RNA-SEQUENCING ANALYSIS IN B-CELL ACUTE LYMPHOBLASTIC LEUKEMIA REVEALS ABERRANT GENE EXPRESSION AND SPLICING ALTERATIONS

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

OLHA KHOLOD

Dr. Kristen Taylor, Thesis Supervisor

MAY 2017

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

RNA-SEQUENCING ANALYSIS IN B-CELL ACUTE LYMPHOBLASTIC LEUKEMIA REVEALS ABERRANT GENE EXPRESSION AND SPLICING ALTERATIONS

Presented by OLHA KHOLOD

A candidate for the degree of Master of Science

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

_____

Kristen Taylor, Ph.D.

_____

Christine Elsik, Ph.D.

_____

Dmitriy Shin, Ph.D.

ACKNOWLEDGEMENTS

First and foremost I would like to acknowledge my academic advisor Dr. Kristen Taylor who gave me the opportunity to be trained in her laboratory. Throughout my study, she contributed to a rewarding graduate school experience by giving me intellectual freedom in research and inspiring me to pursue a career in science. Additionally, I would like to thank my committee members Dr. Christine Elsik and Dr. Dmitriy Shin for their guidance and encouragement. Especially, Dr. Elsik who trained me to perform transcriptome data analysis and to program in Perl.

I also would like to acknowledge the many people I have worked with during the past two years. I want to thank Marianne Emery for assisting me with edgeR analysis and for her valuable advice regarding the processing of RNA-seq data. In addition, I would like to thank Dr. Senthil Kumar for fruitful discussions about cancer epigenetics and guidance in performing cell line treatment experiments. I also would like to acknowledge my laboratory mates Alex Stuckel and Clayton Del Pico for their friendship and support.

I would like to thank the Fulbright Foreign Student Program for providing an opportunity to obtain firsthand research experience in the United States and to meet with amazing people from all over the world. I also want to thank my best friends Sopheak and Xianglei for making me feel like home and for making me a better person. Finally, I would like to express my very profound gratitude to my parents and to my elder sister for providing me with unfailing support and continuous encouragement throughout my life and career.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

5-Aza	5-aza-2-deoxycytidine
AS	Alternative splicing
B-ALL	B-cell acute lymphoblastic leukemia
CGI	CpG island
CLP	Common lymphoid progenitor
DE	Differentially expressed genes
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
eRNA	Enhancer RNA
FISH	Fluorescence in situ hybridization
FPKM	Fragments per kilobase of transcript per million mapped reads
GLM	General linear model
HSC	Hematopoietic stem cell
IPA	Ingenuity pathway analysis
KB	Knowledge Base
LMPP	Lymphoid multipotent progenitor
MDS	Multidimensional scaling
miRNA	MicroRNA
NGS	Next generation sequencing
PCR	Polymerase chain reaction

Pre-BCR	Pre-B cell receptor
RNA-seq	RNA-sequencing
RT-PCR	Reverse transcription polymerase chain reaction
TF	Transcriptional factor
TMM	Trimmed mean of M-values
TR	Transcriptional regulator
TSA	Trichostatin A
UTR	Untranslated region
WBC	While blood cell

PREFACE

B-cell acute lymphoblastic leukemia (B-ALL) is a neoplasm of immature lymphoid progenitors and is the leading cause of cancer-related death in children. The majority of B-ALL cases are characterized by recurring structural chromosomal rearrangements that are crucial for triggering leukemogenesis, but do not explain all incidences of disease. Therefore, other molecular mechanisms, such as alternative splicing and epigenetic regulation may alter expression of transcripts that are associated with the development of B-ALL. It is important to investigate alternatively spliced RNA transcripts that may be affected by aberrant DNA methylation in B-ALL to gain a better understanding of the pathogenesis of this disease.

The goal of this research proposal is to characterize the transcriptome landscape of patients with B-ALL using high throughput RNA-sequencing (RNA-seq) analysis. Specifically, the study aims to identify particular genes and their isoforms that might be controlled by aberrant DNA methylation in B-ALL and contribute to the development of this disease. By analyzing transcriptional patterns between B-ALL patients and healthy cord blood donors differentially expressed and alternatively spliced RNA transcripts have been identified. By examining differentially expressed genes with Ingenuity pathway analysis, the most significant signaling pathways and gene functions have been annotated. By analyzing causative gene networks, novel upstream regulators have been determined for B-ALL patients. Finally, a mechanistic study has been conducted using an

in vitro B-ALL model to investigate if aberrant DNA methylation affects alternatively spliced genes associated with this disease.

In this thesis, chapter 1 will introduce abnormal B-cell development in leukemogenesis and discuss in detail the genetic abnormalities that are hallmarks of B-ALL. Chapter 1 will also introduce aberrant epigenetic modifications including DNA methylation, histone modifications, and non-coding RNAs that have been identified in B-ALL patients to date. Alternative splicing alterations associated with B-ALL will be also described in chapter 1. Chapter 2, the research chapter, investigates the transcriptional regulators and signaling pathways that likely orchestrate the regulation of differentially expressed genes identified in the study. Finally, chapter 2 includes a mechanistic study utilizing the Nalm 6 cell line to determine the role of DNA methylation on the expression of alternatively spliced transcripts.

Our pathway-centric approach may help to explore and characterize novel aberrant gene expression patterns for B-ALL patients, thereby complementing previous research findings aimed at deciphering the pathogenesis of B-ALL. Moreover, identified alternatively spliced transcripts may help better understand the molecular basis of post-transcriptional gene regulation in the context of B-ALL. By inferring a role for DNA methylation in the expression of alternatively spliced isoforms, new avenues might be explored for improved diagnosis, management and treatment of B-ALL patients in the future.

Chapter 1 Literature Review

1.1 B-Cell Acute Lymphoblastic Leukemia

1.1.1 Characteristics of B-ALL

B-cell lymphoblastic leukemia (B-ALL) is a malignant neoplasm derived from B-cell progenitors. B-ALL is common among children, with peak prevalence between the age of 2 and 5 (Pui, Robison, & Look, 2008). The symptoms of B-ALL include fatigue and paleness from anemia, bruising due to thrombocytopenia, and frequent infection caused by neutropenia (Hunger & Mullighan, 2015). Outcome for pediatric cases with B-ALL has significantly improved over the last 2 decades; the 5-year survival rate is greater than 80%. In adults with B-ALL, existing treatments have been less effective, with a disease related mortality of approximately 60% (Redaelli, Laskin, Stephens, Botteman, & Pashos, 2005).

The precise pathogenic events leading to the development of B-ALL are still undetermined. Less than 5% of the cases are associated with inherited, predisposing genetic syndromes, such as Down's syndrome, Bloom's syndrome, ataxia-telangiectasia, and Nijmegen breakage syndrome (Pui et al., 2008). Common genetic events leading to the development of B-ALL include chromosomal translocation, hyperdiploidy and deregulation of proto-oncogenes (Mullighan, 2012). Due to recently developed next-generation sequencing (NGS) technologies, such as transcriptome sequencing, and whole-genome sequencing, the number of genetic alterations identified in B-ALL patients has increased excessively (Roberts & Mullighan, 2015). However, in

experimental models, commonly occurring genetic aberrations do not alone induce leukemia, pointing that additional genetic or epigenetic changes are required. Identification of these additional genetic and epigenetic alterations is crucial for better understanding B-ALL pathogenesis and development.

1.1.2 Abnormal B-Cell Development in Leukemogenesis

B cells are derived from pluripotent hematopoietic stem cells (HSCs) in the bone marrow through sequential stages of cell differentiation, including lymphoid multipotent progenitors (LMPPs), common lymphoid progenitors (CLPs), early pro-B cells, pro-B cells, pre-B cells, and mature B cells (Figure 1). Knowledge of the normal sequence of antigen acquisition is crucial, because B-ALL arises from B-cell progenitors that reflect arrested stages of B-cell maturation. CLPs are characterized by the presence of the cell surface antigens CD34 and CD10. During the transition from CLP to early pro-B cells, CD10 is lost and CD19 is gained; CD34, CD10 and CD19 are positive in pro-B cells and pre-B cells express only CD10 and CD19. In the final transition to immature B-cells, lymphoblasts begin to express CD20 and IgM in addition to CD10 and CD19 markers (Zhou, You, Young, Lin, Lu, Medeiros, & Bueso-Ramos, 2012).

Transcriptional factor E2A triggers early B-lineage development through regulating the downstream transcription factors EBF1 and PAX5. Both EBF1 and PAX5 are critical for maintaining B-lineage maturation, as abscission of PAX5 and reduced *EBF1* expression result in de-differentiation to immature progenitor cells (Pongubala et al., 2008). Enforced expression of CEBPA, a transcription factor crucial for myeloid

development, in progenitor B cells, inhibits B-lineage-specific genes and conversion into macrophages *in vitro* (Bussmann et al., 2009). Dysregulation of some of these transcription factors (TFs) in B-ALL has been long known because their encoding genes are involved in cytogenetic abnormalities, but the broad disruption of B-cell development in more than 40% of B-ALL cases has only been recognized recently by genome-wide genetic analysis (Mullighan et al., 2007).

Early B-lineage development also depends on signal transduction initiated by the interleukin (IL)-7 receptor in pro-B cells and the pre-B-cell receptor (pre-BCR) in pre-B cells. The IL-7 receptor consists of common γ-chain and an IL-7Rα subunit (encoded by *IL7R* gene), while pre-BCR consists of 2 Igμ chains and 2 surrogate light chains. Effects of IL-7R activation are mediated through the JAK-STAT5 pathway (Hennighausen & Robinson, 2008) and in context of this signaling network transcriptional factor STAT5 upregulates *EBF1* and *PAX5* expression (Dias, Silva, Cumano, & Vieira, 2005; Hirokawa, Sato, Kato, & Kudo, 2003) which results in maintaining of pro-B-cell state. When the pro-B-cell stage has been established, B-cell progenitors undergo rearrangement in heavy chain immunoglobulin IgH. After a successful IgH rearrangement, IL-7R acts in combination with other factors, including pre-BCR, to promote expansion of early pre-B cells through an ERK/MAPK-dependent pathway (Fleming & Paige, 2001). Disruption in the pre-BCR component Igμ leads to a complete B-cell developmental block at the pro-B-cell to pre-B-cell transition (Kitamura, Roes, Kuhn, & Rajewsky, 1991). Pre-BCR signaling also activates a negative feedback loop

through suppressing IL-7Rα expression and attenuating STAT5 activation (Marshall, Fleming, Wu, & Paige, 1998). Dysregulation of the signal transduction cascade can be directly oncogenic and likely contributes to poor clinical outcome.

1.1.3 Genetic Alterations in B-ALL

Multiple genetic alterations have been discovered in B-ALL patients and used for risk classification and treatment assignment. Chromosome translocations, such as E2A-PBX1, TEL-AML1 and BCR-ABL1 occur in approximately 80% of children and 60% to 70% of adults with B-ALL. These chromosomal abnormalities can be detected by routine cytogenetic analysis and interphase fluorescence in situ hybridization (FISH). Smaller genetic aberrations, such as *IKZF1*, *PAX5* and *CDKN2A/B* deletions can be determined by polymerase chain reaction (PCR). Combined with high-throughput DNA sequencing and gene expression profiling, genome-wide studies of B-ALL have uncovered remarkable associations between B-ALL and disruptions of B-cell development, loss of tumor suppressor activity, and aberrant signal transduction (Zhang, Mullighan, Harvey, Wu, Chen, Edmonson, & Hunger, 2011; Zhou et al., 2012).

E2A translocations

E2A is a basic helix-loop-helix transcription factor located on chromosome 19p13. E2A is necessary for initiation of B-cell development and is crucial for B-cell differentiation (LeBrun, 2003). The most common translocation involving the E2A gene is t(1;19)(q23;p13). This genetic abnormality appears in approximately 5% of B-ALL cases and is more prevalent among children. The resulting fusion protein consists of

transactivation domains of E2A and the DNA-binding homeodomain of PBX1 (Hunger, 1996). The oncogenic effect of E2A-PBX1 chimeric protein is a result of the upregulation of the *BMI1* gene (Smith et al., 2003), a transcriptional repressor that participates in hematopoietic stem-cell self-renewal (Park et al., 2003). A second E2A associated translocation, t(17;19), occurs rarely among children. This variant consists of transactivation domains of E2A and the leucine zipper dimerization domain of HLF. The aberrant upregulation of *LMO2* and *BCL2* results from the activation of the E2A-HLF fusion protein (De Boer et al., 2011; Hirose et al., 2010). With modern chemotherapy, patients with B-ALL associated with the E2A-PBX1 translocation have a favorable outcome, but B-ALL cases associated with t(17;19) have a poor prognosis (Hu et al., 2016).

BCR-ABL1 (Philadelphia chromosome)

The tyrosine kinase BCR-ABL chimeric protein is the product of the Philadelphia chromosome, which is formed due to the reciprocal translocation t(9;22)(q34;q11) that opposes the *ABL* oncogene 1 on chromosome 9 with the *BCR* gene on chromosome 22 generating the BCR-ABL1 fusion gene (López-Andrade et al., 2015). This protein has constitutive ABL1 kinase activity and localizes in the cell nucleus. It has been shown that BCR-ABL1 alone is sufficient to induce cancerous transformation in pre-B cells in a mouse model and that this process requires the activation of SRC kinase (Huettner, Zhang, Van Etten, & Tenen, 2000). This translocation rarely occurs in children but it is the most common (approximately 25%) cytogenetic abnormality in adults (Moorman,

2016). Depending on location of the breakpoint in the *BCR* gene, BCR-ABL fusion proteins of different molecular weights can be formed. BCR-ABL p210 can be seen in 24% to 50% of adult Philadelphia positive (Ph^+) B-ALL. A shorter form, p190, predominates in pediatric Ph^+ B-ALL and 50% to 76% of adult Ph^+ B-ALL. BCR-ABL p230 usually is not observed in B-ALL. Comparisons of adult Ph^+ B-ALL patients with p210 or p190 variants showed consistency in the presence of additional cytogenetic abnormalities, white blood cell (WBC) count or outcome (Rieder, Banta, Köhrer, McCaffery, & Emr, 1996). B-ALL associated with BCR-ABL1 shows a common immunophenotype, that being $CD34^+$, $CD10^+$, and $CD19^+$; myeloid markers are positive in up to 71% of cases in adults. B-ALL associated with BCR-ABL1 has a very poor outcome with a 5-year overall survival of less than 10% (Moorman et al., 2010).

Mixed lineage leukemia rearrangements (MLL)

The mixed lineage leukemia (MLL) gene is involved in a wide range of leukemia-associated translocations (Meyer et al., 2009). The most common chromosomal rearrangement involving MLL in B-ALL is t(4;11)(q21;q23), which results in an MLL-AF4 fusion gene. This particular translocation is associated with very poor prognosis for infants under 1 year, the vast majority of whom have a relapse and die of progressive disease. However, for children 1-9 years old or those 10 years of age or older t(4;11)(q21;q23) is correlated with more favorable prognosis (Pui et al., 2003). *MLL* gene rearrangements have been diagnosed in approximately two thirds of infantile ALL cases, and MLL-AF4 consists of more than 50% of the rearrangements (Pieters et al., 2007). In

adults, MLL-AF4 occurs in 4% to 8% of ALL in general (Moorman et al., 2010; Wetzler et al., 1999), but it is more frequent (24%) in patients who have received chemotherapy for other malignancies (Tang, Neufeld, Rubin, & Müller, 2001). Pro-B ALL with t(4;11)/MLL rearrangements is most often myeloid antigen-positive disease (including expression of CD15) (Chiaretti, Zini, & Bassan, 2014). Patients with B-ALL associated with MLL-AF4 have a high risk of relapse.

ETV6-RUNX1 (TEL-AML1)

ETV6, located on chromosome 12p13 previously known as TEL is an ETS family transcriptional repressor and is frequently rearranged or fused with other genes in human leukemias of myeloid or lymphoid origins (Zhang et al., 2015). RUNX1, located on chromosome 21q22 and previously known as AML1, is a transcription factor that participates in hematopoietic development at an early embryonic stage as well as B-cell differentiation in adult hematopoiesis (Ichikawa et al., 2004) results in the ETV6-RUNX1fusion protein, t(12;21)(p13;q22), consists of the N-terminal non-DNA-binding region of ETV6 combined with RUNX1. Enforced expression of ETV6-RUNX1 in HSCs results in expansion of multipotent progenitors and partial arrest of B-cell development at the pro-B cell stage (Tsuzuki, Seto, Greaves, & Enver, 2004). ETV6-RUNX1 is the most frequent alteration in pediatric B-ALL, present in approximately 30% of cases, but is rare in adults (Raynaud et al., 1996). Secondary genetic abnormalities including loss of the ETV6 allele and other genes in the B-cell development pathway are frequently identified at the time of diagnosis of B-ALL (Hong et al., 2008; Mullighan et al., 2009). B-ALL

associated with t(12;21) is usually positive for CD10, CD19, CD34, and the myeloid associated antigen CD13. Patients with B-ALL associated with ETV6-RUNX1 have a highly positive outcome.

Immunoglobulin heavy-chain locus (IGH@)

Recurrent translocations of the IGH@ locus in B-ALL are relatively rare but have been well documented (Dyer et al., 2010). Fusions of IGH@ with each of the 5 members of the CEBP family have been reported in B-ALL in children and adults (Akasaka et al., 2007). The fusion with CEBPD, as a result of t(8;14)(q11;q32), is the most common (Lundin, Heldrup, Ahlgren, Olofsson, & Johansson, 2009). This translocation occurs mostly in children, either as a sole acquired abnormality or in conjunction with t(9;22) or Down syndrome. Partners of IGH@ translocation also include ID4 (Russell et al., 2008), erythropoietin receptor (Russell et al., 2009), CRLF2, IL3 (Grimaldi & Meeker, 1989), and miRNA-125-b-1 (Sonoki, Iwanaga, Mitsuya, & Asou, 2005). The IGH-IL3 translocation, t(5;14)(q31;q32), commonly results in eosinophilia. The IGH-MYC rearrangement, t(8;14)(q24;q32), and IGH-BCL2 translocation, t(14;18)(q32;q21) were identified in 7% and 4% of adult patients with B-ALL, respectively. Patients with B-ALL associated with t(8;14)(q24;q32) or t(14;18)(q32;q21) have a very poor outcome (Moorman et al., 2010).

Numerical chromosomal abnormalities

Several chromosome abnormalities have been identified in B-ALL, including hyperdiploidy, hypodiploidy, near-haploidy and complex karyotypes. Hyperdiploidy

occurs predominantly in pediatric B-ALL, accounting for nearly 40% of cases, and is associated with favorable prognosis. Hypodiploidy, near-haploidy, and complex karyotypes are rare in childhood B-ALL, but their frequency increases with age. Together, these abnormalities account for approximately 15% of B-ALL cases in patients older than 60 years. Hypodiploidy, near-haploidy, and a complex karyotype are associated with poor outcome, with less than 20% of patients surviving for 5 years (Moorman et al., 2010).

Intrachromosomal amplification of chromosome 21

Intrachromosomal amplification of chromosome 21 (iAMP21) is defined as the presence of 3 or more copies of the *RUNX1* gene (Harrison, 2011). The 5.1-Mb common region of amplification contains *RUNX1*, *mIR-802*, and genes in the Down syndrome critical region. iAMP21 occurs in approximately 2% of childhood B-ALL, and these malignancies have a common/pre-B immunophenotype (Harewood et al., 2003). B-ALL with iAMP21 occurs with high frequency in B-ALL associated with Down syndrome. Other genetic alterations associated with iAMP21 include deletion of *RB1*, *CDKN2A*, *IKZF1*, and *PAX5* (Rand et al., 2011). Patients with iAMP21 have relatively poor prognosis if not treated with enhanced chemotherapy (Moorman et al., 2007).

IKZF1 deletion

IKZF1, located at 7p13-p11.1, encodes IKAROS, a zinc-finger containing DNA-binding protein. IKAROS isoforms lacking N-terminal zinc-finger domains have

abnormal localization and function as a dominant negative of wild-type IKAROS. Genome-wide single nucleotide polymorphism array analysis has shown that *IKZF1* deletions are among the most common genetic lesions in high-risk B-ALL, present in 75% to 90% of BCR-ABL1^{+} B-ALL (Mullighan et al., 2008) and 29% of pediatric high-risk BCR-ABL1 B-ALL (Mullighan et al., 2009). Deletions of *IKZF1* are predominantly monoallelic and are limited to the gene in approximately 40% cases (Mullighan et al., 2008). Various patterns of deletions occur, but the most frequent deletions involve the N-terminal zinc-finger domain of IKAROS and result in expression of dominant-negative isoforms with cytoplasmic localization and oncogenic activity (Iacobucci et al., 2012). IKZF1 deletion in B-ALL is associated with a high risk of relapse.

PAX5 deletion and translocation

PAX5 encodes a B-lineage specific transcription factor located at chromosome 9p13. PAX5 is among the most frequent targets of genetic alterations in B-ALL, observed in approximately 30% of cases (Dang et al., 2015). There are several genetic aberrations associated with *PAX5* gene, including monoallelic deletions, translocations and point mutations. Deletions are frequently associated with BCR-ABL1, E2A-PBX1, and complex karyotype with secondary genetic changes (Coyaud et al., 2010). PAX5 rearrangements are relatively rare, occurring in 2.5% of B-ALL cases; at least 12 different fusion partners including TFs, structural proteins, and protein kinases have been reported (Nebral et al., 2009). Deletion and mutation of other genes essential in B-cell

development, including *EBF1*, *RAG1*, *RAG2*, *LEF1*, and *BLINK*, are also frequently detected in B-ALL (Mullighan et al., 2007).

CDKN2A/B deletion

CDKN2A and adjacent *CDKN2B* on chromosome 9p21 are tumor suppressor genes that encode p16INK4a/p14ARF and p15INK4b, respectively. The proteins are involved in controlling G1/S cell-cycle progression. In B-ALL, deletion of *CDKN2A/B* is the most frequent genetic abnormality detected by genome-wide copy number alteration and loss of heterozygosity analysis. These deletions are present in 21% to 36% pediatric B-ALL (Mullighan et al., 2008; Kawamata et al., 2008), and nearly 50% of adult and adolescent B-ALL (Paulsson et al., 2008). *CDKN2A/B* deletions are frequently associated with BCR-ABL1 and E2A-PBX1 fusion, and are less frequently present in B-ALL associated with ETV6-RUNX1, MLL translocation, or hyperdiploidy (Sulong et al., 2009). *CDKN2A/B* deletion can be detected at initial diagnosis or acquired at relapse; there is no difference in frequency between diagnosis and relapse, suggesting that *CDKN2A/B* deletion is a secondary genetic event.

Janus kinase mutations

JAK is a protein tyrosine kinase and a key player in the JAK-STAT pathway. Mutations in *JAK1* and *JAK2* were initially identified in B-ALL associated with Down syndrome (Bercovich et al., 2008; Kearney et al., 2009). Heterozygous somatic mutations of JAKs are seen in approximately 10% of non-Down syndrome B-ALL (Mullighan et al., 2009). JAK mutations occur in highly conserved residues in the kinase and

pseudokinase domain and result in constitutive kinase activation. It appears that aberrant kinase signaling requires interaction with a cytokine receptor, because ectopic expression of ALL-associated JAK1 mutant alone fails to trigger STAT activation in the absence of a γ-chain containing cytokine receptor (Hornakova et al., 2009). In fact, JAK mutation is highly associated with aberrant cytokine receptor expression in B-ALL. Moreover, 70% of B-ALL cases carrying a *JAK* mutation have concomitant deletion of *IKZF1* and/or *CDKN2A/B*. Patients with B-ALL associated with JAK mutation tend to have poor outcome.

1.1.4 Epigenetic alterations in B-ALL

Aberrant microRNA expression

MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression at a posttranscriptional level and are involved in many biological processes, such as cell proliferation and apoptosis. It has been shown that alterations in miRNA levels due to genetic changes may be involved in leukemogenesis. For example, miRNA-125b1 (also known as miR-125-1) was the first miRNA documented in B-ALL (Chapiro et al., 2010; Sonoki et al., 2005). The gene encoding miR-125b1 is located at chromosome 11q24.1 but is inserted into rearranged IGH@ at chromosome 14q32 in rare patients with B-ALL. The translocation causes overexpression of miR-125b1. MiR-125b1 is a negative regulator of p53 (Le et al., 2009). The expression of miRNA are also characteristically associated with genetic types of pediatric B-ALL and predict for clinical outcome

(Schotte et al., 2011). Discovery of novel miRNAs in B-ALL is still in progress, and the clinical and biological significance of these miRNAs needs to be clarified.

DNA methylation

Our laboratory has extensively studied DNA methylation patterns in lymphoid malignancies. Taylor and colleagues (2007) identified 262 unique methylated CpG island (CGI) loci in ALL lymphoblasts utilizing CGI microarray technology. By examining the relationship between methylation and expression for 10 genes (*DCC*, *DLC-1*, *DDX51*, *KCNK2*, *LRP1B*, *NKX6-1*, *NOPE*, *PCDHGA12*, *RPIB9*, *ABCB1*, and *SLC2A14*) cell culture treatments were conducted with 5-aza-2-deoxycytidine and trichostatin A followed by subsequent reverse transcription polymerase chain reaction (RT-PCR) analysis. More than a 10 fold increase in mRNA expression was observed for two previously identified tumor suppressor genes (*DLC-1* and *DCC*) and also for *RPIB9* and *PCDHGA12* genes after treating cells with demethylation agents. Bisulfite sequencing of the promoter of *RPIB9* indicated that expression might be inhibited by methylation within SP1 and AP2 transcription factor binding motifs (Taylor et al., 2007). This study was expanded by Burmeister and colleagues (2015) by investigating methylation status of six regions spanning the CpG island in the promoter region of *RUNDC3B* in cancer cell lines. Lymphoid malignancies were found to have higher methylation level and did not express *RUNDC3B* compared with myeloid malignancies and solid tumors, supporting the potential use of DNA methylation in this region as a biomarker for lymphoid malignancies (Burmeister et al., 2017).

To elucidate the role of DNA methylation during B-cell development, genome-wide DNA methylation analysis was also performed in our laboratory. The DNA methylation status of pro-B, pre-BI, pre-BII, and naïve-B-cells was determined using the methylated CpG island recovery assay followed by NGS. 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 (Almamun et al., 2014). Furthermore, integrated methylome and transcriptome analysis was conducted to determine novel regulatory elements for pediatric B-ALL patients. 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 differentially methylated regions (DMRs). 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 (Almamun et al., 2015). Finally, the impact of aberrant intergenic DNA methylation on gene expression was investigated in B-ALL patients. 84% of differentially methylated intergenic loci, determined for B-ALL patients, were also bound by TFs known to play roles in differentiation and B-cell development in a lymphoblastoid cell line. Further, an overall downregulation of enhancer RNA (eRNA) transcripts was observed in pre-B ALL

patients and these transcripts were associated with the downregulation of putative target genes involved in B-cell migration, proliferation, and apoptosis. The identification of novel putative regulatory regions highlights the significance of intergenic DNA sequences and may contribute to the identification of new therapeutic targets for the treatment of B-ALL patients in the future (Almamun et al., 2017).

Other research groups have also investigated a role of DNA methylation in B-ALL pathogenesis. Examining DNA methylation patterns in 69 pediatric B-ALL and 42 control samples Chatterton and colleagues report 325 genes that were hypermethylated and down regulated, and 45 genes that were hypomethylated and upregulated across all B-ALL samples, regardless of subtype (Chatterton et al., 2012). Furthermore, functional annotation of these epigenetically deregulated genes underlined the role of genes involved in cell signaling, cellular development, cell survival and apoptosis. Another study investigating 764 cases of newly diagnosed ALL and 27 cases of relapse, determined 9406 hypermethylated CpG sites with each cytogenetic subtype portraying a unique set of hyper- and hypomethylated sites (Nordlund et al., 2013). These differentially hypermethylated CpG sites were enriched for genes such as *NANOG*, *OCT4*, *SOX2*, and *REST*. MLL-rearranged infant leukemia is one specific ALL subtype that has been shown to display distinct promoter hypermethylation (Schafer et al., 2010). Stumpel and colleagues identified a distinct DNA methylation pattern dependent on the presence and type of MLL-fusion partner in a cohort of 57 newly diagnosed infant ALL patients (Stumpel et al., 2009). In addition, the level of hypermethylation appeared to

correlate with a higher risk of relapse among infants carrying t(4;11) or t(11;19) translocations. In another study of 5 MLL-rearranged infant ALL samples, genes involved in oncogenesis and tumor progression (*DAPK1*, *CCR6*, *HRK*, *LIFR*, and *FHIT*) were differentially methylated suggesting a role in the leukemogenesis of MLL-rearranged ALL (Schafer et al., 2010).

Histone modification

Mutations in epigenetic modifying genes can result in a gain or loss of function of key genes known to regulate histone marks. Jaffe and colleagues have used global chromatin profiling and mass spectrometry to measure levels of histone modifications on bulk chromatin in pediatric ALL cell lines (Jaffe et al., 2013). A novel cluster of cell lines with a specific epigenetic signature was determined and increased dimethylation of histone H3 at lysine 36 (H3K36me2) and decreased unmodified H3K36 have been observed. Approximately half of the cell lines in this cluster harbored the t(4;14) translocation, which can contribute to *NSD2* overexpression (Malgeri et al., 2000). NSD2 is a member of the HKMTs that catalyze the conversion of unmodified H3K36 to mono- and dimethylated states (Kuo et al., 2011). *NSD2* mutations were found to be enriched in ETV6-RUNX1 and TCF3-PBX1 sub-types of pediatric B-ALL, while no mutations were identified in 30 adult ALL samples. These were gain-of function mutations and their overexpression led to a global increase in H3K36me2, with subsequent decrease in H3K27me3. These results show that NSD2 mutation may affect expression of a number of genes involved in normal lymphoid development.

In order to identify novel mutations in relapsed ALL, Mullighan and colleagues performed targeted resequencing of 300 genes in 23 matched relapse-diagnosis B-ALL pairs (Mullighan et al., 2011).The authors determined novel mutations in *CREBBP*, a gene encoding the transcriptional coactivator CREB binding protein with histone acetyltransferase activity. The overall frequencies of these mutations were 18.3% in relapse cases. However, high incidences of somatic *CREBBP* alterations (63%) were found in the high hyperdiploidy relapse cases. The majority of these mutations occurred in the HAT domain (Inthal et al., 2012). Mutations in other important epigenetic regulators such as *NCOR1* (nuclear corepressor complex), *EP300* (a paralog of CREBBP), *EZH2* (histone methyltransferase gene), and *CTCF* (zinc finger protein involved in histone modifications) were less frequently observed (Mullighan et al., 2011). Additionally, transcriptome sequencing has identified relapse-specific mutations in *CBX3* (encoding heterochromatin protein), *PRMT2* (gene encoding protein arginine methyltransferase 2), and *MIER3* (involved in chromatin binding); providing further evidence of aberrant epigenetic mechanisms that play a role at relapse (Meyer et al., 2013).

1.2 Alternative Splicing in B-ALL

1.2.1 Characteristics of Alternative Splicing Events in Cancer

Alternative splicing generates numerous protein isoforms through modifying mRNA precursors. This mechanism is highly regulated under normal conditions in order to generate proteomic diversity sufficient for the functional requirements of complex

tissues. While corrupted, cancer cells take advantage of this mechanism to generate abnormal proteins with added, deleted, or altered functional domains that contribute to carcinogenesis (Zhang & Manley, 2013). Cancer-specific alternative splicing includes all of the five main alternative splicing patterns observed in normal tissues: cassette exons, alternative 5′ splice sites, alternative 3′ splice sites, intron retention, and mutually exclusive exons. The most prevalent pattern is the cassette-type alternative exon, including skipping of one exon, skipping of multiple exons and/or exon inclusion. This alteration results into truncated RNA transcript that may not be translated into functional protein. Additionally, crucial protein domains may be excluded from protein structure that will lead to the inability to interact with variety of protein partners and the deregulation of signaling pathways. Alternative selection of 5′ or 3′ splice sites within exon sequences may lead to subtle changes in the coding sequence, and an additional layer of complexity arises with mutually exclusive alternative exons (Wang et al., 2015). Both mechanisms may lead to alteration of amino acid composition of the protein and an inability to perform its original function. Intron retention is positioned primarily in the untranslated regions (UTRs) (Galante, Sakabe, Kirschbaum-Slager, & de Souza, 2004) and has been associated with weaker splice sites, short intron length and the regulation of cis-regulatory elements (Sakabe & de Souza, 2007). Complex splicing patterns may affect gene expression as well and contribute to the diversity of protein isoforms. Specific examples for each of these alterations are described in Table 1.

1.2.2 Alternative splicing isoforms in B-ALL

There are several studies that investigated alternatively spliced (AS) transcripts in B-ALL. A transcript variant of Beclin 1 gene carrying a deletion of exon 11 has been discovered in human B-cell acute lymphoblastic leukemia cells (Niu et al., 2014). The alternative isoform was assessed by bioinformatics, immunoblotting and subcellular localization. The results showed that this variable transcript is generated by alternative 3' splicing, and its translational product displayed a reduced activity in induction of autophagy by starvation, indicating that the spliced isoform might function as a dominant negative modulator of autophagy and might play important roles in leukemogenesis.

In another study, expression levels of IKAROS have been measured in human bone marrow samples from patients with adult acute lymphoblastic leukemia (Nakase et al., 2000). Overexpression of the dominant negative isoform of IKAROS gene IK-6 was observed in 14 of 41 B-cell ALL patients by RT-PCR, and the results were confirmed by sequencing analysis and immunoblotting. Southern blotting analysis with *PstI* digestion revealed that those patients with the dominant negative isoform IK-6 might have small mutations in the IKAROS locus that may contribute to B-ALL through the dominant negative isoform IK-6.

Different AS variants of activation-induced cytidine deaminase (AID) gene have been identified among 61 adult BCR-ABL1^{+} ALL patients (Iacobucci et al., 2010). AID expression was detected in 36 patients (59%); it correlated with the BCR-ABL1 transcript levels and disappeared after treatment with tyrosine kinase inhibitors. Different AID

splice variants were identified: full-length isoform; AIDΔE4a, with a 30-bp deletion of exon 4; AIDΔE4, with exon 4 deletion; AIDins3, with the retention of intron 3; AIDΔE3-E4 isoform without deaminase activity. AID expression correlated with a higher number of copy number alterations identified in genome-wide analysis using a single-nucleotide polymorphism array. However, the expression of AID at diagnosis was not associated with a worse prognosis.

Alternative PAX5 splicing was observed in 49 out of 100 ALL patients, which comprises 62% of adult and 36% of pediatric ALL cases (Santoro et al., 2009). Different isoforms were detected: PAX5D2 was found in 29 patients, PAX5D8–9 in 14 patients; the novel PAX5D5 isoform was documented in six patients. These results suggests that that altered PAX5 isoform expression may be involved in ALL pathogenesis.

1.3 Rationale for Thesis

To extend the integrated methylome and transcriptome analysis for B-ALL patients reported by Almamun and colleagues (2015), sixteen RNA-seq samples (eight B-ALL patients and eight healthy donors) have been analyzed with the edgeR package for the purpose of obtaining a set of statistically significant transcripts that are differentially expressed between these conditions. Some of the patient samples were excluded from analysis due to high proportion of reads aligned to 5'UTR region (around 80%) reducing the patient sample number to 8. Therefore, this analysis utilized an equal number of patient and control samples improving statistical robustness and providing increased power in determining the differences in variances and means for DE genes.

The Bioconductor package edgeR was utilized to identify DE transcripts due to its advantages over Cuffdiff. For example, edgeR normalizes RNA-seq data according to library size (trimmed mean of M-values, TMM method), while Cuffdiff software normalizes data according to previously annotated genes and their gene coordinates (fragments per kilobase of transcript per million mapped reads, FPKM method). In addition, edgeR and Cuffdiff differ in the calculation of mean and variance of gene expression values. The negative binomial model, implemented into Cuffdiff, assumes that there is no relationship between mean and variance of gene expression values in experimental and control groups. Contrarily, the edgeR algorithm "borrows" information about variances across multiple genes that undergo statistical testing, making this model more robust in determining a set of DE transcripts. Moreover, edgeR implements several modalities to perform statistical test depending on experimental design: the classic edgeR model utilizes Fisher's exact test for pairwaise comparisons, while the generalized linear model (GLM) is more suitable for multigroup experiments. Further, edgeR comprises of wide range of graphic functions that allow the researcher to visualize and plot RNA-seq data in addition to performing statistical tests, such as multidimensional scaling plots (MDS plot) or volcano plots. Finally, R code can be utilized to modify the functions in the edgeR package according to experimental demands. In sum, edgeR analysis has multiple advantages in comparison to Cuffdiff analysis and provides a superior analysis of transcriptome data. Currently, it is one of the best methodologies for RNA-Seq data analysis along with DESeq analysis.

It has been previously shown that alternative splicing is a hallmark of a variety of malignancies, including both solid and soft tissues cancers (Table 1). Prior to NGS, transcriptome-wide analysis of AS genes was limited due to the inability to generate primers (or hybridization probes) for regions, where novel alternative transcripts may be located. Currently, RNA-seq technology allows one to investigate not only differentially expressed (DE) genes across multiple groups, but also provides information about disease-specific gene isoforms. To identify a set of differentially spliced variants common across B-ALL patients, a custom Perl script was designed. This information may shed light on the functional implication of AS isoforms that may be involved in B-ALL pathogenesis.

Finally, to explore the potential role of DNA methylation in transcriptional regulation, an *in vitro* model for B-ALL – Nalm 6 cell line – was utilized. Prior to this study, Taylor and colleagues (2007) examined the relationship between methylation and expression for 10 genes using CpG island microarrays and observed more than a 10 fold increase in mRNA expression for two tumor suppressor genes (*DLC-1* and *DCC*) and also for *RPIB9* and *PCDHGA12* genes. In this study, the Nalm 6 cell line was treated with a demethylating agent followed by NGS analysis. Although *in vitro* models may not reflect the whole complexity of patient transcriptomes, they provide a means to explore potential functional mechanisms responsible for the aberrant transcript expression identified in our computational analysis.

1.4 Experimental Aims and Hypothesis

We hypothesize that DE genes, identified between B-ALL patients and healthy donors, are involved in the development and progression of this malignancy. Further, we hypothesize that leukemic cells will have unique splicing alterations that result in abnormal transcripts which promote the survival and uncontrolled proliferation of malignant cells. Our goal is to conduct genome-wide transcriptome analysis to identify a set of differentially expressed and spliced genes between B-ALL patients and healthy donors and to investigate the functional implications of these alterations using network-based analysis. In addition, a mechanistic study utilizing the Nalm 6 cell line was performed to explore if methylation influences alternative splicing of transcripts in B-ALL. To address our hypotheses the following project objectives were completed:

1. Perform edgeR analysis between B-ALL and healthy donor samples to determine a set of statistically significant DE genes.
2. Utilize Ingenuity Knowledge Base (KB) to annotate functions and enrichment of signaling pathways for DE genes.
3. Identify novel transcriptional regulators that control aberrant expression of genes involved in the development of B-ALL using Ingenuity pathway analysis (IPA).
4. Determine a set of common splicing isoforms for B-ALL patients using custom Perl script.
5. Perform a mechanistic study in the Nalm 6 cell line to explore the impact of DNA methylation upon common AS isoforms.

The complexity of a disease such as B-ALL provides many difficulties to determining diagnosis, prognosis, and appropriate treatment. To date a number of genetic abnormalities have been identified that contribute to B-ALL development but there are still many to be characterized. Therefore, a complete transcriptome analysis to identify DE genes is very important for better understanding B-ALL pathobiology. This research provides a characterization of aberrant gene expression patterns in B-ALL at the whole transcriptome scale in an attempt to improve diagnosis, prognostication and treatment of B-ALL patients in the future.

Chapter 2 RNA-Sequencing Analysis in B-cell Acute Lymphoblastic Leukemia Reveals Aberrant Gene Expression and Splicing Alterations

Abstract

Background: B-cell acute lymphoblastic leukemia (B-ALL) is a neoplasm of immature lymphoid progenitors and is the leading cause of cancer-related death in children. The majority of B-ALL cases are characterized by recurring structural chromosomal rearrangements that are crucial for triggering leukemogenesis, but do not explain all incidences of disease. Therefore, other molecular mechanisms, such as alternative splicing and epigenetic regulation may alter expression of transcripts that are associated with the development of B-ALL. To determine differentially expressed and spliced RNA transcripts in precursor B-cell acute lymphoblastic leukemia patients a high throughput RNA-seq analysis was performed.

Methods: Eight B-ALL patients and eight healthy donors were analyzed by RNA-seq analysis. Statistical testing was performed in edgeR. Each annotated gene was mapped to its corresponding gene object in the Ingenuity KB. Analysis of RNA-seq data for splicing alterations in B-ALL patients and healthy donors was performed with custom Perl script.

Results: Using edgeR analysis, 3877 DE genes between B-ALL patients and healthy donors based on TMM (trimmed mean of M-values) normalization method and false discovery rate, FDR < 0.01, logarithmically transformed fold changes, logFC > 2) were identified. IPA revealed abnormal activation of ERBB2, TGFB1 and IL2 transcriptional factors that are crucial for maintaining proliferation and survival potential of leukemic

cells. B-ALL specific isoforms were observed for genes with roles in important canonical signaling pathways, such as oxidative phosphorylation and mitochondrial dysfunction. A mechanistic study with the Nalm 6 cell line revealed that some of these gene isoforms significantly change their expression upon 5-Aza treatment, suggesting that they may be epigenetically regulated in B-ALL.

Conclusion: Our data provide new insights and perspectives on the regulation of the transcriptome in B-ALL. In addition, we identified transcript isoforms and pathways that may play key roles in the pathogenesis of B-ALL. These results further our understanding of the transcriptional regulation associated with B-ALL development and will contribute to the development of novel strategies aimed towards improving diagnosis and managing patients with B-ALL.

Keywords: B-ALL, RNA-sequencing, differential gene expression, alternative splicing

Introduction

B-cell precursor acute lymphoblastic leukemia (B-ALL), a malignant disease of lymphoid progenitor cells, affects both children and adults, with peak prevalence between the ages of 2 and 5 years (Pui et al., 2008). A number of genetic alterations have been determined in B-ALL (Woo, Alberti, & Tirado, 2014); however, a complete understanding of pathogenic mechanisms underlying B-ALL development is still lacking. To identify genetic alterations in B-ALL, a wide range of methods have been applied including cytogenetic analysis (Mrózek, Harper, & Aplan, 2009), array comparative genomic hybridization (Dawson et al., 2011) and recently whole exome sequencing

(Lilljebjörn et al., 2012). The whole exome sequencing of B-ALL samples has also resulted in the identification of novel recurring mutations in *NRAS, KRAS, FLT3, CREBBP, XBP1, WHSC1,* and *UBA2* genes (Griffith et al., 2016; Lilljebjörn et al., 2012).

To study the whole transcriptome of cells, microarrays have been extensively used, and these studies have determined a number of DE genes (Ross et al., 2003). Unfortunately, microarray techniques have a number of limitations including, cross hybridization of transcripts, limitation in coverage, inability to resolve novel transcripts and falsely higher estimation of low abundance transcripts (Pawitan, Michiels, Koscielny, Gusnanto, & Ploner, 2005). With the development of massive parallel RNA-sequencing (RNA-seq) technology, there have been a growing number of genome-wide studies that have analyzed the complete transcriptome of cells in different malignancies (Eswaran et al., 2012), and non-malignant diseases (Twine, Janitz, Wilkins, & Janitz, 2011). Besides analyzing the expression level of genes, RNA-seq technology has the added advantage of analyzing expression at the exon level and provides detailed information about alternative splicing variations, novel transcripts, fusion genes, differential transcriptional start sites and genomic mutations (Wang et al., 2008). As all the RNA transcripts are being directly sequenced, this technology is ideally suited to study altered splicing patterns which is especially relevant in cancer cells (David & Manley, 2010).

In this study we performed RNA-seq analysis on B-ALL patient samples and healthy donor samples to determine transcriptome differences and splicing variations. A number of DE genes and novel isoforms were identified. These findings may facilitate

the identification of novel prognostic markers, therapeutic targets and altered signaling pathways in B-ALL.

Materials and Methods

Sample isolation and characterization

De-identified patient samples were obtained under full ethical approval of the Institutional Review Board at the University of Missouri. A total of 8 pre-B ALL patient samples were used for this study (Table 2). ALL patient samples contain at least 88% blasts. The age of patients varied between 17 month and 15 years. The blast cells were positive for CD19 and CD10 markers. A half of the B-ALL patients have normal karyotype and the rest of the patients have multiple chromosome abnormalities, including deletions, translocations and presence of derivative chromosome. Patient A19 had been identified with hyperdiploid genotype. Normal control pre-BI and pre-BII cells were isolated from 8 human umbilical cord blood samples as previously described (Almamun et al., 2013) and served as the control group. Briefly, mononuclear cells were isolated by density gradient centrifugation using Ficoll-Paque PLUS (GE Healthcare Bio-Sciences AB; cat. no. 17-1440-03) 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; order no. 130-093-660). Finally, the fluorescently labeled cells were sorted as pre-BI ($CD19^{+}/CD34^{-}/CD45^{low}$) and pre-BII ($CD19^{+}/CD34^{-}/CD45^{med}$). 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 (Almamun et al., 2015).

RNA-seq and library preparation

RNA samples were also obtained from the pre-B ALL patients (8 samples) and from normal precursor B-cells isolated from HCB (8 samples). RNA sequencing libraries were constructed with the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England Biolabs; cat. no. E7420) and sequenced on the Illumina HiSeq 2000 (1×100 bp reads) at the University of Missouri DNA Core Facility. All RNA-seq data were deposited in NCBI Sequence Read Archive (Accession SRP058414). (Almamun et al., 2015).

Primary processing and mapping of RNA-seq reads

100 bp single-end RNA-seq reads were obtained from Illumina HiSeq 2000 sequencing platform. Raw data files were generated in FASTQ format and adaptor sequences had been trimmed. RNA-seq data were processed using an in-house pipeline. The Fred quality score of RNA-seq reads was obtained by using the FastX-Toolkit v. 0.0.13 and the mean value for Fred base calling was 32, indicating a good-quality call in the 100 bp reads (Gordon and Hannon, unpublished). Reads were then processed and aligned to the UCSC *H. sapiens* reference genome (build hg19) using TopHat v1.3.3 (Trapnell, Pachter, & Salzberg, 2009).

Assembly of transcripts and differential expression

The aligned read BAM files were assembled into transcripts, their abundance estimated by Cufflinks v2.0.1 (Trapnell et al., 2012). Cufflinks uses the normalized RNA-seq fragment counts to measure the relative abundances of transcripts. The unit of measurement is fragments per kilobase of exon per million fragments mapped (FPKM). Confidence intervals for FPKM estimates were calculated using a Bayesian inference method. After assembly with Cufflinks, the output files were sent to Cuffmerge along with a reference annotation file. To produce count tables for edgeR analysis, HTSeq v0.6.1 software was utilized (Anders, Pul, & Huber, 2014). The count tables represent the total number of reads aligning to each gene (or other genomic locus). To normalize multiple samples for differential expression analysis, we applied calcNormFactors function in edgeR to find a set of scaling factors for the library sizes that minimize the log-fold changes between the samples for most genes. The default method for computing these scale factors uses a trimmed mean of M-values (TMM) between each pair of samples (Robinson & Oshlack, 2010). For cross-replicate dispersion estimation, a quantile-adjusted conditional maximum likelihood (qCML) method was used to calculate the likelihood by conditioning on the total counts for each tag, using pseudo counts after adjusting for library sizes. qCML common dispersion and tagwise dispersions were estimated using the estimateCommonDisp() and estimateTagwiseDisp() functions (Robinson, McCarthy, & Smyth, 2010). The expression testing was done at the level of transcripts and genes and pairwise comparisons of expression between B-ALL and normal samples. Only the comparisons with p-value and FDR less than 0.01 and

expression fold change greater than two fold in the edgeR output were regarded as showing significant differential expression.

Identification of common gene isoforms

To identify common gene isoforms for B-ALL patients, unique identifiers were assigned to each isoform using a custom Perl script. Briefly, after alignment to the reference hg19 human genome, each patient file was processed using the Cufflinks program and individual transcriptomes were assembled into corresponding transcripts.gtf files. Each transcripts.gtf file consists of eight columns: the first seven columns have standard GTF format, and the last column contains attributes. To create a unique identifier for each transcript the following information was extracted from transcripts.gtf files: transcript ID, chromosome number and exon coordinates. Then, intron coordinates were calculated for each transcript ID using a Perl script. Furthermore, chromosome number and intron coordinates were merged into unique identifier (for example: CUFF.59863.1 transcript has unique ID chr7:156629580-156685621:156626487-156629506:156619439-156626446:156589187-156619298). Then, FPKM values were extracted from the same transcripts.gtf files to obtain relative abundance for transcripts with unique IDs. Finally, identified transcripts were annotated with corresponding genes. PerlDBI module and MySQL quarries were utilized to obtain a set of common unique transcripts with corresponding FPKM values. Overall, 338 common transcripts were identified in B-ALL patients. The corresponding FPKM values were extracted with further logarithmic transformation (base 2) and clustered using the R package

ComplexHeatmap (Gu et al., 2016). By agglomerative hierarchical cluster analysis, Euclidian distances have been determined for each pair of transcripts and plotted as a heatmap to visualize transcripts abundances for B-ALL patients.

Cell line treatment experiment

The pre-B ALL cell line Nalm 6 was grown in RPMI 1640 medium (Gibco®, ThermoFisher) supplemented with 10% fetal bovine serum, L-glutamine, and gentamicin. Cell culture treatments were conducted, as described previously with minor alterations (Taylor et al., 2007). Briefly, Nalm 6 cells were seeded at 3 X 10^6 cells/mL. Based on prior practice, 5-Aza was added at either a 0.3 or 0.4 μmol/L final concentration with acetic acid as the vehicle and was incubated for 78 h, with new medium added every 24 h. Control cells were cultured with acetic acid alone. RNA from the cultured cells was extracted for use in NGS, using the AllPrep DNA/RNA Mini Kit (QIAGEN). High quality RNA was submitted to the University of Missouri DNA Core Facility for library generation using the TruSeq mRNA stranded library preparation kit (Illumina). Paired-end sequences (2 X 75) were generated by the University of Missouri DNA Core Facility using the Illumina HiSeq 2500 platform. Sequence files were generated in FASTQ format and processed as described for B-ALL patients and healthy donors.

Functional annotation of differentially expressed genes

QIAGEN's Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, CA www.qiagen.com/ingenuity) is a powerful analysis and search tool that uncovers the significance of omics data and identifies new targets or candidate biomarkers within the

context of biological systems. IPA was used to categorize genes that were differentially expressed between B-ALL patients and healthy donors. The analysis was run using the following setting in IPA: all defaults setting for the selection of dataset, 2 fold change cutoff, FDR = 0.001 and p-value = 0.001.

The functional analysis in IPA identified the biological functions that were most significant to the analyzed dataset. The significance value associated with functional analysis for a dataset is a measure of the likelihood that the association between a set of DE genes in our dataset and a given process or pathway is due to random chance. The smaller the p-value the less likely that the association is random and the more significant the association. In general, p-values less than 0.05 indicate a statistically significant, non-random association. The p-value is calculated using the right-tailed Fisher exact test. In this method, the p-value for a given function is calculated by considering a) the number of DE genes that participate in that function and b) the total number of genes that are known to be associated with that function in the Ingenuity KB. The more DE genes that are involved, the more likely the association is not due to random chance, and thus the more significant the p-value. Similarly, the larger the total number of DE genes known to be associated with the process, the greater the likelihood that an association is due to random chance, and the p-value accordingly becomes less significant. To sum up, the p-value identifies statistically significant over-representation of DE genes in a given process. Over-represented functional or pathway processes are processes which have more focus molecules than expected by chance.

Canonical pathway analysis identified the pathways from the Ingenuity KB that were most significant to the dataset. DE genes from the dataset that were associated with a canonical pathway in the Ingenuity KB were considered for the analysis. The significance of the association between the data set and the canonical pathway was measured in 2 ways: 1) a ratio of the DE genes that mapped to the pathway divided by the total number of genes that mapped to the canonical pathway; 2) an FDR ≤ 0.05 to calculate a p-value determining the probability that the association between the DE genes and the signaling canonical pathway was explained by chance alone. A simple p-value was also considered and reported in the results.

The IPA upstream regulator analysis was also performed. This analysis is based on prior knowledge of expected effects between transcriptional regulators and their target genes stored in the Ingenuity KB. The analysis examines how many known targets of each transcription regulator are present in the provided dataset, and also compares their direction of change to what is expected from the literature in order to predict likely relevant transcriptional regulators. If the observed direction of change is mostly consistent with a particular activation state of the transcriptional regulator ("activated" or "inhibited"), then a prediction is made about that activation state. IPA's definition of upstream transcriptional regulator is quite broad – any molecule that can affect the expression of other molecules, which means that upstream regulators can be almost any type of molecule, from TFs, to miRNAs, kinases, compound or drug.

For each potential transcriptional regulator (TR) two statistical measures, an overlap p-value and an activation z-score are computed. The overlap p-value calls likely upstream regulators based on significant overlap between dataset genes and known targets regulated by a transcriptional regulator. The activation z-score is used to infer likely activation states of upstream regulators based on comparison with a model that assigns random regulation directions. The purpose of the overlap p-value is to identify transcriptional regulators that are able to explain observed gene expression changes. The overlap p-value measures whether there is a statistically significant overlap between the dataset genes and the genes that are regulated by a transcriptional regulator. It is calculated using Fisher's exact test and significance is generally attributed to p-values < 0.01. Since the regulation direction ("activating" or "inhibiting") of an edge is not taken into account for the computation of overlap p-values the underlying network also includes findings without associated directional attributes, such as protein-DNA binding.

Results

Analysis of RNA-seq data

Normal precursor B-cells from 8 healthy donors (HCB11, HCB12, HCB13, HCB15, HCB16, HCB17, HCB18 and HCB19) and malignant precursor B-cells from 8 B-ALL patients (B-ALL18, B-ALL19, B-ALL20, B-ALL23, B-ALL24, B-ALL26, B-ALL30 and B-ALL36) were subjected to RNA single-end RNA-sequencing. The total number of raw reads in healthy (n = 8) and B-ALL (n = 8) samples ranged from 27 to 52 million reads, and 25 to 51 million reads, respectively (Supplemental Tables 1 and 2). To

assess the quality of mapping reads to the reference genome hg19, some key metrics were extracted from the TopHat2 output, and analyzed using the RNA-seq quality control package RseQC (Wang, Wang, & Li, 2012). The majority of reads (between 76 % and 89.5 %) were uniquely mapped to the reference genome sequence across all samples (Supplemental Tables 1 and 2). The mean mapping percentage for healthy donors and B-ALL patients was 88.9 % and 85.8 %. In addition 2.5% to 4.0% of the reads mapped to known splice junctions in healthy donors and B-ALL patients respectively (Supplemental Tables 3 and 4).

To further examine the read distribution, the uniquely mapped reads were assigned to: exon coding sequence (CDS), 5' and 3' untranslated regions (5'UTR and 3'UTR), introns and intergenic regions. In Figure 2, the distribution of mapped reads is shown across the samples. 28.2 % to 55.0 % of reads mapped to exon coding sequence, 3.0 % to 7.1 % mapped to 5'UTR while 9 % to 19.5 % mapped to 3'UTR. The introns and intergenic regions account for about 30.5 % and 10.1 %, respectively (Supplemental Tables 5 and 6). To further visualize the read distribution percentages in healthy donors and B-ALL patients, mapping data from Figure 2 was averaged and plotted as a pie chart (Figure 3). The exonic reads (CDS) were higher in B-ALL patients (~51%) as compared to healthy donors (~31%) while intronic reads were higher in the healthy donors (~43%), compared to B-ALL patients (18%). The high number of reads mapping to introns have been reported in other RNA-seq analysis (Kapranov et al., 2011) and could be due to

novel exons, or nascent transcription and co-transcriptional splicing as described by Ameur and colleagues.

Analysis of differentially expressed genes

To determine the DE genes between B-ALL patients and healthy donors an edgeR analysis was performed. For this purpose we used the "classic" edgeR model that employs Fisher's exact test for identifying DE genes. After filtering DE genes with a FDR < 0.01, p-value < 0.01 and logFC > 2, there were 3877 DE genes between B-ALL patients and healthy donors. Among these genes, 2601 were upregulated in B-ALL and 1276 genes were downregulated. The top twenty upregulated and twenty downregulated genes are listed in Table 3.

Treatment of a pre-B ALL cell line with a demethylating agent reverses expression of alternatively spliced isoforms *in vitro*

Because alternative isoform usage have been shown to be associated with aberrant DNA methylation in cancer (Bujko et al., 2016), a pre-B ALL cell line Nalm 6 was treated with a demethylating agent (5-aza-2'-deoxycytidine, 5-Aza) and RNA-seq was performed. Differential gene expression was calculated between Nalm 6 samples and healthy donor's samples using edgeR package for each of the 338 common transcripts in B-ALL patients identified by custom Perl script (see section "analysis of differentially expressed genes"). Three pairwise comparisons have been examined: B-ALL versus healthy donors, Nalm 6 (untreated) versus healthy donors and Nalm6 (treated with 5-

Aza) versus healthy donors. Nalm 6 cells treated with 5-aza have higher expression values in comparison to untreated Nalm 6 cells, as expected after treatment with demethylating agent. While analyzing expression values for 338 common transcripts, 295 transcripts have been identified in all three pairwise comparisons, 275 transcripts among them met criteria p-value < 0.05 and 78 common transcripts among them have logFC > 2. Interestingly, we identified 19 common transcripts that have shown significant increase in expression after 5-Aza treatment (Table 4). The associated gene ontology terms for these genes are presented in Table 5.

Functional pathway analysis

Several top bio functions were identified by IPA, including cellular growth and proliferation (1.65E-05 - 8.80E-28), cell death and survival (1.34E-05 - 6.55E-21), cellular movement (1.47E-05 - 5.00E-20), cellular development (1.65E-05 - 6.55E-18) and cell cycle (1.63E-05 - 9.22E-13). The cellular growth and proliferation category describes functions associated with cell expansion and propagation, such as proliferation and outgrowth of cells. This category included 1351 genes, including syndecan 2 (*SDC2*), CD2 molecule (*CD2*), MAM domain-containing protein 1 (*MDGA2*) and Wnt Family Member 10A (*WNT10A*). The cellular development category describes functions associated with the development and differentiation of cells, including maturation and senescence of cells. This category consisted of 1164 genes, including neuritin 1 (*NRN1*), kinesin family member 26A (*KIF26A*), intelectin 1 (*ITLN1*) and uroplakin 2 (*UPK2*) genes. The cell death and survival category (represented by 1155 genes including

baculoviral IAP repeat containing 7 (*BIRC7*), Fc fragment of IgG receptor IIIa (*FCGR3A*), calcium/calmodulin dependent protein kinase II alpha (*CAMK2A*) and nephrin (*NPHS1*)) describes functions associated with cellular death and survival, such as cytolysis, necrosis, apoptosis and recovery of cells. The cellular movement category (represented by 812 genes, including prostaglandin D2 receptor (*PTGDR*), semaphorin 3F (*SEMA3F*) and natriuretic peptide B (*NPPB*)) describes functions associated with movement and localization of cells, including chemotaxis, infiltration, rearrangement, and transmigration of cells. These functions were primarily up-regulated among B-ALL patients.

The IPA software reported several significant canonical pathways, including protein kinase A signaling (p-value ≤ 1.55E-06), interferon signaling (p-value ≤ 3.26E-03), cyclins and cell cycle regulation (p-value ≤ 2.20E-03), phospholipase C signaling (p-value ≤ 1.56-E03) and cell cycle control of chromosomal replication (p-value ≤ 4.39E-05). The result from this part of functional analysis is reported in Table 6. In addition, identified common gene isoforms for B-ALL patients associated with oxidative phosphorylation (p-value ≤ 4.58E-13) and mitochondrial dysfunction pathways (p-value ≤ 4.04E-11).

The upstream regulatory analysis performed by IPA predicted regulators based on the consistency of expression direction changes for DE genes within each pathway. The most important regulators identified in this analysis were Erb-B2 receptor tyrosine kinase 2 (ERBB2), transforming growth factor beta 1 (TGFB1) (Figure 5), interleukin-2 (IL2),

tumor protein P53 (TP53) and cyclin dependent kinase inhibitor 1A (CDKN1A). ERBB2, TGFB1 and IL2 were predicted to be activated in B-ALL group. For TP53 and CDKN1A it was not possible to infer their activation or inactivation based upon DE gene set.

Discussion

On average, more than 38 million unique mapped RNA-seq reads were generated providing genome-wide coverage of the transcriptome in eight pediatric B-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 DE genes. Previous studies in B-ALL have shown an inverse correlation between DNA methylation and gene expression in CpG islands and gene promotors (Busche et al., 2013); however more than 80% of DMRs are located in intronic or intergenic regions (Almamun et al, 2015). The novelty of our study is to investigate how DNA methylation affects alternatively expressed and spliced transcripts unique to B-ALL patients. Since DNA methylation can be used as a biomarkers and as a target for novel therapeutics, we sought to identify B-ALL specific alternate transcript candidates that were the most likely to be regulated by DNA methylation.

The edgeR analysis identified DE genes involved in immune regulation and provide survival advantage to cancer cells. For example, a member of the IAP family of apoptosis inhibitors *BIRC7* was top upregulated gene in B-ALL group. This gene had also been overexpressed 25-fold in ETV6-RUNX1 (also known as TEL-AML1*)* leukemia

(Ross et al., 2003). The top downregulated gene in B-ALL group – *CAMK2A* has been identified as distinctive protein kinase gene at ALL1/AF4 subgroup of adult B-cell acute lymphoblastic leukemia patients (Messina et al., 2010). The product of this gene belongs to the serine/threonine protein kinases family and is involved in calcium signaling. Several novel upregulated genes, including *FAM19A5* (chemokine regulation), *PTGDR* (prostaglandin D receptor activity), *GIMAP6* (regulation of cell survival), *FCN1* (antigen binding activities) and *GZMA* (regulator of apoptosis) also involved in regulation of immune system and cell death. Interestingly, *TSHZ3* gene may play role in epigenetic regulation, because TSHZ3-mediated transcription repression involves the recruitment of histone deacetylases HDAC1 and HDAC2. Furthermore, several novel downregulated genes have been annotated with immune response and signal transduction categories, including *ITLN1* (IL-7 signaling pathway regulator), *CD244* (adaptive immune response regulator) and *ORM1* (immunosuppression process). Moreover, downregulation of *TNS4* gene may disrupt the link between signal transduction pathways and cytoskeleton, which results into apoptosis inhibition. Taken together, these genes may contribute to the immune dysfunction of B-cells and disrupt proper differentiation of B-cells.

Many of biological functions reported by IPA are likely related to the malignant phenotype of cancer cells. The top functional category – cellular growth and proliferation had been comprised of 1351 DE gene, which highlight abnormal propagation of leukemic cells. The cell transformation category (Figure 6) involved upregulated genes, such as *CD4* (regulator of N-RAS pathway), *E2F1* (control of cell cycle), *MYB* (proto-oncogene),

RUNX1 (enhancer activity), *VEGFA* (growth factor activity), *AURKA* and *AURKB* (kinase activity) and downregulated genes *HES1* (transcription factor activity) and *IRF4* (regulator of B-cell receptor pathway). Similarly, proliferation of cancer cells (Figure 7) involved upregulated genes, such as *BIRC5* (negative regulator of apoptosis), *CXCL8* (angiogenic factor), *IL1B* (cell differentiation regulator), *NOTCH1* (transcription factor activity) and downregulated *IL6* (regulator of B-cell maturation) and *IFNG* (cytokine activity) genes. In summary, the B-ALL expression profiles included the upregulation of genes involved in cell proliferation and the downregulation of genes involved in B-cell maturation.

The upstream regulatory analysis performed by IPA, which seeks to identify the upstream transcriptional regulatory cascades that are likely to elucidate the observed changes in gene expression may shed some light on the biological activities that occur in leukemic cells. This analysis predicted the top upstream regulators to include TGFB1 which was predicted to be activated in B-ALL group (Figure 5). The transforming growth factor-β (TGF-β) signaling pathway is an essential regulator of cellular processes, including proliferation, differentiation, migration, and cell survival. During hematopoiesis, the TGF-β signaling pathway is a potent negative regulator of proliferation while stimulating differentiation and apoptosis when appropriate. However, in hematologic malignancies, including leukemias, resistance to the homeostatic effects of TGF-β develops. Mechanisms for this resistance include mutation or deletion of

members of the TGF-β signaling pathway and disruption of the pathway by oncoproteins (Dong & Blobe, 2006).

Protein kinase A signaling was the top canonical pathway based on DE genes between B-ALL patients and healthy donors. Protein kinase A (PKA), as cAMP-dependent protein kinase, mediates signal transduction of G-protein coupled receptors through its activation upon cAMP binding. It is involved in the control of a wide variety of cellular processes from metabolism to ion channel activation, cell growth and differentiation, gene expression and apoptosis. Importantly, since it has been implicated in the initiation and progression of many tumors, PKA has been proposed as a novel biomarker for cancer detection, and as a potential molecular target for cancer therapy (Sapio et al., 2014).

The process of generating novel cancer-specific isoforms leads to structural changes in coding regions and consequently, alter functionality of the resulting proteins. It is crucial to distinguish isoforms that are generated due to natural transcriptomic dynamics from the ones that occur in malignant cells. Perhaps the most intriguing finding of this study was the identification of common AS transcripts for the B-ALL cohort. By custom Perl script we elucidate 338 common gene isoforms that may play role in oxidative phosphorylation and mitochondrial dysfunction pathways. Cancer cells prefer glycolysis over oxidative phosphorylation to fulfill their energy demand, suggesting that they have adapted to survive and proliferate in the absence of fully functional mitochondria. In addition to that, dysfunctional mitochondria cannot neutralize effect

from reactive forms of oxygen (ROS), which may lead to oxidative stress inside cells and alter crucial cellular processes, including regulation of gene transcription and alternative splicing. Thus, leukemic cells may generate abnormal proteins with added, deleted, or altered functional domains that contribute to pathogenesis of B-ALL.

Furthermore, the mechanistic study utilizing the Nalm 6 cell line revealed that nineteen common gene isoforms significantly change their expression level after 5-Aza treatment. Five genes among them – *TK1*, *SNN*, *PLCG2*, *CYTIP* and *SDF2L1* – showed consistent gene expression patterns in both comparisons: B-ALL versus healthy donors and Nalm 6 versus healthy donors. *TK1*, *SNN*, *PLCG2* and *CYTIP* genes were downregulated in B-ALL and Nalm 6 groups, while *SDF2L1* gene was upregulated. Interestingly, *SNN* downregulation has been shown in monocytic cell populations in chronic lymphocytic leukemia patients (Maffei et al., 2013) and might be regulated by the TNFα-PKCε signaling pathway, which implies a role for SNN in cell death and cell cycle regulation (Billingsley et al., 2006). In addition, downregulation of the *PLCG2* gene may alter B-cell receptor signaling and lead to the disruption of the B-cell maturation process (Ramsay & Rodriguez-Justo, 2013). Surprisingly, we do not observe upregulation of *TK1* gene, which is a well-known marker for ALL patient response to therapy and reflects the aggressiveness of leukemic cells (O'Neill, Zhang, Li, Fuja, & Murray, 2007). *CYTIP* upregulation was also reported in metastatic renal cancer (Vanharanta et al., 2013), but it is not consistent with our findings for B-ALL patients. To

sum up, a mechanistic study with the Nalm 6 cell line, suggests that some of the common gene isoforms may undergo epigenetic regulation in B-ALL.

Conclusions

The main strength of RNA sequencing data is that besides providing expression analysis it can be further mined for a number of other genetic abnormalities, including splicing alterations, fusion transcripts, alternate transcription start sites, point mutations, novel transcripts, fusion genes that will provide novel insights in B-ALL. Our data provide new insides and perspectives on the transcriptome regulation in B-ALL. We identified transcript isoforms and pathways that play key role in pathogenesis of B-ALL. These results improve our understanding of the transcriptional regulation underplaying B-ALL development and will help develop strategies for better diagnosis and managing patients with B-ALL in the future.

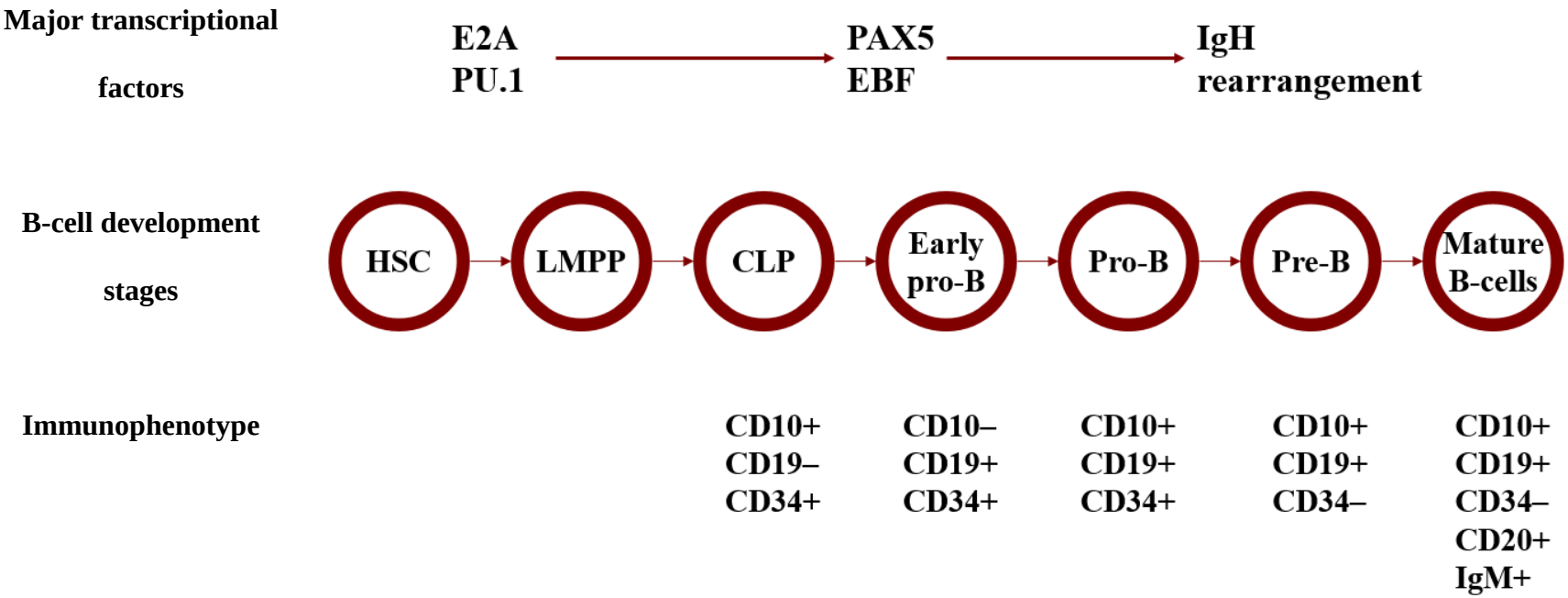


Figure 1. Schematic diagram of B-cell development stages, immunophenotype and major transcription factors (from Zhou et al. with changes, 2008). HSC – hematopoietic stem cell, LMPP – lymphoid multipotent progenitor, CLP – common lymphoid progenitor.

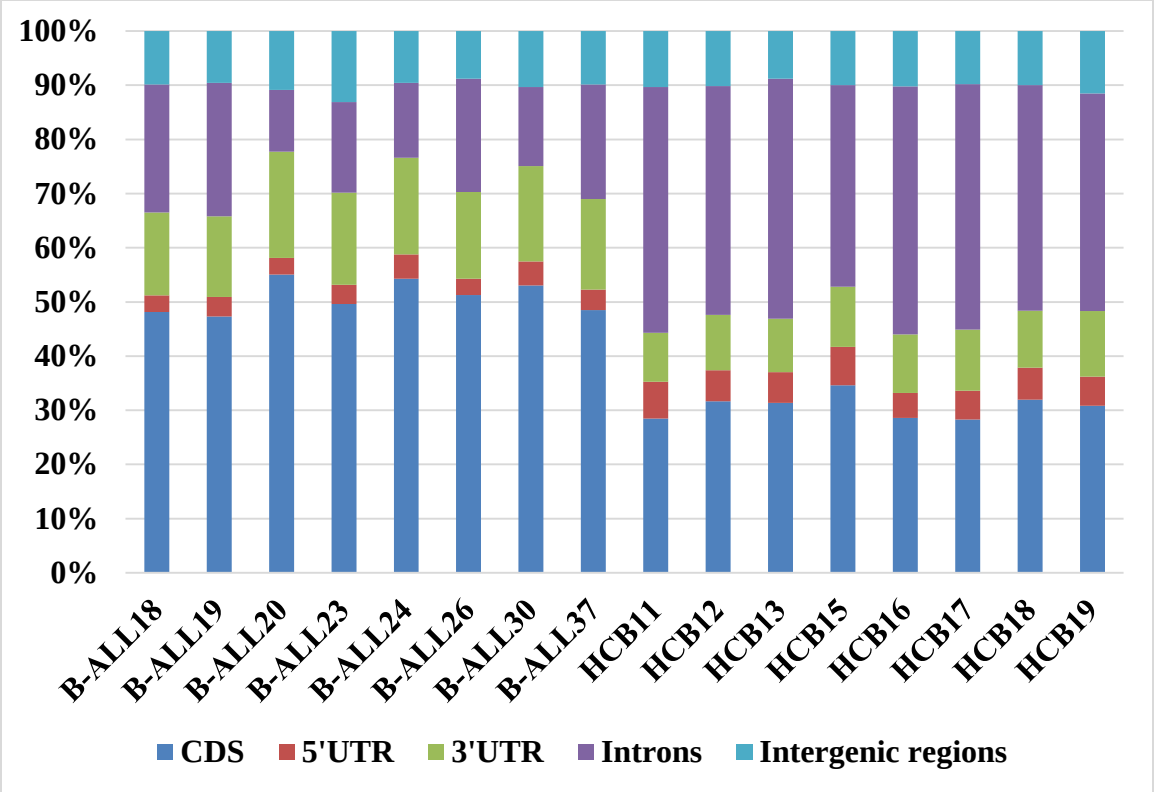
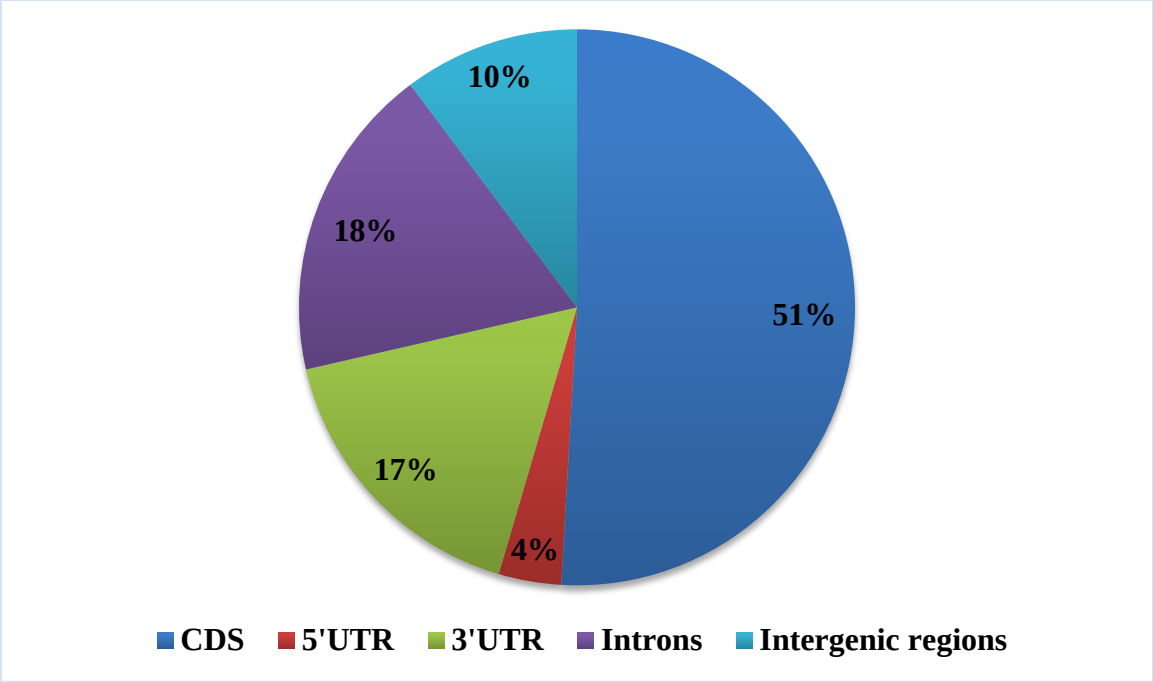


Figure 2. Bar diagram represents distribution of uniquely mapped reads to human genome UCSC hg19 (GRCh37). Each bar depicts the percentage of reads from individual samples (8 B-ALL patients and 8 healthy donors) mapped to coding sequence exon (CDS), 5’ and 3’ untranslated regions (5’ and 3’UTR), introns and intergenic regions.



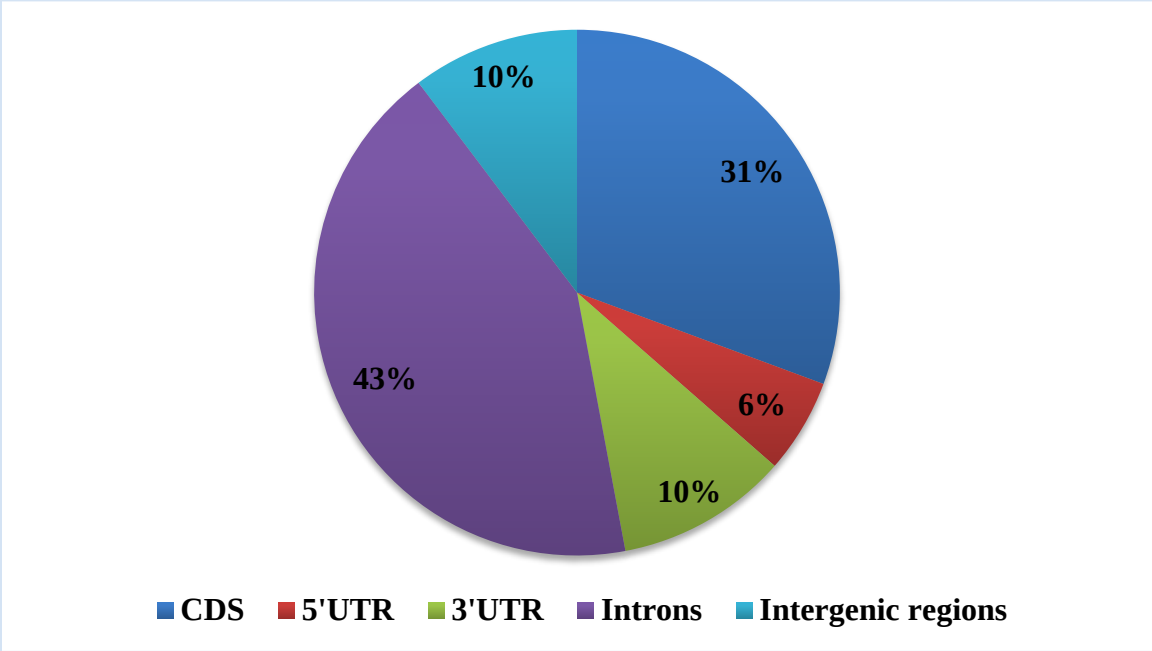


Figure 3. Average percentage of sequencing reads from 8 B-ALL (top) and 8 healthy donors (bottom) that map to coding sequence exon (CDS), 5' and 3' untranslated regions (5' and 3'UTR), introns and intergenic regions.

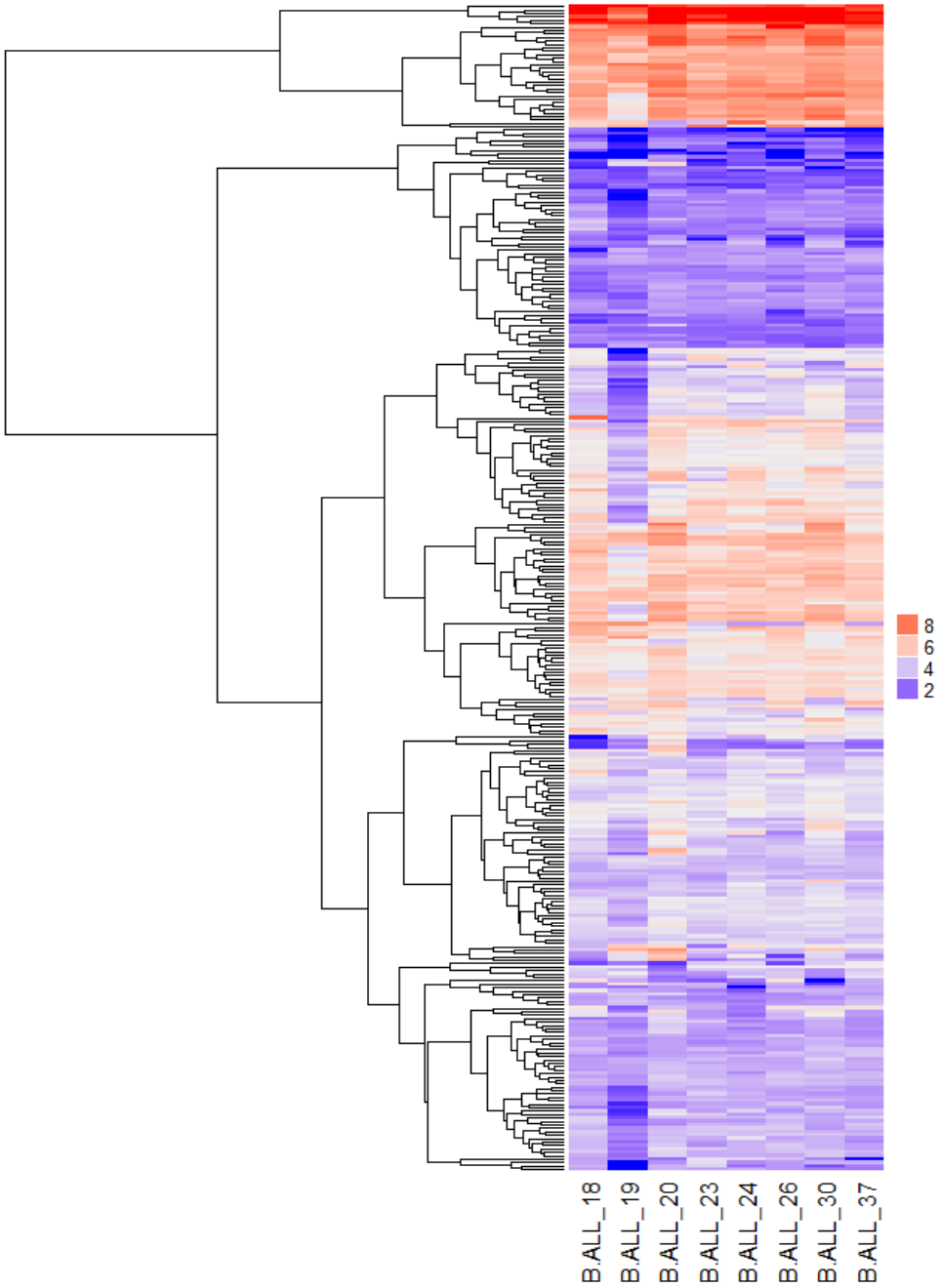


Figure 4. The heatmap representing common gene isoforms for B-ALL patients identified by custom Perl script. The heatmap representing common gene isoforms for B-ALL patients identified by custom Perl script. High-abundance transcripts in B-ALL patients represented in red. Low-abundance transcripts in B-ALL patients represented in blue. The intensity of color is related to level of transcripts abundances.

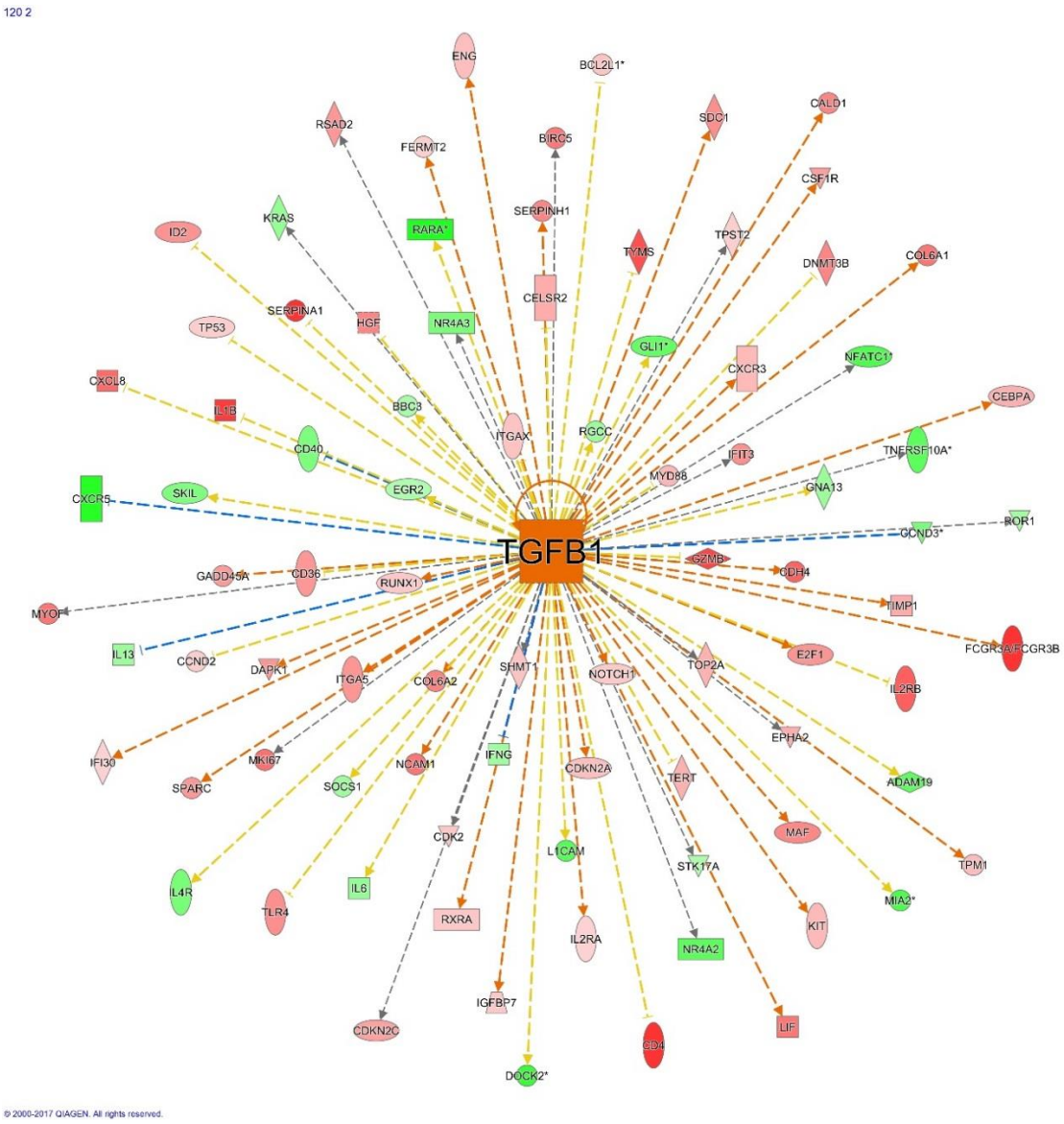


Figure 5. The mechanistic network of the inferred upstream regulator TGFB1. Genes presented in red are related to genes that up-regulated in B-ALL dataset. The mechanistic network of the inferred upstream regulator TGFB1. Genes presented in red are related to genes that up-regulated in B-ALL dataset. Genes presented in green are related to genes that down-regulated in B-ALL. The intensity of the colors is related to fold change estimates. Arrows presented in orange, gray and yellow indicate activation, effect not predicted and inconsistency, respectively.

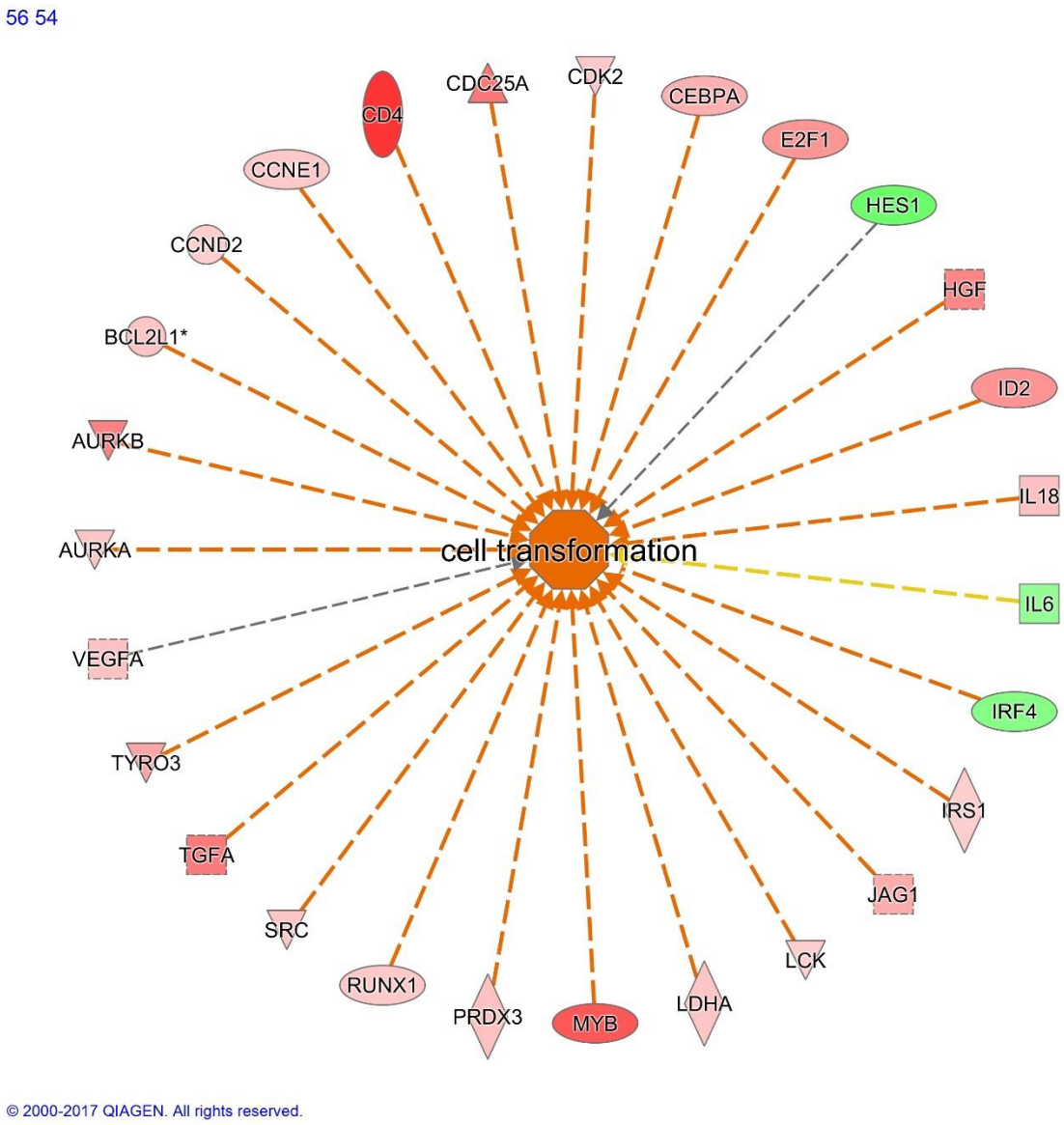


Figure 6. The differentially expressed gene network with function in cell transformation. Genes represented in red are upregulated in B-ALL group. The differentially expressed gene network with function in cell transformation. Genes represented in red are upregulated in B-ALL group. Genes presented in green are downregulated in B-ALL. The intensity of the colors is related to fold change estimates. Arrows presented in orange, gray and yellow indicate activation, effect not predicted and inconsistency, respectively.

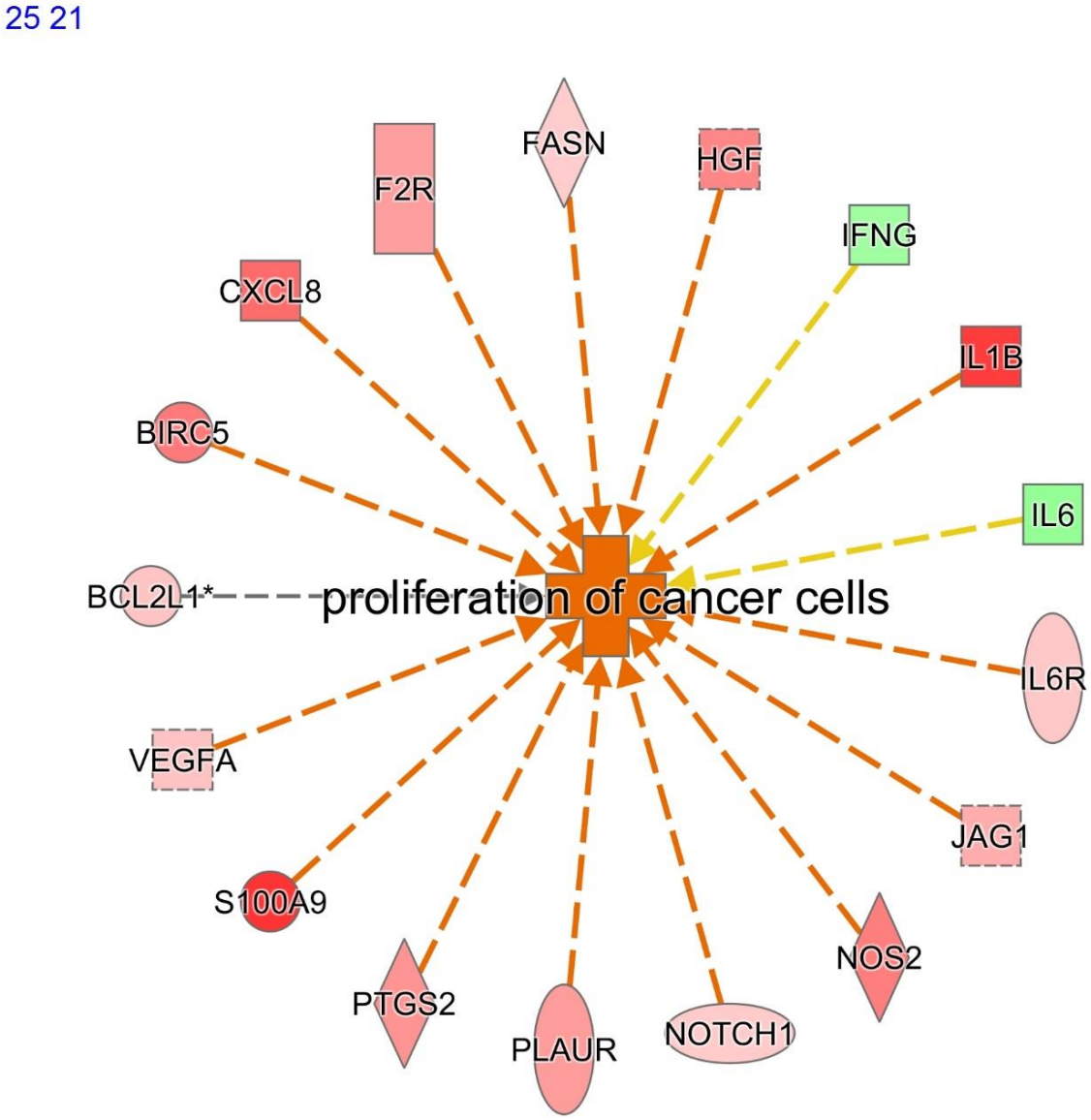


Figure 7. The differentially expressed gene network with function in proliferation of cancer cells. The differentially expressed gene network with function in proliferation of cancer cells. Genes represented in red are upregulated in B-ALL group. Genes presented in green are downregulated in B-ALL. The intensity of the colors is related to fold change estimates. Arrows presented in orange, gray and yellow indicate activation, effect not predicted and inconsistency, respectively.

Table 1

Alternative splicing events in cancer

Type of splicing	Gene	Spliced isoform	Type of cancer	Citation
Cassette exons (skipping one exon)	*RON*	ΔRON (lacks exon 11)	Breast and colon tumors	Ghigna et al., 2005
Cassette exons (skipping on multiple exons)	*BRAF*	Skipping of exon 4-8 in BRAFV600E	Melanoma	Poulikakos et al., 2011
Cassette exons (exon inclusion)	*SYK*	SYK(L) includes exon 9	T-cell lymphomas, chronic leukemias, head and neck carcinomas	Feldman et al. 2008; Buchner et al., 2009; Luangdilok et al., 2007
Alternative 5′ splice sites	*BCL2L1*	BCL-XL	Hepatocellular carcinoma, colorectal cancer	Takehara et al., 2001; Scherr et al., 2016
Alternative 3′ splice sites	*VEGF*	VEGFxxx	Osteosarcoma	Kaya et al., 2000
Intron retention	*HER2*	Herstatin (results from intron 8 retention) and p100 (results from intron 15 retention)	Breast cancer	Jackson et al., 2013
Mutually exclusive exons	*ACTN1*	Mutially exclusive exons - 19a and 19b	Colon cancer	Gardina et al., 2006

Complex splicing patterns	*MDM2*	more than 40 different splice variants	Breast carcinoma, ovarian and bladder cancers, glioblastoma	Bartel et al., 2002

Table 2

Patient characteristics

Patient ID	Blast rate (%)	Age (month)	WBC, $10^3/\mu l$	Sex	Immunophenotype	Cytogenetics
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
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)
A26	91	48	4.3	M	19;10	47, XY
A30	94	24	3.7	F	19;10;20wk	46, XX
A36	91	72	2.7	F	19;10;20	46, XX

Table 3

Top twenty upregulated and down-regulated genes in B-ALL patients versus healthy donors

Upregulated genes				Downregulated genes			
Gene	Description	logFC	FDR	Gene	Description	logFC	FDR
BIRC7	baculoviral IAP repeat containing 7	12.58774792	1.79E-22	CAMK2A	calcium/calmodulin dependent protein kinase II alpha	-11.33305197	1.35E-31
FAM69C	family with sequence similarity 69 member C	12.57422628	7.50E-32	CDH22	cadherin 22	-10.80719677	8.20E-65
NOL4	nucleolar protein 4	12.42793145	3.68E-17	ARSI	arylsulfatase family member I	-9.368798926	7.94E-16
NRN1	neuritin 1	12.21760376	7.21E-23	APLP1	amyloid beta precursor like protein 1	-8.874598807	1.50E-38
NKAIN4	Sodium/potassium transporting ATPase interacting 4	12.07076779	7.33E-20	ITLN1	intelectin 1	-8.322284896	1.56E-54
PTGDR	prostaglandin D2 receptor	11.78288729	4.21E-51	WNT10A	Wnt family member 10A	-7.92759092	1.96E-24
SDC2	syndecan 2	11.71967708	4.55E-35	NPHS1	NPHS1, nephrin	-7.712620956	1.06E-36
BMP2	bone morphogenetic protein 2	11.63391213	3.27E-26	CELA2A	chymotrypsin like elastase family member 2A	-7.115075267	2.05E-34
CD2	CD2 molecule	11.60059176	2.56E-21	CLLU1OS	chronic lymphocytic leukemia up-regulated 1 opposite strand	-7.106790442	1.13E-07

RGMA	repulsive guidance molecule family member a	11.55426857	1.82E-23	UPK2	uroplakin 2	-6.924019587	6.40E-28
FCN1	ficolin 1	11.52363369	1.08E-17	CD244	CD244 molecule	-6.80438577	2.65E-32
GIMAP6	GTPase, IMAP family member 6	11.27208145	3.85E-50	SEMA3F	semaphorin 3F	-6.785844561	7.67E-38
CLIC5	chloride intracellular channel 5	11.26816477	5.10E-21	NPPB	natriuretic peptide B	-6.77218967	3.03E-30
MDGA2	MAM domain containing glycosylphosphatidylinositol anchor 2	11.22689521	2.85E-20	ORM1	orosomucoid 1	-6.756404377	6.00E-50
FAM19A5	family with sequence similarity 19 member A5, C-C motif chemokine like	11.22294095	8.46E-21	CHADL	chondroadherin like	-6.740609261	2.00E-45
FCGR3A	Fc fragment of IgG receptor IIIa	11.181703	1.07E-17	TRPC5	transient receptor potential cation channel subfamily C member 5	-6.732986733	8.19E-36
LOXHD1	lipoxygenase homology domains 1	11.11395696	4.14E-18	LRRC18	leucine rich repeat containing 18	-6.641037404	3.59E-27
KIF26A	kinesin family member 26A	11.0719285	6.04E-50	SLC36A3	solute carrier family 36 member 3	-6.519546438	1.76E-28
GZMA	granzyme A	10.91223894	2.40E-18	ODF3L1	outer dense fiber of sperm tails 3 like 1	-6.379525998	4.72E-28
TSHZ3	teashirt zinc finger homeobox 3	10.86466044	1.60E-15	TNS4	tensin 4	-6.355462382	1.15E-37

Table 4

Common transcripts that affected by DNA methylation

Gene name	logFC (untreated Nalm 6)	logFC (treated Nalm 6)	logFC (B-ALL patients)
CYTIP	-11.86241096	-10.58620701	-3.320701712
TK1	-16.6756522	-15.03359745	-2.395212117
PLCG2	-16.38554599	-15.1004864	-2.32810002
SNN	-10.95532228	-9.653917529	-2.110926604
PRDX5	-9.712780965	-8.174896096	2.038034986
COX8A	-5.530015836	-4.317455614	2.063655717
UBQLN4	-10.21190755	-8.729855393	2.066030835
RNF181	-3.72386729	-2.09967035	2.115850312
TEX261	-9.351320334	-8.003411693	2.141855285
OSTC	-11.92871196	-8.737987284	2.254406726
DAD1	-9.46691059	-5.870977984	2.261850469
PITPNC1	-5.764856109	-4.208504433	2.458969425
LDHA	-13.23369912	-10.34328818	2.669660908
SDF2L1	2.551866594	3.711258169	2.671980184
IDH2	-6.207710831	-3.276596892	3.081108162
GAPDH	-11.06919991	-9.238478066	3.301257076

S100A6	-8.971244012	-5.824001533	3.612659686
ISG15	-9.14616871	-4.3756427	3.698909038
PRDX1	-11.30244531	-7.869383782	4.400595351

Table 5

Gene ontology terms for common transcripts that affected by DNA methylation

Gene name	GO term
CYTIP	protein binding
TK1	thymidine kinase activity
PLCG2	signal transducer activity and phosphatidylinositol phospholipase C activity
SNN	endosomal maturation
PRDX5	receptor binding and protein dimerization activity
COX8A	cytochrome-c oxidase activity
UBQLN4	identical protein binding and damaged DNA binding
SUMO2	poly(A) RNA binding and SUMO transferase activity
RNF181	ligase activity and ubiquitin-protein transferase activity
TEX261	COPII adaptor activity
OSTC	dolichyl-diphosphooligosaccharide-protein glycotransferase activity
DAD1	dolichyl-diphosphooligosaccharide-protein glycotransferase activity and oligosaccharyl transferase activity
PITPNC1	lipid binding and phosphatidylinositol transporter activity
LDHA	oxidoreductase activity and L-lactate dehydrogenase activity
SDF2L1	chaperone binding and misfolded protein binding
IDH2	magnesium ion binding and oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
GAPDH	identical protein binding and NAD binding
S100A6	calcium ion binding and calcium-dependent protein binding

ISG15	protein tag
PRDX1	poly(A) RNA binding and identical protein binding

Table 6

Top canonical pathways identified by IPA

Pathway	P-value	Overlap
Protein kinase A signaling	1.55E-06	28.4 % (105/370)
Interferon signaling	3.26E-03	38.8 % (14/36)
Cyclins and cell cycle regulation	2.20E-03	32.5 % (25/77)
Phospholipase C signaling	1.56E-03	26.7 % (58/217)
Cell cycle control of chromosomal replication	4.39E-05	47.4 % (18/38)

GENERAL DISCUSSION

Alternative splicing plays a crucial role in numerous cellular and developmental processes (Chen & Manley, 2009). In recent years, alternative splicing has been recognized as a mechanism involved in many human disorders, including cancer (Singh & Cooper, 2012). Changes in splicing patterns occur widely in cancer cells and has been shown to be associated with resistance to therapeutic treatments (David & Manley, 2010).

Despite decades of leukemia research, there is still a need for reliable cancer biomarkers for B-ALL diagnostics. The majority of pediatric B-ALL cases harbor gross numerical and structural chromosomal alterations, but they do not explain all incidences of disease. Therefore other molecular mechanisms likely contribute to B-ALL development, including alternative splicing. RNA-seq analysis allows one to effectively and efficiently evaluate the entire transcriptome by analyzing aberrant transcriptional patterns and splicing alterations that are crucial for B-ALL pathogenesis. In combination with pathway analysis, alternatively spliced transcripts may help better understand the molecular basis of post-transcriptional gene regulation in the context of B-ALL.

Here, we employed a pathway-centered approach that allows one to characterize the functional implications of differentially expressed and alternately spliced RNA transcripts in pediatric B-ALL patients. A custom Perl script was designed to obtain a set of common gene isoforms across individual B-ALL patients along with their corresponding transcript abundances. The functional annotation and enrichment analyses in IPA identified aberrant activation of cancer-related signaling pathways and transcriptional regulators associated with a B-ALL malignant phenotype, such as ERBB2, TGFB1 and IL2. A distinctive feature of the common

gene isoforms which were identified, is their implication in oxidative phosphorylation and mitochondrial dysfunction pathways. It has been shown, that mitochondrial damage modulates alternative splicing in neuronal cells leading to changes in the abundance of certain isoforms (Maracchioni et al., 2007). Therefore, mitochondrial dysfunction, a notable feature of cancer, may also be the mechanism underlying the changes in alternative splicing patterns observed in B-ALL patients.

Future directions for our research will integrate these findings with whole-genome DNA methylation studies on B-ALL patients previously analyzed in our research group. Furthermore, the leukemia-associated alternative splicing variants identified in this study may be utilized as novel tools for the diagnosis and classification of leukemias and could also be the targets for innovative therapeutical interventions based on highly selective splicing correction approaches.

BIBLIOGRAPHY

Akasaka, T., Balasas, T., Russell, L. J., Sugimoto, K. J., Majid, A., Walewska, R., … & Dyer, M. J. (2007). Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood*, *109*, 3451-3461. http://doi.org/10.1182/blood-2006-08-041012

Almamun, M., Kholod, O., Stuckel, A. J., Levinson, B. T., Johnson, N. T., Arthur, G. L., … & Taylor K. H. (2017). Inferring a role for methylation of intergenic DNA in the regulation of genes aberrantly expressed in precursor B-cell acute lymphoblastic leukemia. *Leuk Lymphoma*, *17*, 1-12. http://doi.org/10.1080/10428194.2016.1272683

Almamun, M., Levinson, B. T., Gater, S. T., Schnabel, R. D., Arthur, G. L., Davis, J. W., and Taylor, K. H. (2014). Genome-wide DNA methylation analysis in precursor B-cells. *Epigenetics*, *9*(12), 1588-1595. http://doi.org/10.4161/15592294.2014.983379

Almamun, M., Levinson, B. T., van Swaay, A. C., Johnson, N. T., McKay, S. D., Arthur, G. L., … & Taylor, K. H. (2015). Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia. *Epigenetics*, *10*(9), 882-890. http://doi.org/10.1080/15592294.2015.1078050

Almamun, M., Schnabel, J. L., Gater, S. T., Ning, J., & Taylor, K. H. (2013). Isolation of precursor B-cell subsets from umbilical cord blood. *Journal of Visualized Experiments*, (74), 50402. http://doi.org/10.3791/50402

Ameur, A., Wetterbom, A., Feuk, L., & Gyllensten, U. (2010). Global and unbiased detection of splice junctions from RNA-seq data. *Genome Biology*, *11*(3), R34. http://doi.org/10.1186/gb-2010-11-3-r34

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. http://doi.org/10.1093/bioinformatics/btu638

Bartel, F., Taubert, H., & Harris, L. C. (2002). Alternative and aberrant splicing of MDM2 mRNA in human cancer. *Cancer Cell*, *2*(1), 9-15.

Bercovich, D., Ganmore, I., Scott, L. M., Wainreb, G., Birger, Y., Elimelech, A., … & Izraeli, S. (2008). Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. *Lancet*, *372*(9648), 1484-92. http://doi.org/10.1016/S0140-6736(08)61341-0

Billingsley, M. L., Yun, J., Reese, B. E., Davidson, C. E., Buck-Koehntop, B. A., & Veglia, G. (2006). Functional and structural properties of stannin: Roles in cellular growth, selective toxicity, and mitochondrial responses to injury. *J Cell Biochem.*, *98*(2), 243-50. http://doi.org/10.1002/jcb.20809

Buchner, M., Fuchs, S., Prinz, G., Pfeifer, D., Bartholomé, K., Burger, M., … & Zirlik, K. (2009). Spleen tyrosine kinase is overexpressed and represents a potential therapeutic target in chronic lymphocytic leukemia. *Cancer Res.*, *69*(13), 5424-32. http://doi.org/10.1158/0008-5472.CAN-08-4252

Bujko, M., Kober, P., Rusetska, N., Wakuła, M., Goryca, K., Grecka, E., … & Siedlecki, J. A. (2016). Aberrant DNA methylation of alternative promoter of DLC1 isoform 1 in meningiomas. *Journal of Neuro-Oncology*, *130*(3), 473-484. http://doi.org/10.1007/s11060-016-2261-3

Burmeister, D. W., Smith, E. H., Cristel, R. T., McKay, S. D., Shi, H., Arthur, G. L., … & Taylor, K. H. (2017). The expression of RUNDC3B is associated with promoter

methylation in lymphoid malignancies. *Hematol Oncol*., *35*(1), 25-33. http://doi.org/10.1002/hon.2238

Busche, S., Ge, B., Vidal, R., Spinella, J. F., Saillour, V., Richer, C., … & Pastinen, T. (2013). Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia. *Cancer Res*., *73*(14), 4323-36. http://doi.org/10.1158/0008-5472.CAN-12-4367

Bussmann, L. H., Schubert, A., Vu Manh, T. P., De Andres, L., Desbordes, S. C., Parra, M., … & Graf, T. (2009). A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell*, *5*(5), 554-66. http://doi.org/10.1016/j.stem.2009.10.004

Chapiro, E., Russell, L. J., Struski, S., Cavé, H., Radford-Weiss, I., Valle, V. D., … & Nguyen-Khac, F. (2010). A new recurrent translocation t(11;14)(q24;q32) involving IGH@ and miR-125b-1 in B-cell progenitor acute lymphoblastic leukemia. *Leukemia*, *24*(7), 1362-4. http://doi.org/10.1038/leu.2010.93

Chatterton, Z., Morenos, L., Saffery, R., Craig, J. M., Ashley, D., & Wong, N. C. (2012). DNA methylation and miRNA expression profiling in childhood B-cell acute lymphoblastic leukemia. *Epigenomics*, *2*(5), 697-708. http://doi.org/10.2217/epi.10.39

Chen, M., & Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nature Reviews. *Molecular Cell Biology*, *10*(11), 741–754. http://doi.org/10.1038/nrm2777

Chiaretti, S., Zini, G., & Bassan, R. (2014). Diagnosis and subclassification of acute lymphoblastic leukemia. *Mediterranean Journal of Hematology and Infectious Diseases*, *6*(1), e2014073. http://doi.org/10.4084/MJHID.2014.073

Coyaud, E., Struski, S., Prade, N., Familiades, J., Eichner, R., Quelen, C., … & Broccardo, C. (2010). Wide diversity of PAX5 alterations in B-ALL: a Groupe Francophone de Cytogenetique Hematologique study. *Blood*, *115*(15), 3089-97. http://doi.org/10.1182/blood-2009-07-234229

Dang, J., Wei, L., de Ridder, J., Su, X., Rust, A. G., Roberts, K. G., … & Mullighan, C. G. (2015). PAX5 is a tumor suppressor in mouse mutagenesis models of acute lymphoblastic leukemia. *Blood*, *125*(23), 3609-3617. http://doi.org/10.1182/blood-2015-02-626127

David, C. J., & Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & Development*, *24*(21), 2343-2364. http://doi.org/10.1101/gad.1973010

Dawson, A. J., Yanofsky, R., Vallente, R., Bal, S., Schroedter, I., Liang, L., & Mai, S. (2011). Array comparative genomic hybridization and cytogenetic analysis in pediatric acute leukemias. *Current Oncology*, *18*(5), e210-e217.

De Boer, J., Yeung, J., Ellu, J., Ramanujachar, R., Bornhauser, B., Solarska, O., … & Brady, H. J. (2011). The E2A-HLF oncogenic fusion protein acts through Lmo2 and Bcl-2 to immortalize hematopoietic progenitors. *Leukemia*, *25*, 321-30. http://doi.org/10.1038/leu.2010.253.pmid:21072044

Dias, S., Silva, H., Cumano, A., & Vieira, P. (2005). Interleukin-7 is necessary to maintain the B cell potential in common lymphoid progenitors. *The Journal of Experimental Medicine*, *201*(6), 971-979. http://doi.org/10.1084/jem.20042393

Dong, M., & Blobe, G. C. (2006). Role of transforming growth factor-β in hematologic malignancies. *Blood, 107*(12), 4589-4596. http://doi.org/10.1182/blood-2005-10-4169

Dyer, M. J., Akasaka, T., Capasso, M., Dusanjh, P., Lee, Y. F., Karran, E. L., … & Siebert, R. (2010). Immunoglobulin heavy chain locus chromosomal translocations in B-cell precursor acute lymphoblastic leukemia: rare clinical curios or potent genetic drivers? *Blood, 115*(8), 1490-9. http://doi.org/10.1182/blood-2009-09-235986

Eswaran, J., Cyanam, D., Mudvari, P., Reddy, S. D. N., Pakala, S. B., Nair, S. S., … & Kumar, R. (2012). Transcriptomic landscape of breast cancers through mRNA sequencing. *Scientific Reports*, *2*, 264. http://doi.org/10.1038/srep00264

Feldman, A., Sun, D., Law, M., Novak, A., Attygalle, A., Thorland, E., … & Dogan, A. (2008). Overexpression of Syk tyrosine kinase in peripheral T-cell lymphomas. *Leukemia, 22*(6), 1139-1143. http://doi.org/10.1038/leu.2008.77

Fleming, H. E., & Paige, C. J. (2001). Pre-B cell receptor signaling mediates selective response to IL-7 at the pro-B to pre-B cell transition via an ERK/MAP kinase-dependent pathway. *Immunity*, *15*(4), 521-31.

Galante, P. A. F., Sakabe, N. J., Kirschbaum-Slager, N., & de Souza, S. J. (2004). Detection and evaluation of intron retention events in the human transcriptome. *RNA*, *10*(5), 757-765. http://doi.org/10.1261/rna.5123504

Gardina, P. J., Clark, T. A., Shimada, B., Staples, M. K., Yang, Q., Veitch, J., … & Turpaz, Y. (2006). Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics*, *7*, 325. http://doi.org/10.1186/1471-2164-7-325

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P. M., … & Biamonti, G. (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol Cell*, *20*(6), 881-90. http://doi.org/10.1016/j.molcel.2005.10.026

Gordon, A., & Hannon, G. J. (n.d.) *"FASTX-Toolkit", FASTQ/A short-reads pre-processing tools*. Unpublished manuscript. http://hannonlab.cshl.edu/fastx_toolkit/.

Griffith, M., Griffith, O. L., Krysiak, K., Skidmore, Z. L., Christopher, M. J., Klco, J. M., … & Ley, T. J. (2016). Comprehensive genomic analysis reveals FLT3 activation and a therapeutic strategy for a patient with relapsed adult B-lymphoblastic leukemia. *Exp Hematol*., *44*(7), 603-13. http://doi.org/10.1016/j.exphem.2016.04.011

Grimaldi, J. C., & Meeker, T. C. (1989). A novel translocation, t(14;19)(q32;p13), involving IGH@ and the cytokine receptor for erythropoietin. *Leukemia*, *23*(3), 614-7. http://doi.org/10.1038/leu.2008.250

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847-9. http://doi.org/10.1093/bioinformatics/btw313

Harewood, L., Robinson, H., Harris, R., Al-Obaidi, M. J., Jalali, G. R., Martineau, M., … & Harrison, C. J. (2003). Amplification of AML1 on a duplicated chromosome 21 in acute lymphoblastic leukemia: a study of 20 cases. *Leukemia*, *17*(3), 547-53. http://doi.org/10.1038/sj.leu.2402849

Harrison, C. (2011). New genetics and diagnosis of childhood B-cell precursor acute lymphoblastic leukemia. *Pediatric Reports*, *3*(Suppl 2), e4. http://doi.org/10.4081/pr.2011.s2.e4

Hennighausen, L., & Robinson, G. W. (2008). Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. *Genes & Development*, *22*(6), 711-721. http://doi.org/10.1101/gad.1643908

Hirokawa, S., Sato, H., Kato, I. & Kudo, A. (2003). EBF-regulating Pax5 transcription is enhanced by STAT5 in the early stage of B cells. *Eur J Immunol.*, *33*(7), 1824-9. http://doi.org/10.1002/eji.200323974

Hirose, K., Inukai, T., Kikuchi, J., Furukawa, Y., Ikawa, T., Kawamoto, H., … & Sugita, K. (2010). Aberrant induction of LMO2 by the E2A-HLF chimeric transcription factor and its implication in leukemogenesis of B-precursor ALL with t(17;19). *Blood*, *116*(6), 962-70. http://doi.org/10.1182/blood-2009-09-244673

Hong, D., Gupta, R., Ancliff, P., Atzberger, A., Brown, J., Soneji, S., … & Enver, T. (2008). Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science*, *319*(5861), 336-9. http://science.sciencemag.org/content/319/5861/336

Hornakova, T., Chiaretti, S., Lemaire, M. M., Foà, R., Ben Abdelali, R., Asnafi, V., … & Knoops, L. (2009). ALL-associated JAK1 mutations confer hypersensitivity to the antiproliferative effect of type I interferon. *Blood*, *115*(16), 3287-95. http://doi.org/10.1182/blood-2009-09-245498

Hu, Y., Zhang, Z., Kashiwagi, M., Yoshida, T., Joshi, I., Jena, N., … & Georgopoulos, K. (2016). Superenhancer reprogramming drives a B-cell-epithelial transition and high-risk leukemia. *Genes Dev*., *30*(17), 1971-90. http://doi.org/10.1101/gad.283762.116

Huettner, C. S., Zhang, P., Van Etten, R. A. & Tenen, D. G. (2000). Reversibility of acute B-cell leukaemia induced by BCR-ABL1. *Nat Genet*., *24*(1), 57-60. http://doi.org/10.1038/71691

Hunger, S. P. (1996). Chromosomal translocation involving the E2A gene in acute lymphoblastic leukemia: clinical features and molecular pathogenesis. *Blood, 87,* 1211-1224.

Hunger, S. P., & Mullighan, C. G. (2015). Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine. *Blood, 125*(26), 3977-3987. http://doi.org/10.1182/blood-2015-02-580043

Iacobucci, I., Iraci, N., Messina, M., Lonetti, A., Chiaretti, S., Valli, E., … & Martinelli, G. (2012). IKAROS Deletions Dictate a Unique Gene Expression Signature in Patients with Adult B-Cell Acute Lymphoblastic Leukemia. *PLoS ONE, 7*(7), e40934. http://doi.org/10.1371/journal.pone.0040934

Iacobucci, I., Lonetti, A., Messa, F., Ferrari, A., Cilloni, D., Soverini, S., … & Martinelli, G. (2010). Different isoforms of the B-cell mutator activation-induced cytidine deaminase are aberrantly expressed in BCR-ABL1-positive acute lymphoblastic leukemia patients. *Leukemia, 24*(1), 66-73. http://doi.org/10.1038/leu.2009.197

Ichikawa, M., Asai, T., Saito, T., Seo, S., Yamazaki, I., Yamagata, T., … & Kurokawa, M. (2004). AML-1 is required for megakaryocytic maturation and lymphocytic

differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med*., *10*, 299-304.

Inthal, A., Zeitlhofer, P., Zeginigg, M., Morak, M., Grausenburger, R., Fronkova, E., … & Panzer-Grümayer, R. (2012). CREBBP HAT domain mutations prevail in relapse cases of high hyperdiploid childhood acute lymphoblastic leukemia. *Leukemia*, *26*(8), 1797-1803. http://doi.org/10.1038/leu.2012.60

Jackson, C., Browell, D., Gautrey, H., & Tyson-Capper, A. (2013). Clinical Significance of HER-2 Splice Variants in Breast Cancer Progression and Drug Resistance. *International Journal of Cell Biology*, 973584. http://doi.org/10.1155/2013/973584

Jaffe, J. D., Wang, Y., Chan, H. M., Zhang, J., Huether, R., Kryukov, G. V., … & Stegmeier, F. (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nature Genetics*, *45*(11), 1386-1391. http://doi.org/10.1038/ng.2777

Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C. P., Sorensen, P. H., … & Triche, T. (2011). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is "dark matter" un-annotated RNA. *BMC Biology*, *9*, 86. http://doi.org/10.1186/1741-7007-9-86

Kawamata, N., Ogawa, S., Zimmermann, M., Kato, M., Sanada, M., Hemminki, K., … & Koeffler, H. P. (2008). Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood*, *111*(2), 776-784. http://doi.org/10.1182/blood-2007-05-088310

Kaya, M., Wada, T., Kawaguchi, S., Nagoya, S., Yamashita, T., Abe, Y., … & Ishil, S. (2002). Increased pre-therapeutic serum vascular endothelial growth factor in patients with early clinical relapse of osteosarcoma. *British Journal of Cancer*, *86*(6), 864–869. http://doi.org/10.1038/sj.bjc.6600201

Kearney, L., Gonzalez De Castro, D., Yeung, J., Procter, J., Horsley, S. W., Eguchi-Ishimae, M., … & Greaves, M. (2009). Specific JAK2 mutation (JAK2R683) and multiple gene deletions in Down syndrome acute lymphoblastic leukemia. *Blood, 113*(3), 646-8. http://doi.org/10.1182/blood-2008-08-170928

Kitamura, D., Roes, J., Kuhn, R., & Rajewsky, K. (1991). A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin μ chain gene. *Nature, 350*(6317), 423-6. http://doi.org/10.1038/356154a0

Krämer, A., Green, J., Pollard, J., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, *30*(4), 523-530. http://doi.org/10.1093/bioinformatics/btt703

Kuo, A. J., Cheung, P., Chen, K., Zee, B. M., Kioi, M., Lauring, J., … & Gozani, O. (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Molecular Cell*, *44*(4), 609-620. http://doi.org/10.1016/j.molcel.2011.08.042

Le, M. T. N., Teh, C., Shyh-Chang, N., Xie, H., Zhou, B., Korzh, V., … & Lim, B. (2009). MicroRNA-125b is a novel negative regulator of p53. *Genes & Development*, *23*(7), 862-876. http://doi.org/10.1101/gad.1767609

LeBrun, D. P. (2003). E2A basic helix-loop-helix transcription factors in human leukemia. *Front Biosci*, 8, 206-22.

Lilljebjörn, H., Rissler, M., Lassen, C., Heldrup, J., Behrendtz, M., Mitelman, F., … & Fioretos, T. (2012). Whole-exome sequencing of pediatric acute lymphoblastic leukemia. *Leukemia, 26*(7), 1602-7. http://doi.org/10.1038/leu.2011.333

López-Andrade, B., Sartori, F., Gutiérrez, A., García, L., Cunill, V., Durán, M. A., … & Martínez-Serra, J. (2015). Acute lymphoblastic leukemia with e1a3 BCR/ABL fusion protein. A report of two cases. *Experimental Hematology & Oncology*, *5*, 21. http://doi.org/10.1186/s40164-016-0049-y

Luangdilok, S., Box, C., Patterson, L., Court, W., Harrington, K., Pitkin, L., … & Eccles, S. (2007). Syk tyrosine kinase is linked to cell motility and progression in squamous cell carcinomas of the head and neck. *Cancer Res*., *67*(16), 7907-16. http://doi.org/10.1158/0008-5472.CAN-07-0331

Lundin, C., Heldrup, J., Ahlgren, T., Olofsson, T., & Johansson, B. (2009). B-cell precursor t(8;14)(q11;q32)-positive acute lymphoblastic leukemia in children is strongly associated with Down syndrome or with a concomitant Philadelphia chromosome. *Eur J Haematol*., *82*(1), 46-53. http://doi.org/10.1111/j.1600-0609.2008.01166.x

Maffei, R., Bulgarelli, J., Fiorcari, S., Bertoncelli, L., Martinelli, S., Guarnotta, C., … & Marasca, R. (2013). The monocytic population in chronic lymphocytic leukemia shows altered composition and deregulation of genes involved in phagocytosis and inflammation. *Haematologica*, *98*(7), 1115-1123. http://doi.org/10.3324/haematol.2012.073080

Malgeri, U., Baldini, L., Perfetti, V., Fabris, S., Vignarelli, M. C., Colombo, G., … & Neri, A. (2000). Detection of t(4;14)(p16.3;q32) chromosomal translocation in multiple myeloma

by reverse transcription-polymerase chain reaction analysis of IGH-MMSET fusion transcripts. *Cancer Res*., *60*(15), 4058-61. Retrieved from http://cancerres.aacrjournals.org/content/60/15/4058.short

Maracchioni A, Totaro A, Angelini DF, Di Penta A, Bernardi G, Carri MT, & Achsel T. (2007). Mitochondrial damage modulates alternative splicing in neuronal cells: implications for neurodegeneration. *J Neurochem*., *100*(1):142–53. doi: 10.1111/j.1471-4159.2006.04204.x.

Marshall, A. J., Fleming, H. E., Wu, G. E., & Paige C. J. (1998). Modulation of the IL-7 dose-response threshold during pro-B cell differentiation is dependent on pre-B cell receptor expression. *J Immunol*., *161*(11), 6038-45.

Messina, M., Chiaretti, S., Tavolaro, S., Peragine, N., Vitale, A., Elia, L., … & Foà, R. (2010). Protein kinase gene expression profiling and in vitro functional experiments identify novel potential therapeutic targets in adult acute lymphoblastic leukemia. *Cancer*, *116*(14), 3426-37. doi: 10.1002/cncr.25113.

Meyer, C., Hofmann, J., Burmeister, T., Gröger, D., Park, T. S., Emerenciano, M., Pombo de Oliveira, M., … & Marschalek, R. (2013). The MLL recombinome of acute leukemias in 2013. *Leukemia*, *27*(11), 2165-2176. http://doi.org/10.1038/leu.2013.135

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., … & Marschalek, R. (2009). New insights to the MLL recombinome of acute leukemias. *Leukemia*, *23*(8), 1490-9. http://doi.org/10.1038/leu.2009.33

Moorman, A. V. (2016). New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica*, *101*(4), 407-416. http://doi.org/10.3324/haematol.2015.141101

Moorman, A. V., Ensor, H. M., Richards, S. M., Chilton, L., Schwab, C., Kinsey, S. E., … & Harrison, C.J. (2010). Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. *Lancet Oncol*., *11*(5), 429-38. http://doi.org/10.1016/S1470-2045(10)70066-8.

Moorman, A. V., Richards, S. M., Robinson, H. M., Strefford, J. C., Gibson, B. E., Kinsey, S. E., … & Harrison, C. J. (2007). Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromosome 21 (iAMP21). *Blood*, *109*(6), 2327-30. http://doi.org/10.1182/blood-2006-08-040436

Mrózek, K., Harper, D. P., & Aplan, P. D. (2009). Cytogenetics and Molecular Genetics of Acute Lymphoblastic Leukemia. *Hematology*, *23*(5), 991–v. http://doi.org/10.1016/j.hoc.2009.07.001

Mullighan, C. G. (2012). Molecular genetics of B-precursor acute lymphoblastic leukemia. *J Clin Invest*., *122*(10), 3407-15. http://dx.doi.org/10.1172/JCI61203

Mullighan, C. G., Collins-Underwood, J. R., Phillips, L. A. A., Loudin, M. L., Liu, W., Zhang, J., … & Rabin, K. R. (2009). Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia. *Nature Genetics*, *41*(11), 1243-1246. http://doi.org/10.1038/ng.469

Mullighan, C. G., Goorha, S., Radtke, I., Miller, C. B., Coustan-Smith, E., Dalton, J. D., … & Downing, J. R. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*, *446*(7137), 758-64. http://doi.org/10.1038/nature05690

Mullighan, C. G., Miller, C. B., Radtke, I., Phillips, L. A., Dalton, J., Ma, J., … & Downing, J. R. (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature*, *453*(7191), 110-4. http://doi.org/10.1038/nature06866

Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A. A., Miller, C. B., … & Downing, J. R. (2009). Deletion of IKZF1 and Prognosis in Acute Lymphoblastic Leukemia. *The New England Journal of Medicine*, *360*(5), 470–480. http://doi.org/10.1056/NEJMoa0808253

Mullighan, C. G., Zhang, J., Kasper, L. H., Lerach, S., Payne-Turner, D., Phillips, L. A., … & Downing, J. R. (2011). CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature*, *471*(7337), 235-239. http://doi.org/10.1038/nature09727

Nakase, K., Ishimaru, F., Avitahl, N., Dansako, H., Matsuo, K., Fujii, K., … & Harada, M. (2000). Dominant negative isoform of the Ikaros gene in patients with adult B-cell acute lymphoblastic leukemia. *Cancer Res.*, *60*(15), 4062-5.

Nebral, K., Denk, D., Attarbaschi, A., König, M., Mann, G., Haas, O. A., & Strehl, S. (2009). Incidence and diversity of PAX5 fusion genes in childhood acute lymphoblastic leukemia. *Leukemia*, *23*(1), 134-43. http://doi.org/10.1038/leu.2008.306

Niu, Y. N., Liu, Q. Q., Zhang, S. P., Yuan, N., Cao, Y., Cai, J. Y., … & Wang, J-R. (2014). Alternative messenger RNA splicing of autophagic gene Beclin 1 in human B-cell acute

lymphoblastic leukemia cells. *Asian Pac J Cancer Prev*., *15*(5), 2153-8. Retrieved from https://pdfs.semanticscholar.org/4f80/23a4239e516109264c383f5dcfe139129d9f.pdf

Nordlund, J., Bäcklin, C. L., Wahlberg, P., Busche, S., Berglund, E. C., Eloranta, M.-L, … & Syvänen, A-C. (2013). Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biology*, *14*(9), r105. http://doi.org/10.1186/gb-2013-14-9-r105

O'Neill, K. L., Zhang, F., Li, H., Fuja, D. G., & Murray, B. K. (2007). Thymidine kinase 1 – a prognostic and diagnostic indicator in ALL and AML patients. *Leukemia*, *21*(3), 560-3. http://doi.org/10.1038/sj.leu.2404536

Park, I. K., Qian, D., Kiel, M., Becker, M. W., Pihalja, M., Weissman, I. L., … & Clarke, M. F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature*, *423*(6937), 302-5. http://doi.org/10.1038/nature01587

Paulsson, K., Forestier, E., Lilljebjörn, H., Heldrup, J., Behrendtz, M., Young, B. D., & Johansson, B. (2010). Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21719-21724. http://doi.org/10.1073/pnas.1006981107

Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A., & Ploner A. (2005). False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics*, *21*(13), 3017-24. http://doi.org/10.1093/bioinformatics/bti448

Pieters, R., Schrappe, M., De Lorenzo, P., Hann, I., De Rossi, G., Felice, M., … & Valsecchi, M. G. (2007). A treatment protocol for infants younger than 1 year with acute lymphoblastic

leukaemia (Interfant-99): an observational study and a multicentre randomised trial. *Lancet*, *370*, 240-250.

Pongubala, J. M., Northrup, D. L., Lancki, D. W., Medina, K. L., Treiber, T., Bertolino, E., … & Singh, H. (2008). Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nat Immunol*., *9*(2), 203-15. http://dx.doi.org/10.1038/ni1555

Poulikakos, P. I., Persaud, Y., Janakiraman, M., Kong, X., Ng, C., Moriceau, G., … & Solit, D. B. (2011). RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature*, *480*(7377), 387-390. http://doi.org/10.1038/nature10662

Pui, C. H., Chessells, J. M., Camitta, B., Baruchel, A., Biondi, A., Boyett, J. M., … & Schrappe, M. (2003). Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. *Leukemia*, *17*(4), 700-6. http://doi.org/10.1038/sj.leu.2402883

Pui, C. H., Robison, L. L., &Look A. T. (2008). Acute lymphoblastic leukaemia. *Lancet*, *371*(9617), 1030-43. http://dx.doi.org/10.1016/S0140-6736(08)60457-2

Ramsay, A. D., & Rodriguez-Justo M. (2013). Chronic lymphocytic leukaemia – the role of the microenvironment pathogenesis and therapy. *Br J Haematol*., *162*(1), 15-24. http://doi.org/10.1111/bjh.12344

Rand, V., Parker, H., Russell, L. J., Schwab, C., Ensor, H., Irving, J., … & Harrison, C. J. (2011). Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia. *Blood*, *117*(25), 6848-55. http://doi.org/10.1182/blood-2011-01-329961

Raynaud, S., Cave, H., Baens, M., Bastard, C., Cacheux, V., Grosgeorge, J., … & Grandchamp, B. (1996). The 12;21 translocation involving TEL and deletion of the other TEL allele: two frequently associated alterations found in childhood acute lymphoblastic leukemia. *Blood*, *87*(7), 2891-9. Retrieved from http://www.bloodjournal.org/content/87/7/2891.short

Redaelli, A., Laskin, B. L., Stephens, J. M., Botteman, M. F. & Pashos, C. L. (2005). A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL). *Eur J Cancer Care*, *14*(1), 53-62. http://doi.org/10.1111/j.1365-2354.2005.00513.x

Rieder, S. E., Banta, L. M., Köhrer, K., McCaffery, J. M., & Emr, S. D. (1996). Multilamellar endosome-like compartment accumulates in the yeast vps28 vacuolar protein sorting mutant. *Molecular Biology of the Cell*, *7*(6), 985-999.

Roberts, K. G., & Mullighan, C. G. (2015). Genomics in acute lymphoblastic leukaemia: insights and treatment implications. *Nat Rev Clin Oncol*., *12*(6), 344-57. http://dx.doi.org/10.1038/nrclinonc.2015.38

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. http://doi.org/10.1186/gb-2010-11-3-r25

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140. http://doi.org/10.1093/bioinformatics/btp616

Ross, M. E., Zhou, X., Song, G., Shurtleff, S. A., Girtman, K., Williams, W. K., … & Downing, J. R. (2003). Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood*, *102*(8), 2951-9. http://doi.org/10.1182/blood-2003-01-0338

Russell, L. J., Akasaka, T., Majid, A., Sugimoto, K. J., Loraine Karran, E., Nagel, I., … & Harrison, C. J. (2008). t(6;14)(p22;q32): a new recurrent IGH@ translocation involving ID4 in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood*, *111*(1), 387-91. http://doi.org/10.1182/blood-2007-07-092015

Russell, L. J., De Castro, D. G., Griffiths, M., Telford, N., Bernard, O., Panzer-Grümayer, R., … & Harrison, C. J. (2009). A novel translocation, t(14;19)(q32;p13), involving IGH@ and the cytokine receptor for erythropoietin. *Leukemia*, *23*(3), 614-7. http://doi.org/10.1038/leu.2008.250

Sakabe, N. J., & de Souza, S. J. (2007). Sequence features responsible for intron retention in human. *BMC Genomics*, *8*, 59. http://doi.org/10.1186/1471-2164-8-59

Santoro, A., Bica, M. G., Dagnino, L., Agueli, C., Salemi, D., Cannella, S., … & Basso, G. (2009). Altered mRNA expression of PAX5 is a common event in acute lymphoblastic leukaemia. *Br J Haematol*., *146*(6), 686-9. http://doi.org/10.1111/j.1365-2141.2009.07815.x

Sapio, L., Di Maiolo, F., Illiano, M., Esposito, A., Chiosi, E., Spina, A., & Naviglio, S. (2014). Targeting protein kinase A in cancer therapy: an update. *EXCLI Journal*, *13*, 843–855.

Schafer, E., Irizarry, R., Negi, S., McIntyre, E., Small, D., Figueroa, M. E., … & Brown, P. (2010). Promoter hypermethylation in MLL-r infant acute lymphoblastic leukemia:

biology and therapeutic targeting. *Blood*, *115*(23), 4798-4809. http://doi.org/10.1182/blood-2009-09-243634

Scherr, A.-L., Gdynia, G., Salou, M., Radhakrishnan, P., Duglova, K., Heller, A., … & Koehler, B. C. (2016). Bcl-xL is an oncogenic driver in colorectal cancer. *Cell Death & Disease*, *7*(8), e2342–. http://doi.org/10.1038/cddis.2016.233

Schotte, D., Akbari Moqadam, F., Lange-Turenhout, E. A., Chen, C., van Ijcken, W. F., Pieters, R., & den Boer, M. L. (2011). Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia. *Leukemia*, *25*(9), 1389-99. http://doi.org/10.1038/leu.2011.105

Singh, R. K., & Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends in Molecular Medicine*, *18*(8), 472-482. http://doi.org/10.1016/j.molmed.2012.06.006

Smith, K. S., Chanda, S.K., Lingbeek, M., Ross, D. T., Botstein, D., van Lohuizen, M., & Cleary M. L. (2003). Bmi-1 regulation of INK4A-ARF is a downstream requirement for transformation of hematopoietic progenitors by E2a-Pbx1. *Mol Cell*, *12*(2), 393-400.

Sonoki, T., Iwanaga, E., Mitsuya, H., & Asou, N. (2005). Insertion of microRNA-125b-1, a human homologue of lin-4, into a rearranged immunoglobulin heavy chain gene locus in a patient with precursor B-cell acute lymphoblastic leukemia. *Leukemia*, *19*, 2009-2010. http://doi.org/10.1038/sj.leu.2403938

Stumpel, D. J., Schneider, P., van Roon, E. H., Boer, J. M., de Lorenzo, P., Valsecchi, M. G., … & Stam, R. W. (2009). Specific promoter methylation identifies different subgroups of MLL-rearranged infant acute lymphoblastic leukemia, influences clinical outcome, and

provides therapeutic options. *Blood*, *114*(27), 5490-8. http://doi.org/10.1182/blood-2009-06-227660

Sulong, S., Moorman, A. V., Irving, J. A., Strefford, J. C., Konn, Z. J., Case, M. C., … & Harrison, C. J. (2009). A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood*, *113*(1), 100-7. http://doi.org/10.1182/blood-2008-07-166801

Takehara, T., Liu, X., Fujimoto, J., Friedman, S. L., & Takahashi, H. (2001). Expression and role of Bcl-xL in human hepatocellular carcinomas. *Hepatology*, *34*(1), 55-61. http://doi.org/10.1053/jhep.2001.25387

Tang, A. H., Neufeld, T. P., Rubin, G. M., & Müller, H. A. (2001). Transcriptional regulation of cytoskeletal functions and segmentation by a novel maternal pair-rule gene, lilliputian. *Development*, *128*(5), 801-13.

Taylor, K. H., Pena-Hernandez, K. E., Davis, J. W., Arthur, G. L., Duff, D. J., Shi, H., … & Caldwell, C. W. (2007). Large-scale CpG methylation analysis identifies novel candidate genes and reveals methylation hotspots in acute lymphoblastic leukemia. *Cancer Res.*, *67*(6), 2617-25. http://doi.org/10.1158/0008-5472.CAN-06-3993

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105-1111. http://doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with

TopHat and Cufflinks. *Nature Protocols, 7*(3), 562-578. http://doi.org/10.1038/nprot.2012.016

Tsuzuki, S., Seto, M., Greaves, M., & Enver, T. (2004). Modeling first-hit functions of the t(12;21) TEL-AML1 translocation in mice. *Proceedings of the National Academy of Sciences of the United States of America, 101*(22), 8443-8448. http://doi.org/10.1073/pnas.0402063101

Twine, N. A., Janitz, K., Wilkins, M. R., & Janitz, M. (2011). Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS ONE, 6*(1), e16266. http://doi.org/10.1371/journal.pone.0016266

Vanharanta, S., Shu, W., Brenet, F., Hakimi, A. A., Heguy, A., Viale, A., … & Massagué, J. (2013). Epigenetic expansion of VHL-HIF signal output drives multi-organ metastasis in renal cancer. *Nature Medicine*, *19*(1), 50-56. http://doi.org/10.1038/nm.3029

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., … & Burge, C. B. (2008). Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature*, *456*(7221), 470-476. http://doi.org/10.1038/nature07509

Wang, E. T., Ward, A. J., Cherone, J. M., Giudice, J., Wang, T. T., Treacy, D. J., … & Burge, C. B. (2015). Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Research*, *25*(6), 858–871. http://doi.org/10.1101/gr.184390.114

Wang, L., Wang, S., & Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, *28*(16), 2184-5. http://doi.org/10.1093/bioinformatics/bts356

Wetzler, M., Dodge, R. K., Mrózek, K., Carroll, A. J., Tantravahi, R., Block, A. W., … & Bloomfield, C. D. (1999). Prospective karyotype analysis in adult acute lymphoblastic leukemia: the cancer and leukemia Group B experience. *Blood*, *93*(11), 3983-93. Retrieved from http://www.bloodjournal.org/content/93/11/3983?variant=long

Woo, J. S., Alberti, M. O., & Tirado, C. A. (2014). Childhood B-acute lymphoblastic leukemia: a genetic update. *Experimental Hematology & Oncology*, *3*, 16. http://doi.org/10.1186/2162-3619-3-16

Zhang, J., & Manley, J. L. (2013). Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discovery*, *3*(11), 10.1158/2159–8290.CD–13–0253. http://doi.org/10.1158/2159-8290.CD-13-0253

Zhang, J., Mullighan, C. G., Harvey, R. C., Wu, G., Chen, X., Edmonson, M. & Hunger, S. P. (2011). Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood*, *118*(11), 3080-3087. http://doi.org/10.1182/blood-2011-03-341412

Zhang, M. Y., Churpek, J. E., Keel, S. B., Walsh, T., Lee, M. K., Loeb, K. R., … & Shimamura, A. (2015). Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nature Genetics*, *47*(2), 180-185. http://doi.org/10.1038/ng.3177

Zhou, Y., You, M. J., Young, K. H., Lin, P., Lu, G., Medeiros, L. J. & Bueso-Ramos, C. E. (2012). Advances in the molecular pathobiology of B-lymphoblastic leukemia. *Hum. Pathol*., *43*(9), 1347-62. http://dx.doi.org/10.1016/j.humpath.2012.02.004

APPENDIX

	Healthy donors							
	HCB11	HCB12	HCB13	HCB15	HCB16	HCB17	HCB18	HCB19
Raw reads	45370007	52383718	50788574	38083710	37026414	38475933	27610245	44975265
Total Reads aligned	40,061,716	44,788,079	43,627,385	32,333,070	33,434,852	35,013,099	25,263,374	42,006,898
Reads QC failed	0	0	0	0	0	0	0	0
Optical/PCR duplicate	0	0	0	0	0	0	0	0
Non Primary Hits	5,779,454	6,237,085	5,977,381	4,304,285	4,249,987	4,238,068	3,088,355	5,334,181
Unmapped reads	0	0	0	0	0	0	0	0
Multiple mapped reads	2,087,020	2,200,116	2,082,332	1,675,683	1,407,004	1,423,610	1,021,579	1,754,035
Uniquely mapped	37,974,696	42,587,963	41,545,054	30,657,387	32,027,848	33,589,490	24,241,795	40,252,862
% Uniquely mapped	83.7	81.3	81.8	80.5	86.5	87.3	87.8	89.5
Read-1	0	0	0	0	0	0	0	0
Read-2	0	0	0	0	0	0	0	0
Reads map to '+'	15,638,595	17,838,159	17,606,893	12,770,649	13,607,785	14,201,953	10,145,281	16,880,920
Reads map to '-'	18,966,339	21,061,997	20,303,875	15,277,217	15,688,415	16,501,251	11,904,178	19,862,999
Non-splice reads	31,228,571	34,479,087	33,782,113	24,502,591	26,414,387	27,618,743	19,577,648	32,658,679
Splice reads	3,376,363	4,421,069	4,128,655	3,545,275	2,881,813	3,084,461	2,471,811	4,085,240

Supplementary Table 1

	B-ALL patients							
	B-ALL20	B-ALL23	B-ALL24	B-ALL26	B-ALL30	B-ALL37	B-ALL18	B-ALL19
Raw reads	50032958	54614253	56376631	42722037	52718310	63454534	42183675	53607546
Total Reads aligned	42,528,014	47,186,715	46,567,097	36,826,396	42,438,240	51,905,809	36,362,328	45,191,161
Reads QC failed	0	0	0	0	0	0	0	0
Optical/PCR duplicate	0	0	0	0	0	0	0	0
Non Primary Hits	5,019,186	6,256,098	5,684,306	5,044,960	5,000,916	6,601,771	4629827	5,119,497
Unmapped reads	0	0	0	0	0	0	0	0
Multiple mapped reads	2,301,516	2,785,327	2,480,572	2,007,936	2,372,324	3,045,818	1,856,082	2,358,732
Uniquely mapped	40,226,498	44,401,388	44,086,525	34,818,460	40,065,916	48,859,991	34,506,246	42,832,429
% Uniquely mapped	80.4	81.3	78.2	81.5	76	77	81.8	79.9
Read-1	0	0	0	0	0	0	0	0
Read-2	0	0	0	0	0	0	0	0
Reads map to '+'	18,450,843	20,359,828	20,270,776	15,929,453	18,472,492	22,521,427	15807483	19,730,537
Reads map to '-'	18,314,087	20,250,024	20,167,337	15,897,181	18,351,714	22,401,926	15770496	19,602,882
Non-splice reads	27,604,708	32,109,365	30,959,479	24,861,531	28,199,870	35,917,416	25273538	31,629,680
Splice reads	9,160,222	8,500,487	9,478,634	6,965,103	8,624,336	9,005,937	6304441	7,703,739

Supplementary Table 2

	Healthy donors							
	HCB11	HCB12	HCB13	HCB15	HCB16	HCB17	HCB18	HCB19
Total splicing events	3589751	4699749	4389391	3772446	3064685	3279562	2626804	4351674
Known splicing events	3404878	4424894	4166434	3584675	2895880	3096573	2497134	4141947
% Known splicing events	94.849977	94.151709	94.920548	95.022566	94.49193	94.420322	95.06358	95.18054
Partial novel splicing events	130700	190516	155412	135738	121659	134038	96866	149544
% Partial novel splicing events	3.6409210	4.0537484	3.5406278	3.598143	3.9697065	4.0870702	3.687599	3.436470
Novel splicing events	53286	83273	66588	51459	46477	48421	32362	59159
% Novel splicing events	1.4843926	1.7718605	1.5170213	1.3640752	1.5165343	1.4764472	1.231991	1.359453
Total splicing junctions	138558	153466	149824	134191	127733	133619	124285	138593
Known splicing junctions	109887	115208	115759	106629	101652	104470	101095	109554
%Known splicing junctions	0.27429429	0.2572291	0.2653356	0.3297831	0.3040301	0.298374	0.400164	0.2608
Partial novel splicing junctions	21469	28066	25157	20874	19566	21695	17600	21281
Novel splicing junctions	7202	10192	8908	6688	6515	7454	5590	7758

Supplementary Table 3

	B-ALL patients							
	B-ALL20	B-ALL23	B-ALL24	B-ALL26	B-ALL30	B-ALL37	B-ALL18	B-ALL19
Total splicing events	9851538	9144385	10181969	7488233	9277409	9675573	6784995	8263922
Known splicing events	9385328	8712925	9745204	7115059	8893958	9190657	6384977	7783632
% Known splicing events	95.267642	95.2816947	95.710407	95.016528	95.866831	94.9882451	94.104373	94.188110
Partial novel splicing events	289698	286133	298921	249899	264558	333204	263620	338216
% Partial novel splicing events	2.9406372	3.12905679	2.9357878	3.3372225	2.8516367	3.44376504	3.8853381	4.0926814
Novel splicing events	175497	144090	136867	122425	117416	150800	135572	141449
% Novel splicing events	1.7814172	1.57572106	1.3442096	1.6348984	1.265612	1.55856403	1.9981149	1.7116449
Total splicing junctions	189195	193244	188890	190806	171837	202554	192485	201765
Known splicing junctions	143697	141811	140489	141994	132994	144514	137141	143879
%Known splicing junctions	0.3378878	0.30053162	0.30169155	0.38557669	0.31338246	0.27841585	0.3771513	0.3183786
Partial novel splicing junctions	31236	35723	34680	35192	27770	42664	40063	44575
Novel splicing junctions	14262	15710	13721	13620	11073	15376	15281	13311

Supplementary Table 4

	Number of reads							
Group feature	HCB11	HCB12	HCB13	HCB15	HCB16	HCB17	HCB18	HCB19
CDS_Exons	11530604	14581064	14007722	11639516	9755829	10127011	8275239	13354468
% CDS_Exons	28.44807942	31.6562045	31.3933102	34.599826	28.6016	28.278345	31.92736	30.84546
5'UTR_Exons	2767789	2630653	2525129	2392435	1571717	1909197	1540451	2311559
% 5'UTR_Exons	6.828634588	5.71127658	5.65917556	7.1117935	4.607873	5.3311813	5.9433369	5.3391196
3'UTR_Exons	3651463	4705182	4404389	3723887	3693864	4040499	2719995	5260576
% 3'UTR_Exons	9.008817703	10.2151807	9.87086623	11.069691	10.829466	11.282562	10.49423	12.150607
Introns	18400918	19451595	19768486	12532754	15610704	16223904	10792210	17367931
% Introns	45.39838301	42.230366	44.3040069	37.25508	45.766599	45.303116	41.638287	40.11555
% Intergenic	10.31608528	10.1869723	8.77264109	9.9636095	10.194462	9.8047956	9.9967869	11.549264
Total Tags	40532100	46060683	44620086	33640389	34109382	35811894	25918958	43294760

Supplementary Table 5

	Number of reads							
Group feature	B-ALL20	B-ALL23	B-ALL24	B-ALL26	B-ALL30	B-ALL37	B-ALL18	B-ALL19
CDS_Exons	27045723	26161597	28923849	21298072	25760423	28015015	19459838	23753628
% CDS_Exons	55.0504871	49.5986743	54.293404	51.290235	53.053297	48.4858147	48.130223	47.289245
5'UTR_Exons	1517741	1882774	2396034	1250153	2159858	2189243	1257368	1826434
% 5'UTR_Exons	3.08930108	3.56947225	4.4976325	3.0106312	4.4482029	3.7889407	3.1098615	3.6361049
3'UTR_Exons	9624804	8968305	9510804	6632308	8544238	9673994	6174012	7487293
% 3'UTR_Exons	19.5909035	17.0026333	17.852877	15.971992	17.596761	16.7428602	15.270249	14.905867
Introns	5610895	8816111	7334080	8702890	7075890	12207271	9557997	12372353
% Introns	11.4207523	16.714095	13.766915	20.958389	14.572715	21.1272234	23.639895	24.631152
% Intergenic	10.848556	13.1151251	9.5891711	8.7687534	10.329025	9.85516102	9.8497713	9.53763
Total Tags	49128944	52746565	53273228	41524614	48555744	57779817	40431639	50230508

Supplementary Table 6

VITA

Olha Kholod was born on May 19th, 1993 in Kyiv, Ukraine. Olha grew up with her parents Volodymyr (Father) and Halyna (Mother) and her elder sister Mariia. Olha attended Taras Shevchenko National University of Kyiv during 2010-2016. In June, 2014, Olha graduated with a bachelor degree in Biology. In June, 2016, Olha graduated with master degree in Genetics from the same institution. In May, 2015, Olha awarded Fulbright Graduate Scholarship. In August 2015, Olha came to the University of Missouri, Columbia and joined Dr. Taylor's laboratory to pursue a Master of Science degree in Pathology. After she completes her M.S. degree in May 2017, Olha will continue her post-academic training in Dr. Nathan Sheffield's laboratory at University of Virginia in Charlottesville.