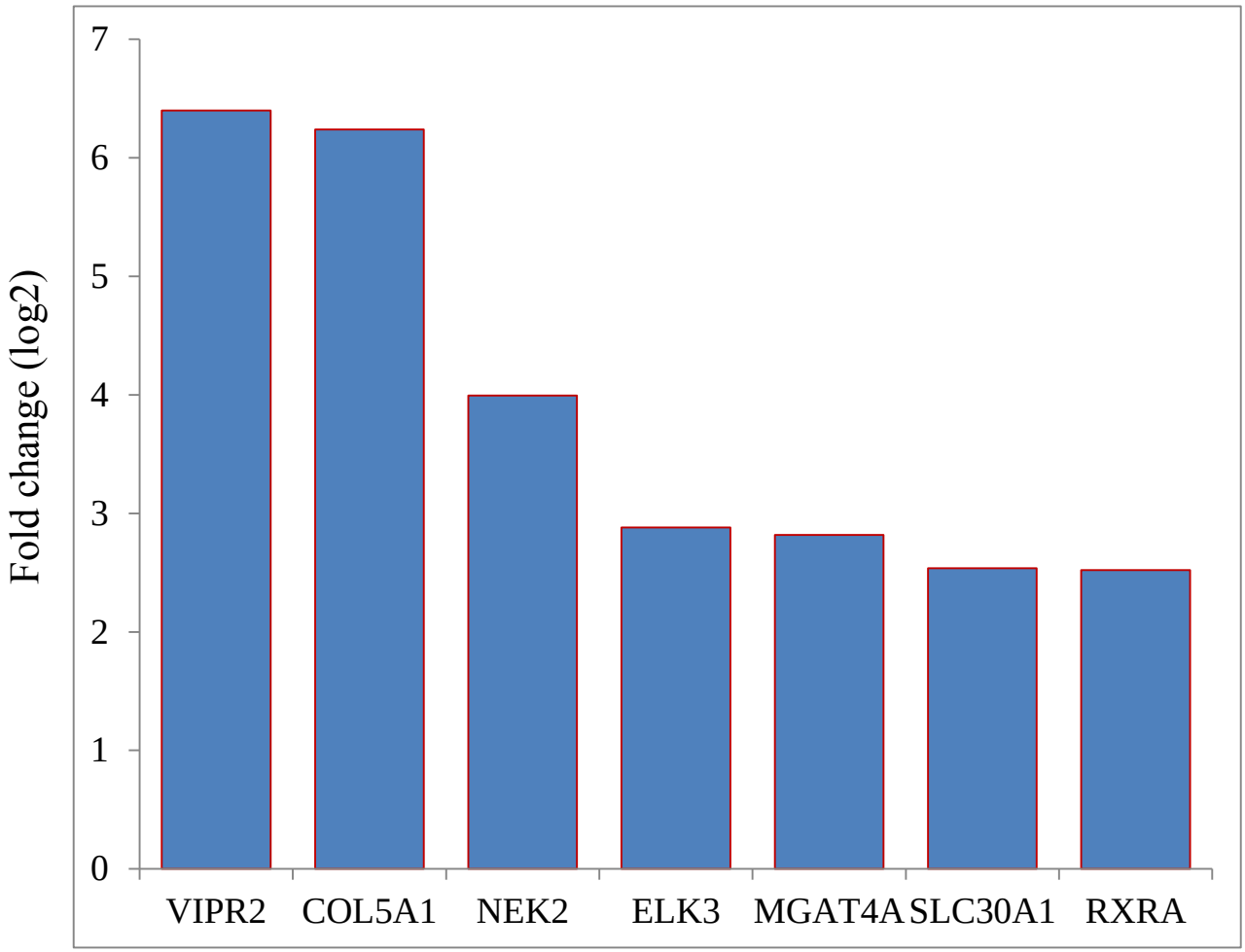


Figure 6.9: Up-regulation of potential target genes associated with the higher eRNA expression in ALL.

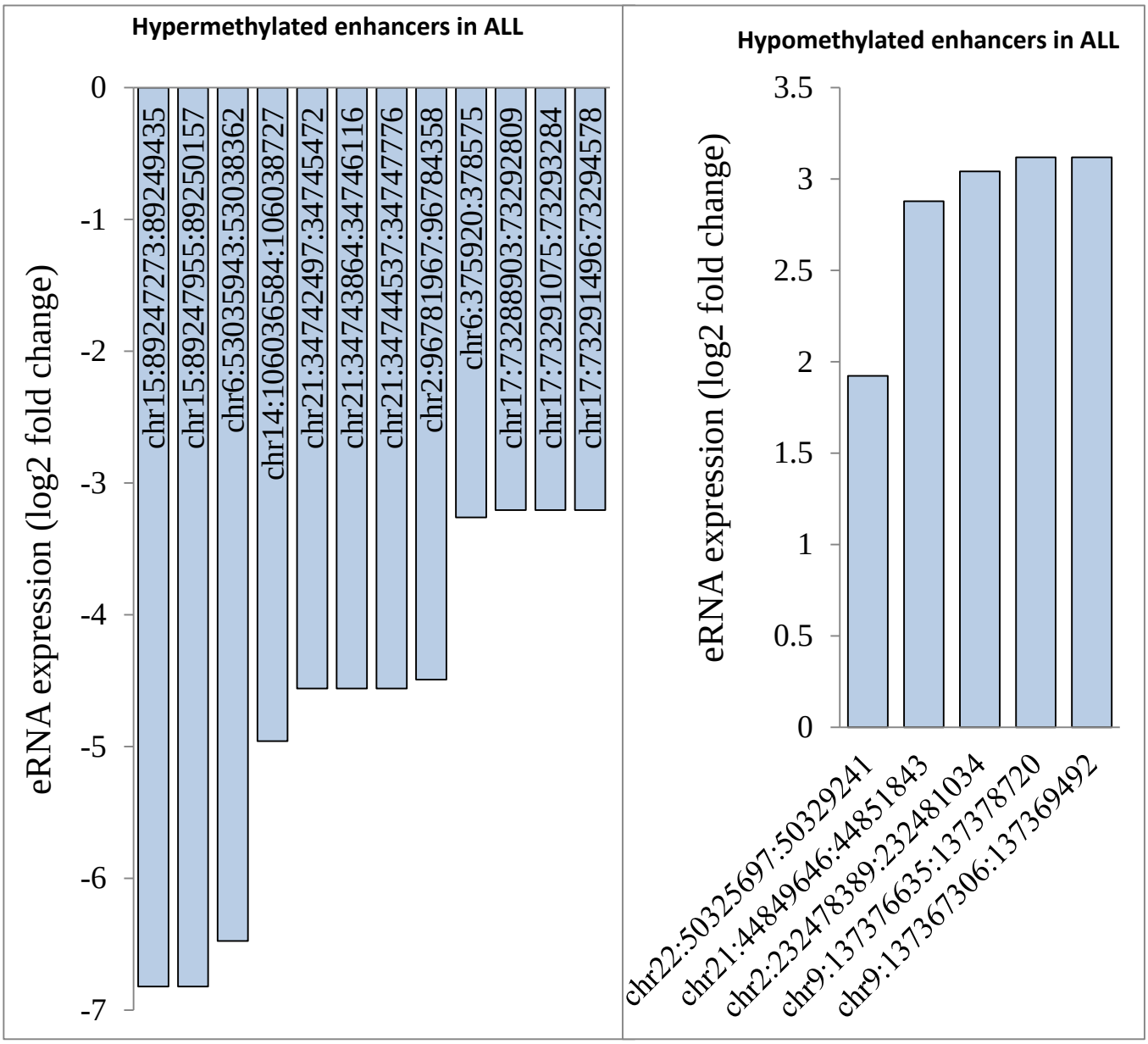


Figure 6.10: Differential methylation significantly affected eRNA expression in ALL.

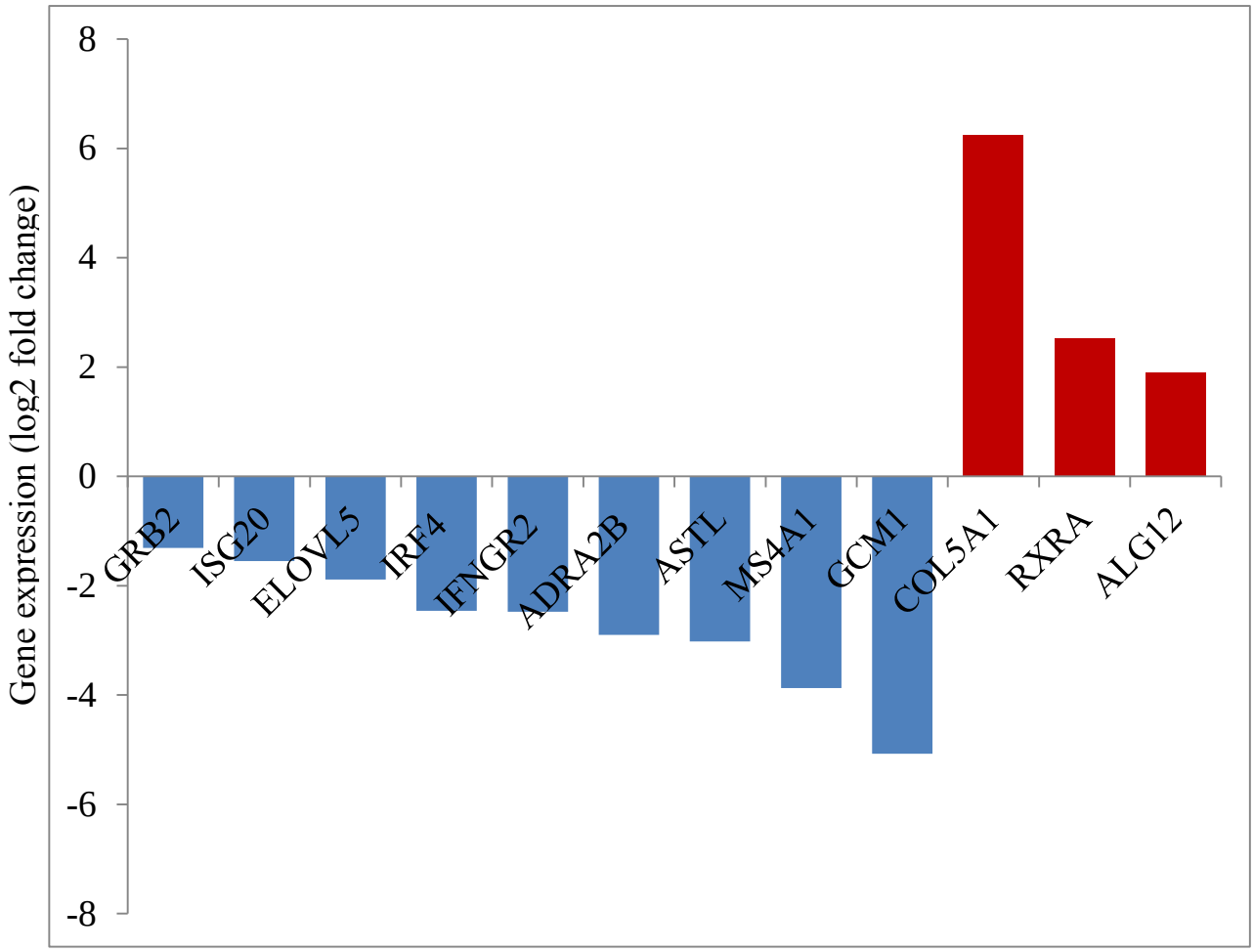


Figure 6.11: Differential enhancer methylation and eRNA expression affected the expression of nearby target genes in ALL. Blue bar=enhancer hypermethylation and decreased eRNA associated with the significant down-regulation of nearby target genes. Red bar=enhancer hypomethylation and increased eRNA associated with the significant up-regulation of nearby target genes

CH6: REFERENCES

Ackermann JA, Radtke D, Maurberger A, Winkler TH, Nitschke L. Grb2 regulates B-cell maturation, B-cell memory responses and inhibits B-cell Ca2+ signalling. EMBO J. 2011 Apr 20;30(8):1621-33.

Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJ, Gómez-Skarmeta JL. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. 2012 Oct;22(10):2043-53.

Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011 Feb 4;144(3):327-39.

Cheon DJ, Tong Y, Sim MS, Dering J, Berel D, Cui X, Lester J, Beach JA, Tighiouart M, Walts AE, Karlan BY, Orsulic S. A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer. Clin Cancer Res. 2014 Feb 1;20(3):711-23.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21931-6.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 2010 May 11;8(5):e1000384.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74.

Engels N, König LM, Heemann C, Lutz J, Tsubata T, Griep S, Schrader V, Wienands J. Recruitment of the cytoplasmic adaptor Grb2 to surface IgG and IgE provides antigen receptor-intrinsic costimulation to class-switched B cells. Nat Immunol. 2009 Sep;10(9):1018-25.

Gabriele L, Phung J, Fukumoto J, Segal D, Wang IM, Giannakakou P, Giese NA, Ozato K, Morse HC 3rd. Regulation of apoptosis in myeloid cells by interferon consensus sequence-binding protein. J Exp Med. 1999 Aug 2;190(3):411-21.

Gu Z, Zhou W, Huang J, Yang Y, Wendlandt E, Xu H, He X, Tricot G, Zhan F. Nek2 is a novel regulator of B cell development and immunological response. Biomed Res Int. 2014;2014:621082.

Gui R, Li D, Qi G, Suhad A, Nie X. Inhibition of Grb2-mediated activation of MAPK signal transduction suppresses NOR1/CB1954-induced cytotoxicity in the HepG2 cell line. Oncol Lett. 2012 Sep;4(3):566-570.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007 Mar;39(3):311-8.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010; 38:576-589.

Herz HM, Hu D, Shilatifard A. Enhancer malfunction in cancer. Mol Cell. 2014 Mar 20;53(6):859-66.

Holtschke T, Löhler J, Kanno Y, Fehr T, Giese N, Rosenbauer F, Lou J, Knobeloch KP, Gabriele L, Waring JF, Bachmann MF, Zinkernagel RM, Morse HC 3rd, Ozato K, Horak I. Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. Cell. 1996 Oct 18;87(2):307-17.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010 May 13;465(7295):182-7.

Korostowski L, Raval A, Breuer G, Engel N. Enhancer-driven chromatin interactions during development promote escape from silencing by a long non-coding RNA. Epigenetics Chromatin. 2011 Nov 15;4:21.

Lai F, Shiekhattar R. Enhancer RNAs: the new molecules of transcription. Curr Opin Genet Dev. 2014 Apr;25:38-42.

Lam MT, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M, Lee CY, Watt A, Grossman TR, Rosenfeld MG, Evans RM, Glass CK. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature. 2013 Jun 27;498(7455):511-5.

Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci. 2014 Apr;39(4):170-82.

Lan X, Witt H, Katsumura K, Ye Z, Wang Q, Bresnick EH, Farnham PJ, Jin VX. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. Nucleic Acids Res. 2012 Sep;40(16):7690-704.

Li LY, Li EM, Wu ZY, Cao HH, Shen JH, Xu XE, Chen B, Wu JY, Xu LY. Overexpression of GRB2 is correlated with lymph node metastasis and poor prognosis in esophageal squamous cell carcinoma. Int J Clin Exp Pathol. 2014 May 15;7(6):3132-40.

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013 Jun 27;498(7455):516-20.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May;28(5):495-501.

Melgar MF, Collins FS, Sethupathy P. Discovery of active enhancers through bidirectional expression of short transcripts. Genome Biol. 2011 Nov 14;12(11):R113.

Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA, Elkon R, Melo SA, Léveillé N, Kalluri R, de Laat W, Agami R. eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell. 2013 Feb 7;49(3):524-35.

Natoli G, Andrau JC. Noncoding transcription at enhancers: general principles and functional models. Annu Rev Genet. 2012;46:1-19.

Orom UA, Shiekhattar R. Noncoding RNAs and enhancers: complications of a long-distance relationship. Trends Genet. 2011 Oct;27(10):433-9.

Pearce G, Angeli V, Randolph GJ, Junt T, von Andrian U, Schnittler HJ, Jessberger R. Signaling protein SWAP-70 is required for efficient B cell homing to lymphoid organs. Nat Immunol. 2006 Aug;7(8):827-34.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2.

Romanuik TL, Ueda T, Le N, Haile S, Yong TM, Thomson T, Vessella RL, Sadar MD. Novel biomarkers for prostate cancer including noncoding transcripts. Am J Pathol. 2009 Dec;175(6):2264-76.

Royo JL, Hidalgo C, Roncero Y, Seda MA, Akalin A, Lenhard B, Casares F, Gómez-Skarmeta JL. Dissecting the transcriptional regulatory properties of human chromosome 16 highly conserved non-coding regions. PLoS One. 2011;6(9):e24824.

Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012 Sep 6;489(7414):109-13.

Seda V, Mraz M. B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. Eur J Haematol. 2015 Mar;94(3):193-205.

Sérandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Métivier R, Salbert G, Eeckhoute J. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res. 2011 Apr;21(4):555-65.

Shukla V, Lu R. IRF4 and IRF8: Governing the virtues of B Lymphocytes. Front Biol (Beijing). 2014 Aug;9(4):269-282.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schübeler D. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011 Dec 14;480(7378):490-5.

Su CL, Deng TR, Shang Z, Xiao Y. JARID2 inhibits leukemia cell proliferation by regulating CCND1 expression. Int J Hematol. 2015 May 5. [Epub ahead of print].

Shukla V, Lu R. IRF4 and IRF8: Governing the virtues of B Lymphocytes. Front Biol (Beijing). 2014 Aug;9(4):269-282.

Tamura T, Kong HJ, Tunyaplin C, Tsujimura H, Calame K, Ozato K. ICSBP/IRF-8 inhibits mitogenic activity of p210 Bcr/Abl in differentiating myeloid progenitor cells. Blood. 2003 Dec 15;102(13):4547-54.

Tripathy S, Jump DB. Elovl5 regulates the mTORC2-Akt-FOXO1 pathway by controlling hepatic cis-vaccenic acid synthesis in diet-induced obese mice. J Lipid Res. 2013 Jan;54(1):71-84.

Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature 2009;461:199-205.

Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu XD. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature. 2011 May 15;474(7351):390-4.

Wang F, Yang Y, Fu Z, Xu N, Chen F, Yin H, Lu X, Shen R, Lu C. Differential DNA methylation status between breast carcinomatous and normal tissues. Biomed Pharmacother. 2014 Jul;68(6):699-707.

Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ, Thurman RE, Stamatoyannopoulos JA, Hager GL. DNA methylation status predicts cell type-specific enhancer activity. EMBO J. 2011 Jun 24;30(15):3028-39.

Yamada K, Hayashi M, Madokoro H, Nishida H, Du W, Ohnuma K, Sakamoto M, Morimoto C, Yamada T. Nuclear localization of CD26 induced by a humanized monoclonal antibody inhibits tumor cell growth by modulating of POLR2A transcription. PLoS One. 2013 Apr 29;8(4):e62304.

Yoo EJ, Cooke NE, Liebhaber SA. An RNA-independent linkage of noncoding transcription to long-range enhancer function. Mol Cell Biol. 2012 May;32(10):2020-9.

CHAPTER 7: SUMMARY AND FUTURE DIRECTIONS

A large number of genes are mis-regulated (overexpressed, underexpressed) during carcinogenesis without harboring a driver gene mutation. These aberrantly expressed genes, if regulated by epigenetic modifications providing a selective growth advantage to the pre-neoplastic cells are termed as Epi-driver genes (Vogelstein et al. 2013). Of 294,881 reported mutations in 3,284 tumors, only 125 were identified as a mutation driver genes which confer a selective growth advantage (Vogelstein et al. 2013). Additionally, the identification of mutation driver genes has reached a plateau as the same mutation driver genes have been rediscovered in different tumor types. Despite rigorous efforts, reliable genetic deregulation that distinguishes between metastasized cancers and those that have not yet metastasized still remains to be recognized. Therefore, a complete understanding of the gene regulation network during malignant transformation is critical but still far beyond our achievement. Although genetic origins of cancer are well recognized, it seems that epigenetic modifications including DNA methylation are early events in carcinogenesis (Jones and Baylin 2007; Feinberg and Tycko 2004). The overall objective of this study was to determine the role of altered DNA methylation in the pathogenesis of acute lymphoblastic leukemia (ALL). Therefore, we have identified differentially methylated loci in the regulatory regions of the genome that result in differential gene expression in ALL.

Our first experiment was conducted to isolate immature B-cells at four different stages of their development (pro-B, pre-BI, pre-BII, and naïve B-cell). The objective of this study was to get the appropriate healthy control precursor B-cell from the easily available human umbilical cord blood (HCB) which correspond to the cells affected by ALL. We had to optimize several

technical challenges in order to get sufficient numbers of cells in each subset. For example, contaminating debris had great impact on the success of cell sorting by flow cytometry. This contaminating debris was minimized by using fresh cord blood (getting cord blood sample as early as possible after collection) and gently handling the cells during the entire cell separation procedure. Unlike adult peripheral blood separations, the mononuclear cell layer separated from HCB is often contaminated with red blood cells (Kanof et al. 1996) and platelets. Our protocol included an additional PBS wash step to remove contaminating platelets. Although higher numbers of lymphocytes were present in HCB samples classified as lymphocytosis, inadequate numbers of precursor-B cells were obtained limiting the samples that could be acquired for our studies. For the enrichment of precursor B-cells, all non B-cells were first depleted by negative selection through column purification. Since the protocol provided with the antibodies (Miltenyi Biotec) was optimized for peripheral blood samples, it was necessary to modify the protocol for successful sorts resulting in the use much lower amounts of the biotin antibody cocktail. Although different sorting strategies use different cell surface markers depending on the cell subtype of interest, our protocol uses only three antibodies against CD34, CD19 and CD45 which greatly reduced the complexity and the cost of performing cell sorting experiments. After all optimization was complete, 4 subsets of precursor B-cells, [pro-B (CD34+/CD19+); pre-BI (CD34-/CD19+/CD45low); pre-BII (CD34-/CD19+/CD45med); and naïve B-cell (CD34-/CD19+/CD45high)], were isolated at sufficient quantity and quality for downstream experiments. This developed protocol offers an approach for obtaining rare populations of precursor B-cells and this protocol can be modified to isolate any rare population present in cord blood including hematopoietic stem cells and immature T-cells.

Epigenetic modifications including DNA methylation play important roles in normal cellular differentiation and hematopoiesis (Hogart et al. 2012; Hodges et al. 2011; Bröske et al. 2009). Therefore, our second experiment was designed to identify the establishment of DNA methylation during B-cell development. Using MIRA-seq, we generated whole-genome DNA methylation profiles for 4 subsets of B-cells in 10 biological replicates. On average, 182 – 200 million sequences reads were generated for each precursor B-cell subset, which was sufficient to determine the methylated peaks in individual subset, and differentially methylated regions (DMRs) in between two consecutive stages of B-cell development. More than 201,000 peaks were common across all subsets and less than 5% of the peaks were unique to any individual subset. The majority of the methylation peaks present in intergenic and intronic regions in all four subsets, and a total of 3-4% peaks overlapped with a CpG island (CGI), 5-6% were found within a CpG shore, and 3-4% was found within a CpG shelf. Differential methylation analysis revealed that 14,294 DMRs were hypomethylated and 4,210 DMRs were hypermethylated (FDR $\leq$ 5%) during the transition from pro-B to pre-BI cells. No DMRs were observed at FDR $\leq$ 5% between more similar pre-BI to pre-BII, and pre-BII to naïve B-cell transition. Although the loss of methylation is predominant during the pro-B to pre-BI transition, a higher percentage of gene promoters were hypermethylated in pre-BI (3% hypermethylated versus 1% hypomethylated DMRs on promoter regions). Although no functional groups were significantly enriched using DAVID functional annotation analysis, genes involved in ubiquitin-dependent protein catabolic process (*USP17L6P*, *CDC27*, *USP17L1P*, *USP17L3*, *USP17L4*), and GTPase regulatory activity (*TBC1D3C, AGAP8, TBC1D3H, SIPA1L1*, *TBC1D3, RAB3IP*) were hypomethylated at the promoter regions during pro-B to pre-BI transition.

The vast majority of DMRs were present in intergenic and intronic loci, known to be involved in maintaining genomic stability (Jones 2012). However, intergenic and intronic regions also harbor vital regulatory sequences like enhancers that are affected by methylation (Bulger et al., 2011; Shore et al., 2010). Further analysis identified 467 potential enhancer-like regions within the intergenic and intronic DMRs, based on coinciding with enhancer specific histone marks (H3K4me1 and H3K27ac). Of these, 197 were hypomethylated and 270 were hypermethylated in pre-BI cells compared to pro-B cells. Several important genes such as *EGR1*, *FOXP1*, *KLF6* involved in lymphocyte differentiation, and *CARD11*, *LILRB3*, *CD81* involved in BCR signaling were located immediately upstream or downstream of putative enhancer regions that become hypermethylated during the transition from pro-B to pre-BI cells. In contrast, *VAV3, GRB2, PPP3CA,* and *NFATC1,* genes involved in BCR signaling pathway, were in proximity to hypomethylated enhancer regions during the transition from pro-B to pre-B cells. Therefore, it is reasonable to infer that differential methylation on cell-type specific enhancers regulate the expression of genes required for B-cell differentiation. Finally the established genome-wide methylation patterns may be used to gain a better understanding of the contribution of aberrant DNA methylation in the development of hematological disorders associated with precursor B-cells.

Next, the methylation profiles of 20 pre-B ALL samples were determined using MIRA-seq. Then these profiles were compared to healthy pre-B cell isolated from umbilical cord blood (HCB), the healthy counterparts of malignant precursor B-cells. Methylation peaks were more abundant in HCB samples (305,736) compared to ALL samples (162,832), revealing a global reduction of methylation in ALL. DMRs were determined between ALL and HCB samples to recognize the alteration of DNA methylation in ALL (FDR $\leq$5% and at least 2 fold change

difference). A total of 15,492 regions lost methylation (hypomethylation) and 9,790 regions gained methylation (hypermethylation) in ALL compared to the normal HCB samples. More DMRs were hypermethylated at the 5' regulatory regions of genes than hypomethylated, and the majority of the DMRs overlapped with intergenic and intronic genomic regions. A total of 1,252 gene promoters were hypermethylated and 240 gene promoters were hypomethylated in ALL. Our identified DMRs could have applicability as ALL specific biomarkers and may also play regulatory roles in the expression of genes that are involved in the pathogenesis of ALL. In order to determine the regulatory potential of DMRs, whole genome transcriptomes were generated in both ALL and HCB samples using RNA-seq. Integration of differential methylation and differential gene expression revealed that hypermethylation of 62 gene promoters had significantly decrease expression in ALL and these mis-regulated genes were involved in several functional groups including regulation of transcription and apoptosis as well as tumor suppressor genes *MTSS1, PAWR,* and *EXT1*. On the other hand, 37 gene promoters were hypomethylated and significantly increased expression in ALL, and were significantly enriched for genes involved in GTPase activation, regulation of cell proliferation, and genes involved in protein complex assembly. Therefore, we sought to document the potential Epi-driver gene in ALL. From the integration analysis, we found many genes including *SYNE1* (cytokinesis), *PTPRS* (signaling molecule), *PAWR* (pro-apoptotic gene), *HDAC9* (downstream target of KRAS), *RGCC* (cell-cycle regulator) and *MCOLN2* (unknown function) were hypermethylated at the promoter regions and down-regulated in ALL. Interestingly, these genes have also been shown to be hypermethylated and/or down-regulated in other malignancies (Nagai et al., 2010; Okudela et al., 2014; Vlaicu et al., 2008), signifying their potential tumor suppressor activity and supporting the role of DNA methylation as a regulator of gene expression.

Promoter hypermethylation in combination with mounting gene body methylation has a stronger repressive effect on gene expression during normal B-cell development compared to promoter methylation alone (Lee et al., 2012). Several mechanisms are suggested for how gene body methylation leads to defective gene expression, such as the regulation of transcriptional elongation (Lorincz et al., 2004), cell-type specific selection of alternative promoters (Maunakea et al., 2010), modulating alternative RNA splicing (Maunakea et al., 2013), or defining alternative polyadenylation sites (Wood et al. 2008). However, the cause and effect of DNA methylation within gene bodies is not fully understood. We observed that gene body hypermethylation was associated with an increase in expression of some genes and a decrease in expression of other genes in ALL, suggesting that the functional effect of gene body methylation may be exerted in a gene specific manner depending on the presence or absence of other regulatory events.

In addition to protein coding gene promoters, differential methylation was also observed in 69 miRNA, 65 lincRNA and 55 pseudogenes. Since, the miRNA, lincRNAs and pseudogenes have the potential to epigenetically regulate the expression of multiple target genes and cellular processes, their mis-regulation may play important roles in the pathogenesis of ALL. Additionally, many intergenic DMRs in ALL were associated with transposable elements including centromeric alpha satellite repeats. Interestingly, transposable elements are often silenced by DNA methylation (Maksakova et al. 2008), and transcriptional activation of these elements results in transposable element mediated insertional mutagenesis and chromosomal rearrangements in many cancers (Lee et al. 2012). The hypomethylation of centromeric alpha satellite repeats may block the binding of centromere specific histone protein CENP-A and result in the inactivation of the centromere.

We also identified a total of 3,699 significantly up-regulated genes and 2,734 down-regulated genes in ALL compared to healthy HCB samples (q-value ≤0.05). The up-regulated genes were significantly enriched for many functional groups including cell cycle processes, mitosis, DNA replication, alternative splicing, phosphoprotein, cytoskeleton, chromatin assembly, kinetochore, mismatch repair, histone folding and more. On the other hand, the down-regulated genes were significantly enriched for several other functional groups including B cell receptor signaling pathway, transcription regulation, serine/threonine-protein kinase, translational elongation, phosphorylation, MAPK signaling pathway, GTPase regulator activity, lymphocyte activation and differentiation, positive regulation of apoptosis, and regulation of B cell proliferation. Remarkably, many genes known to be involved in B-cell development and epigenetic modifier genes were mis-regulated in ALL. For example, *DNMT1, DNMT3A*, and *DNMT3B* are responsible for DNA methylation and were significantly up-regulated in ALL, along with down-regulation of actively demethylating genes *TET2* and *TET3*. Histone proteins are wrapped by DNA to form a nucleosome, and 22 genes encoding histone proteins were significantly up-regulated in ALL. However, chromatin activating histone lysine acetyltransferases (*MGEA5, CDYL, CREBBP, EP300,* and *NCOA3*) were down-regulated and the chromatin inactivating histone deacetylases (*HDAC11* and *SIRT2*) were up-regulated in ALL. Additionally, we also identified 197 deregulated lincRNA and 131 deregulated pseudogenes in ALL. Due to sequence similarity, pseudogene transcripts interact with their parent genes and found to be differentially expressed in different types of cancer (Han et al. 2014; Xiao-Jie et al. 2015). Interestingly, we also found a large number of protein-coding parent genes are mis-regulated corresponding with differential expression of their pseudogenes.

Enhancers are regulatory sequences that can be located upstream, downstream, or within their target gene(s) and can control gene expression through physical interactions with gene promoters. Due to the availability of active enhancer specific histone marks data, we identified potential enhancer regions in a lymphoblastoid cell line to serve as a surrogate for healthy precursor B-cells. The identified enhancer regions were further analyzed with our differential methylation and differential gene expression data between ALL and HCB samples. Overall 245 intergenic and 545 intronic enhancer regions became hypermethylated in ALL. Enhancer hypermethylation in ALL was associated with the down-regulation of 204 nearby potential target genes, and transcriptional silencing of those genes by enhancer hypermethylation may provide survival and growth advantages to the leukemia cell in ALL.

Recent studies provided evidence that enhancer regions are actively transcribed to RNA (eRNA) and that eRNA plays a crucial role in regulating gene expression in normal development. In our study, a total of 1,903 intergenic enhancers were found to be transcribed in HCB and ALL samples where 319 eRNA were significantly differentially expressed in ALL (q-value <0.05). The majority of differentially expressed eRNA were significantly down-regulated in ALL compared to healthy pre-B cell (284- vs 35 eRNAs). This result is expected since the enhancers identified in the lymphoblastoid cell line represent active enhancers in normal cells. The gaining of DNA methylation in ALL may lead to transcriptional inactivation of target genes during ALL pathogenesis. Further, we analyzed the effect of differential eRNA expression on the expression of potential target genes. We found the significant down–regulation of 89 potential target genes were associated with lower levels of eRNA transcripts, and the up-regulation of seven potential targets genes were associated with higher levels of eRNA transcripts. Several of the potential target genes are involved in normal B-cell differentiation and activation, and

enhancer mediated mis-regulation of those genes may contribute to the pathogenesis of leukemia. We also identified DNA methylation mediated mis-regulation of eRNA transcription in ALL. Thirteen hypermethylated enhancers were associated with the significant down-regulation of eRNA transcripts (q-value ≤ 0.05), whereas five hypomethylated enhancers were associated with the up-regulation of eRNA transcripts in ALL. Additionally, enhancer hypermethylation and down-regulation of eRNA transcripts was also associated with the significant down-regulation of nearby *GRB2, ISG20, ELOVL5, IRF4, IFNGR2, ADRA2B, ASTL, MS4A1, GCM1* target genes in ALL. On the other hand, enhancer hypomethylation and up-regulation of eRNA transcripts was associated with the up-regulation of nearby *COL5A1, RXRA, ALG12* genes. Collectively, our results shed new light on the enhancer mediated mis-regulation of nearby target genes in ALL, and that is commendable for further research.

This study has highlighted many features of altered DNA methylation and gene regulation in ALL that are just beginning to be elucidated. Although genome-wide sequencing studies detected hundreds of thousands of point mutations in different tumors, only a small number are considered to be mutation driver genes. Our studies have revealed many epi-mutations that may be epi-driver mutations or may only be passengers. It is imperative to sort out the driver epi-mutations from passenger epi-mutations so that the best therapeutic targets may be identified. In addition, we report many differentially methylated loci which may be used as biomarkers in ALL in order to aid in diagnosis, prognosis and response to treatment. To identify the utility of the differentially methylated loci as suitable biomarkers, one of the most important next steps would be to assess the sensitivity and specificity of the differentially methylated regulatory regions. For example, promoter hypermethylation of tumor suppressor

genes *MTSS1, PAWR, EXT1*; *SYNE1,* involved in cytokinesis; the signaling molecule, *PTPRS*; the pro-apoptotic gene, *PAWR*; the downstream target of KRAS, *HDAC9*; the cell-cycle regulator, *RGCC*; and *MCOLN2* was associated with down-regulation of gene expression in ALL. Validation of these genes in large number of patients with ALL could be useful for early detection of ALL. Methylation-specific PCR (MSP) validation of five genes MGMT, p16INK4a, RASSF1A, DAPK, and RAR-β in two hundreds patients demonstrated that when DNA methylation was positive in at least one gene, specificity and predictive value for diagnosis of lung cancer was 85% and 75%, respectively. This study also found that the odds ratio of having lung cancer is 5.28 for patients with methylation in one gene, and 5.89 for patients with promoter methylation in two or more of those genes (Fujiwara K. et al., 2005). A similar study (Müller et al., 2004) used MethyLight to validate the sensitivity and specificity of *SFRP2* hypermethylation in fecal DNA from patients with colorectal cancer (77% sensitivity and specificity). Specific DNA methylation marks can also be used for the prediction of treatment outcome of many tumors. Further studies are required to correlate the altered DNA methylation pattern and prognostic outcomes in ALL.

Additional work should be focused on identifying the phenotypic effect of transcriptional restoration of genes that are inactivated by DNA methylation in ALL. Recently, it was shown that tumor suppressor gene ASC/TMS1 is hypermethylated and down regulated in primary human renal cell carcinoma (RCC) and in RCC cell lines (Liu et al. 2015). Treatment of RCC cell lines with demethylating agent 5-Aza-2′-deoxycytidine restored the expression of ASC/TMS1 at the mRNA and protein level. The restoration of ASC/TMS1 inhibited cell viability, colony formation, induced apoptosis, arrested cell cycle, and repressed tumorigenicity in immunodeficient SCID mice. Another study demonstrated that methylation of *DLX4* gene was

associated with its down-regulation in chronic myeloid leukemia, and the removal of methylation marks using 5-Aza-2′-deoxycytidine restored the expression of *DLX4* (Zhou et al. 2015). Phenotypic evaluation of our results may provide useful biomarkers for designing upgraded diagnostic and therapeutic approaches for ALL.

Moving forward, it would be valuable to examine how the mis-regulated lincRNA contributes to the development of ALL. We have identified a large number of lincRNA are differentially methylated in ALL. Furthermore, hypomethylation was associated with the up-regulation of lincRNA *AC002398.5, DIO3OS* and *LINC00642* in ALL, but their regulatory mechanism in ALL remain unknown. LincRNAs can control transcriptional activation and their differential expression profiles are strongly associated with cancer progression (Calin et al. 2007). Our finding of mis-regulated lincRNA could be worthwhile to investigate the possible connection with ALL development.

Our study identified the putative enhancer regions in a lymphoblastoid cell line, as a surrogate of healthy pre-B cell. We also determined the methylation mediated mis-regulation of enhancer activation, which was associated with the deregulation of nearby genes. Determining the physical interaction between regulatory enhancer and gene promoter is critical for identifying the target genes. Many techniques such as 4C (circular chromosome conformation capture), 5C (chromosome conformation capture carbon copy), and Hi-C can be used to identify the physical interaction between regulatory enhancers and target genes. It is vitally important to test the functionality of the identified putative enhancer regions. Functional validation of enhancers can also be performed using luciferase reporter assays (Heintzman et al. 2009). Enhancers function as a key determinant of cell type specificity in gene expression; therefore further ChIP-seq

experiments may be beneficial in healthy precursor B-cells and ALL samples to identify the tissue specific active and poised enhancers.

Epigenetic modifications like DNA methylation and histone modification interact with each other to regulate the complex gene regulation mechanism, and cancer cells possess distinct epigenomes. Therefore, further advancement in understanding the cancer specific epigenetic alterations along with the development of more target specific drugs may hold the key to effectively reset the abnormal cancer epigenome.

CH7: REFERENCES

Bröske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, et al. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 2009; 41:1207-15.

Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell 2011;144:327-39.

Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, Fabbri M, Cimmino A, Lee EJ, Wojcik SE, Shimizu M, Tili E, Rossi S, Taccioli C, Pichiorri F, Liu X, Zupo S, Herlea V, Gramantieri L, Lanza G, Alder H, Rassenti L, Volinia S, Schmittgen TD, Kipps TJ, Negrini M, Croce CM. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell. 2007 Sep;12(3):215-29.

Feinberg AP, Tycko B. The history of cancer epigenetics. Nat Rev Cancer. 2004 Feb;4(2):143-53.

Fujiwara K, Fujimoto N, Tabata M, Nishii K, Matsuo K, Hotta K, Kozuki T, Aoe M, Kiura K, Ueoka H, Tanimoto M. Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. Clin Cancer Res. 2005 Feb 1;11(3):1219-25.

Han L, Yuan Y, Zheng S, Yang Y, Li J, Edgerton ME, et al. The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. Nat Commun. 2014; 5:3963.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009 May 7;459(7243):108-12.

Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell. 2011; 44:17-28.

Hogart A, Lichtenberg J, Ajay SS, Anderson S; NIH Intramural Sequencing Center, Margulies EH, Bodine DM. Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. Genome Res. 2012 Aug;22(8):1407-18.

Jones PA. Functions of DNA methylation: islands, strat sites, gene bodies and beyond. Nature Reviews Genetics 2012. 13:484-492.

Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007 Feb 23;128(4):683-92.

Kanof ME, Smith PD, Zola H.. Isolation of whole mononuclear cells from peripheral blood and cord blood. Curr Protoc Immunol. 2001 May;Chapter 7:Unit 7.1.

Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK, Zheng S, Dou X, de Smith A, Chokkalingam A, Buffler P, Ma X, Wiemels JL. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. Nucleic Acids Res 2012; 40(22):11339-51.

Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, et al. Landscape of somatic retrotransposition in human cancers. Science 2012; 337:967-971.

Liu Q, Jin J, Ying J, Cui Y, Sun M, Zhang L, Fan Y, Xu B, Zhang Q. Epigenetic inactivation of the candidate tumor suppressor gene ASC/TMS1 in human renal cell carcinoma and its role as a potential therapeutic target. Oncotarget. 2015 Jun 5. [Epub ahead of print].

Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 2004; 11:1068-1075.

Maksakova IA, Mager DL, Reiss D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cell Mol Life Sci. 2008; 65:3329-3347.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 2010; 466:253-257.

Maunakea AK, Chepelev I, Cui K, Zhao K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 2013; 23:1256-1269.

Müller HM, Oberwalder M, Fiegl H, Morandell M, Goebel G, Zitt M, Mühlthaler M, Ofner D, Margreiter R, Widschwendter M. Methylation changes in faecal DNA: a marker for colorectal cancer screening? Lancet. 2004 Apr 17;363(9417):1283-5.

Nagai MA, Gerhard R, Salaorni S, Fregnani JH, Nonogaki S, Netto MM, et al. Down-regulation of the candidate tumor suppressor gene PAR-4 is associated with poor prognosis in breast cancer. Int J Oncol 2010; 37:41-49.

Okudela K, Mitsui H, Suzuki T, Woo T, Tateishi Y, Umeda S, et al. Expression of HDAC9 in lung cancer--potential role in lung carcinogenesis. Int J Clin Exp Pathol 2014; 7:213-220.

Shore A, Karamitri A, Kemp P, Speakman JR, Lomax MA. Role of Ucp1 enhancer methylation and chromatin remodelling in the control of Ucp1 expression in murine adipose tissue. Diabetologia 2010; 53(6):1164-73.

Vlaicu SI, Cudrici C, Ito T, Fosbrink M, Tegla CA, Rus V, et al. Role of response gene to complement 32 in diseases. Arch Immunol Ther Exp (Warsz) 2008; 56:115-122.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013 Mar 29;339(6127):1546-58.

Wood AJ, Schulz R, Woodfine K, Koltowska K, Beechey CV, Peters J, et al. Regulation of alternative polyadenylation by genomic imprinting. Genes Dev 2008; 22:1141-1146.

Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X. Pseudogene in cancer: real functions and promising signature. J Med Genet 2015; 52:17-24.

Zhou JD, Wang YX, Zhang TJ, Yang DQ, Yao DM, Guo H, Yang L, Ma JC, Wen XM, Yang J, Lin J, Qian J. Epigenetic inactivation of DLX4 is associated with disease progression in chronic myeloid leukemia. Biochem Biophys Res Commun. 2015 Jun 15. pii: S0006-291X(15)30145-5.

VITA

Md Almamun was born in Mymensingh, Bangladesh in 1980. Almamun started his college education in Veterinary Medicine in 1998 at the Bangladesh Agricultural University in Mymensingh, Bangladesh. After receiving his Doctor of Veterinary Medicine (DVM) degree in 2003, he joined in a pharmaceutical company and worked as a product executive. He moved to the USA in 2007, and worked as a Research Assistant with Dr. OJ Ginther. There he studied on reproductive physiology of cattle and horse. In April of 2009, Almamun moved to the Division of Animal Sciences at the University of Missouri to work on a Master of Science degree in developmental epigenetics under the supervision of Dr. Rocío M. Rivera. He began work towards a Doctor of Philosophy in the MU Pathobiology Area Program in August 2011 under the supervision of Dr. Kristen H. Taylor. During his graduate studies, he has served as a lab mentor for four undergraduate students, and awarded an Outstanding Graduate Student Research award 2015 from the Department of Pathology and Anatomical Science. He has also given several poster presentations and won first place at the Mizzou Epigenetics symposium 2014. After receiving his PhD, he will start working as a postdoctoral research fellow at the Boston Children Hospital/Harvard Medical School.