5-2009

Pax5 Haploinsufficiency Cooperates with BCR-ABL1 to Induce Acute Lymphoblastic Leukemia

Christopher B. Miller
University of Tennessee Health Science Center

Pax5 Haploinsufficiency Cooperates with BCR-ABL1 to Induce Acute Lymphoblastic Leukemia

DOI
10.21007/etd.cghs.2009.0212

Pax5 HAPLOINSUFFICIENCY COOPERATES WITH *BCR-ABL1* TO INDUCE ACUTE LYMPHOBLASTIC LEUKEMIA

A Dissertation
Presented for
The Graduate Studies Council
The University of Tennessee
Health Science Center

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy
From The University of Tennessee

By
Christopher B. Miller
May 2009

"But seek first His kingdom and His righteousness,
and all these things will be added to you."
Matthew 6:33

"Call to Me and I will answer you,
and I will tell you great and mighty things, which you do not know."
Jeremiah 33:3

"The secret of my success? It is simple. It is found in the Bible."
George Washington Carver

DEDICATION

I would like to dedicate this dissertation to my wife Stephanie, my son Josiah, and my Jesus. Thank you for your love and patience. May our Lord and Savior bless our lives beyond measure and allow us to be a blessing to others as you have been to me.

ACKNOWLEDGEMENTS

I would like to start by thanking my research advisor, Dr. James R. Downing, for selecting me to join his laboratory five years ago and for his commitment to my training. He has helped me to develop the ability to determine "what is the question?" That is, what is the relevant, defining, and most important question that we are trying to answer for a given experiment or set of experiments. Dr. Downing has also been instrumental in nurturing my desire to pursue clinical oncology, and for that I am very grateful.

I would like to continue by thanking the members of my graduate committee, Dr. Suzanne Baker, Dr. Dario Campana, Dr. Lawrence Pfeffer, and Dr. Gerard Zambetti, for their commitment to my education and development as an independent investigator.

The members of the Downing laboratory have been an integral part of my pre-doctoral education. It was from them that I learned the ins and outs of laboratory culture and etiquette. I would like to extend my thanks to Drs. Charles Mullighan, Noel Lenny, Jinjun Dang, Masami Ishii, Salil Goorha, Ina Radtke, Joy Nakitandwe and to Claire Boltz, and Zhongling Cai for their continued support over the past five years. I would like to offer my support and encouragement to Jinjun Cheng, a current graduate student in Dr. Downing's lab; may he be blessed as I have been.

During my final year as a graduate student I received a T32 grant from the CTSI (Clinical Translational Science Institute) at the University of Tennessee. This grant provided funding to help support my salary as a graduate student. I would like to thank the CTSI committee members for selecting me to receive this grant; I hope that the NIH rewards your hard work and dedication to the training of translational scientists.

I would also like to thank Dr. Richard Williams and the members of his laboratory for their help in teaching me the *BCR-ABL1* transduction method and for helping me to establish a murine model of *BCR-ABL1* leukemia in the Downing lab. I would like to offer special thanks to Dr. Williams, whom I consider an honorary committee member, for the many constructive conversations regarding *Pax5*, *Arf*, *BCR-ABL1* and life in general.

Finally, I would like to thank my family, my wife Stephanie, my son Josiah and my child on the way. What is life worth if it cannot be shared with a family like you? Thank you, Stephanie for being patient and waiting nine years instead of four for me to complete medical school. This training period has tested our family and we have come a long way and I would not trade this time for anything, thank you, God be with us!

Chapter 2 acknowledgements are as follows: I would like to thank Z. Cai and L. Phillips for technical assistance; B. Schulman for modeling of PAX5 mutations; M. Roussel, M. Busslinger, and J. Hagman for providing reagents; the Tumor Processing Laboratory of SJCRH for providing tumor samples; and C Li for discussions and modifications of dChipSNP. This work was supported by grants from the National

Cancer Institute (to J.R.D. and W.E.E.), the National Institute of General Medical Sciences (to M.V.R.), the National Health and Medical Research Council (Australia) (to C.G.M.), the Hematology Society of Australia and New Zealand (to C.G.M.), the Royal Australasian College of Physicians (to C.G.M.), and the American Lebanese and Syrian Associated Charities (ALSAC) of SJCRH. Author contributions were as follows: C.G.M. performed the SNP analysis including all bioinformatic studies, I.R. summarized the sequencing analysis, and C.B.M. performed experiments delineating the extent of the deletions identified by the SNP analysis, identified and cloned all *PAX5* mutants and *PAX5* translocations, performed western blots, and assessed the function of the various PAX5 mutants using the *in vitro* assays described in this chapter.

Chapter 3 acknowledgements are as follows: I would like to thank Dr. Brenda Schulman for her assistance in modeling the Pax5 point mutations, Dr. Ina Radtke for help with the DNA sequence analysis, the members of the St. Jude Functional Genomics and Flow Cytometry laboratories for their technical assistance, Dr. Kelli Boyd for histological analysis, Dr. M. Busslinger for providing *Pax5*$^{-/-}$ mice, Dr. C. Sherr for providing *Arf*$^{-/-}$ mice, and Dr. O. Witte for providing the MSCV-GFP-IRES-p185 construct. This study was supported by a Cancer Center Core Grant (2 P30CA021765-30) from the National Cancer Institute; a Leukemia and Lymphoma Society Specialized Center of Research grant (LLS7015-09); the American Lebanese Syrian Associated Charities (ALSAC) of St. Jude Children's Research Hospital. Author contributions were as follows: C.B. M. performed all experiments, analyzed all data, and wrote the manuscript with J.R.D. X. S. established a bioinformatic pipeline for all aCGH and gene expression data, A. A., helped in the analysis and interpretation of gene expression data, J.D. performed ENU the experiment, and all other authors provided valuable insights during the development of this project.

Chapter 4 acknowledgements are as follows: I would like to thank Z. Cai for technical help, K. Rakestraw and J. Armstrong for assistance with sequencing, R. Williams and C. Sherr for the provision of Arf null hematopoietic cells and *BCR–ABL1* retroviral vectors, O. Heidenreich for providing the SKNO-1 cell line, and D. Campana for providing the OP1 cell line. This study was supported by the American Lebanese Syrian Associated Charities of St Jude Children's Research Hospital. C.G.M. was supported by grants from the National Health and Medical Research Council (Australia), the Royal Australasian College of Physicians, and the Haematology Society of Australasia. Author contributions were as follows: C.G.M. performed the SNP analysis including all bioinformatic studies, C.B.M. performed experiments delineating the extent of the deletions identified by the SNP analysis, identified and cloned all *IKAROS* mutants, performed western blots, and aided in the assessment of RAGs role in the deletion of *IKAROS*.

ABSTRACT

Acute lymphoblastic leukemia (ALL) is the commonest pediatric malignancy and comprises several distinct subtypes each with its own unique pathogenesis, clinical behavior, and response to therapy. Chromosomal aberrations are a hallmark of ALL but alone fail to induce leukemia. Pediatric ALLs can be divided into several categories based on the expression of several genetically conserved chromosomal translocations including the t(9,22)[*BCR-ABL1*], t(1,19)[*TCF3-PBX1*], t(12,21)[*ETV6-RUNX1*], *MLL* rearranged leukemia's, hyperdiploid and hypodiploid karyotypes, and T-lineage leukemia. Each translocation confers a characteristic transforming phenotype within the cell in which it originates but is alone insufficient to induce overt leukemia. In order to identify oncogenic lesions that cooperate with the aforementioned initiator lesions, we have performed genome-wide analysis of leukemia cells from 242 pediatric ALL patients using high resolution, single-nucleotide polymorphism microarrays. Our analysis revealed deletion, amplification, point mutation, and structural rearrangements in genes encoding principal regulators of B lymphocyte development and differentiation in 40% of B-progenitor ALL cases. The *PAX5* gene was the most frequent target of somatic mutation, being altered in 31.7% of cases. The identified *PAX5* mutations resulted in reduced levels of PAX5 protein or the generation of hypomorphic alleles. Deletions were also detected in *TCF3* (also known as *E2A*), *EBF1*, *LEF1*, *IKZF1* (IKAROS), and *IKZF3* (AIOLOS). These findings suggest that direct disruption of pathways controlling B-cell development and differentiation contributes to B-progenitor ALL pathogenesis. Moreover, these data demonstrate the power of high-resolution, genome-wide approaches to identify new molecular lesions in cancer.

The Philadelphia chromosome, a chromosomal abnormality that encodes *BCR-ABL1*, is the defining lesion of chronic myelogenous leukemia (CML) and a subset of ALL. To specifically define oncogenic lesions that cooperate with *BCR-ABL1* to induce ALL, we subsequently performed genome-wide analysis of diagnostic leukemia samples from 304 individuals with ALL, including 43 *BCR-ABL1* B-progenitor ALLs and 23 CMLcases. *IKZF1* (encoding the transcription factor Ikaros) was deleted in 83.7% of *BCR-ABL1* ALL, but not chronic-phase CML. Deletion of *IKZF1* was also identified as an acquired lesion at the time of transformation of CML to ALL (lymphoid blast crisis). The *IKZF1* deletions resulted in haploinsufficiency, expression of a dominant-negative Ikaros isoform, or the complete loss of Ikaros expression. Sequencing of *IKZF1* deletion breakpoints suggested that aberrant RAG-mediated recombination is responsible for the deletions. These findings suggest that genetic lesions resulting in the loss of Ikaros function are an important event in the development of *BCR-ABL1* ALL.

In order to assess the contribution of the loss of B-cell developmental regulatory genes with *BCR-ABL1* we performed bone marrow transplant assays. *Pax5* haploinsufficiency was shown to cooperate with *BCR-ABL1* during leukemogenesis. Furthermore, as seen in human ALL, both *Pax5* and *p19Arf* haploinsufficiency further cooperate during leukemogenesis. Pathological analysis of the leukemias revealed a B-lymphoid phenotype suggesting that this model results in the development of ALL.

Secondary transplant studies confirmed that this was ALL and not a lymphoproliferative disorder. Immunophenotypic analysis confirmed the B-ALL phenotype and further revealed striking differences between *Pax5*$^{+/+}$, *Pax5*$^{+/-}$, and *Pax5*$^{-/-}$ leukemias. The leukemias that have lost either one or both *Pax5* alleles revealed a more immature immunophenotype that was most pronounced in those with bi-allelic *Pax5* loss. The wild-type leukemias were consistent with a Hardy fraction C immunophenotype while the Pax5 null leukemias were akin to Hardy fraction A with no expression of any definitive B-cell surface antigens except B220 and CD43.

The *Pax5*$^{+/+}$ and *Pax5*$^{+/-}$ leukemias were monoclonal while the *Arf*$^{+/-}$ and the compound heterozygous (*Pax5*$^{+/-}$ *Arf*$^{+/-}$) leukemias were oligoclonal suggesting that the loss of p19Arf confers greater leukemogenic properties to a cell than does Pax5 loss. This is substantiated by the data that Arf heterozygous animals are tumor prone alone while Pax5 heterozygous animals live full normal lives.

Genomic analysis of 50 murine leukemias (15 WT, 25 *Pax5*$^{+/-}$ and 10 *Arf*$^{+/-}$) revealed that some of the same genomic abnormalities found in human ALL also develop in our murine ALL model. This finding suggests that our model is accurately recapitulating the development of human ALL. Furthermore, we have found through extensive gene expression studies that our mouse BCR-ABL1 leukemias share a gene expression profile similar to human BCR-ABL1 leukemias further supporting our murine model of ALL.

The gene expression studies that we have performed have also given us insight into the role that *Pax5* haploinsufficiency may be playing during leukemogenesis. Recently D.J Wong *et al* performed an exhaustive study of embryonic and adult stem cell gene expression profiling. He found that there are modules of genes that are common to either embryonic or adult stem cells. He went further to delineate a group of genes that are commonly expressed in both murine and human embryonic stem cells and call this the core ESC-like module. We performed gene set enrichment analysis (GSEA) using this core ESC-like module and found that this geneset is significantly enriched in our murine leukemias. In addition, we found that this core ESC-like module was significantly enriched in normal Hardy fraction B cells, but not in A, or C-F. We also found, using a principal component analysis method, that *Pax5*$^{+/+}$ leukemias are most similar to Hardy fraction C while leukemias that have lost either one or both *Pax5* alleles are most similar to Hardy fraction B suggesting that the loss of Pax5 blocks B-cell development and results in cells that are more similar to a stage that shares similar expression of embryonic stem cell genes. Taken together this data suggests that by losing Pax5 the leukemia becomes more stem cell like and may gain advantages that other B-cells do not because they continue down the road of differentiation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
GENERAL INTRDUCTION

Acute lymphoblastic leukemia (ALL) is the commonest pediatric malignancy accounting for 75% to 85% of all leukemia's[1] and occurs in roughly 31 per million children under the age of 15 each year.[2-4] In contrast, adult ALL comprises 15% of leukemia's with an annual incidence estimated at 22 per million in individual's ≥25 years old (www.seer.cancer.gov). ALL is distinguished from lymphoblastic lymphoma by the presence of ≥25% lymphoid blasts in the bone marrow. Currently, the World Health Organization (WHO) uses an integrated morphologic, immunologic, and cytogenetic disease classification system. The WHO also incorporates current knowledge of molecular lesions associated with different subtypes to help properly classify and risk stratify ALLs.[5] According to the WHO classification system, ALLs can be divided in to several genetically distinct subtypes including B-progenitor leukemia's with the translocations t(9,22)[*BCR-ABL1*], t(1,19)[*TCF3-PBX1*], t(12,21)[*ETV6-RUNX1*], *MLL* rearranged leukemia's, hyperdiploid and hypodiploid karyotypes, and T-lineage leukemia (Figure 1-1). These genetic lesions have been shown to be important in leukemia initiation[6] but cooperating oncogenic lesions are required for leukemia development.

In an effort to identify cooperating oncogenic lesions, our laboratory has recently performed a genome-wide analysis of pediatric ALL using single nucleotide polymorphism microarrays (SNP).[7] The SNP arrays allow us to investigate copy number, that is, the number of alleles that a leukemia has at a given genomic locus. The SNP arrays also reveal information about genotype and copy number loss of heterozygosity (LOH). LOH is a phenomenon that occurs when one allele undergoes gene conversion into another allele. For example, if a leukemia had the following genotype at the *PAX5* locus AB; then one allele is A and the other is B. When the leukemia develops, the SNP data may reveal that the leukemia shows a genotype of AA or BB at the *PAX5* locus and, as a result, one can conclude that LOH has occurred at this locus. Chapter 2 of this dissertation will outline the results of the SNP analysis. Briefly, the SNP analysis identified mutations in genes encoding regulators of B-cell development and differentiation in 40% of B-progenitor ALL cases. Of these genetic lesions, deletion of the transcription factor *PAX5* was the most common mutation occurring in 32% of cases. The result of the *PAX5* mutations was a reduction in the level of the transcription factor as a result of gene deletion/point mutation or generation of an altered form of PAX5 protein.

Pax5, which is also known as B-cell specific activator protein (BSAP), is a transcription factor that is only expressed in the B lymphoid lineage of the hematopoietic system.[8] Pax5 is a major regulator of B-cell development controlling the expression of many B-cell specific genes including *Cd19*, *Blnk*, *Lef1*, *Ebf1*, *Cd79a*, and *Mb1*.[9-14] Pax5 also controls the expression of alternate lineage genes such as *Flt3*, *Notch1*, *Lmo2*, and *Cd28* by down regulating their expression.[15] The expression of Pax5 is so important for proper B-cell development that loss of Pax5 results in an arrest in B-cell development at an early B-cell progenitor stage in *Pax5*$^{-/-}$ mice.[16] Interestingly, Pax5 is so important for

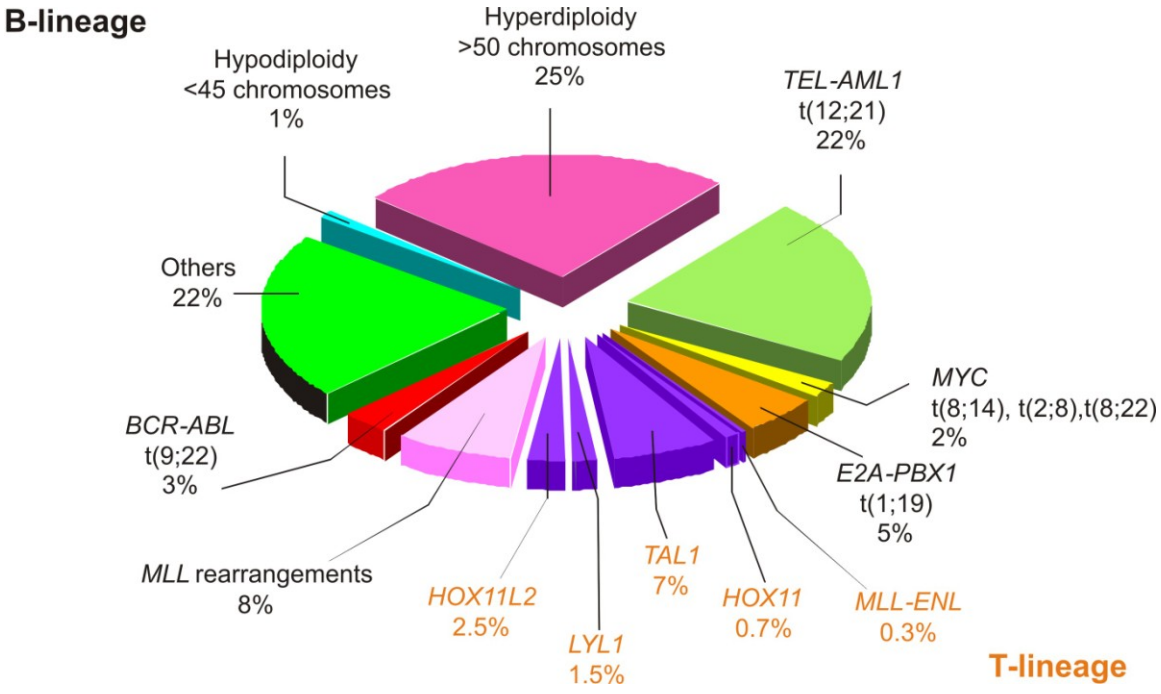


Figure 1-1 Subtypes of pediatric acute leukemia

Acute leukemia can be sub-divided into B-lineage and T-lineage leukemia's based on cell surface antigen immunophenotyping. B-lineage leukemia's are more common than T-lineage in children and can be subdivided into groups based on chromosomal abnormalities such as whole chromosome deletion, amplification, and chromosomal translocations. These chromosomal abnormalities often have prognostic significance i.e. *BCR-ABL* and *MLL* rearranged leukemia's carry a poor prognosis while hyperdiploid and *TEL-AML1* leukemia's are associated with better outcomes. Adapted with permission. Pui, C, Relling, M, and Downing J. Acute lymphoblastic leukemia. *New England Journal of Medicine* **350**, 1535-1548 (2004).[17]

B-cell fate determination that B-cells from *Pax5*$^{-/-}$ mice can dedifferentiate into alternate lineages.[18] Chapter 2 of this dissertation will discuss the effects that *Pax5* deletions and mutations have on the ability of Pax5 to transactivate the known Pax5 targets *Cd19* and *Cd79a*.

The genome-wide SNP study that we performed revealed that regulators of B-cell development are deleted in 40% of ALLs with *PAX5* being the most common target of mutation at 32% of cases. The frequency and type of mutations varied by leukemia subtype, specifically, *PAX5* was deleted in 13/47 (28%) *ETV6-RUNX1*, 7/17 (41%) *TCF3-PBX1*, and 4/9 (44%) *BCR-ABL1* ALL cases.[7] Because the number of *BCR-ABL1* cases analyzed in the study was small, we subsequently analyzed 43 *BCR-ABL1* ALLs including 21 pediatric and 22 adult cases[19] (chapter 4). The, now more robust, analysis revealed that 22/43 (51%) of *BCR-ABL1* ALL cases harbored mutations in the *PAX5* gene. Specifically, 12/21 (57%) of pediatric ALL and 10/22 (45%) of adult ALL cases harbored copy numbers abnormalities (CNA) at the *PAX5* locus. The *PAX5* mutations resulted in the functional loss of one *PAX5* allele, suggesting that *PAX5* haploinsufficiency may cooperate with *BCR-ABL1* to induce ALL.

The *BCR-ABL1* fusion, also known as, the Philadelphia chromosome (Ph), is the pathognomonic lesion in chronic myelogenous leukemia (CML).[20] Ph positive leukemia also accounts for 5% of B-progenitor pediatric ALL and 40% of adult ALL, in addition to CML.[21,22] The *BCR-ABL1* fusion arises from the t(9,22)(q34;q11.2) translocation which leads to the fusion of the *ABL* proto-oncogene from chromosome 9 with the breakpoint cluster region (*BCR*) gene on chromosome 22.[23,24] The generation of the Ph chromosome can lead to the formation of alternative products depending on the context of the leukemia. The $p210^{BCR-ABL1}$ and the $p185^{BCR-ABL1}$ fusion proteins are typically found in CML and Ph+ ALL, respectively.[25-27] The BCR-ABL1 protein encodes a constitutively active tyrosine kinase that is important for cell transformation.[26,28] The Ph chromosome has also been shown to be an important negative prognostic indicator in both pediatric and adult ALL.[29-31]

Chapter 3 will describe the generation of a $p185^{BCR-ABL1}$ mouse model of ALL which will also incorporate the additional lesions that we have detected using the SNP analysis. Specifically, both *Pax5* and *p19Arf* haploinsufficiency will be modeled on the *BCR-ABL1* background individually and together in order to assess their ability to cooperate with the Ph chromosome to induce ALL. This model will be analyzed for additional cooperating events, as well as, assessed for its similarity to human Ph+ ALL using gene set enrichment analysis (GSEA). The GSEA analysis will bring this dissertation full circle; from the identification of lesions in human ALL to mouse modeling of those lesions and the comparison of the effects that these lesions have on transcriptional networks in humans and in mice. This analysis will allow us to better understand the human disease and will help us to move closer to developing therapeutics that will eventually cure children with ALL.

CHAPTER 2
GENOME-WIDE ANALYSIS OF GENETIC ALTERATIONS IN ACUTE LYMPHOBLASTIC LEUKEMIA*

2.1 INTRODUCTION

Pediatric acute lymphoblastic leukaemia (ALL) consists of genetically distinct subtypes including B-progenitor leukaemia's with t(9;22)[*BCR-ABL1*], t(1;19)[*TCF3-PBX1*], t(12;21)[*ETV6-RUNX1*], rearrangements of *MLL*, hyper- and hypodiploid karyotypes, and T-lineage leukaemia (T-ALL). These genetic lesions are important in leukemia initiation,[6] but alone are insufficient to generate a full leukemic phenotype, suggesting that cooperating oncogenic lesions are required. Although additional mutations have been identified in a subset of cases, the full complement of cooperating lesions and their distribution within the known genetic subtypes of ALL remains to be defined.

To obtain a comprehensive registry of genetic lesions in ALL, we examined DNA from the leukemic blasts of 242 cases of pediatric ALL using Affymetrix SNP arrays that interrogate over 350,000 loci and permit identification of copy number changes at an average resolution of less than 5kb. In addition, paired copy number and LOH analysis was performed for 228 ALL cases with matched remission samples. Our analyses identified a mean of 6.46 somatic copy number alterations per case, with deletions outnumbering amplification almost 2:1 (Table 2-1). The frequency of genomic deletions and amplifications varied significantly between ALL subtypes (Table 2-1). Genomic gains were frequent in B-progenitor ALL with high hyperdiploidy, but were rare in other B-ALL subtypes. By contrast, deletions were more frequent, ranging from 6 deletions per case in some genetic subtypes (*ETV6-RUNX1* and hypodiploid ALL) to only a single deletion per case in *MLL*-rearranged ALL.

Fifty-four recurrent somatic regions of deletion were identified, with the minimal deleted regions typically measuring less than 1 Mb in size, and 24 of the deletions containing only a single gene. None were present in the germline samples. Although technical aspects of the methodology used could theoretically lead to false positive or negative results, FISH and/or qPCR confirmatory studies described below validated each of the examined lesions. The recurring deletions included 3p14.2 (*FHIT*),[32] 6q16.2-3 (including CCNC),[33] 9p21.3 (two regions involving *CDKN2A*[34] and *MLLT3*), 12p13.2 (*ETV6*),[35] 11q23 (including *ATM*),[36] 13q14.2 (*RB1*)[37] and 13q14.2-3 (including mir-16-1 and mir-15a).[38] In addition, deletions of other tumor associated genes not previously implicated in ALL were identified including *LEF1*, *BTG1*, and *ERG*. The most striking

---

* Adapted with permission. Mullighan C, Goorha S, Radtke I, Miller C, Coustan-Smith E, Dalton J, Girtman K, Mathew S, Ma J, Pounds S, Su X, Pui C, Relling M, Evans W, Shurtleff S, Downing J. Genome-wide analysis of genetic alterations in acute lymphoblastic leukemia. *Nature* **446**, 758-764 (2007).[7]

Table 2-1 Frequency of genomic amplification and deletions in pediatric ALL

Group	Subtype	N	Amplifications mean ± SD (range)	Deletions mean ± SD (range)	All lesions mean ± SD (range)
B-ALL	Hyperdiploidy with >50 chromosomes	39	9.56 ± 3.59 (5-20)	1.59 ± 2.49 (0-11)	11.13 ± 5.0 (5-27)
	TCF3-PBX1	17	1.59 ± 0.62 (1-3)	2.12 ± 1.17 (1-4)	3.7 ± 1.53 (2-7)
	ETV6-RUNX1	47	0.89 ± 1.51 (0-8)	6.0 ± 4.63 (1-21)	6.68 ± 4.8 (0-21)
	MLL rearranged	11	0.09 ± 0.3 (0-1)	0.91 ± 1.81 (0-6)	1 ± 1.79 (0-6)
	BCR-ABL1	9	4 ± 5.3 (0-12)	4.2 ± 4.15 (0-12)	6.8 ± 4.52 (0-13)
	Hyperdiploidy with 47-50 chromosomes	23	1.70 ± 1.55 (0-7)	3.5 ± 3.12 (0-12)	5.1 ± 4.31 (0-15)
	Hypodiploid	10	1.1 ± 1.91 (0-6)	6.0 ± 4.42 (3-18)	7.1 ± 6.12 (3-24)
	Other	36	1.06 ± 3.21 (0-19)	4.64 ± 5.14 (0-20)	5.58 ± 6.57 (0-23)
	Total	192	2.97 ± 4.28 (0-20)	3.83 ± 4.2 (0-21)	6.63 ± 5.56 (0-27)
T-ALL		50	0.9 ± 1.98 (0-9)	4.9 ± 6.21 (0-30)	5.8 ± 7.12 (0-39)
All cases		242	2.54 ± 4.0 (0-20)	4.06 ± 4.69 (0-38)	6.46 ± 5.90 (0-39)

Lesions exclude inherited copy number variations,[39] as well as focal deletions arising from antigen receptor gene rearrangements at 2p11.2 (*IGKL*), 7p14.1 (*TRGV*), 7q34 (*TRBV*), 14q11.2 (*TRAV*, *TRDV*, *TRDJ*, *TRDC*, and *TRAJ*), 14q32.33 (*IGHV*) and 22q11.22 (*IGLL*). Homozygous deletions are scored as two lesions. There is significant variation in the mean number of amplifications (ANOVA P<0.0001), deletions (P<0.0001), and all lesions (P<0.0001) between B-progenitor ALL subtypes. Amplifications are more common in B-ALL than T-ALL (t test P=0.001).

observation, however, was the identification of genomic alterations in genes that regulate B-lymphocyte differentiation in 40% of B-progenitor ALL cases.

2.2 MATERIALS AND METHODS

2.2.1 Patients and Samples

Two hundred and forty two patients with acute lymphoblastic leukemia (ALL) treated at St Jude Children's Research Hospital (SJCRH) between 1993 and 2005 were studied. These included precursor-B ALL with high hyperdiploidy (greater than 50 chromosomes on karyotyping, HD>50, n=39), *TCF3-PBX1* positive (*E2A-PBX1*, n=17), *ETV6-RUNX1* positive (*TEL-AML1*, n=47), *BCR-ABL1* positive (n=9), *MLL* rearranged (n=11), low hyperdiploidy (47-50 chromosomes, HD47-50, n=23), hypodiploidy (n=10), B-ALL with pseudodiploidy, near haploid karyotype, normal cytogenetics or non-recurring cytogenetic abnormalities (n=36), and T-lineage ALL (T-ALL, n=50).

Informed consent for the use of leukemic cells for research was obtained from patients, parents or guardians in accordance with the Declaration of Helsinki, and study approval was obtained from the SJCRH institutional review board.

Mononuclear cells were purified from the diagnostic bone marrow or peripheral blood samples by density gradient centrifugation and were cryopreserved in liquid nitrogen. Diagnostic samples were characterized by conventional cytogenetics, reverse transcriptase–polymerase chain reaction (RT-PCR) assays for *BCR-ABL1*, *TCF3-PBX1 [E2A-PBX1]*, and *ETV6-RUNX1 [TEL-AML1]*, and for the presence of *MLL* chimeric fusion genes by at least two of the following methods: cytogenetics, 11q23 fluorescence *in situ* hybridization (FISH), or RT-PCR for t(9;11)[*MLL-AF9*], t(11;19)[*MLL-ENL*], t(11;19)[*MLL-ELL*], t(10;11)[*MLL-AF10*], or t(4;11)[*MLL-AF4*]. Remission samples were obtained from peripheral blood collected into EDTA vacutainer tubes at a median time of 200 days (range <1-3009) from diagnosis.

DNA was extracted from diagnostic leukemia samples using the DNA Blood Mini Kit (Qiagen, Valencia, CA). DNA was quantified using either a Nanodrop Spectrophotometer or PicoGreen. Quality was assessed using the Nanodrop and by agarose gel electrophoresis. RNA was extracted from diagnostic samples using TRIzol (Invitrogen, Carlsbad, CA). Cells for FISH analysis were stored in Carnoy's fixative.

2.2.2 Affymetrix Mapping 100K and 500K Single Nucleotide Polymorphism Arrays

Samples were genotyped with Affymetrix 50K GeneChip Human Mapping 50K Hind, 50K Xba, and 250K Sty arrays (Affymetrix, Santa Clara, CA). DNA was restriction enzyme digested, PCR-amplified, purified, labeled, fragmented and hybridized to the arrays according to the manufacturer's instructions. Briefly, 250ng of DNA was

digested with *Xba*I, *Hind*III or *Sty*I (New England Biolabs, Boston, MA). Digested DNA was adaptor-ligated and PCR-amplified using AmpliTaq Gold (Applied Biosystems, Foster City, CA) in four 100µl PCR reactions for each enzyme-digested sample. PCR products from each set of four reactions were pooled, concentrated and fragmented using DNase I. Fragmented PCR products were then labeled, denatured and hybridized to the arrays. Arrays were then washed using Affymetrix fluidics stations, and scanned using the Gene Chip Scanner 3000. CEL files were generated using either Affymetrix GeneChip Operating Software v 3.0 or Affymetrix GeneChip Genotyping Analysis Software (GTYPE) v 4.0. SNP calls were generated using GTYPE. Affymetrix CEL and GTYPE-generated SNP call text files for the 242 blast samples, corresponding remission samples in those cases with *PAX5* or *EBF1* mutations, and the 62 remission samples used as diploid reference samples for copy number analysis have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE5511.

2.2.3 Fluorescence *In Situ* Hybridization (FISH) Analysis

Dual-color FISH was performed on archived bone marrow cells obtained at presentation, treated with Carnoy's fixative and dried onto slides. Probes were derived from bacterial artificial chromosome (BAC) clones (Children's Hospital Oakland Research Institute, Oakland, CA.; Invitrogen; Open Biosystems, Huntsville, AL). BACs used were RP11-586B19 (*EBF1* deletion), RP11-160B5 (*EBF1* deletion, case Hypodip-SNP-#5), RP11-96F2 (*PAX5* deletion), CTD-2535J16 (*PAX5* deletion), RP11-614P24 and RP11-1136K1 (*PAX5-ETV6* fusion), RP11-652D9 and RP11-79P21 (*PAX5-FOXP1* fusion), RP11-652D9 and RP11-73L12 (*PAX5-ZNF521* fusion) RP11-652D9 (*PAX5* amplification in case Hypodip-SNP-#7), CTC-600N23 (*IKZF3* deletion). BACs labeled for control probes were CTD-2194L12 (5p13.2), RP11-235C23 (9q31.2) and RP11-4F24 (17p13.3).

BAC clone identity was verified by T7 and SP6 BAC-end sequencing and by hybridization of fluorescently labeled BAC DNA with normal human metaphase preparations. BACs were labeled with either fluorescein isothiocyante or rhodamine fluorochromes. Target probes were paired with control probes from the opposite chromosomal arm where possible. All probe mixtures were diluted 1:50 in DenHyb buffer (Insitus Biotechnologies, Albuquerque, NM) and co-denatured with the target cells on a hotplate at 90°C for 1 minute. The slides were incubated overnight at 37°C on a slide moat and then washed in 50% formamide/1xSSC at 25°C for 5 minutes. Nuclei and metaphases were counterstained with DAPI (200ng/ml) (Insitus Bio.) for viewing on either an Olympus BX60 or a Nikon Eclipse E800 fluorescence microscope equipped with a 100 watt mercury lamp; FITC, Rhodamine, and DAPI filters; a 100X PlanApo (1.40) oil objective; and a COHU CCD or Photometrics SenSys camera. Images were captured and processed with an exposure time ranging from 0.5-2 seconds for each fluorochrome using Cytovision v3.6 software from Applied Imaging (San Jose, CA). Images were captured and enhanced using Applied Imaging's MacProbe v4.3 software.

2.2.4 Fluorescence Activated Cell Sorting

Sorting of leukemic blasts according to level of CD10 expression was performed using a BD FACSAria Cell-Sorting System and a phycoerythrin-labeled anti-CD10 antibody (BD Pharmingen, BD, Franklin Lakes, NJ).

2.2.5 Genomic Sequencing

All 16 exons of *EBF1* were sequenced using genomic DNA from 8 blast samples with *EBF1* deletions and an additional 106 B-progenitor ALL cases without *EBF1* deletions. DNA was amplified using Accuprime *Pfx* DNA polymerase (Invitrogen) or Accuprime GC-rich DNA polymerase (Invitrogen). Exon sequences were determined aligning the reference *EBF1* mRNA and DNA sequences (Genbank accessions NM_024007.2 and NC_000005.8). The coding regions of all 11 exons of *PAX5* (exons 1A, 1B, and 2-10) and the promoter region of exon 1B were sequenced in all 242 blast samples. Remission samples for patients with blast samples harboring *PAX5* mutations were also sequenced. Primers were designed using Primer 3.[40] PCR amplification was performed according to the manufacturer's instructions using Eppendorf Mastercyclers (Eppendorf North America, Westbury, NY). Thermal cycling parameters for Accuprime *Pfx* DNA polymerase were 95°C for 2 minutes followed by 35 cycles of 95°C for 15 seconds, 60°C for 30 seconds, and 68°C for 60 seconds. Thermal cycling parameters for Accuprime GC-rich DNA polymerase were 95°C for 3 minutes followed by 35 cycles of 95°C for 30 seconds, 60°C (or annealing temperature as indicated in Supplementary Tables 3-5) for 30 seconds, and 72°C for 60 seconds, with a final extension of 10 minutes at 72°C. PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). PCR products were sequenced directly using Big Dye Terminator (v.3.1) chemistry on 3730xl DNA Analyzers (Applied Biosystems, Foster City, CA). Mutations detected by direct sequencing were confirmed by cloning PCR products into either pCR2.1-TOPO (Invitrogen) or pGEM-T Easy (Promega, Madison, WI) vectors, and sequencing multiple clones in both directions using M13 primers.

2.2.6 Modeling of *PAX5* Paired Domain Mutations

A structural view of the PAX5 paired domain was generated by PyMOL v0.99 (http://pymol.sourceforge.net/) using the coordinates of the X-ray structure of PAX5 interacting with ETS1 on DNA, and PAX6 deposited with the Brookhaven Data Bank (PDB: 1K78 and 6pax; http://www.rcsb.org/pdb/).[41,42] The protein sequences of the paired domains of PAX6 and PAX5 are 70.1% identical.

2.2.7 Cell Culture

The human pre-B ALL cell lines Kasumi-2 and REH, the Burkitt lymphoma cell line Raji, and the T-lineage ALL cell lines Jurkat and MOLT-4 (all obtained from the

Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ, Braunschweig, Germany) were grown in RPMI-1640 containing 100 units/ml penicillin, 100 µg/ml streptomycin, 2 mM glutamine and 10% fetal bovine serum (20% for MOLT-4). 293T cells were maintained in DMEM containing 100 units/ml penicillin, 100 µg/ml streptomycin, 2 mM glutamine and 10% fetal bovine serum.

2.2.8 Quantitative RT-PCR and Genomic PCR

PAX5 gene expression of ALL samples was quantitated using Taqman® Assays-on-demand Hs00277134_m1 (specific for *PAX5* exons 4-5) and Hs00172001_m1 (*PAX5* exons 7-8) (Applied Biosystems, Foster City, CA). RNA was extracted using TRIzol (Invitrogen), and reverse transcribed using random hexamer primers and Superscript III Reverse Transcriptase (Invitrogen). Taqman® assays were performed using a 7500 Real-Time PCR system and 7500 System Software (Applied Biosystems, Foster City, CA), using the 7500 universal cycling conditions: 50°C for 2 minutes, followed by 95°C for 10 minutes, then 40 cycles of 95°C for 1 minute and 60°C for 1 minute.

Standard curves for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and *PAX5* gene expression were generated from 10-fold serial dilutions of the human t(12;21) *[ETV6-RUNX1]* REH cell line (American Type Culture Collection, Manassas, VA), which expressed high levels of *PAX5* as determined by gene expression profiling using Affymetrix U133A chips (Affymetrix, Santa Clara, CA) (data not shown). These normalized ratios were compared to each other for differences in overall levels of *PAX5* expression.

Primers for genomic quantitative PCR were designed using Primer Express 3.0 (Applied Biosystems, Foster City, CA). Taqman® RNase P primers (Applied Biosystems, Foster City, CA) were used for control amplification. 200ng of leukemic blast DNA or control human DNA was amplified using the same conditions as those described for RNA real-time PCR. Standard curves for each *PAX5* exon and RNase P were generated using normal human DNA. Assays were performed in duplicate. *PAX5* exon-specific copy numbers values were normalized by dividing the value obtained for *PAX5* by the paired value obtained for RNase P for each sample. Cutoffs of 0.7 and 0.3 were used to identify hemizygous and homozygous *PAX5* deletions, respectively.

2.2.9 Detection of *PAX5* in Leukemic Blasts by Flow Cytometry

We characterized PAX5 expression in bone marrow mononuclear cells obtained at diagnosis from 16 patients with B lineage ALL. After density gradient separation, cells were labeled with anti-CD34 (of IgG2a class) conjugated to allophycocyanin (APC; Miltenyi Biotech, Auburn, CA) and anti-human CD19 (of IgM class; Research Diagnostics, Concord, MA) followed by goat-anti mouse IgM conjugated to phycoerythrin (PE; Jackson ImmunoResearch Laboratories, West Grove, PA). After cell permeabilisation with 8E, a paraformaldehyde-based reagent developed in the laboratory

of Dario Campana (SJCRH), cells were labeled with anti-human PAX5 (of IgG1 class; BD Transduction Labs, San Jose, CA) followed by goat anti-mouse IgG1 conjugated to fluorescein isothiocyanate (FITC; Southern Biotechnology Associates, Birmingham, AL). In parallel tests, anti-CD19 was omitted and cells were also labeled with a goat F(ab')2-anti-human IgM PE (Southern Biotechnology Associates, Birmingham, AL) after permeabilisation. Isotype-matched non-reactive antibodies were used as controls. We used the cell lines Raji and Kasumi-2 as positive controls for PAX5 staining; Molt-4 and Jurkat were used as negative controls. We analyzed 20,000 cells for each antibody combination using a FACSCalibur flow cytometer (BD Biosciences) and either CellQuestPro (BD Biosciences) or FlowJo (Treestar Inc, Ashland, OR) software.

To show the B-cell specificity of the PAX5 antibody, we examined PAX5 expression in peripheral blood mononuclear cells from a healthy donor. After density gradient separation, cells were labeled with anti-CD3 (of IgG2a class) conjugated to Pacific Blue (Invitrogen) and goat F(ab')2-anti -human IgM PE. After cell permeabilisation with 8E, cells were labeled with the anti-human PAX5 antibody followed by goat-anti-mouse IgG1 FITC. In a parallel tube, we used an isotype-matched non-reactive control antibody (DakoCytomation, Carpinteria, CA) in place of PAX5 followed by goat-anti-mouse IgG1 FITC (Southern Biotechnology Associates). We analyzed 10,000 cells using an LSR II flow cytometer (BD Biosciences) and CellQuestPro. The PAX5 antibody stained B-lymphocytes, but not T-lymphocytes.

To demonstrate the PAX5-specificity of the anti-PAX5 antibody, we synthesized 25 12mer overlapping peptides corresponding to the immunogen used to raise the PAX5 antibody (residues 151-306). Peptides were synthesized using an Aapptec 396 Multiple Organic Synthesizer (Aapptec, Louisville, KY). Peptide purity and quality was assessed by high performance liquid chromatography (HPLC) and matrix assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry. Peptide sequences are available upon request. PAX5, CD79A (Dako Cytomation) or mouse IgG1 antibody were preincubated with 15ng of either the PAX5 peptide pool or a non-PAX5 negative control peptide (corresponding to an internal sequence of SMAC) and then used to stain either peripheral blood mononuclear cells (PBMNCs) obtained from a normal donor or the Raji cell line. The PAX5 peptide pool specifically abolished staining of B lymphocytes by PAX5 but not CD79A. No blocking was seen with the control peptide and no reactivity with mouse IgG1 antibody was seen. These results demonstrate the PAX5 specificity of the anti-PAX5 antibody used for quantitation of leukemic blast intracellular PAX5 levels.

2.2.10 Cloning of *PAX5* and *EBF1* Wild-Type and Mutant Alleles

Total RNA was extracted from leukemia blast samples using TRIzol (Invitrogen) and 1μg of total RNA was reverse transcribed using Superscript III (Invitrogen). The coding region of the exon 1a isoform of *PAX5* was amplified using primers C282 and C302, the exon 1b isoform was amplified using primers C317 and C302, and the entire coding region of EBF1 was amplified using primers C503 and C504 using 1μl of cDNA and the Advantage 2 PCR Kit (Clontech, Mountain View, CA). Thermal cycling

conditions were 5 cycles of 94°C for 30 sec and 72°C 3 min, followed by 5 cycles of 94°C for 30 sec, 70°C for 30 sec and 72°C 3 min, followed by 25 cycles of 94°C for 30 sec, 68°C for 30 sec and 72°C 3 min. PCR products were sequenced directly and after cloning into either pCR2.1-TOPO (Invitrogen) or pGEM-T Easy (Promega, Madison, WI) vectors.

2.2.11 Identification, Detection, and Cloning of *PAX5* Translocations

The *PAX5-ETV6 [PAX5-TEL]* translocation was detected using primers PAX5ex3-F1 and ETV6ex3-R1 as previously described.[43] The coding region of *PAX5-ETV6* mRNA was amplified using primers C282 and C283. 3' rapid amplification of cDNA ends (RACE) was performed using the BD Smart RACE™ amplification kit (Clontech, Mountain View, CA) according to the manufacturer's instructions, using the supplied universal primer mix and gene specific primers C294 (*PAX5* exon 3) and C300 (exon 1a). Nested PCR was performed using the supplied nested universal primer, and nested gene-specific primer C295. PCR products were gel-purified and sequenced directly and after cloning into pCR2.1-TOPO (Invitrogen). Following identification of *PAX5-FOXP1* and *PAX5-ZNF521 [PAX-EVI3]* fusions by RACE, translocation-specific RT-PCR was performed using primers C303 and C304 for *PAX5-FOXP1*, and C303 and C309 for *PAX5-ZNF521.* The coding region of the *PAX5-FOXP1* mRNA was amplified using primers C334 and C335, and *PAX5-ZNF521* with primers C326 and C327, using Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) and the following thermal cycling parameters: 98°C for 60 seconds, followed by 35 cycles of 98°C for 10 seconds, 72°C for 90 sec (*PAX5-FOXP1*) or 3 minutes (*PAX5-ZNF521)* followed by a final extension step of 72°C for 10 min.

2.2.12 Reporter Assays

The coding regions of the exon 1a isoforms of wild type *PAX5* and *PAX5* mutations, deletions and fusions were cloned into the *Xho*I site of the MSCV-IRES-mRFP (MIR) vector by blunt-ended ligation. This vector was created by replacing the green fluorescent protein (GFP) cassette of the MSCV-IRES-GFP vector[44] with a monomeric red fluorescent protein cassette[45] (kindly provided by Martine Roussel of St Jude Children's Research Hospital, Memphis, TN). Twenty-four hours after plating, $2x10^5$ 293T cells were transfected with 1 μg of wild-type MIR-PAX5, mutant MIR-PAX5 or MIR without PAX5 plasmid DNA, 1 μg of *luc*-CD19 reporter plasmid DNA (kindly provided by Meinrad Busslinger, Vienna, Austria),[46] and 0.1 μg of pRL-TK *Renilla* luciferase plasmid DNA (Promega) using FuGENE 6 (Roche Diagnostics, Alameda, CA). Forty-eight hours post-transfection, cell lysis and measurement of firefly and *Renilla* luciferase activity was performed using the Dual-Luciferase® Reporter Assay System (Promega) according to the manufacturer's instructions. Transfections were performed in triplicate. The firefly luciferase activity was normalized according to corresponding *Renilla* luciferase activity, and luciferase activity was reported as mean (± s.e.m.) relative to the *luc*-CD19/PAX5 WT transfection. For competition assays, in which

increasing amounts of either *PAX5-ETV6* or *PAX5-FOXP1* vector was transfected with fixed amounts of *PAX5* wild type vector, "empty" MSCV-IRES-mRFP vector was also used to maintain a constant mass of expression vector in each experiment.

2.2.13 Electrophoretic Mobility Gel-Shift Assays

Nuclear extracts of 293T cells transfected with wild type *PAX5* or *PAX5* mutant alleles were prepared using the method of Andrews and Faller.[47] Equivalent expression of each PAX5 variant was confirmed by western blotting of nuclear extracts (data not shown). Two micrograms of protein was incubated with [γ-^{32}P]ATP end-labeled double-stranded oligonucleotides containing a CD19 promoter PAX5 binding site[10] using the Gel Shift Assay System (Promega). A mutated CD19 binding site was used as a non-specific competitor. Complexes were supershifted using a PAX5 N-terminus specific rabbit polyclonal antibody (Chemicon, Temecula, CA) and were resolved using 6% DNA retardation gels (Invitrogen).

2.2.14 Western Blotting

Four million leukemic blasts were washed with PBS and lysed with 4x LDS sample buffer (Invitrogen). For nuclear extracts, 10µl of nuclear lysate was blotted. Lysates were separated using NuPAGE 10% Bis-Tris gels, transferred to nitrocellulose membranes, and after blocking were incubated with N-terminal (Chemicon) or C-terminal PAX5 antibodies (Santa Cruz Biotechnology, Santa Cruz, CA), C-terminal Actin (Santa Cruz), and as a control for blotting of nuclear extracts, N-terminal DEK antibody (BD Biosciences). Nuclear extracts of 558LµM cells were incubated with PCNA antibody (Santa Cruz). Following incubation with secondary antibody, blots were developed using the SuperSignal West Femto Max Sensitivity Substrate chemiluminescent reagent (Pierce, Rockford, IL) and exposed to film or captured using a ChemiDoc XRS gel imaging system (Bio-Rad Laboratories, Hercules, CA).

2.2.15 Transduction and Analysis of IgM Expression by 558LµM Cells

The terminally differentiated B-cell line 558LµM is a derivative of the J558L plasmacytoma cell line stably transfected with a construct expressing three of four components required for surface IgM expression: IgM heavy chain, immunoglobulin lambda light chain, and Ig-β. The cells do not express mb-1, hence the cells do not produce Ig-α and surface IgM is not expressed. mb-1 expression is dependent on Pax5, thus following transduction of 558LµM cells with ecotropic retroviruses expressing PAX5, surface IgM expression serves as a useful readout of the transactivating activity of PAX5 variant alleles.

558LµM cells were grown in RPMI 1640 media (Invitrogen) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT), 2mM L-glutamine (Invitrogen), 50mg/ml

gentamicin (Invitrogen), 0.3μg/ml Xanthine (Sigma, St Louis, MO), and 1μg/ml mycophenolic acid (Sigma) as previously described.[48] The Phoenix packaging system was used to generate ecotropic retrovirus expressing PAX5 cloned into either the MSCV-IRES-mRFP or MSCV-IRES-YFP constructs, as described above. 500,000 558LμM cells were transduced in six-well dishes with 3 ml of retroviral supernatant and 8μg/ml polybrene. One day post transduction cells were transferred to 10ml of fresh media. Cells were harvested for flow cytometric analysis and fluorescence activated cell sorting (FACS) day 3 post transduction.

Samples were stained with fluorescein isothiocyanate conjugated anti-mouse IgM (Southern Biotechnology Associates, Birmingham, AL) and a rat IgG2aκ isotype control antibody (BD Pharmingen) on a FACSVantage SE (BD Biosciences). 100,000 viable events were collected for analysis. IgM expression was analyzed in the RFP or YFP positive populations. Two million RFP-positive cells were flow sorted for western blotting.

2.2.16 Methylation Analysis

Methylation status of the promoter regions of *EBF1*, *PAX5* exons 1a and 1b, and *IKZF1* (Ikaros) was analyzed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) of PCR-amplified, bisulfite-modified leukemic blast DNA, as previously described (Sequenom, San Diego, CA).[49,50] This method allows semiquantitative, high-throughput analysis of methylation status of multiple CpG units in each amplicon generated by base-specific cleavage.

Ninety-six samples were examined. PCR reactions were designed using annotated CpG island data obtained from the University of California Santa Cruz genome browser. The *PAX5* gene contains several CpG islands, one of which (Chr 9:37024136-37028341) lies upstream of the coding region of *PAX5* exon 1a. Two amplicons were used to examine this region: X014_PAX5 lies immediately upstream of the *PAX5* exon 1a coding region in the known regulatory region of the *PAX5* promoter;[51,52] X019_PAX5 lies is further upstream in the same CpG island. Amplicon X021_PAX5 is located in the CpG island (Chr 9 37016223-37018014) upstream of the coding region of *PAX5* exon 1b. Methylation data for each amplicon was viewed in GeneMaths XT v 1.5 (Applied Maths, Austin, TX). To compare methylation levels between ALL subtypes, the mean methylation levels of CpG units in each amplicon was calculated for each patient. Mean methylation levels for each amplicon were then compared across ALL subtypes by one-way ANOVA, and Dunn's post-hoc test.

2.2.17 Gene Set Enrichment Analysis

GSEA[53] considers the genome-wide expression profiles of two classes of samples (here, *PAX5*-mutated and *PAX5*-wild-type). Genes are ranked based on correlation between expression and class distinction. GSEA then determines if the members of a

gene set *S* are randomly distributed in the ranked gene list *L*, or primarily found at the top or bottom. An enrichment score *ES* is calculated that reflects the degree to which a gene set is overrepresented at the top or bottom of the entire ranked list *L*. The *ES* is a running sum, Kolmogorov-Smirnov like statistic calculated by walking down list *L* and increasing the statistic when a gene in *S* is encountered, and decreasing it when it is not. The magnitude of the increment depends on the strength of association with phenotype, and the *ES* is the maximum deviation from zero encountered in the random walk. The significance level of *ES* is calculated by phenotype-based permutation testing, and when a database of gene sets are evaluated, as in this analysis, the significance level is adjusted for multiple hypothesis testing by calculation of a false discovery rate FDR.

2.3 RESULTS

2.3.1 Focal Deletions of *EBF1* in B-ALL

Eight B-progenitor ALL cases harbored mono-allelic deletions of *EBF1* (early B-cell factor), with deletions limited to this gene in 6 cases (Figure 2-1A-B). EBF1 is required for the development of B cells, and with E2A regulates the expression of B-lineage specific genes. Mice null for *Ebf1* arrest B cell development at the pro-B cell stage, whereas *Ebf1*$^{+/-}$ mice have a 50% reduction in the number of mature B cells but a normal number of pro-B cells.[54] These observations suggest that *EBF1* haploinsufficiency may contribute to leukemogenesis. In support of this interpretation, FISH analysis of one case revealed two populations of blasts, one with mono-allelic and the other with bi-allelic *EBF1* deletions (Figure 2-1C). Flow cytometric analysis of this case demonstrated two distinct blast populations, an immature pro-B cell like CD10^{dim}CD22^{dim} population and a more mature CD10^{bright}CD22^{+} population (Figure 2-1D). These populations significantly differed in expression of CD79A (MB-1), a transcriptional target of EBF1 (Figure 2-1E). Remarkably, FISH on sorted blasts revealed homozygous deletion of *EBF1* in the immature CD10^{dim} fraction, and hemizygous *EBF1* deletions in the more mature CD10^{bright} fraction (Figure 2-1F). Thus, these data suggest that deletion of *EBF1* in these cases may contribute to the block in differentiation that is a hallmark of ALL.

In ALL cases with mono-allelic deletion of *EBF1*, expression of the wild-type *EBF1* allele was detected by RT-PCR consistent with haploinsufficiency. Moreover, no evidence of promoter methylation or point mutations of *EBF1* was detected (data not shown).[49]

2.3.2 A High Frequency of Mono-Allelic *PAX5* Deletions in B-ALL

We next examined the cohort for abnormalities in other genes within the B-cell development pathway. Remarkably, we identified copy number changes in *PAX5* in 57 of 192 (29.7%) B-progenitor ALL cases (mono-allelic loss in 53, biallelic loss in 3, and an

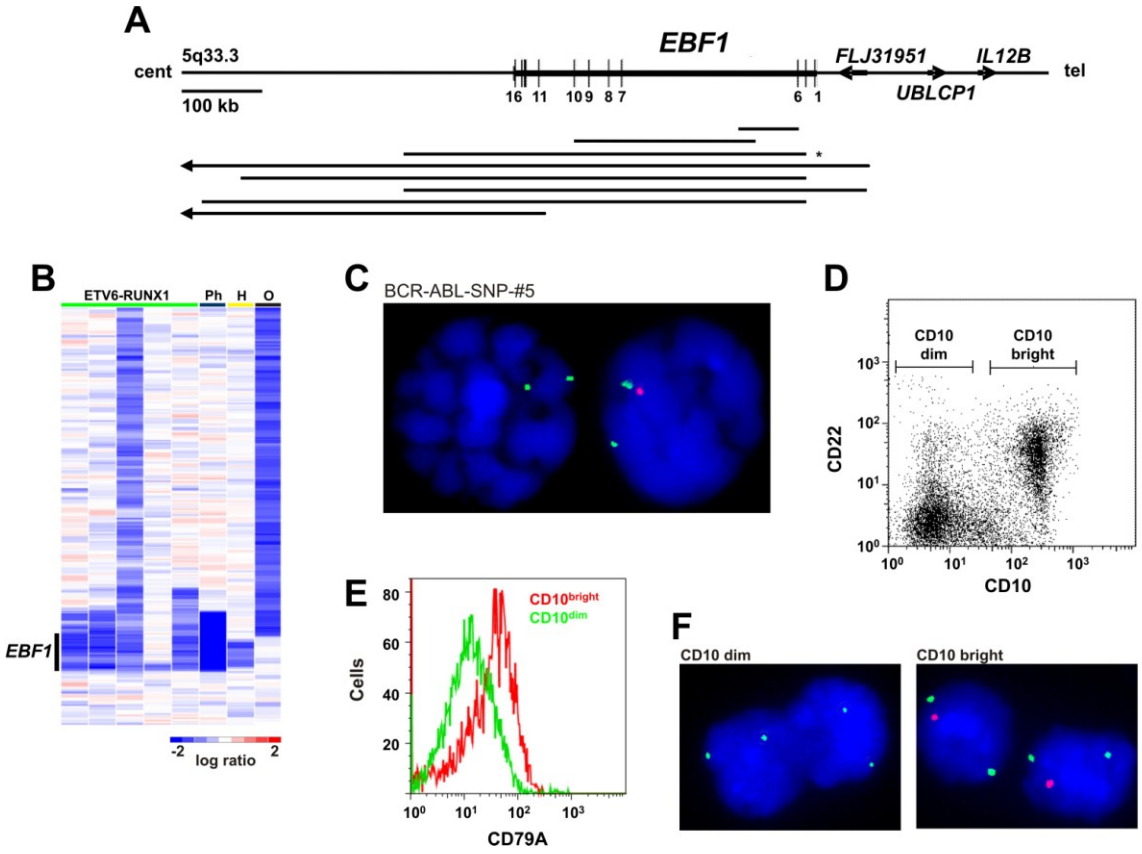


Figure 2-1 *EBF1* deletions in B-progenitor ALL

A, *EBF1* exons are indicated by vertical lines. The extent of *EBF* deletion in each case is shown by a horizontal line, with large deletions denoted by arrowed lines. *FISH analysis for case BCR-ABL-SNP#5 is shown in **C**. **B**, Copy number heatmap showing *EBF1* deletion in 8 B-ALL cases. Ph, BCR-ABL1 positive; H, hypodiploid; O, other. **C,** *EBF1* FISH of case BCR-ABL-SNP#5 identifies two distinct blast populations, with either mono-allelic or bi-allelic *EBF1* deletion; the *EBF1* probe is red and the control probe is green. **D**, Two distinct blast populations differ in surface expression of CD22 and CD10. **E**, Correlation of blast CD79A and CD10 expression. **F**, FISH analysis of cells sorted by CD10 expression.

internal amplification in 1, Figure 2-2). PAX5 is essential for B-lineage commitment and maintenance[55] and acts downstream of E2A and EBF1 to activate expression of B-lineage specific genes such as *CD19, CD79A, BLNK* and *CD7*, and to repress expression of alternate lineage genes.[56,57] Four patterns of *PAX5* deletion were identified (Figure 2-2B): (1) focal deletions involving only *PAX5* (25 cases), (2) broader deletions involving *PAX5* and a variable number of flanking genes (7 cases), (3) large 9p deletions involving the 3' portion of *PAX5*, (5 cases) and (4) deletion of all of chromosome 9 or 9p (19 cases). The *PAX5* deletions and amplification were confirmed by FISH and/or genomic qPCR. Importantly, FISH demonstrated that, with rare exception, the *PAX5* deletions were present in a dominant leukemic clone, consistent with a role in leukemogenesis.

Of 25 B-progenitor ALLs with deletions confined to *PAX5*, 23 deleted only a subset of *PAX5* exons, resulting in the expression of internally deleted transcripts. In nine cases, in-frame splicing across the deletions would be predicted to encode an internally deleted PAX5 protein that either lacks the DNA-binding domain and/or transcriptional regulatory domains. Western blot analysis in cases with sufficient material revealed expression of the predicted size mutant PAX5 protein (Figure 2-2 D-E and data not shown). The other transcripts are predicted to encode prematurely truncated proteins lacking key PAX5 functional domains. The focal amplification of *PAX5* exons 2-5 in case Hypodip-SNP-#7 is predicted to abolish expression of normal *PAX5* from the amplified allele. Thus, of the 57 cases with *PAX5* copy number changes, the majority either lack expression of *PAX5* from the altered allele (27 cases), produce proteins that lack the DNA-binding domain (22 cases), or transcriptional regulatory domain (7 cases), or have a bi-allelic promoter mutations (one case), suggesting that these genetic alterations primarily result in loss-of-function.

Flow cytometry performed on leukemic blasts in 17 cases demonstrated an approximately twofold reduction in the level of wild-type PAX5 protein in cases with mono-allelic deletions, suggestive of haploinsufficiency. Consistent with this interpretation, *PAX5* mRNA levels were quantitated in 46 B-ALL cases (26 wild-type, 20 mutated), and were found to be reduced in cases with *PAX5* deletions. A small number of *PAX5* wild-type cases also had low PAX5 mRNA levels, raising the possibility of epigenetic *PAX5* silencing in these cases.

2.3.3 Cryptic Translocations and Point Mutations of PAX5 in B-ALL

Five B-progenitor ALL cases had large 9p deletions extending into the 3' portion of *PAX5*. Two had rearrangements of chromosomes 9 and 12, either as a dic(9;12) or a t(9;12), and were shown by FISH and RT-PCR to harbor the previously described *PAX5-ETV6* fusion[43,58] (Figure 2-3A-B). Rapid amplification of cDNA ends identified novel *PAX5* chimeric transcripts in two of the remaining three cases: fusion of *PAX5* exon 6 to exon 7 of the forkhead box P1 gene (*FOXP1*), and fusion of *PAX5* exon 7 to exon 4 of the zinc finger protein 521 gene (*ZNF521,* the human homolog of murine *Evi3*) (Figure 2-3A). These were confirmed by FISH (Figure 2-3C-D), RT-PCR, and sequencing. The *PAX5-ETV6* fusion was also detected as a 65kDa protein in both cases containing this

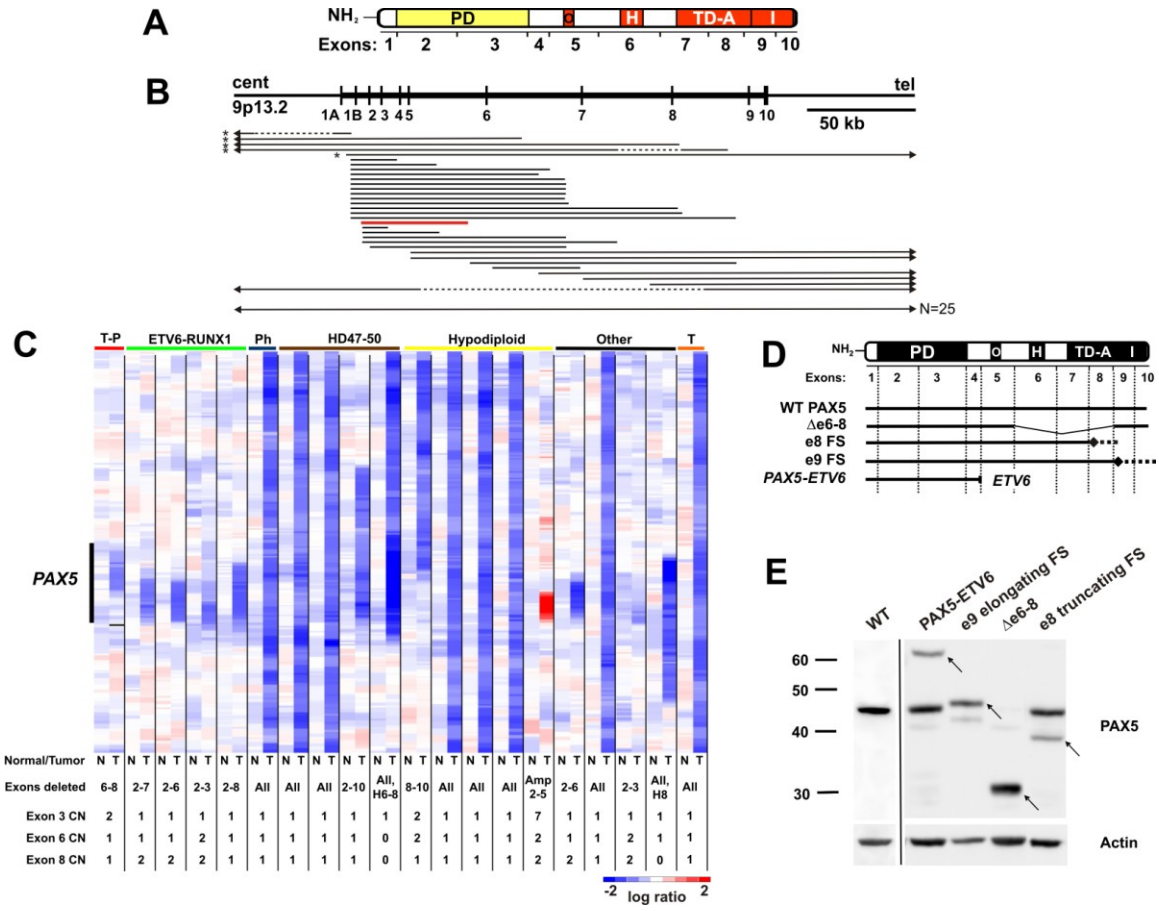


Figure 2-2 *PAX5* deletions in ALL

A, *PAX5* exons and domains. PD, paired domain; O, octapeptide; H, homeodomain-like; TD-A, transactivation, activating; I, transactivating, inhibitory. **B**, Extent of *PAX5* deletions: hemizygous, solid line; homozygous, dashed; amplification, red; beyond region shown, arrowed; *deletion confined to *PAX5*. **C**, *PAX5* copy number heatmap for 20 ALL cases and corresponding germline samples. *PAX5* exon-specific copy number as defined by quantitative genomic PCR is also shown. Amp, amplified; CN, copy number; T-P, *TCF3-PBX1*; H, homozygous; N, normal (remission sample); Ph, *BCR-ABL1*; T, T-ALL. **D-E**, Schematic of *PAX5* mutations and western blot of primary leukemic blasts (mutant PAX5 proteins indicated by arrows).

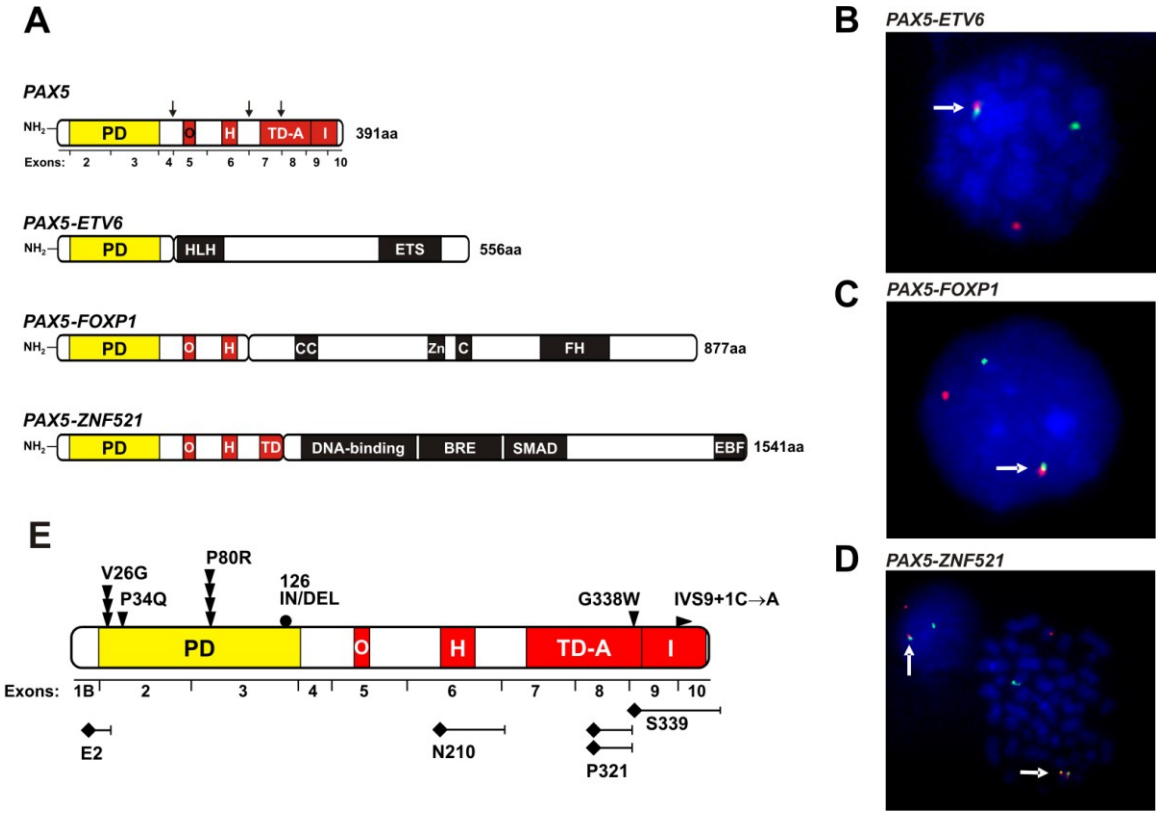


Figure 2-3 *PAX5* translocations in B-progenitor ALL

A, Schematic of PAX5 translocation-encoded fusion proteins. Breakpoints are indicated by (↓); bp, base pairs; BRE, BMP2 response element domain; EBF, EBF-interaction domain; ETS, Ets domain; C and CC, coiled-coil domain; FH, forkhead domain; H, homeodomain-like; HLH, helix-loop-helix; O, octapeptide domain; PD, paired domain; SMAD, SMAD-interacting domain; Zn, zinc finger domain. **B-D**, Interphase FISH for *PAX5-ETV6* (B), *PAX5-FOXP1* (C) and *PAX5-ZNF521* (D). The *PAX5*-specific probe is red, and the partner genes green; arrows denote the fusions. **e**, Location of missense (▼), insertion/deletion (●), frameshift (♦) and splice-site (►) *PAX5* mutations.

rearrangement (Figure 2-2E); primary patient material was not available for western blot analysis for the *PAX5-FOXP1* and *PAX-ZNF521* cases. An additional 50 cases were screened by RT-PCR for the three translocations and no additional cases were identified.

Each of these *PAX5* chimeric genes fuse the DNA-binding paired domain of *PAX5* to the DNA-binding and transcriptional regulatory domains of the partner proteins (Figure 2-3A). Thus, the fusion proteins are predicted to retain the ability to bind to PAX5 transcriptional targets, but would no longer provide normal transcriptional regulatory functions. The fusion proteins may also influence the expression of genes normally regulated by the partner protein, each of which has been implicated in B cell development or hematopoietic malignancies.[59-63]

Sequencing of *PAX5* identified point mutations in 14 cases that clustered in exons encoding the DNA-binding or transcriptional regulatory domains (Figure 2-3E). In 13 cases with available germline DNA, the mutations were shown to be somatically acquired. Moreover, a quantitative analysis suggested that the mutations were present in all blasts in 12 of 14 cases, and in a major subclone in the remaining two cases. The point mutations were hemizygous except in one case where a homozygous splice-site mutation (IVS9+1) of *PAX5* was detected within an extensive region of 9p LOH that included a homozygous deletion of *CDKN2A*.

Nine *PAX5* mutations were identified within exons encoding the paired DNA-binding domain: V26G, P34Q, P80R, and the insertion/deletion NDTVP126RA (Figure 2-3E). Modeling studies using the PAX5 crystal structure[41] suggests that each mutation should impair DNA-binding. Other mutations consist of frame-shift, splice site, or missense mutations that affect the transactivation domain, and a single case with an exon 1B frame-shift mutation that results in a prematurely truncated 10 residue polypeptide (Figure 2-3E). Collectively, the identified *PAX5* mutations are predicted to result in lost or altered DNA-binding or transcriptional regulatory function.

We also assessed whether methylation-induced silencing of *PAX5* occurs in B-progenitor ALL using mass spectrometry.[10] These data revealed high-level *PAX5* promoter methylation in T-ALL, but minimal methylation in B-progenitor ALLs, irrespective of *PAX5* mutational status. Thus, epigenetic silencing due to high-level methylation does not appear to be a prominent mechanism of *PAX5* silencing in B-progenitor ALL.

2.3.4 Functional Consequences of *PAX5* Mutations

To assess the DNA-binding and transcriptional activity of a subset of the identified *PAX5* mutants, luciferase-based reporter assays were performed using the PAX5-dependent reporter plasmid, *luc*-CD19.[46] Each of the PAX5 mutants tested displayed significantly reduced transcriptional activation compared to wild-type PAX5 (Figure 2-4 A-B and Figure 2-5). Moreover, transfection of increasing amounts of *PAX5-ETV6* or *PAX5-FOXP1* together with a fixed amount of wild-type PAX5 demonstrated

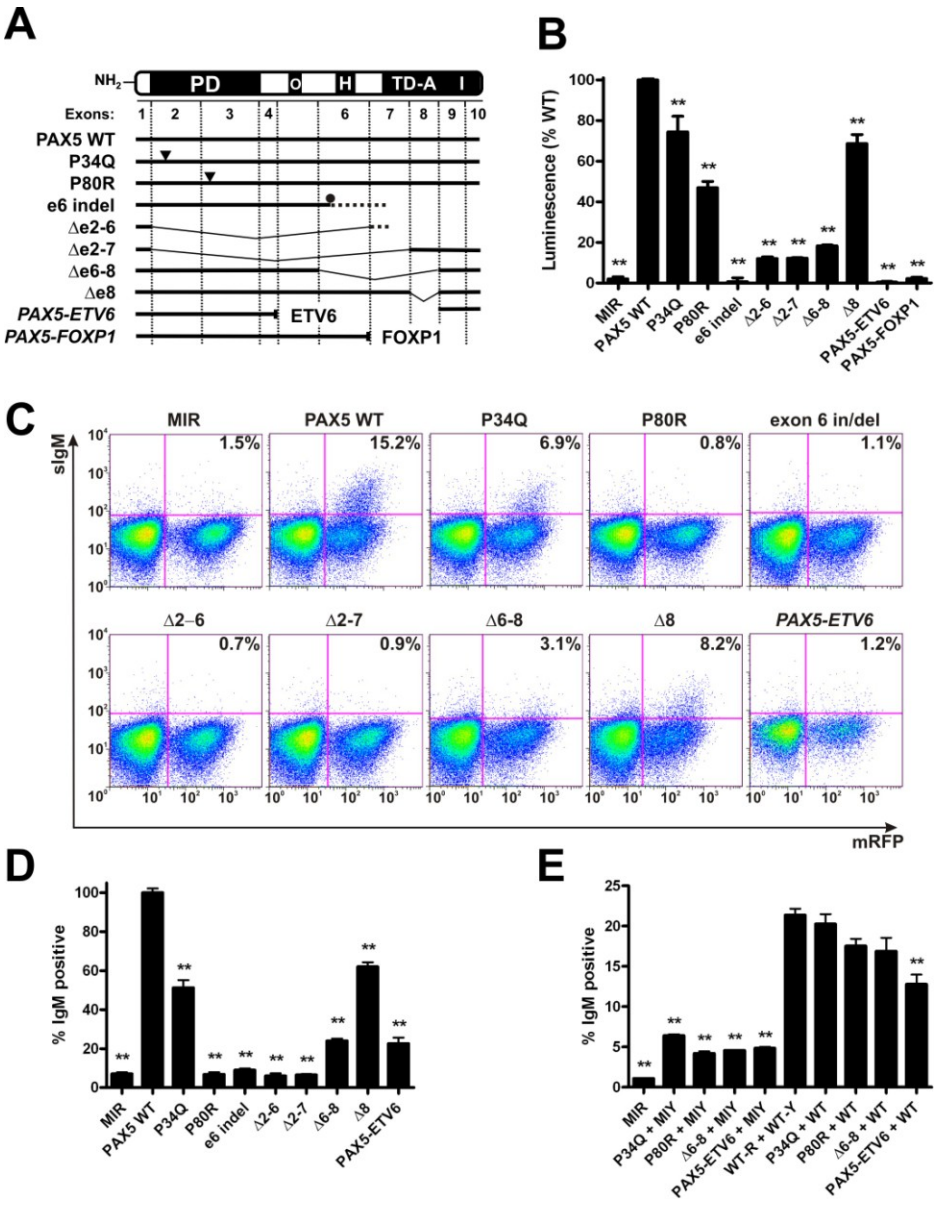


Figure 2-4 Impaired function of PAX5 mutants

A, Structure of PAX5 variants. **B**, transcriptional activity of PAX5 variants using the *luc*-CD19 reporter. Bars show mean (± s.e.m.) luciferase activity of triplicate experiments normalized to luciferase activity of wild-type (WT) PAX5. **ANOVA with Dunnett's test P<0.01 c.f. WT PAX5. **C**, representative flow plots of 558LμM cells transduced with PAX5-expressing retrovirus, showing reduced sIgM expression of PAX5 mutants c.f. WT PAX5. Percents indicate proportion of mRFP+ cells that are sIgM+. **D,** mean (± s.e.m.) 558LμM sIgM expression of triplicate experiments, normalized to sIgM expression of PAX5 WT transduced 550LμM cells. **E**, Impaired sIgM expression is corrected following co-infection with WT PAX5 virus (WT) for the PAX5 deletion and point mutants, but not *PAX5-ETV6.*

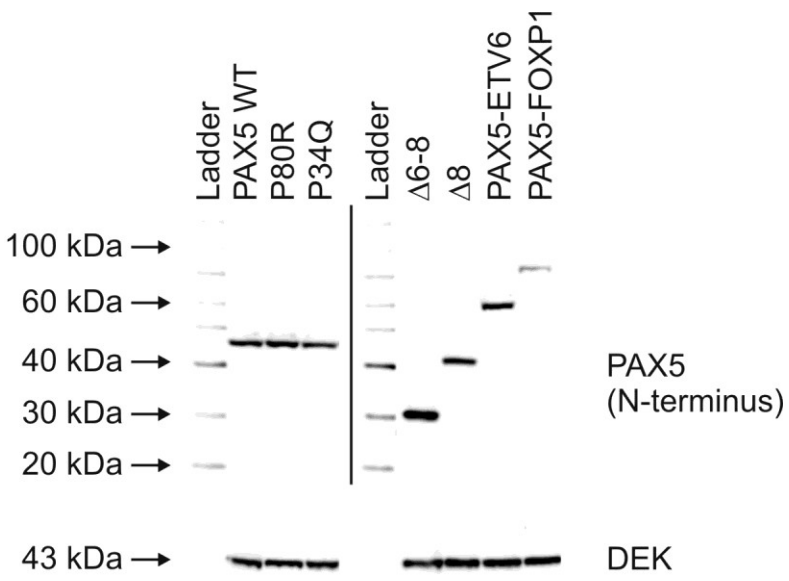


Figure 2-5 PAX5 western blots of 293T cells used for CD19-*luc* reporter assays

Western blots using an N-terminus PAX5 specific antibody (Chemicon) were performed on nuclear extracts of transiently transfected 293T cells. Neither this antibody, nor any of the alternative available PAX5 specific antibodies detect the PAX5 Δ2-6 or Δ2-8 internal truncation mutants.

that the fusion proteins competitively inhibit the transcriptional activation of wild-type PAX5 (Figure 2-6). Analysis of DNA-binding activity revealed a marked reduction in binding activity for PAX5 variants with paired domain mutations or deletions (Figure 2-7).

We next examined the effect of PAX5 mutations on transcriptional activation of the PAX5 target *Cd79a* (mb-1, Ig-α) in the murine plasmacytoma cell line 558Lμ M.[23] This cell line expresses three of four components of the surface IgM (sIgM) receptor complex (μ heavy chain, λ light chain, and Ig-β), but not Cd79a or Pax5, and consequently does not express sIgM. Following transduction with *PAX5*-expressing retrovirus, Cd79a and thus sIgM are expressed, and quantitation of sIgM serves as a sensitive measure of the *Cd79a* transactivating activity of PAX5.[48] This assay confirmed the reduced transcriptional activity of the PAX5 mutants (Figure 2-4C-D and Figure 2-8). Co-transduction with wild-type and mutant *PAX5* retroviruses showed that *PAX5* wild-type virus restored sIgM expression in the presence of the P34Q, P80R or Δ6-8 *PAX5* variants, but incompletely with *PAX5-ETV6* (Figure 2-4E and Figure 2-9). This suggests that the PAX5 DNA-binding and internal deletion mutants act as hypomorphic alleles with weak competitive activity, where as *PAX5-ETV6* acts as a stronger competitive inhibitor of wild-type PAX5.

As the *PAX5* mutations are predicted to reduce or inhibit normal PAX5 functional activity, we next examined expression of PAX5 target genes in leukemic blasts. We did not observe a correlation between *PAX5* mutation status and expression of the PAX5 targets CD19 and CD79A.[11,57] However, using Affymetrix HG-U133A gene expression profiling data,[64] we identified a 42 gene expression signature for *PAX5*-mutated *ETV6-RUNX1* B-ALL that included both up-regulated (PAX5 repressed) and down-regulated (PAX5 stimulated) genes (Figure 2-10A). PAX5-stimulated genes included the known PAX5 target *CD72*[57] as well as genes with known roles in oncogenesis (e.g. *DAPK1*,[65] *TACC1*[66]). We then performed cross-subtype gene set enrichment analysis (GSEA)[53] to test for enrichment of these putative PAX5-regulated genes in non-*ETV6-RUNX1* B-ALL cases. This analysis demonstrated highly significant enrichment of the PAX5 stimulated genes in *PAX5* wild-type B-progenitor ALL cases that lacked recurrent cytogenetic abnormalities, and in *PAX5* wild-type cases from the entire non-*ETV6-RUNX1* B-progenitor ALL cohort (Figure 2-10B-C). This suggests that the identified *PAX5* mutations have a significant effect on the intracellular transcriptional network within primary leukemic cells.

2.3.5 Pattern of B-Cell Development Gene Mutations in ALL

Deletions of additional genes encoding regulators of B-cell development were identified, including *IKZF1* (Ikaros, 17 B-progenitor ALL cases), *IKZF3* (Aiolos, N=3), *LEF1* (N=3) *TCF3* (N=1) and *BLNK* (N=2). Genomic sequencing of *IKZF1* was performed in all cases and no mutations were identified.

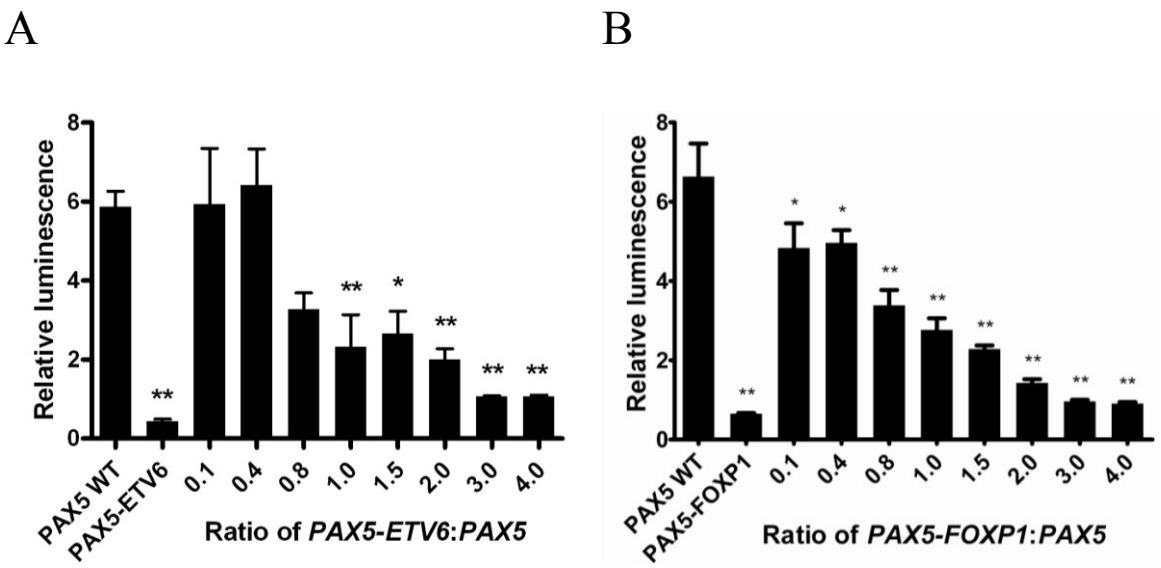


Figure 2-6 *PAX5* translocations competitively inhibit the transcriptional activity of wild-type PAX5

Relative luminescence of 293T cells transfected with *luc*-CD19, fixed amount of PAX5 WT and increasing amounts of *PAX5*-ETV6, **A**, or *PAX5-FOXP1*, **B,** plasmids. Empty MSCV-IRES-mRFP (MIR) expression vector was used to maintain a constant mass of DNA in each transfection experiment. Luminescence is normalized to *Renilla* luciferase activity and is shown relative to 293T cells infected with empty MIR vector. Bars show means ± SEM of triplicate experiments. *ANOVA with Dunnett's test $P<0.05$; **$P<0.01$. Importantly, expression of the various PAX5 mutants did not inhibit the transcription of the pRL-TK *Renilla* luciferase plasmid used as an internal control for transfection efficiency.

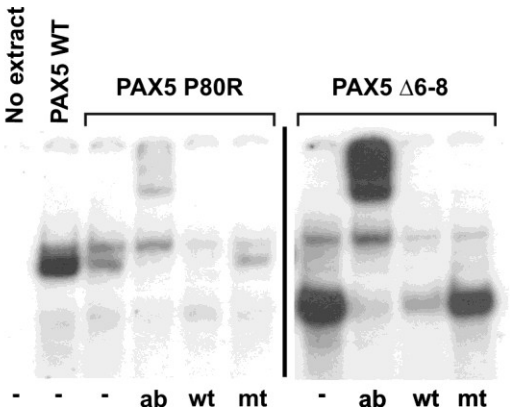


Figure 2-7 DNA-binding of PAX5 mutant alleles

Gel-shift assay showing reduced (P80R) and normal (Δ6-8) binding of PAX5 variants to a CD19 promoter binding site. ab, supershift with PAX5 antibody; wt, wild-type and mt, mutant competitor oligonucleotides; -, no antibody or competitor oligonucleotides.

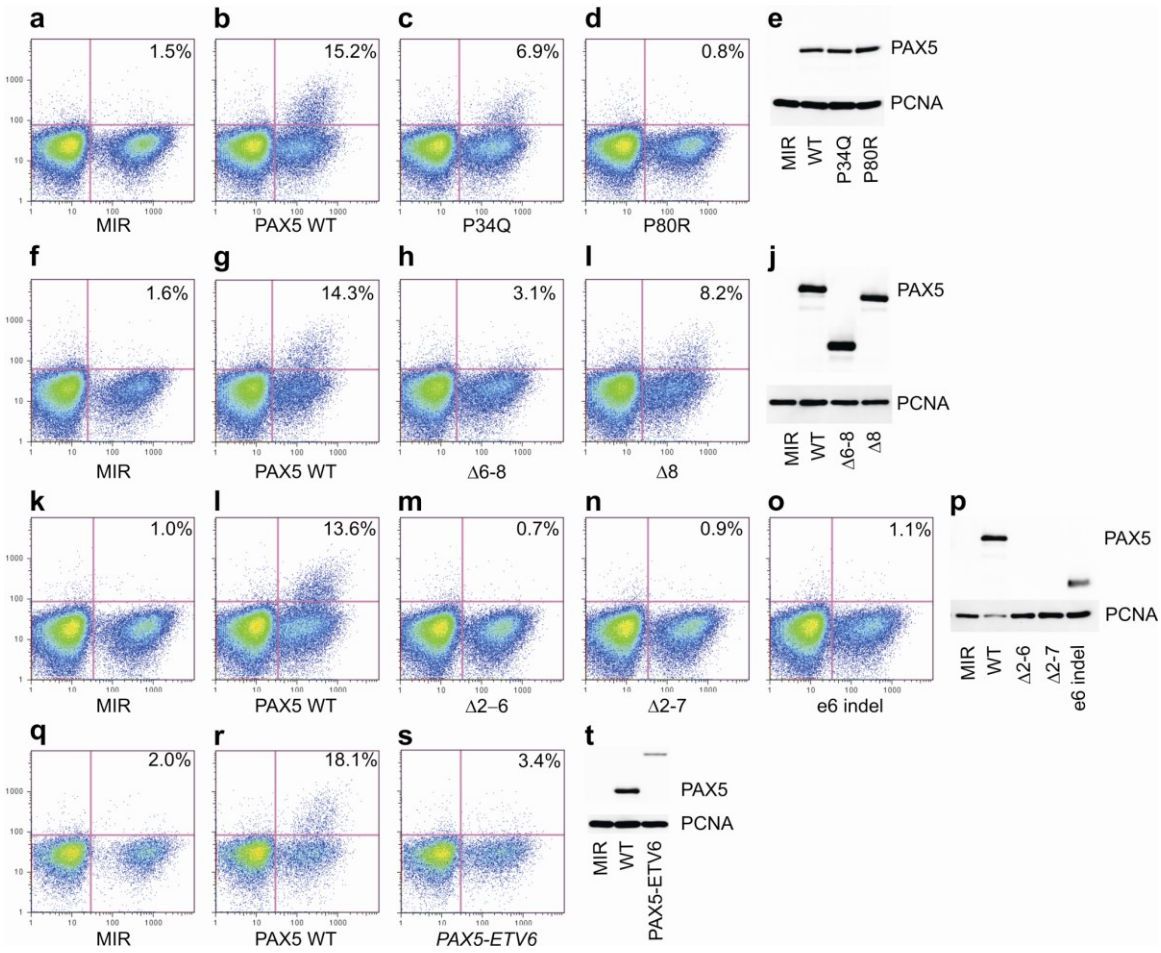


Figure 2-8 *PAX5* mutations impair CD79A transactivation of sIgM expression in the 558LμM cell line

558LμM cells were transduced with MSCV-PAX5-IRES-mRFP retroviruses expressing wild type or mutant PAX5. sIgM expression of the mRFP-positive population was measured. Four experiments examining different PAX5 variants were performed (**A-E**, **F-J**, **K-P** and **Q-T**). Empty vector (MIR) and wild-type PAX5 (PAX5 WT) were included as controls in each experiment. All transductions were performed in triplicate, and representative flow plots are shown. Percentages refer to the proportion of RFP+ cells expressing sIgM. Two million RFP+ cells were sorted from a transduction experiment for each PAX5 variant, and western blotting for PAX5 and the control antigen PCNA performed on nuclear extracts of the RFP+ population. Sort purities were >90%. Western blot analysis confirmed comparable protein expression of the PAX5 variants analyzed, with the exception of PAX5-ETV6, which is larger than the other variants (approx. 65kDa *c.f.* 45 kDa for WT PAX5), and is consistently expressed at lower protein levels than the other variants. The PAX5 antibody used (BD Transduction Labs) was raised against an internal PAX5 epitope and does not recognize the Δ2-6 or Δ2-7 *PAX5* internal deletion mutants shown in **P**.

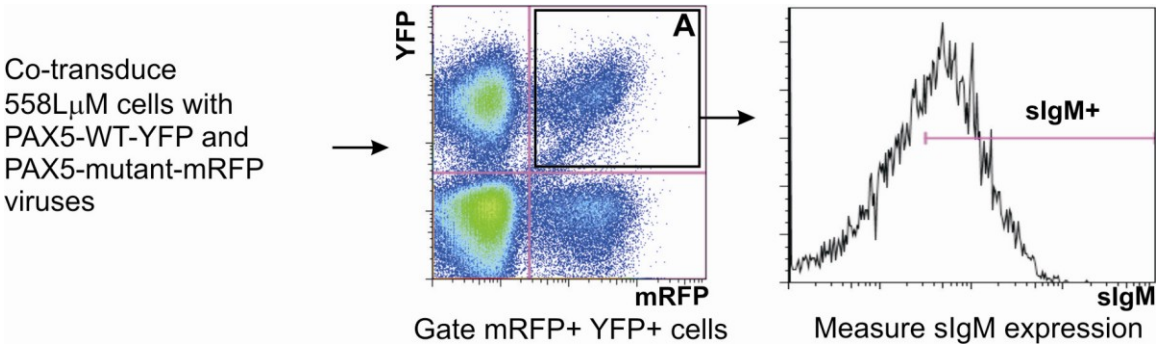


Figure 2-9 Design of 558LμM PAX5 WT and mutant co-transduction experiments

558LμM cells were co-transduced with bicistronic MSCV retroviruses co-expressing wild type PAX5 with YFP, and mutant PAX5 alleles with mRFP. Surface IgM expression was then quantitated on the dual positive population.

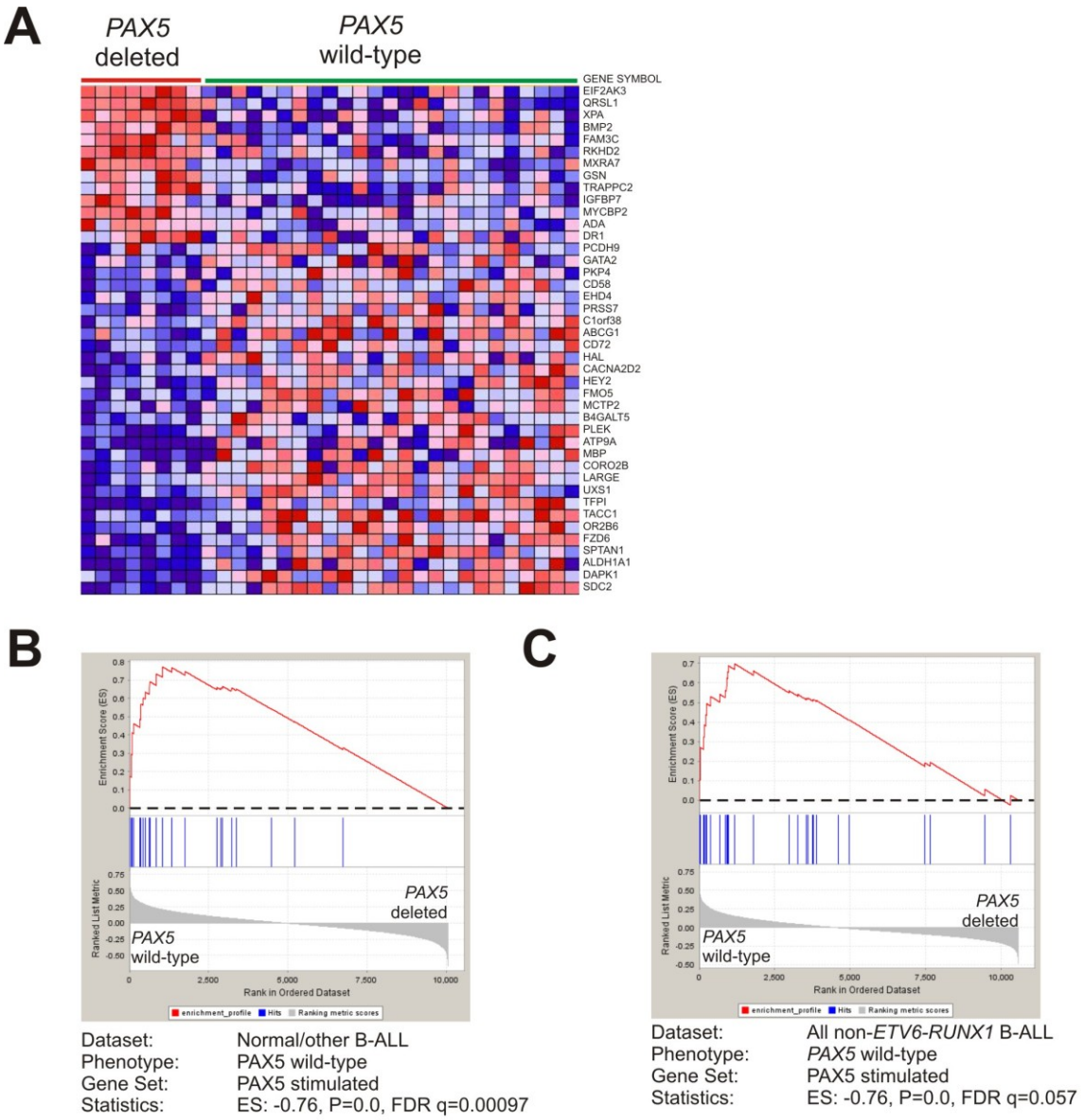


Figure 2-10 Cross-subtype gene set enrichment analysis of PAX5-regulated genes in B-ALL

A, heatmap of 42 differentially expressed genes (at FDR <0.3) in *ETV6-RUNX1 PAX5*-deleted *v. ETV6-RUNX1 PAX5* wild-type ALL. **B**, GSEA showing significant enrichment of the PAX5-stimulated genes shown in B-ALL samples lacking recurring cytogenetic abnormalities. The ranked gene list of *PAX5*-wild-type *v. PAX5*-deleted cases is shown in the lower, grey plot. Vertical blue lines indicated where probe sets in the PAX5-stimulated gene set fall in the ranked gene list. The top, red line indicates the running enrichment score (*ES*) that becomes more positive as probe sets are encountered at the top of the list. **C**, significant enrichment for PAX5-stimulated genes in all non-*ETV6-RUNX1* B-progenitor ALL. Enrichment for the PAX5-repressed gene set in PAX5-mutated cases was also observed in both non-*ETV6-RUNX1* cohorts, but was not significant (i.e. FDR>0.25) after correction for multiple hypothesis testing.

Importantly, there were marked differences in the frequency and type of mutations among genetic subtypes of ALL. Specifically, all 10 hypodiploid cases had one null *PAX5* allele, and six of these also had either a point mutation (N=5) or translocation (*PAX5-ZNF521*, N=1) involving the other *PAX5* allele. Five of these cases also contained deletions in other B-cell development genes, with up to three different genes being mutated within a single case. In contrast, 28% of the *ETV6-RUNX1* cases contained focal mono-allelic *PAX5* deletions but lacked evidence of mutations in the retained *PAX5* allele. Three of these cases also had deletions of a single allele of *EBF1*, and another two cases had deletions of only *EBF1*. At the other end of the spectrum, mutations in B-cell development genes were uncommon in hyperdiploid B-progenitor ALL (13% of cases).

2.4 DISCUSSION

Understanding the molecular pathogenesis of cancer requires a detailed cataloguing of all genetic lesions within a cancer cell. The data presented here provide a rational roadmap for approaching such a task. We have shown that genomic copy number analysis using high-density SNP arrays can pinpoint altered genes and pathways for further analysis. This approach led to the unexpected finding of mutations in genes encoding regulators of B cell development and differentiation in 40% of B-progenitor ALLs. The identified genetic alterations are specific, pathogenic, and somatically acquired. In addition, the majority of *EBF1* and *PAX5* deletions affect only these genes, thus conclusively identifying them as the target of the deletions. Moreover, the average number of deletions within an individual B-ALL case was 3.83, and focal amplifications were uncommon, suggesting global genomic instability is not an underlying mechanism. Lastly, although somatic point mutations were identified in *PAX5*, sequencing of *EBF1* and *Ikaros* failed to reveal any evidence of mutations, ruling out a high mutational rate.

The most common targets of these genetic alterations (*EBF1*, *PAX5*, and *Ikaros*) play central roles in the development of normal B-cells.[56] In mice the complete absence of *Ebf1* or *Pax5* results in the arrest of B-cell development at the early pro-B or pro-B-cell stage,[54,55] and loss of Ikaros leads to an arrest at an even earlier stage of lymphoid development.[67] Moreover, haploinsufficiency of *Ebf1* leads to a partial block in B cell development,[54] a phenotype that is further accentuated in mice haploinsufficient for both *Ebf1* and *E2a*.[68] In addition, loss of *Ikaros* through the expression of a dominant negative *Ikaros* isoform predisposes mice to the development of T-lineage malignancies.[69]

The overall consequence of the identified lesions is to reduce the level of the specific transcription factor either as a result of mono-allelic deletion or the generation of altered forms of the specific protein. Whether some mutations result in dominant negative forms of PAX5, or alternatively alter transcriptional activity in a promoter specific manner requires further investigation. It is important to note that during the normal development of B-cells, *PAX5* is subjected to allele-specific regulation, with only a single *PAX5* allele transcribed during the earliest phase of B-cell commitment and then a switch to bi-allelic expression as B-cells begin to differentiate.[70] The loss of a wild-type *PAX5* allele from the identified mutations would eliminate the ability to turn on normal bi-

allelic transcription, which may directly contribute to the differentiation arrest seen in ALL.

Members of the PAX family of transcription factors including *PAX3, PAX7*, and *PAX8* have previously been identified as targets of tumor-associated translocations[71] that result in the over-expression of PAX fusion proteins. In addition, over-expression of wild-type *PAX5* occurs as a result of its rearrangement into the *IGH* locus in non-Hodgkin lymphoma.[72] In sharp contrast to over-expression, however, our results implicate subtle changes in the dosage of PAX5 and other key regulators of B-cell development in the pathogenesis of ALL. Although heterozygous alterations of *PAX5* may have only subtle effects on their own, these effects are likely to contribute directly to leukemogenesis in combination with other oncogenic lesions.

The identified high frequency of alterations of B-cell development genes in pediatric B-progenitor ALL represents a lower limit of the true frequency. Direct copy number analysis of those genes with low density SNP coverage along with full sequence analysis of all genes involved in controlling B-cell development and differentiation will be required to define the overall frequency. Nevertheless, our results demonstrate an unexpectedly high frequency with distinct patterns of mutations among the various genetic subtypes of pediatric ALL. Experiments to directly assess the effect of co-expressing *PAX5* mutants and fusion proteins such as *ETV6-RUNX1* in murine models should provide valuable insights into the ability of these lesions to collaborate in leukemogenesis. Moreover, attempts to determine whether small molecule inducers of differentiation can bypass the block resulting from the identified lesions, and whether these molecules in turn would trigger a leukemia cell-specific apoptotic response, could lead to new therapeutic approaches for pediatric ALL.

CHAPTER 3
Pax5 HAPLOINSUFFICIENCY COOPERATES WITH *BCR-ABL1* TO INDUCE ACUTE LYMPHOBLASTIC LEUKEMIA*

3.1 INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy and is comprised of several distinct genetic subtypes including B-progenitor leukemia's with t(9,22)[*BCR-ABL1*], t(1,19)[*TCF3-PBX1*], t(12,21)[*ETV6-RUNX1*], rearrangements of *MLL*, and hyperdiploid or hypodiploid karyotypes, and T-lineage leukemias.[1] These defining genetic lesions play a role in disease initiation,[6] but alone are insufficient to generate overt leukemia. In an effort to identify cooperating oncogenic lesions, we recently performed genome-wide copy number and loss-of-heterozygosity (LOH) analyses on over 400 pediatric ALL leukemias.[7] These analyses identified somatic mutations in genes encoding regulators of B-cell development and differentiation in over 40% of B-progenitor ALLs. The most common targets of these loss-of-function mutations were the transcription factors *PAX5*, *IKZF1*, and *EBF1*, with *PAX5* being mutated in over 30% of patients. Notably, *BCR-ABL1* ALL was characterized by a high frequency of mutations of *IKZF1* (85%), *PAX5* (55%), and *CDKN2A/B* (encoding INK4A/ARF and p15, 55%), suggesting that these lesions cooperate with *BCR-ABL1* in lymphoid leukemogenesis.[19]

The mutations of B cell development genes are hypothesized to contribute to leukemogenesis by inducing a block in the differentiation of the leukemic cells. Consistent with this hypothesis, the most frequent target of these mutations, the PAX5 transcription factor, is essential for normal B cell differentiation with loss of the gene resulting in a complete arrest at an early pro-B cell stage of development in the bone marrow of *Pax5*-deficient mice.[16] PAX5 normally functions by restricting the multilineage potential of early lymphoid progenitors to the B cell pathway by repressing lineage-inappropriate genes and activating B cell–specific genes such as *CD79a*, *CD72*, *CD19*, and *BLNK*.[15,73,74] Moreover, continued expression of PAX5 is required for maintenance of cells to the B cell lineage. For example, conditional deletion of *Pax5* in committed murine Pro-B cells results in their reversion to a multipotent state from which T lymphocytes and macrophages can be generated.[18] Thus, PAX5 serves as a master regulator controlling the normal commitment and differentiation of hematopoietic progenitors to the B cell lineage.

The genomic data, supporting a role for altered *PAX5* expression in the pathogenesis of pediatric ALL, are diverse and include mono-allelic deletions of the entire gene, focal deletions leading to the expression of hypomorphic *PAX5* isoforms, chromosomal translocations producing PAX5 fusion proteins, and *PAX5* DNA sequence

---

* Adapted with permission. Miller C, Su X, Andersson A, Dang J, Mullighan C, Ma J, Zhang J, Wang M, Williams R, Downing J. *Pax5* Haploinsufficiency Cooperates with the *BCR-ABL1* Oncogene to Induce Acute Lymphoblastic Leukemia. Manuscript in preparation.

alterations that result in missense mutations of conserved residues in the DNA-binding domain, or nonsense, frame shift, or splice site mutations in the transcriptional regulatory domains.[7] Despite these strong human genomic data, mice heterozygous for a *Pax5* null allele show no significant immune system abnormalities and fail to develop B lineage leukemias during their lifetime.[16] The absence of an overt phenotype within mice raises the possibility that the *PAX5* alterations detected in pediatric ALL are bystander or passenger mutations, and are not mechanistically involved in the development or maintenance of the leukemic clone. To directly assess the contribution of a reduction in Pax5 expression to leukemogenesis, we assessed the susceptibility of $Pax5^{+/-}$ mice to *BCR-ABL1*-induced lymphoid leukemia. Our data demonstrates that haploinsufficiency of *Pax5* renders mice susceptible to the development of ALL, and directly cooperates with the loss of the tumor suppressor gene *Arf* in *BCR-ABL1*-induced lymphoid leukemogenesis.

3.2 MATERIALS AND METHODS

3.2.1 Mice

Wild type, $Pax5^{+/-}$, and $Arf^{+/-}$, as well as intercrosses of these genotypes were maintained on a C57BL/6 background. All animal experiments were performed according to St. Jude Children's Research Hospital Institutional Animal Care and Use Committee-approved protocols.

3.2.2 Bone Marrow Transduction and Transplantation Analysis

All bone marrow transplant experiments were performed as previously described.[75] The Murine Stem Cell Virus (MSCV) retroviral expression construct containing the p185 isoform of *BCR-ABL1* (MSCV-GFP-IRES-p185 was provided by O. Witte (University of California, Los Angeles). The p185 *BCR-ABL1* cDNA cassette was removed from this virus and recloned into a variant version of this vector such that the *BCR-ABL1* gene was cloned 3' of the *GFP* gene with an internal ribosomal entry site (IRES) positioned between the cDNAs (MSCV-*GFP-IRES-p185 BCR-ABL1*). Replication-incompetent, ecotropic retrovirus was made by transient transfection of Phoenix-Eco packaging cells (G. Nolan, Stanford University, Stanford CA) using previously described methods. Bone marrow (BM) cells were harvested from $Pax5^{+/+}$, $Pax5^{+/-}$, $Arf^{+/-}$, and $Pax5^{+/-}Arf^{+/-}$ mice and transduced in suspension for 3 hours with MSCV-*GFP-IRES-p185 BCR-ABL1* retrovirus. Cells were washed and injected (2×10^6 per mouse) into lethally irradiated (11Gy in two doses) wild-type C57BL/6 recipient mice. Mice were observed daily for signs of disease and necropsies were performed on all moribund animals.

3.2.3 Histology

Tissues were fixed in 10% buffered formalin, embedded in paraffin, and sections were stained with hematoxylin and eosin. Cytospins were prepared by centrifuging 1 x 10^5 BM cells onto a microscope slide using a Wescor cytocentrifuge 7620 (Logan, UT). Cytospins were stained with Wright's stain.

3.2.4 FACS Analysis and Cell Sorting

All antibodies were obtained from BD Biosciences (San Jose, CA). Flow cytometric analysis was performed using a BD LSR II (BD Biosciences) and the following antibodies: B220-APC, CD43-biotin/SA-PE-Texas Red, CD19-APC-Cy7, BP1-PE, and IgM-PE-Cy7. Leukemias that were less than 80% GFP positive in the BM were subject to sorting for GFP positive cells using a FACSAria cell sorter (BD Biosciences).

3.2.5 Genomic DNA Extraction and Southern Blot Analysis of Proviral Integrations

Genomic DNA was isolated from leukemia cells by proteinase K digestion followed by phenol:chloroform:isoamylalcohol extraction and ethanol precipitation as previously described. DNA was quantified using a NanoDrop spectrophotometer and its quality assessed using agarose gel electrophoresis. The number of proviral integrations in selected tumors was determined by Southern Blot Analysis. Briefly, 3 micrograms of genomic DNA was digested with HindIII, separated on a 0.7% agarose gel, and transferred to a Hybond N+ membrane (Amersham Biosciences). The membrane was hybridized with a ^{32}P-radiolabeled 724bp NcoI/NotI DNA fragment DNA probe, corresponding to the GFP cDNA.

3.2.6 Western Blotting

Five million leukemias cells were washed with PBS and lysed with 4x LDS sample buffer (Invitrogen). 15ul of lysate was separated on NuPAGE 4-12% Bis-Tris gels, transferred to nitrocellulose membranes, and after blocking were incubated with either an N-terminal Pax5 antibody (Chemicon, now Millipore, Billerica, MA) or a p19 Arf antibody (Abcam, Cambridge, MA). Following incubation with secondary antibodies, blots were developed using the SuperSignal West Femto Max Sensitivity Substrate chemiluminescent reagent (Pierce, Rockford, IL), and images were captured using the ChemiDoc XRS gel imaging system (Bio-Rad Laboratories, Hercules, CA). Membranes were subsequently stripped, blocked and reprobed with a C-terminal Actin antibody (Santa Cruz Biotechnology, Santa Cruz, CA) as a loading control.

3.2.7 Array-based Comparative Genomic Hybridization

The Agilent (Santa Clara, CA, USA) eArray 5.0 (https://earray.chem.agilent.com/earray/) platform was used to design microarrays for CGH analysis. Two separate arrays were designed: For the first array, we used Agilent's Mouse Genome CGH Microarray 244A and removed all probes located to chromosomes X and Y. The vacated space was filled in with 17,573 custom 60-mer oligonucleotide probes that covered at a 250 bp resolution, twenty genes that are the most frequent targets of CNAs in human *BCR-ABL1* ALL. The coverage of these 20 genes included 100kb upstream and downstream of their annotated exons. For the second array, we incorporated 238,382 60-mer oligonucleotide features from across the mouse genome (excluding features chromosomes X and Y) along with an additiona 5,099 control features selected from the eArray library. The combination of 476,623 features provides a median feature spacing of 4,453 bp (the number of features and median spacing for each individual chromosome are listed in Supplementary Table 8). The arrays were printed using the Agilent Sureprint technology using *in situ* oligonucleotide synthesis.

3.2.8 Quantitative Genomic Real-time PCR

Confirmation of copy number abnormalities using genomic quantitative real-time PCR (gqPCR) was performed for *Pax5*, *Cdkn2a/b*, *Ikzf3*, *Ebf1*, *Zfp456*, and *V1rf2*.

3.2.9 Genomic Resequencing of *Pax5*

Genomic resequencing of all the coding exons of *Pax5* was performed on the 50 leukemia samples analyzed by array CGH. Genomic DNA was extracted as previously described and each exon was PCR amplified using 1U of AccuPrime GC-Rich DNA polymerase (Invitrogen, Carlsbad, CA), and 0.2µM forward and reverse primers. Primer sequences are available upon request. PCR products were purified and then sequenced directly using Big Dye Terminator (v3.1) chemistry on a 3730xl DNA Analyzers (Applied Biosystems, Foster City, CA). Mutations detected by direct sequencing were validated by cloning PCR products in pGEM-T easy (Promega, Madison, WI) and multiple clones sequenced in both directions with M13 primers.

3.2.10 Structural Modeling of *Pax5* Mutations

A structural view of the PAX5 paired domain was generated by PyMOL v0.99 (http://pymol.sourceforge.net/) using the coordinates of the X-ray structure of PAX5 interacting with ETS1 on DNA, and PAX6 deposited with the Brookhaven Data Bank (PDB: 1K78 and 6pax; http://www.rcsb.org/pdb/).[41,42] The protein sequences of the paired domains of PAX6 and PAX5 are 70.1% identical.

3.2.11 Gene Expression Profiling

RNA was extracted as previously described[76] from the leukemic bone marrow of 10 *Pax5*$^{+/+}$, 5 *Pax5*$^{+/-}$ mice, as well as 5 *Pax5*$^{+/-}$ mice that later lost their second *Pax5* allele. Additionally, all leukemias with a bone marrow % GFP less than 80% were flow sorted to enrich the sample to greater than 95% GFP positive cells. In addition, RNA was extracted from flow sorted normal hematopoietic cell subpopulations representing Hardy fractions A-F (at least 5 individual animals were used as biological replicates for each fraction) from the bone marrow of *Pax5*$^{+/+}$ and *Pax5*$^{+/-}$ mice. Flow sorted cells from Hardy fractions A and B were also obtained from *Pax5*$^{-/-}$ mice (5 samples from each fraction). The antibodies and sorting strategy used are outlined in detail in a previous methods section. The quality of the RNA was assessed using the Agilent Bioanalyzer (Agilent, Santa Clara, CA). Samples were hybridized to Affymetrix 430v2.0 microarrays (Affymetrix, Santa Clara, CA) as previously described.[77]

Probe intensities were generated using the MAS 5.0 algorithm, probe sets called absent in all samples were excluded, and expression data was log-transformed. To define the gene expression signature of mouse *BCR-ABL1*, we used *Limma* (Linear Models for Microarray Analysis)[78], the empirical Bayes t-test implemented in Bioconductor[79] and the Benjamini-Hochberg method of false discovery rate (FDR) estimation[80] to identify probe sets differentially expressed between mouse *BCR-ABL1* and normal B-cells.

Principal component analyses were performed using Gene Expression Explorer v1.1 (Qlucore AB, Lund, Sweden). In addition to conventional PCA analyses, this program enables PCA to be performed on a selected set of samples (in our case Hardy fractions A-F) and then applying these principal components on another set of samples (in our case the leukemias), without letting the second set of samples contribute to the calculation of the three principal components. This allows diverse samples to be analyzed together and provides a powerful tool for investigating similarities between normal and malignant cells that cannot be revealed otherwise.[81]

3.2.12 Gene Set Enrichment Analysis

To search for enrichment of specific transcriptional programs in the various normal Hardy fractions and leukemias, gene set enrichment analyses[53] were performed. Gene sets of the top up- and down-regulated genes in the signatures of mouse *BCR-ABL1* and Core ESC-like Module were created and added to the collection of currated gene sets available at the Molecular Signatures Database (http://www.broad.mit.edu/gsea/msigdb/). GSEA of mouse leukemia and human *BCR-ABL1* were then performed using this expanded collection of gene sets. GSEA then determines if the members of a gene set (here a Core ESC-like Module) are randomly distributed in the ranked gene list, or primarily found at the top or bottom. Occurrences of members of the gene set in the ranked gene list are shown as vertical black lines above the ranked signature. An enrichment score (ES) is calculated that reflects the degree to which a gene set is overrepresented at the top or bottom of the entire ranked list. The ES is a running sum,

Kolmogorov-Smirnov like statistic calculated by walking down list L and increasing the statistic when a gene in S is encountered, and decreasing it when it is not. The magnitude of the increment depends on the strength of association with phenotype, and the ES is the maximum deviation from zero encountered in the random walk, and is depicted as a red curve. The "leading edge" genes are those members of the gene set responsible for the observed enrichment, and are those hits occurring to the left of the vertical dotted red line. The significance level of ES is calculated by phenotype-based permutation testing, and when a database of gene sets are evaluated, as in this analysis, the significance level is adjusted for multiple hypothesis testing by calculation of a false discovery rate (FDR).

3.3 RESULTS

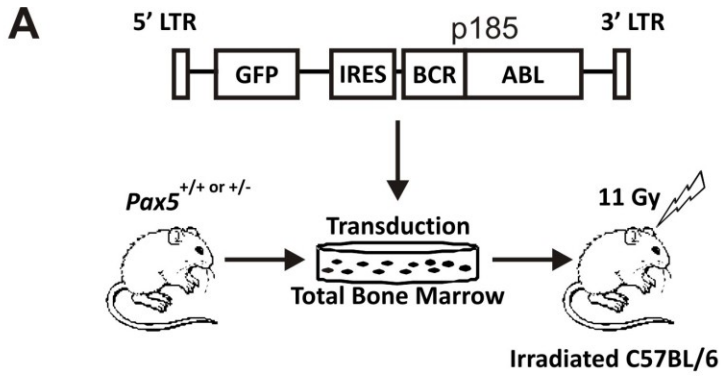
3.3.1 Haploinsufficiency of *Pax5* Renders Mice Susceptible to B-ALL

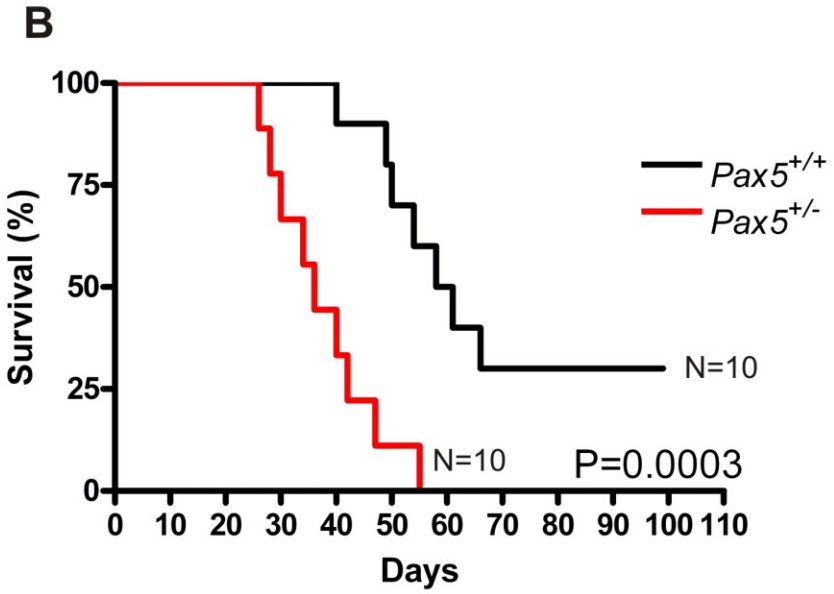
Mice hemizygous for a *Pax5* null allele ($Pax5^{+/-}$) have normal adult B cell development and fail to spontaneously develop leukemia. To directly assess whether the loss of one allele of *Pax5* can cooperate with a known oncogenic lesion in B cell transformation, we directly assessed the ability of *Pax5* haploinsufficiency to cooperate with the expression of the t(9;22) encoded *BCR-ABL1* oncogene, a combination of lesions that is detected in greater than 50% of pediatric *BCR-ABL1* ALL cases. Bone marrow cells from untreated wild-type ($Pax5^{+/+}$) or $Pax5^{+/-}$ mice were transduced with a MSCV-*GFP-IRES-BCR-ABL1* (p185) retrovirus designed to express a low level of the *BCR-ABL1* oncogene from an internal ribosomal entry site. Two million unsorted transduced cells were transplanted after minimal *ex vivo* manipulation into cohorts of lethally irradiated C57BL/6J recipient mice (Figure 3-1A). Expression of *BCR-ABL1* in bone marrow cells from wild type mice induced a pre-B cell lymphoblastic leukemia ($B220^+$, $CD43^+$, $CD19^+$ and $BP1^+$) in recipient animals with a median survival of 60 days post transplant (Figures 3-1B-C; Figure 3-2A). By contrast, mice transplanted with *BCR-ABL1* expressing $Pax5^{+/-}$ cells developed leukemia with significantly decreased latency (median survival of 36 days) and increased penetrance (Figure 3-1B). Interestingly, the leukemias that developed from $Pax5^{+/-}$ cells had a more immature B progenitor phenotype, with the majority containing a mixture of Pre-B cells and Pro-B cells characterized by reduced or absent BP1 expression (Figure 3-2B; Figure 3-3). Moreover, 19% of the *BCR-ABL1* $Pax5^{+/-}$ leukemias had a very immature Pre-Pro-B cell phenotype ($B220^+$, $CD43^+$, $CD19^-$, $BP1^-$) (Figure 3-2C; Figure 3-3), a phenotype that was observed in only a rare *BCR-ABL1* $Pax5^{+/+}$ leukemia.

The primary leukemias that developed from $Pax5^{+/+}$ and $Pax5^{+/-}$ cells were transplantable into sublethally irradiated secondary recipients and resulted in the rapid development of overt leukemia with a phenotype identical to that seen in the donors. Southern blot analysis to assess the clonality of retroviral integrations revealed the vast majority of both the $Pax5^{+/+}$ and $Pax5^{+/-}$ leukemias to be monoclonal, irrespective of their relative stage of B cell differentiation (Figure 3-2D; Figure 3-4; Table 3-1).

Figure 3-1 *Pax5* haploinsufficiency cooperates with *BCR-ABL1* to induce ALL

A, Structure of the retroviral vector used to transduce bone marrow and a schematic representation of the transduction protocol. Briefly, $Pax5^{+/+}$ and $Pax5^{+/-}$ bone marrow was harvested and transduced with the MSCV-*GFP-IRES-BCR-ABL1* retroviruses and transplanted into lethally irradiated C57BL/6J animals. **B**, Kaplan-Meier survival curves from a representative experiment of mice transplanted with MSCV-*GFP-IRES-BCR-ABL1* transduced $Pax5^{+/+}$ or $Pax5^{+/-}$ bone marrow. Leukemias developed in the $Pax5^{+/-}$ mice with a decreased latency when compared to $Pax5^{+/+}$ tumors P=0.0003 (Breslow-Gehan-Wilcoxon test). Similar results were observed in 4 independent retroviral transductions and transplantations experiments with between 10 and 20 mice per group. **C-D**, Bone marrow from $Pax5^{+/+}$ (**C**) and $Pax5^{+/-}$ (**D**) tumors, stained with Hematoxylin-Eosin. Insets, cytospins of bone marrow stained with Wright Giemsa stain.


A
5' LTR
p185
3' LTR
GFP
IRES
BCR
ABL
Pax5+/+ or +/-
Transduction
Total Bone Marrow
11 Gy
Irradiated C57BL/6


B
100
75
50
25
0
Survival (%)
0 10 20 30 40 50 60 70 80 90 100 110
Days
Pax5+/+
Pax5+/-
N=10
N=10
P=0.0003

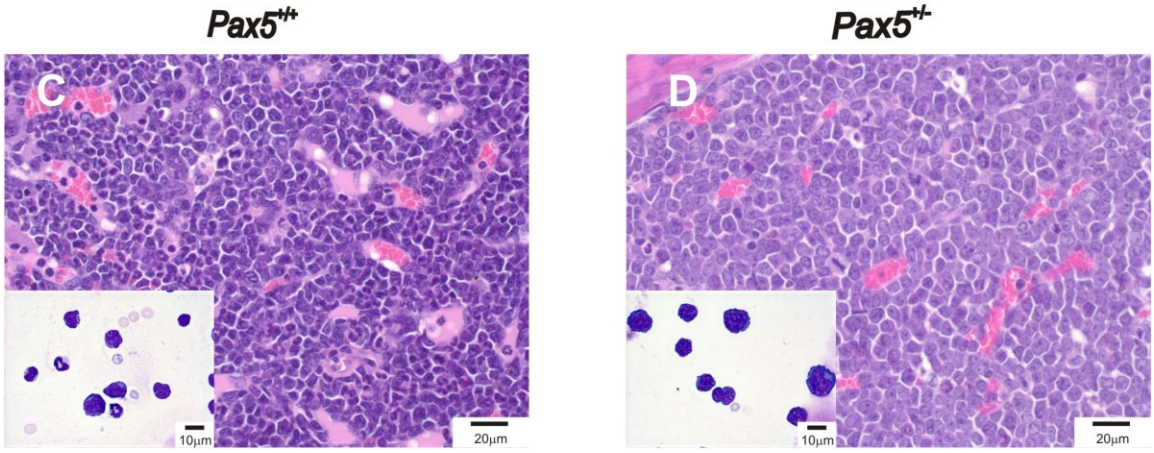
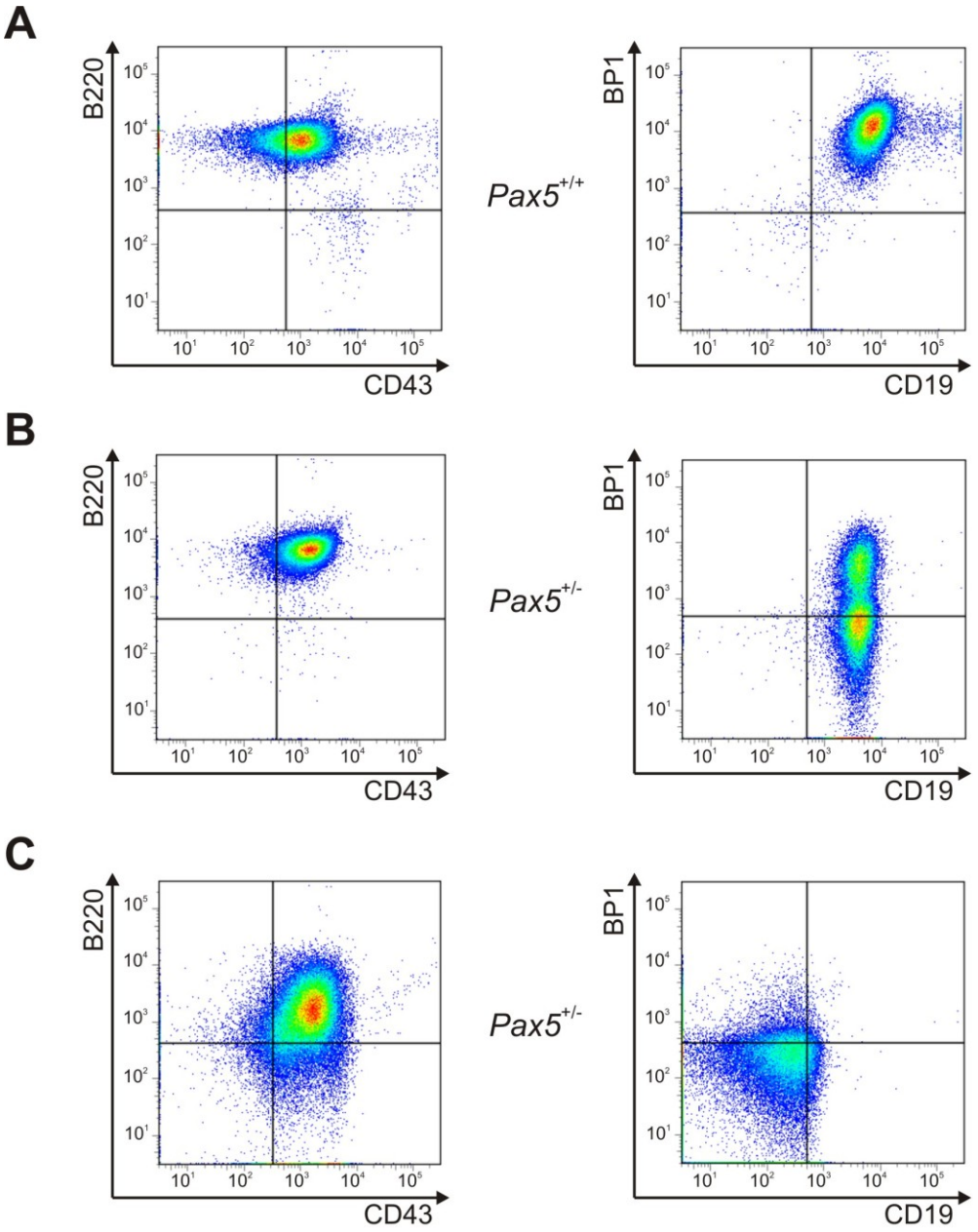

Pax5+/+
Pax5+/-
C
D
10μm
20μm
10μm
20μm

Figure 3-2 Immunophenotype and clonality analysis of *BCR-ABL1* induced leukemias

A, Representative FACS analysis of *BCR-ABL1*/*Pax5*$^{+/+}$ leukemias using antibodies against B220, CD43, CD19, and BP1. **B**, Representative FACS analysis demonstrating the predominant immunophenotype seen in the *BCR-ABL1*/*Pax5*$^{+/-}$ leukemias. **C**, Representative FACS analysis of the immature pre-pro-B cell leukemias identified in 19% of the *BCR-ABL1*/*Pax5*$^{+/-}$ leukemias. **D**, Southern blot analysis of *Hind*III digested genomic DNA from *Pax5*$^{+/+}$and *Pax5*$^{+/-}$ leukemias hybridized with a 724bp GFP probe.


A
B220
CD43
Pax5+/+
BP1
CD19
B
B220
CD43
Pax5+/-
BP1
CD19
C
B220
CD43
Pax5+/-
BP1
CD19
D
Pax5+/+
Pax5+/-
Pre-B
Mixed Pre-B/Pro-B
Pre-Pro B

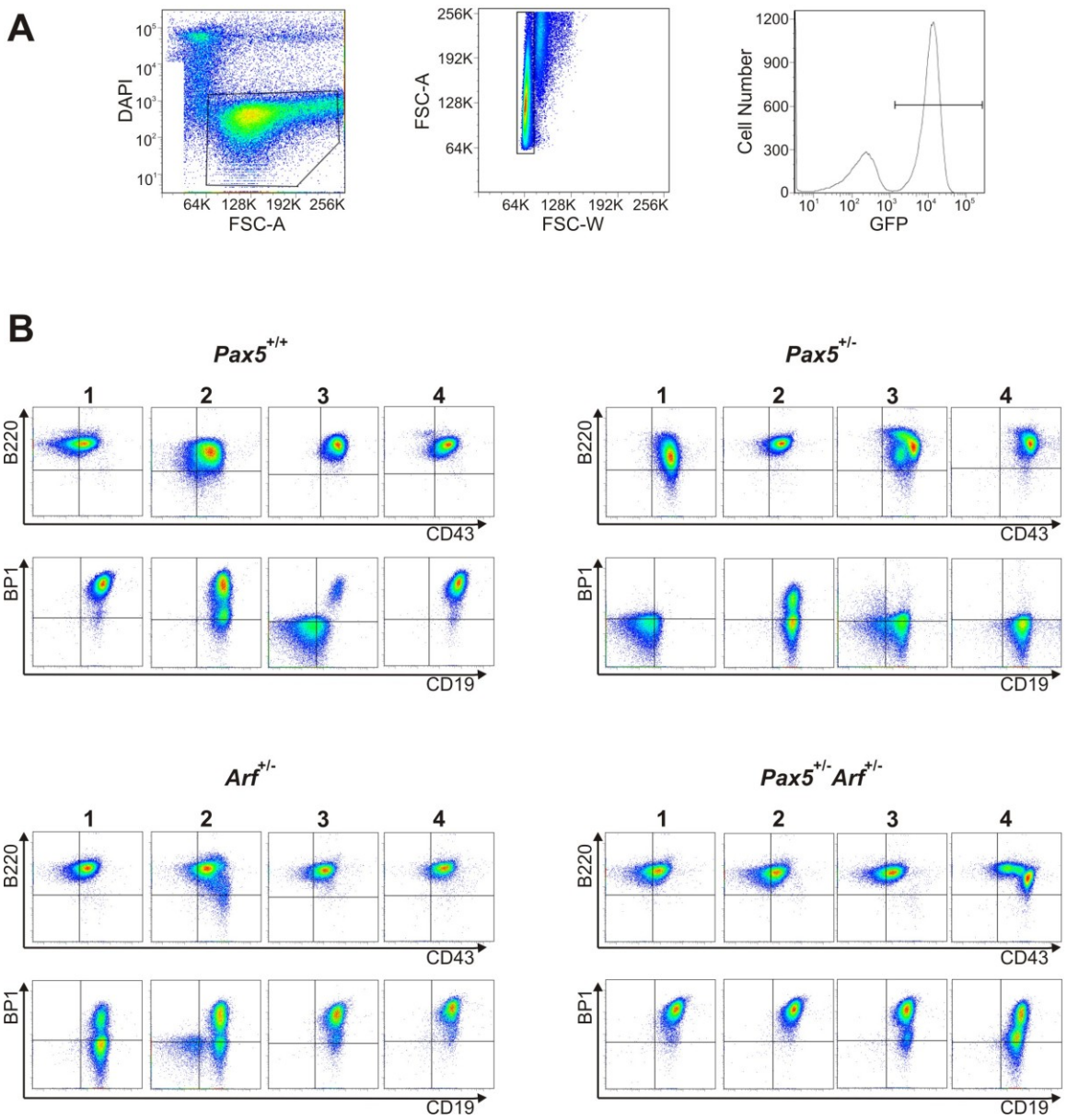
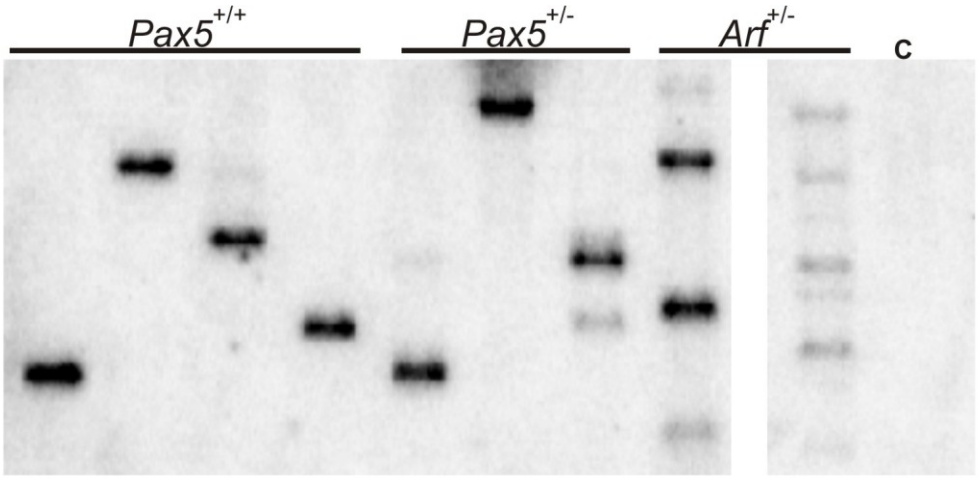


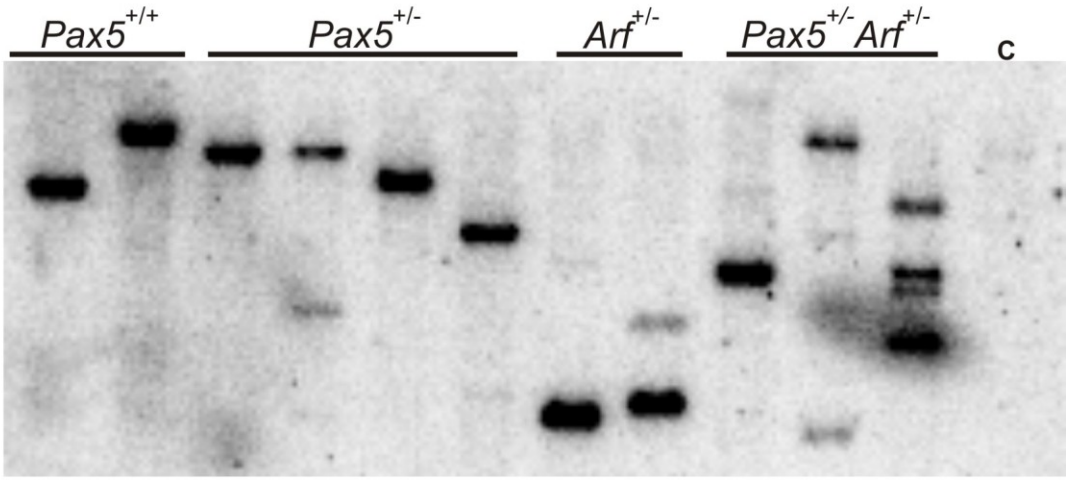
Figure 3-3 Leukemia immunophenotypes

A, Schematic representation of the pathway followed for flow analysis of Pax5 tumors. A live cell gate was drawn using Forward Scatter-Area (FSC-A) and DAPI. A gate was drawn around single cells. Only cells that are GFP positive are analyzed by creating a GFP gate. **B**, *Pax5*$^{+/+}$, *Pax5*$^{+/-}$, *p19Arf*$^{+/-}$, and *Pax5*$^{+/-}$*p19Arf*$^{+/-}$ representative tumors are shown revealing the various immunophenotypes identified in each of the tumor groups.

Figure 3-4 Assessment of leukemia clonality

Southern Blot Analysis of 3 independent experiments is represented. These blots reveal that *Pax5*$^{+/+}$ and *Pax5*$^{+/-}$ tumors are monoclonal while *p19Arf*$^{+/-}$ and *Pax5*$^{+/-}$*p19Arf*$^{+/-}$ tumors are oligoclonal. C=control.


Pax5+/+
Pax5+/-
Arf+/-
C


Pax5+/+
Pax5+/-
Arf+/-
Pax5+/- Arf+/-
C

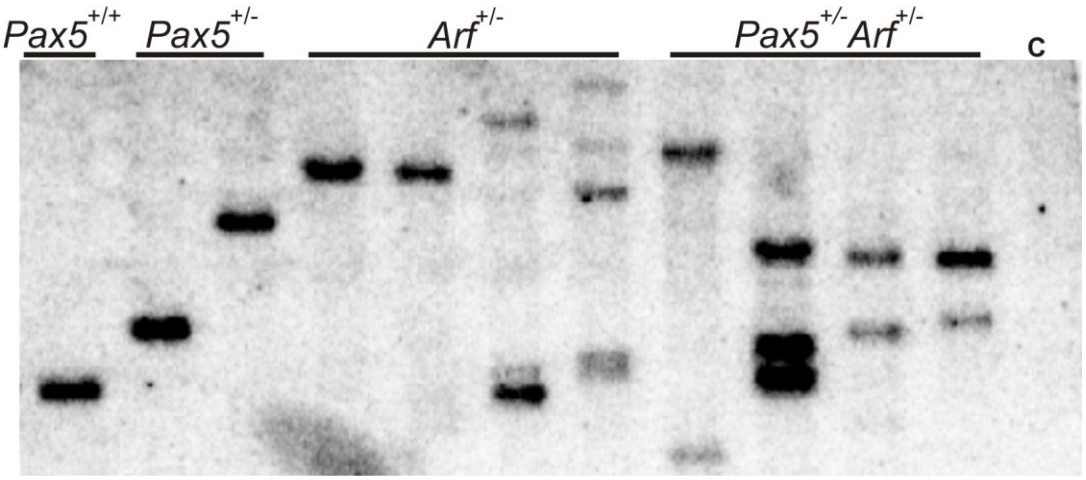

Pax5+/+
Pax5+/-
Arf+/-
Pax5+/- Arf+/-
C

Table 3-1 Clonality analysis of *BCR-ABL1*-induced leukemias

Leukemia genotype (N)	Mono*	Oligo†
Pax5$^{+/+}$ (7)	7	0
Pax5$^{+/-}$ (14)	12	2
Arf$^{+/-}$ (17)	5	12
Pax5$^{+/-}$*Arf*$^{+/-}$ (7)	0	7

*Mono signifies one band by Southern blotting of genomic from bone marrow cells that were either ≥90% GFP+ or if less than 90% was sorted for GFP+ cells;
†Oligo signifies ≥2 bands by Southern blot analysis.

To ensure that the increased susceptibility of *Pax5*$^{+/-}$ mice to leukemia was not restricted to *BCR-ABL1*-induced disease, a cohort of young *Pax5*$^{+/+}$ (N=20) and *Pax5*$^{+/-}$ (N=25) mice were thymectomized and then treated with a single 100mg/kg dose of the alkylating agent N-ethyl-S-nitrosourea (ENU). *Pax5*$^{+/-}$ mice exhibited a markedly increased frequency of ALL in comparison to wild type animals with 24 of 25 *Pax5*$^{+/-}$ animals succumbing to leukemia (median latency 246 days), compared to only 3 of 20 *Pax5*$^{+/+}$ animals (P<0.0001**,** Figure 3-5). Furthermore, 20 of 22 evaluable *Pax5*$^{+/-}$ tumors were of B-progenitor cell lineage, in contrast to the *Pax5*$^{+/+}$ leukemias which were either myeloid (N=2) or T-lymphoid (N=1). Taken together these data demonstrate that haploinsufficiency of *Pax5* results in an increased susceptibility to the development of B progenitor acute lymphoblastic leukemia.

3.3.2 Mono-allelic Loss of *Pax5* Cooperates with *Arf* Loss in *BCR-ABL1* Induced ALL

In addition to *PAX5* deletions, human *BCR-ABL1* ALLs frequently harbor a deletion of *INK4A/ARF* (*CDKN2A*), with over 20% of cases containing deletions of both genes. Expression of *BCR-ABL1* in primary murine pre-B cells induces the Arf checkpoint and triggers an apoptotic response; thus, loss of *Arf* directly enhances the oncogenicity of *BCR-ABL1* to induce ALL in mice. To assess whether mono-allelic deletions of *Pax5* can cooperate with *Arf* loss in *BCR-ABL1* induced leukemia, bone marrow cells from mice heterozygous for null alleles of *Pax5*, *Arf*, or both genes were transduced with MSCV-*GFP-IRES-BCR-ABL1* and transplanted into recipient animals. As shown in Figure 3-6, mice transplanted with *BCR-ABL1* transduced *Pax5*$^{+/-}$ or *Arf*$^{+/-}$ cells developed leukemias with a reduced latency when compared to wild type cells. No significant difference was observed in the latency of disease onset between *Pax5*$^{+/-}$ and *Arf*$^{+/-}$ cells. Transduction of bone marrow cells from compound heterozygous mice (*Pax5*$^{+/-}$ *Arf*$^{+/-}$) followed by transplantation into recipient mice resulted in a further decrease in disease latency over that observed with wild type or single mutant cells (Figure 3-6), thus demonstrating a direct cooperativity of these two lesions in the development of *BCR-ABL1*-induced leukemia. Importantly, the immunophenotype of the *BCR-ABL1*-induced leukemias were primarily dictated by the *Pax5* gene status, with a shift toward a more immature B cell phenotype with the loss of one *Pax5* allele (Figure 3-3). Interestingly, in contrast to the clonal leukemias that developed in wild type or *Pax5*$^{+/-}$ mice, Southern blot analysis of the *Arf*$^{+/-}$ leukemias revealed the majority to be oligoclonal in nature (Figure 3-6; Figure 3-4).

3.3.3 Mutations of Pax5 and Ink4a/Arf are Spontaneously Selected For during Leukemogenesis

Our data demonstrate that mono-allelic loss of *Pax5* not only cooperates with *BCR-ABL1* in leukemogenesis, but further enhance the known cooperativity of *Arf* loss. These data raise the possibility that *BCR-ABL1*-induced leukemias may spontaneously

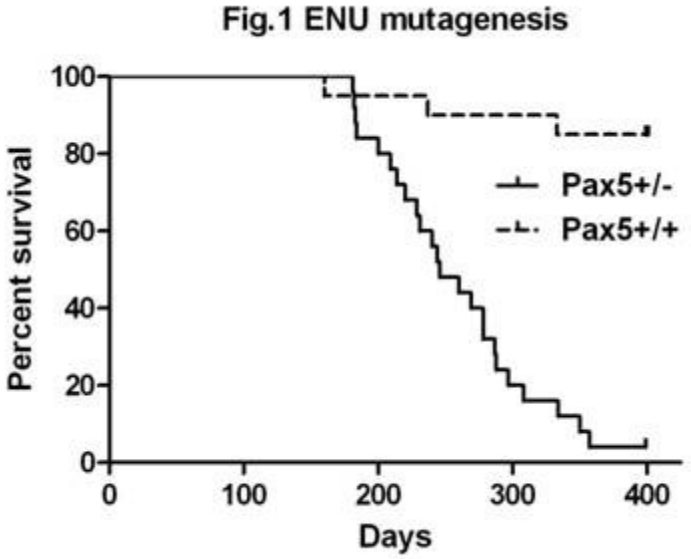
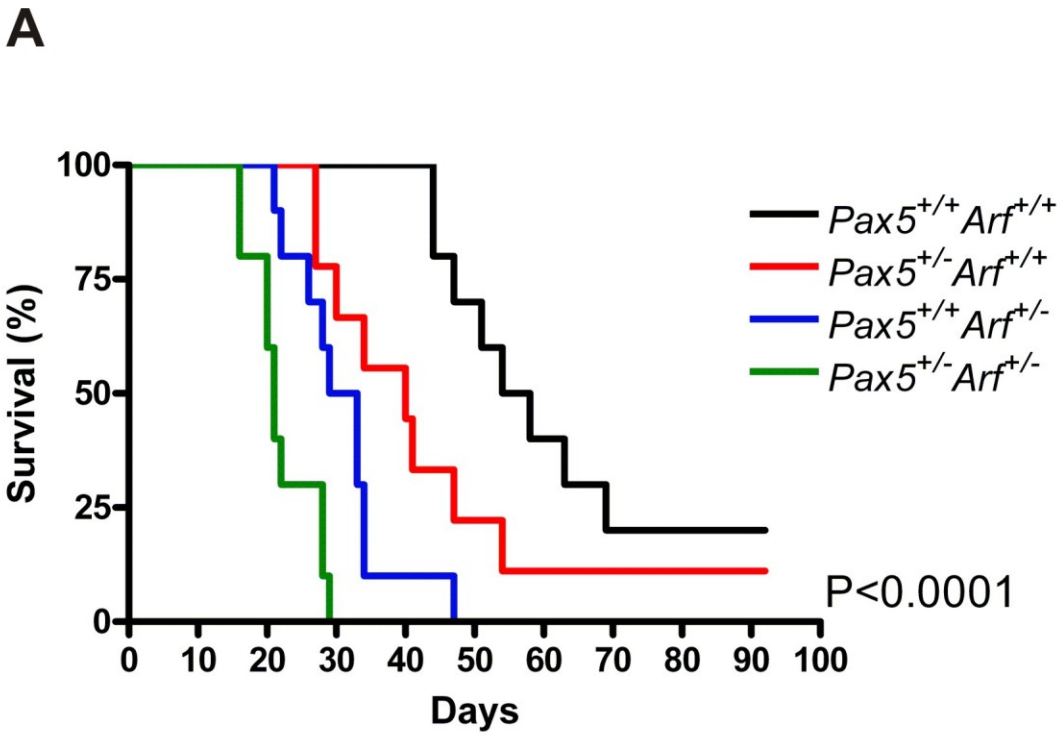


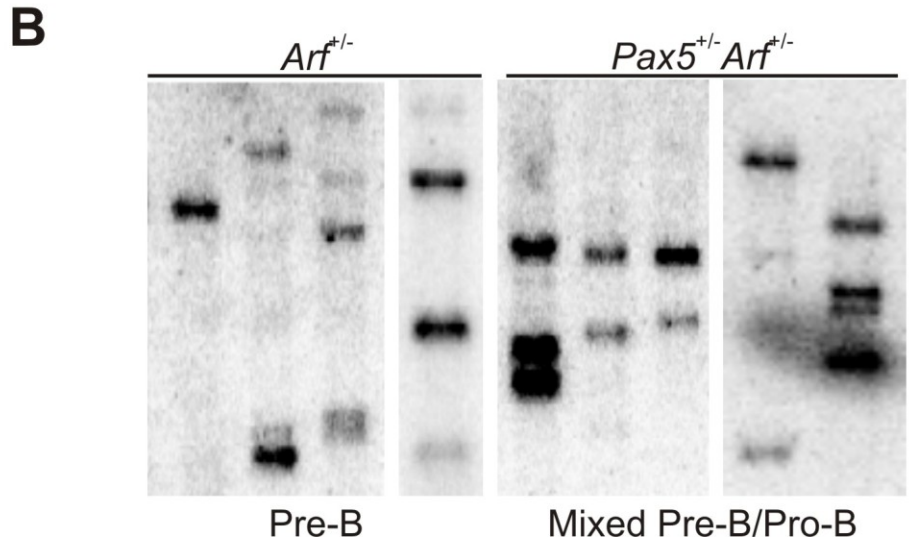
Figure 3-5 Kaplan-Meier survival curves of mice treated with ENU

Kaplan-Meier survival curve of mice that were injected with one dose of 100mg/kg ENU (N-ethyl-S-nitrosourea). Twenty four of twenty five *Pax5*$^{+/-}$ animals succumbed to disease with a median survival of 246 days while only three of twenty *Pax5*$^{+/+}$ animals developed leukemia ($P<0.001$). These data suggest that *Pax5* haploinsufficiency cooperates during leukemogenesis in multiple animal models.

Figure 3-6 *Pax5* and *Ink4a/Arf* haploinsufficiency cooperate with *BCR-ABL1* to induce ALL

A, Kaplan-Meier plot showing survival of mice transplanted with MSCV-*GFP-IRES-BCR-ABL1* transduced $Pax5^{+/+}$, $Pax5^{+/-}$, $Arf^{+/-}$ or $Pax5^{+/-}$ $Arf^{+/-}$ bone marrow cells ($p<0.0001$). $Pax5^{+/-}$ and $Arf^{+/-}$ leukemias develop with a decreased latency when compared to $Pax5^{+/+}$ leukemias, $p = 0.01$ and $p < 0.001$, respectively. $Pax5^{+/-}$ and $Arf^{+/-}$ tumors develop with no significant change in latency when compared to each other $p = 0.077$. $Pax5^{+/-}$ $Arf^{+/-}$ tumors develop with a decreased latency when compared to $Pax5^{+/-}$ or $Arf^{+/-}$ tumors $p = 0.007$ and $p = 0.0045$, respectively. All p values calculated using Breslow-Geahn-Wilcoxon test. N=10 in all experimental arms. **B**, Southern Blot Analysis of *Hind*III digested genomic DNA from $Arf^{+/-}$ and $Pax5^{+/-}$ $Arf^{+/-}$ leukemias hybridized with a 724bp GFP probe.


A
100
75
50
25
0
Survival (%)
0 10 20 30 40 50 60 70 80 90 100
Days
Pax5+/+ Arf+/+
Pax5+/- Arf+/+
Pax5+/+ Arf+/-
Pax5+/- Arf+/-
P<0.0001


B
Arf+/-
Pax5+/- Arf+/-
Pre-B
Mixed Pre-B/Pro-B

select for mutations in these genes. Further supporting this possibility was the occurrence of a Pre-Pro-B cell immunophenotype in a subset of *Pax5*$^{+/-}$ leukemias, and in rare wild type and *Arf*$^{+/-}$ cases, suggesting that this immature phenotype may results from additional lesions in B cell development genes. To explore these hypotheses, we performed array-based comparative genomic hybridization (aCGH) on 50 *BCR-ABL1*-induced leukemias (15 *Pax5*$^{+/+}$, 25 *Pax5*$^{+/-}$, and 10 *Arf*$^{+/-}$) to identify regions of copy number alterations (CNAs). For this analysis we used Agilent (Santa Clara, CA, USA) murine oligonucleotide microarrays containing over 450,000 probes with a resolution across the genome of approximately 4.5kb. To further enhance the resolution of this platform, we customized the arrays by tiling at a 250 bp resolution through the 20 genes that are most commonly mutated by CNAs in human *BCR-ABL1* ALL. Included in this list of genes are *Ink4a/Arf*, *Pax5*, *Ebf1*, and the *Ikaros* gene family *Ikzf1*, *Ikzf2* and *Ikzf3*. We also sequenced all coding exons of *Pax5* in the leukemia cells from this cohort.

The aCGH data revealed a surprisingly low number of CNAs/leukemia with a mean of only 0.96 CNAs/case (range 0-5), with losses (0.62, range 0-3) twice as common as gains (0.34, range 0.3) (Table 3-2). The frequency of CNAs differed across the various genotypes, with a mean of 1.73 CNAs/leukemia (range 0-5) in wild type mice compared to only 0.60 and 0.70 in *Pax5*$^{+/-}$ and *Arf*$^{+/-}$ leukemias, respectively. Despite the paucity of CNAs, a recurrent gain of chromosome 5 or 5q was identified in 7/15 BCR-ABL *Pax5*$^{+/+}$ and 1/10 BCR-ABL1 *Arf*$^{+/-}$ leukemias, but never in *BCR-ABL1 Pax5*$^{+/-}$ leukemias (Figure 3-7A). In addition, a number of other recurrent lesions were identified, with the most common abnormalities targeting *Pax5*, *Ikzf3*, and *Ink4a/Arf* (Figure 3-7; Figure 3-8; Table 3-3).

Three leukemias (2 *Pax5*$^{+/+}$ and 1 *Pax5*$^{+/-}$) had focal deletions targeting the *Ink4a/Arf Ink4b* locus, including a bi-allelic deletion in one of the *Pax5*$^{+/+}$ leukemias that included the entire *Ink4a/Arf Ink4b* locus (Figure 3-7B-C). The second *Pax5*$^{+/+}$ leukemia had a deletion that included the common exon of the *Ink4a* locus, exon 2, but did not include *Ink4b* (Figure 3-7B-C). The *Pax5*$^{+/-}$ leukemia had a deletion that extended from exon 1β through the end of *Ink4b* (Figure 3-7B-C). Taken together, the common region of deletion in the three leukemias was *Arf*, thereby definitively identifying *Arf* as the target of the mutations. Thus, loss of *Ink4a/Arf* is spontaneously selected for in *BCR-ABL1*-induced leukemia. To assess whether *Arf* was inactivated in the absence of genomic alterations, we performed western blot analysis on protein lysates from a subset of the leukemias and observed only a rare leukemia (two *Pax5*$^{+/-}$) with markedly reduced expression of p19^{ARF} consistent with silencing. By contrast, expression of p19^{Arf} was detected in *Arf*$^{+/-}$ leukemias, consistent with haploinsufficiency.

The most common targets of CNAs in the murine leukemias were genes involved in normal B cell development including *Pax5, Ebf1*, and *Ikzf3* (Figure 3-8; Figure 3-9; Table 3-3). In addition 4 leukemias were found to have point mutations of *Pax5* effecting conserved residues in the paired DNA-binding domain, or resulting in a frame-shift mutation in the transcriptional regulatory domain (Figure 3-10; Table 3-4). Collectively, mutations targeting these genes were identified in 30% of analyzed leukemias that arose from *Pax5*$^{+/+}$ or *Arf*$^{+/-}$ cells. Even more surprising was the identification of a high

Table 3-2 Mean number of CNA in murine leukemias

Subtype	N	Gains (range)	Losses (range)	Total (range)
$Pax5^{+/+}$	15	0.93 (0-3)	0.80 (0-3)	1.73 (0-5)
$Pax5^{+/-}$	25	0.04 (0-1)	0.56 (0-1)	0.60 (0-2)
$Arf^{+/-}$	10	0.20 (0-1)	0.50 (0-2)	0.70 (0-2)
All cases	50	0.34 (0-3)	0.62 (0-3)	0.96 (0-5)

Figure 3-7 Detection of recurrent CNAs in murine *BCR-ABL1* leukemias

A, Heatmap of $\log_2$ ratio of array CGH copy number data (median smoothed with a window of five markers) for 50 murine *BCR-ABL1* leukemias. The leukemias are grouped according to their genotype ($Pax5^{+/+}$, $Pax5^{+/-}$, and $Arf^{+/-}$) and the probes are group by chromosomes as listed on the left. Blue depicts regions of loss and red regions of gain with the relative scale shown in the lower right. **B**, Heatmap of $\log_2$ ratio of array CGH copy number data of the region on chromosome 4 that contains *Ink4a/Arf* and *Ink4b* for 3 *BCR-ABL1* leukemias that have acquired CNAs involving this locus. **C**, Genomic map of the *Ink4a/Arf* and *Ink4b* locus showing the location of the CGH oligonucleotide probes (blue lines at the top) and the position of the individual genes contained within this genomic region (bottom). * indicates a gap in the UCSC genome build. The $\log_2$ copy number data for the three *BCR-ABL1* leukemias are depicted. The red arrows indicate the genomic location of exons analyzed by quantitative PCR for copy number validation.

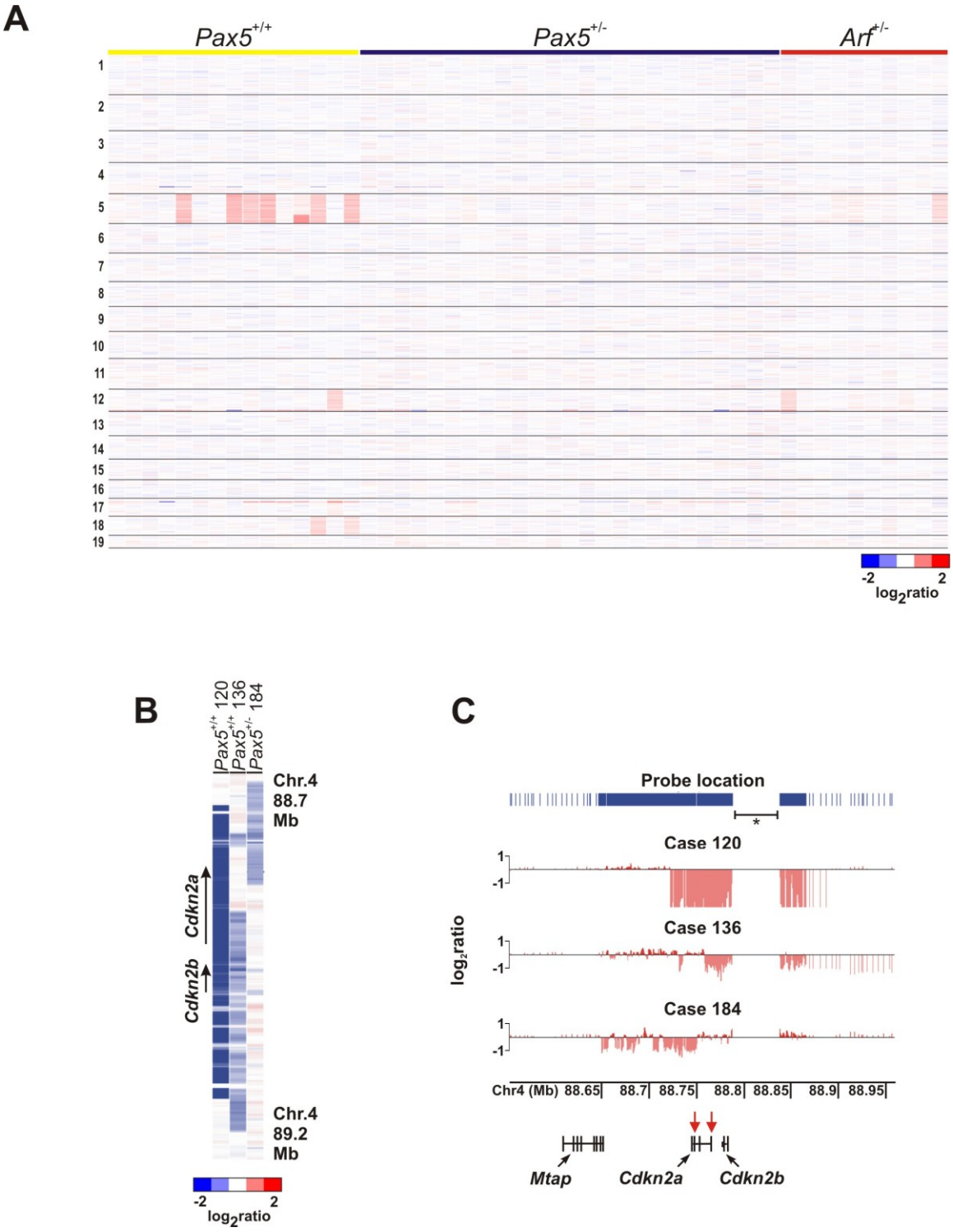

A
Pax5+/+
Pax5+/-
Arf+/-
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
-2
2
log2ratio
B
Pax5+/+ 120
Pax5+/+ 136
Pax5+/- 184
Chr.4
88.7
Mb
Cdkn2a
Cdkn2b
Chr.4
89.2
Mb
-2
2
log2ratio
C
Probe location
*
Case 120
Case 136
Case 184
log2ratio
1
-1
Chr4 (Mb) 88.65 88.7 88.75 88.8 88.85 88.9 88.95
Mtap
Cdkn2a
Cdkn2b

Figure 3-8 *Pax5* is a common target of copy number alterations

A, Heatmap of $\log_2$ ratio of array CGH copy number data of the region on chromosome 4 that contains *Pax5* for 2 *BCR-ABL1 Pax5*$^{+/-}$ leukemias that have acquired deletions of the non-targeted, wild type *Pax5* allele. **B**, Genomic map of the *Pax5* locus showing the location of the CGH oligonucleotide probes (blue lines at the top) and the position of the gene contained within this genomic region (Bottom). The $\log_2$ ratio data for two *BCR-ABL1* leukemias are depicted. The red arrows indicate the genomic location of the exons analyzed by quantitative PCR for copy number validation. **C**, Heatmap of $\log_2$ ratio of array CGH copy number data of the region containing *Pax5* exon 2 (represented by a single probe within this region). *Pax5* exon 2 was deleted in the targeted *Pax5* null allele. A deletion involving the single exon 2 probe was detected in 7 *Pax5*$^{+/-}$ leukemias, suggesting gene conversion leading to bi-allelic targeted alleles. **D**, Genotyping PCR confirmed loss of the wild type allele in leukemias that had deleted the exon 2 probe (designated by an *). Western blot analysis also demonstrated an absence of Pax5 protein expression in these leukemias. WT, wild type band; KO, knock out band.

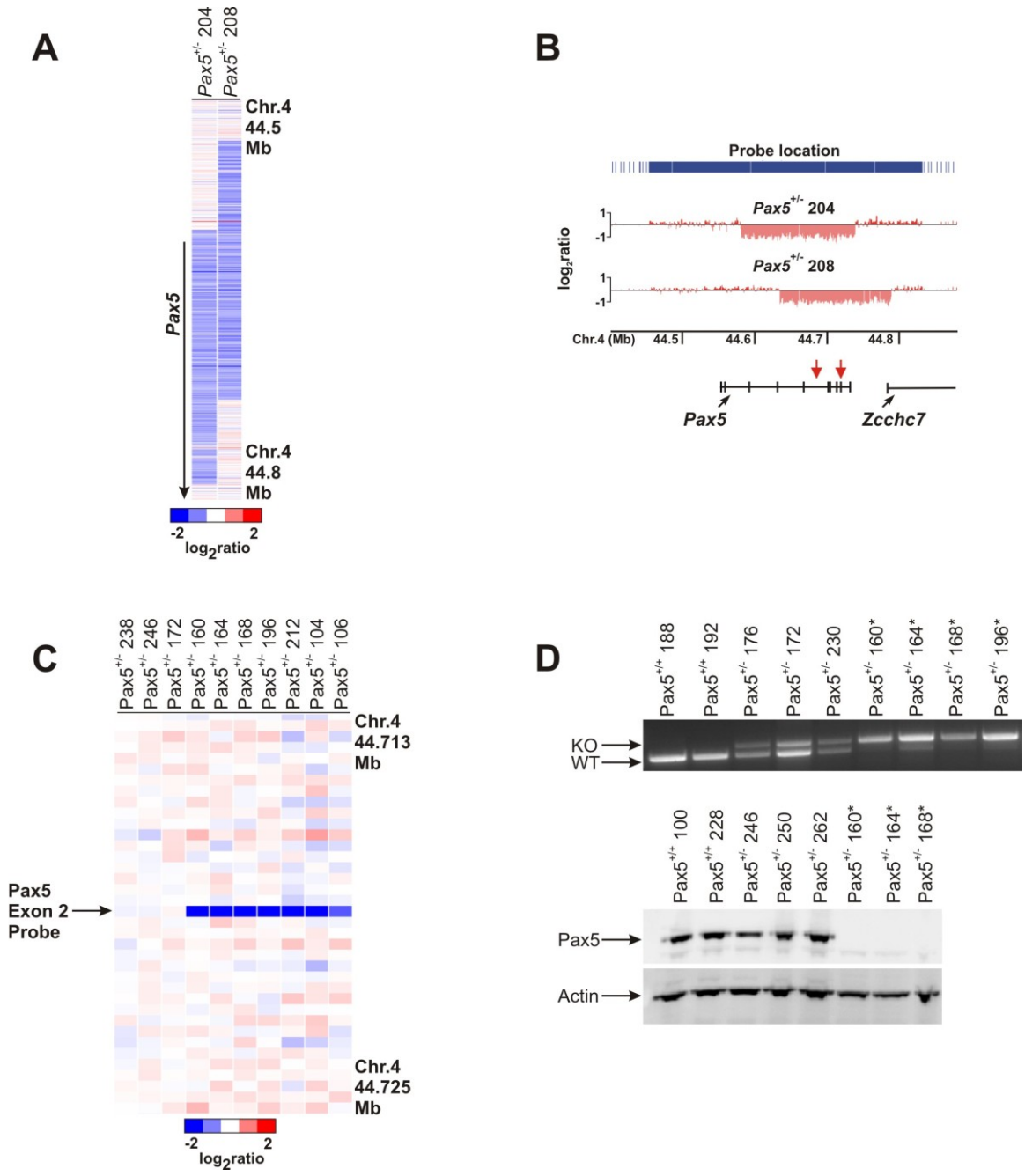

A
Pax5+/- 204
Pax5+/- 208
Chr.4
44.5
Mb
Pax5
Chr.4
44.8
Mb
-2
2
log2ratio
B
Probe location
Pax5+/- 204
Pax5+/- 208
log2ratio
Chr.4 (Mb)
44.5
44.6
44.7
44.8
Pax5
Zcchc7
C
Pax5+/- 238
Pax5+/- 246
Pax5+/- 172
Pax5+/- 160
Pax5+/- 164
Pax5+/- 168
Pax5+/- 196
Pax5+/- 212
Pax5+/- 104
Pax5+/- 106
Chr.4
44.713
Mb
Pax5
Exon 2
Probe
Chr.4
44.725
Mb
-2
2
log2ratio
D
Pax5+/+ 188
Pax5+/+ 192
Pax5+/- 176
Pax5+/- 172
Pax5+/- 230
Pax5+/- 160*
Pax5+/- 164*
Pax5+/- 168*
Pax5+/- 196*
KO
WT
Pax5+/+ 100
Pax5+/+ 228
Pax5+/- 246
Pax5+/- 250
Pax5+/- 262
Pax5+/- 160*
Pax5+/- 164*
Pax5+/- 168*
Pax5
Actin

Table 3-3 Recurring CNA (rank ordered by frequency)

Genes	Loss/ Gain	*Pax5*$^{+/+}$ N=15	*Pax5*$^{+/-}$ ‡ N=25	*Arf*$^{+/-}$ N=10	# total lesions
*Pax5**	Loss†	2 (13%)	9 (36%)	1 (10%)	12
Ikzf3	Both	3 (20%)	3 (12%)	3 (30%)	10
+5	Gain	7 (47%)	0	1 (10%)	8
Ink4a/Arf Ink4b	Loss	2 (13%)	1 (4%)	0	3
+12	Gain	1 (7%)	0	1 (10%)	2
+18	Gain	2 (13%)	0	0	2

*Includes loss of heterozygosity (LOH) in 7 Pax5$^{+/-}$ leukemias.
†Eleven of twelve CNA were deletions while one was amplification.
‡*Pax5*$^{+/-}$ leukemias reveal a significant increase in acquired Pax5 lesions, however, the cohort has been biased by the selection of leukemias that harbor an immature immunophenotype.

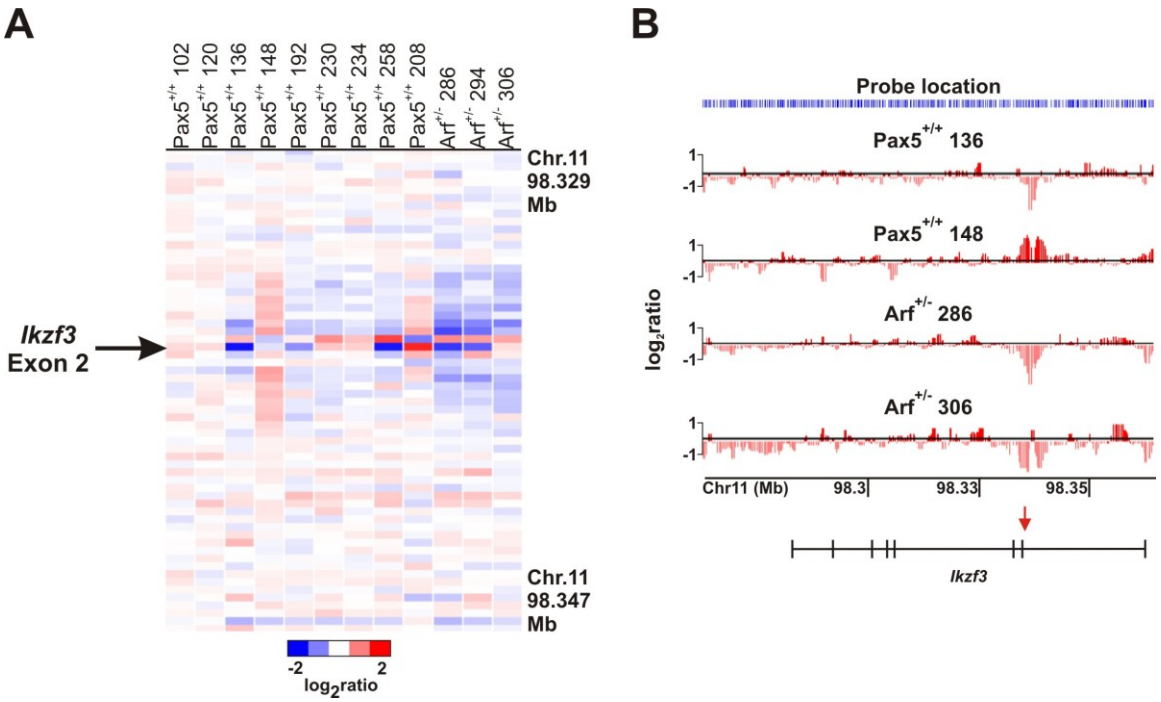


Figure 3-9 *Ikzf3* copy number analysis

A, Log_2 ratio aCGH copy number data of chromosome 11 flanking *Ikzf3* exon 2 for 12 representative leukemias (9 with CNA and 3 without) showing deletion or amplification of *Ikzf3* exon 2. **B**, Coverage of the *Ikzf3* locus for four of the leukemias shown in A. Each vertical line represents the genomic position and log ratio of an individual marker. A Red arrow indicates the genomic location of a quantitative PCR assay.

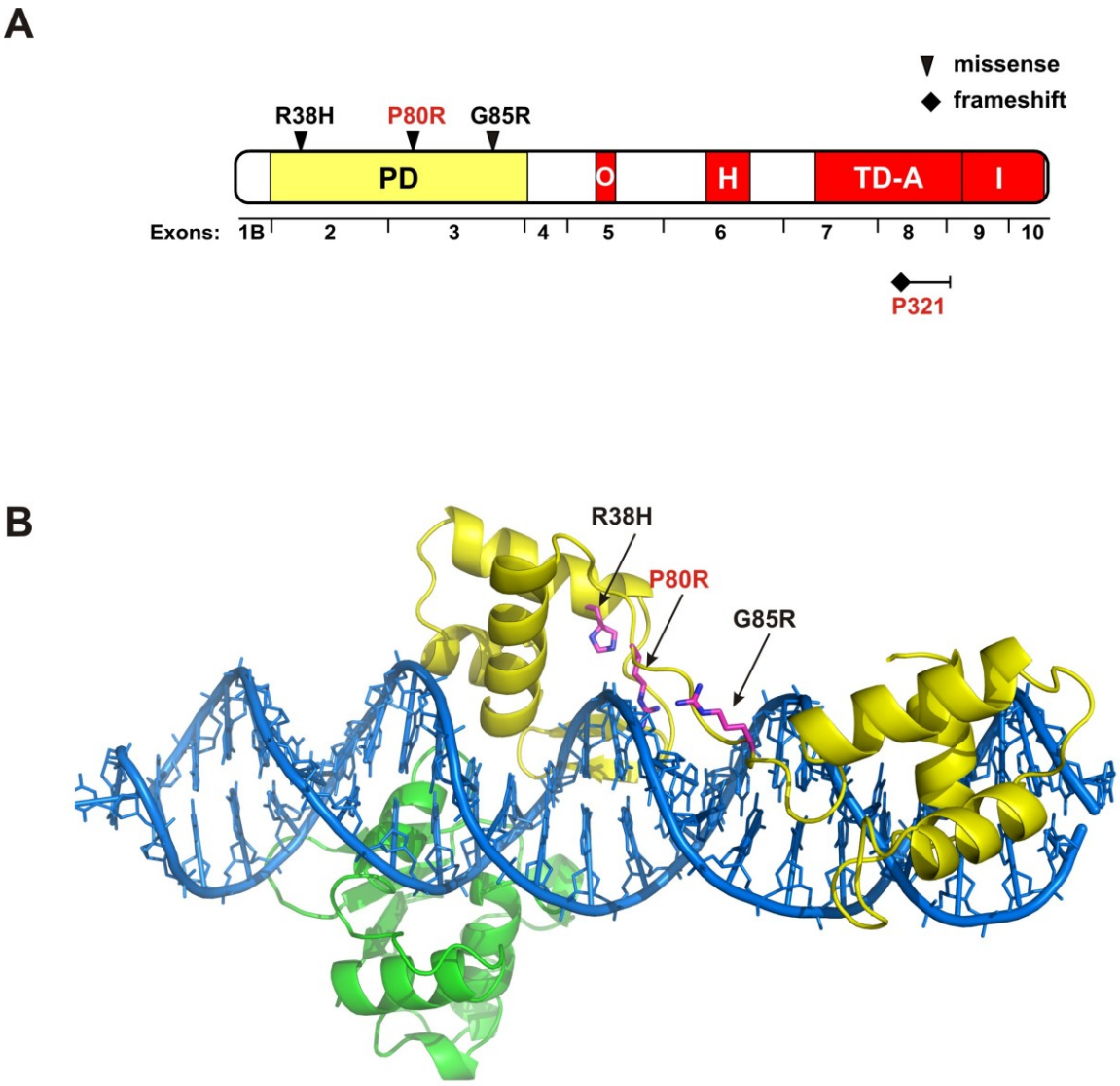


Figure 3-10 *Pax5* point mutations

A, Schematic of the functional domains of Pax5 showing the location of missense (▼) and frameshift (♦) mutations detected in this study. The mutations in red correspond to the most frequent Pax5 mutation detected in human pediatric ALLs. PD, paired domain, O, octapeptide domain, H, homeodomain, TD, transactivating domain, A, activating, I, inhibitory. **B**, Structural modeling of the position of the Pax5 paired domain mutations and their relationship to DNA-binding and Ets protein interaction. The DNA double helix is blue, the paired Pax5 DNA binding domain is yellow, and ETS-1, which binds to Pax5 resulting in an increase in the latter's DNA-binding activity, is green. Pax5 mutations that have been previously identified in human pediatric ALL are shown in red. Each mutation is predicted to either disrupt DNA-binding or alter the transcriptional regulatory domains of the protein.

Table 3-4 Mutation analysis (including CNAs and re-sequencing of *Pax5*)

ID	Genotype	IP*	*Pax5* Mutation†	Other CNA
100	*Pax5*$^{+/+}$	Pre-B		
102	*Pax5*$^{+/+}$	Pre-B		
116	*Pax5*$^{+/+}$	Pre-pro B	homo G85R	
120	*Pax5*$^{+/+}$	Pre-B		*Arf* deln
124	*Pax5*$^{+/+}$	Pre-B		*Ebf1* Δ6-7/+5
128	*Pax5*$^{+/+}$	Pre-B		
228	*Pax5*$^{+/+}$	Pre-B		
136	*Pax5*$^{+/+}$	Pre-B	intron 5 deln	*Arf* deln/*Ikzf3* deln/+5
140	*Pax5*$^{+/+}$	Pre-B		+5
144	*Pax5*$^{+/+}$	Pre-B		+5
148	*Pax5*$^{+/+}$	Pre-B	intron 5 amp	*Ikzf3* amp
152	*Pax5*$^{+/+}$	Pre-B		
156	*Pax5*$^{+/+}$	Pre-B		+5/+18
188	*Pax5*$^{+/+}$	Pre-B		+12
192	*Pax5*$^{+/+}$	Pre-B		*Ikzf3* deln/+5
176	*Pax5*$^{+/-}$	Pre-B		
184	*Pax5*$^{+/-}$	Pre-B		*Arf* deln
172	*Pax5*$^{+/-}$	Pro-B	P321fs	
180	*Pax5*$^{+/-}$	Pro-B	P80R	
230	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		*Ikzf3* deln
234	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		*Ikzf3* deln
238	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		
242	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		
246	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		
250	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		
254	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		
258	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		*Ikzf3* deln
262	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		
266	*Pax5*$^{+/-}$	Mixed Pre-B/Pro-B		
104	*Pax5*$^{+/-}$	Pre-pro B	LOH	
106	*Pax5*$^{+/-}$	Pre-pro B	LOH	
160	*Pax5*$^{+/-}$	Pre-pro B	LOH	
164	*Pax5*$^{+/-}$	Pre-pro B	LOH	
168	*Pax5*$^{+/-}$	Pre-pro B	LOH	
196	*Pax5*$^{+/-}$	Pre-pro B	LOH	
200	*Pax5*$^{+/-}$	Pre-pro B	R38H	
204	*Pax5*$^{+/-}$	Pre-pro B	Δ1-6	
208	*Pax5*$^{+/-}$	Pre-pro B	Δ1-8	
212	*Pax5*$^{+/-}$	Pre-pro B		
216	*Pax5*$^{+/-}$	Pre-pro B	LOH	

Table 3-4 continued

ID	Genotype	IP*	Pax5 Mutation†	Other CNA
270	*Arf*$^{+/-}$	Pre-B		+12
274	*Arf*$^{+/-}$	Pre-B		
278	*Arf*$^{+/-}$	Pre-B		
282	*Arf*$^{+/-}$	Pre-B		
286	*Arf*$^{+/-}$	Pre-B	intron 5 deln	*Ikzf3* deln
290	*Arf*$^{+/-}$	Pre-B		
294	*Arf*$^{+/-}$	Pre-B		*Ikzf3* deln
298	*Arf*$^{+/-}$	Pre-B		
302	*Arf*$^{+/-}$	Pre-B		
306	*Arf*$^{+/-}$	Pre-B		*Ikzf3* deln

*IP, immunophenotype; LOH, loss of heterozygosity; Δ, deletion
†Pax5 mutations include deletions, amplification, LOH, and sequence mutations

frequency of mutations (CNAs and point mutations) in the non-targeted *Pax5* allele in *BCR-ABL1* $Pax5^{+/-}$ leukemias. Specifically, two $Pax5^{+/-}$ leukemias had focal deletions affecting a subset of the *Pax5* coding exons, three had *Pax5* point mutations, and 7 had undergone a gene conversion (copy neutral loss-of-heterozygosity) event leading to loss of the wild type allele, duplication of the targeted *Pax5* null allele, and an absence of Pax5 expression (Figure 3-8C-D; Table 3-4). Thirteen of the $Pax5^{+/-}$ leukemias were selected for genomic analysis because of a very immature Pro-B or Pre-Pro-B immunophenotype. Remarkably, 12 of these 13 leukemias had mutations of the wild type Pax5 allele. In addition, a single leukemia that arose in $Pax5^{+/+}$ cells had a very immature Pre-Pro-B cell immunophenotype and was found to have a homozygous G85R Pax5 DNA-binding domain mutation (Figure 3-10; Table 3-4). Thus, spontaneous inactivation of the wild type *Pax5* allele is selected for at a relatively high frequency and correlates with a more immature phenotype. Taken together these data demonstrate a strong selective pressure for the acquisition of loss of function mutations in genes that regulate the normal transcriptional cascade required for early B cell differentiation.

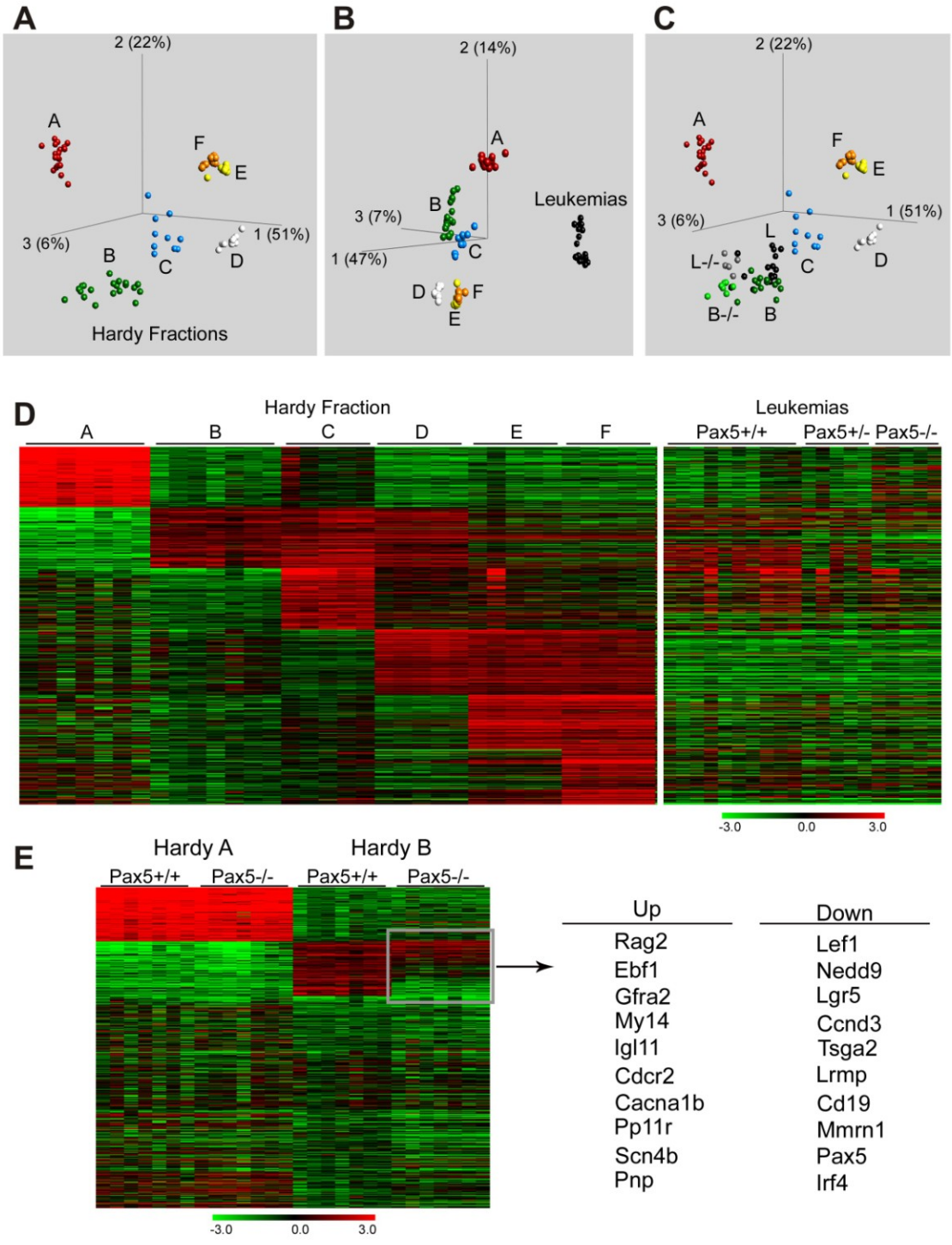
3.3.4 Gene Expression Analysis Reveals Similarities between the Leukemias and Their Normal Counterparts

Due to the importance of *Pax5* during normal B-cell development, gene expression profiling of flow sorted cells from distinct stages of B-cell development i.e., Hardy fractions A-F from $Pax5^{+/+}$ and $Pax5^{+/-}$ mice were performed. In addition, to investigate the effect of complete loss of *Pax5* on the gene expression pattern during early B-cell development, cells from Hardy fraction A and B, were flow sorted from $Pax5^{-/-}$ mice. In our first analysis, we used principal component analysis (PCA) and multi-group comparison to explore the gene expression changes that occur during normal development. This analysis revealed that cells from the different Hardy Fractions have distinct gene expression patterns ($q = 7.2 \times 10^{-19}$) and the PCA accurately captures the temporal changes that occur during normal differentiation (Figure 3-11A). This analysis also showed that subtle gene expression changes occur with regard to a heterozygous loss of *Pax5*. In Hardy fraction A, no difference in the gene expression pattern could be detected with PCA or other supervised methods in bone marrow cells obtained from $Pax5^{+/+}$, $Pax5^{+/-}$, or $Pax5^{-/-}$ mice. As expected, in Hardy fraction B, both PCA and supervised analyses revealed that BM from $Pax5^{-/-}$ mice has a distinct gene expression profile. However, the difference that could be seen between $Pax5^{+/+}$ and $Pax5^{+/-}$ cells was very subtle using this algorithm.

Next, we investigated the relationship between normal flow sorted B-cells and the leukemias from $Pax5^{+/+}$, $Pax5^{+/-}$, and $Pax5^{+/-}$ mice that later lost their second *Pax5* allele. PCA showed that all leukemias, irrespective of their *Pax5* status, were distinct from their normal counterparts (Figure 3-11B). To investigate if the leukemias shared more subtle transcriptional programs with any of the normal Hardy fractions, we applied PCA obtained from the normal cells only (Figure 3-11) on the leukemias. In this analysis, the leukemias do not contribute to the principal components, allowing us to investigate the

Figure 3-11 Gene expression analysis

A, Principal Component Analysis of Hardy fraction development on Hardy stages A-F. All three *Pax5* genotypes are included. L, leukemias; L-/-, $Pax5^{-/-}$ leukemias; B-/-, $Pax5^{-/-}$ Hardy fraction B. **B**, PCA of Hardy fractions and *Pax5* leukemias reveals that leukemias harbor a gene expression profile that is unique when compared to normal B-cells. **C**, The PCA analysis from part A was locked and the leukemias were overlaid in order to determine which Hardy stage of B-cell development the leukemias most closely resemble. This analysis revealed that the $Pax5^{+/+}$ leukemias most closely resemble Hardy fraction C. The leukemias that have lost one *Pax5* allele are most similar to Hardy fraction B, while the leukemias that have lost both functional *Pax5* alleles, L-/-, most closely resemble Hardy fraction B cells from $Pax5^{-/-}$ animals, depicted here as B-/-. **D**, The top 100 genes that define Hardy Fractions A through F are represented along with 20 mouse leukemias (10 $Pax5^{+/+}$ and 5 $Pax5^{+/-}$ and 5 $Pax5^{+/-}$ that have undergone LOH and are effectively $Pax5^{-/-}$). These gene expression profiles reveal that the *Pax5* leukemias harbor gene expression profiles that are relatively immature; sharing the most similarity with the gene expression profile of Hardy fractions B and C. **E**, Gene expression analysis assessing changes that occur in Hardy fraction B when *Pax5* is lost. $Pax5^{-/-}$ Hardy fraction B cells harbor a completely dysregulated gene expression profile when compared to wild-type Hardy fraction B cells. Several dysregulated *Pax5* targets are included in the top 100 genes that define Hardy fraction B, including *Rag2*, *Lef1*, and *CD19*.


A
2 (22%)
A
F
E
3 (6%)
B
C
D
1 (51%)
Hardy Fractions
B
2 (14%)
A
B
3 (7%)
C
1 (47%)
D
F
E
Leukemias
C
2 (22%)
A
F
E
3 (6%)
1 (51%)
L
L-/-
C
D
B-/-
B
D
Hardy Fraction
A
B
C
D
E
F
Leukemias
Pax5+/+
Pax5+/-
Pax5-/-
-3.0
0.0
3.0
E
Hardy A
Hardy B
Pax5+/+
Pax5-/-
Pax5+/+
Pax5-/-
-3.0
0.0
3.0
Up
Rag2
Ebf1
Gfra2
My14
Igl11
Cdcr2
Cacna1b
Pp11r
Scn4b
Pnp
Down
Lef1
Nedd9
Lgr5
Ccnd3
Tsga2
Lrmp
Cd19
Mmrn1
Pax5
Irf4

similarities between the normal and the malignant cells (Figure 3-11C). This analysis showed that most of the leukemias shared transcriptional programs with Hardy fraction B cells. Interestingly, as can clearly be seen in Figure 3-11C, leukemias developed in *Pax5*$^{+/-}$ mice that, as demonstrated by ArrayCGH had lost their other *Pax5* allele, clustered close to the Hardy fraction B cells from *Pax5*$^{-/-}$ mice. In addition, several of the leukemias that developed in *Pax5*$^{+/+}$ mice shared transcriptional programs with Hardy fraction C cells suggesting that loss of one Pax5 allele is enough to impair leukemia development past Hardy fraction B.

To further explore the similarities between normal B-cells and the leukemias, the top 100 uniquely expressed genes in Hardy fractions A-F were extracted and investigated in the murine leukemias (Figure 3-11D). This analysis showed, as above, that the leukemias share the expression of some genes with Hardy fractions B and C. Collectively; the above analyses indicate that the murine leukemias maintain the transcription of certain genes that are important during the normal differentiation, likely reflecting the cell of origin of the leukemia. In addition, leukemias from *Pax5*$^{+/-}$ mice that lost their second *Pax5* allele during leukemia progression appear to reactivate genes that normally are expressed in early B-cells, in agreement with their immature immunophenotype. Importantly, although the leukemias maintain the activation of some genes that are expressed during normal differentiation, they exhibit profoundly different gene expression patterns as compared to their normal counterparts. Interestingly, N-Myc is the top differentially expressed gene between the leukemias and normal cells. Importantly, 10 wild-type and 10 Pax5 heterozygous leukemias were analyzed and half of the Pax5 hets were selected because of their immature immunophenotype. Thus, it is important to consider that the *Pax5* het group is not representative of the majority of *Pax5* heterozygous leukemias.

3.3.5 Gene Set Enrichment Analysis Reveals Murine Ph+ Leukemias Harbor Stem Cell-like Gene Expression Profile and Significantly Resemble Human Ph+ ALL

Our gene expression data demonstrate that as *BCR-ABL1* leukemias lose a *Pax5* allele, the leukemia becomes more immature and consequently harbors a gene expression profile most similar to Hardy fraction B. In addition, the majority of B-progenitor ALLs arrested at a stage of development very similar to Hardy fraction B. Taken together these data suggest that further characterization of Hardy fraction B cells may yield valuable insight into the mechanism by which *Pax5* haploinsufficiency cooperates during *BCR-ABL1* induced leukemogenesis.

In an effort to better understand the gene expression program in *BCR-ABL1* leukemia cells we performed Gene Set Enrichment Analysis (GSEA). GSEA is a statistical algorithm that is used to interrogate a defined gene expression for the common expression of a set of *a priori* identified genes. We first generated a rank ordered gene list by comparing the gene expression profiles of our murine *BCR-ABL1* leukemias versus all normal B-cell fractions. Next, we walked along our rank ordered gene list with a series of currated genesets from the Molecular Signatures Database. Our GSEA revealed that

the core ESC-like geneset[82] was significantly enriched in our murine *BCR-ABL1* leukemias (FDR=0.0869; Figure 3-12A). The core ESC-like gene set is a set of genes that are commonly expressed in both murine and human embryonic stem cells. This geneset is comprised of 335 genes many of which are regulated by C-Myc. Interestingly, the gene with the highest expression in our leukemias was N-Myc. Taken together these data suggest that our *BCR-ABL1* leukemias harbor a gene expression signature that is similar embryonic stem cells. Furthermore, our leukemias reveal high expression of N-Myc which may explain the similarities in expression. Next we performed GSEA on normal B-cells by comparing Hardy fraction A to all others and B to all others, for all Hardy fractions. We then assessed the enrichment of the core ESC-like geneset in any of the Hardy fractions. This GSEA revealed that the core ESC-like geneset was only significantly enriched in Hardy fraction B. Taken together, these data suggest that the stem cell signature identified in the leukemias results from an inherent signature retained by Hardy fraction B cells and further suggests that Pax5 loss may function to arrest B-cell development in such a way that the cell becomes more stem cell like.

We next used GSEA to explore similarities between human and mouse *BCR-ABL1* ALLs. We generated a rank ordered gene list by comparing the gene expression profiles of human *BCR-ABL1* ALL to all other B-progenitor ALL subtypes. Next, we generated a murine *BCR-ABL1* geneset by taking the top 100 genes that are expressed highly in our murine leukemias. Subsequently, we found that the top 100 genes from our murine *BCR-ABL1* leukemias were significantly enriched in human *BCR-ABL1* ALLs suggesting that our murine model of *BCR-ABL1* ALL is accurately recapitulating the human malignancy (FDR=0.161; Figure 3-12B).

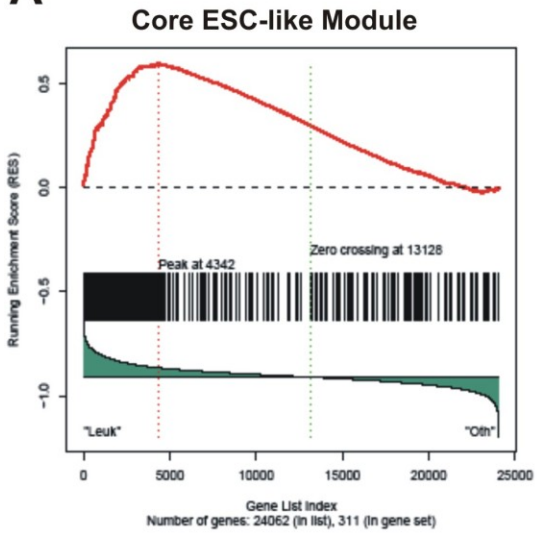
3.4 DISCUSSION

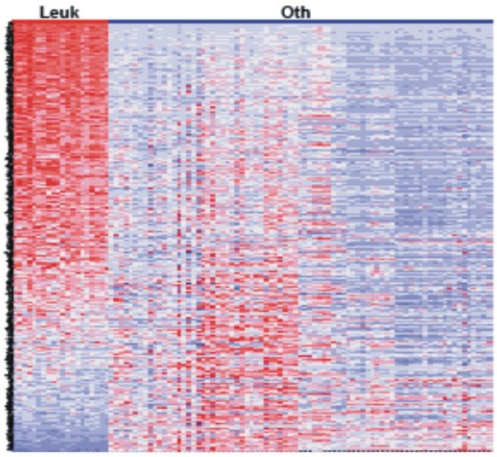
In summary, we have shown that *Pax5* haploinsufficiency cooperates with *BCR-ABL1* during leukemogenesis. We observed that *Pax5*$^{+/-}$ *BCR-ABL1* leukemias exhibit a more immature immunophenotype when compared to wild-type leukemias, including 19% which harbored a Pre-pro B-cell immunophenotype. Furthermore, we observed that our murine *BCR-ABL1* leukemias acquire many of the same cooperating oncogenic lesions that are seen in human Ph+ ALLs, including recurrent copy number alterations of *Pax5*, *Ink4a/Arf*, and *Pax5* sequence mutations. We have also shown that *Arf* haploinsufficiency further cooperates with *Pax5* haploinsufficiency during *BCR-ABL1* induced leukemogenesis suggesting that these two pathways may further cooperate during leukemogenesis. Importantly, in leukemias that started as *Arf*$^{+/-}$, we were able to show that these leukemias continue to express Arf protein suggesting that the wild-type *Arf* allele is intact and *Arf* is functioning as a haploinsufficient tumor suppressor. However, there is the possibility that *Pax5* loss may somehow impact the *Arf*/*Mdm2*/*p53* tumor suppressor pathway and may act as a second hit in this pathway, thereby eliminating the need to inactivate the wild-type *Arf* allele. Future studies assessing the interplay between *Pax5* status and *Arf* expression will need to be performed.

Figure 3-12 Gene set enrichment analysis

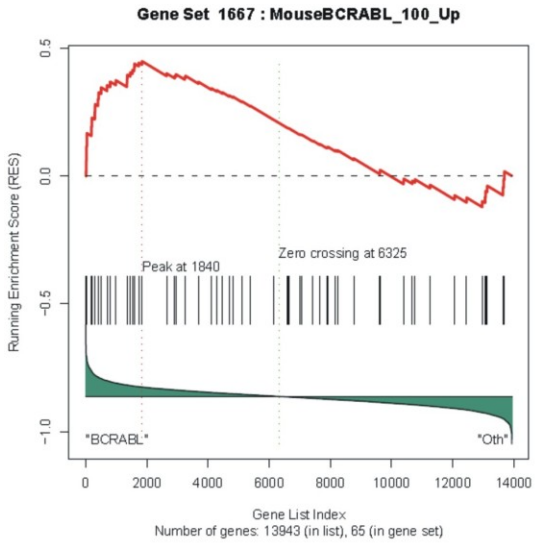
A, Gene Set Enrichment Analysis (GSEA) was used to evaluate enrichment of a host of genesets in murine *BCR-ABL1* leukemias. A rank ordered gene list was generated by comparing murine *BCR-ABL1* ALLs with normal Hardy fraction B cells (A-F). GSEA was then performed and revealed that the geneset "Core-ESC-like module" was significantly enriched in the leukemias (FDR=0.0869), suggesting that the murine *BCR-ABL1* leukemias share similar expression of core embryonic stem cell genes. **B**, Gene Set Enrichment Analysis (GSEA) reveals that the top 100 overexpressed genes from *Pax5*$^{+/+}$ leukemias enrich in *BCR-ABL1* signature of human B-ALL. A rank ordered gene list was generated by comparing *BCR-ABL1* ALLs with other B-lymphoid ALLs to define the *BCR-ABL1* signature. GSEA was then performed and the top 100 overexpressed genes from the *Pax5*$^{+/+}$ leukemias were found to be enriched in the human *BCR-ABL1* signature (FDR=0.161).

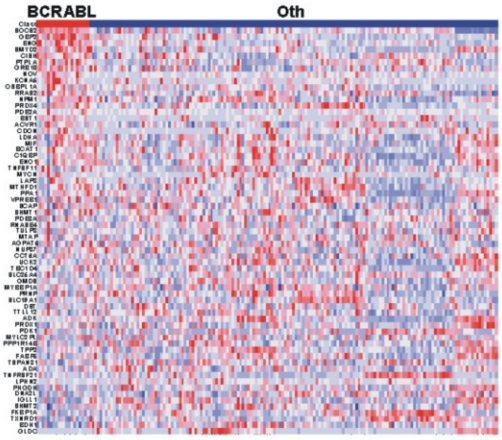
A


Core ESC-like Module
Running Enrichment Score (RES)
0.5
0.0
−0.5
−1.0
Peak at 4342
Zero crossing at 13128
"Leuk"
"Oth"
0
5000
10000
15000
20000
25000
Gene List Index
Number of genes: 24062 (in list), 311 (in gene set)


Leuk
Oth

B


Gene Set 1667 : MouseBCRABL_100_Up
Running Enrichment Score (RES)
0.5
0.0
−0.5
−1.0
Peak at 1840
Zero crossing at 6325
"BCRABL"
"Oth"
0
2000
4000
6000
8000
10000
12000
14000
Gene List Index
Number of genes: 13943 (in list), 65 (in gene set)


BCRABL
Oth

In an effort to gain mechanistic insight into the contribution that Pax5 loss may play during leukemogenesis we performed gene expression analysis and compared the expression profiles of Ph+ leukemias to their normal counterparts. Preliminary supervised gene expression analysis revealed that normal Hardy fraction stages of B-cell development harbor unique gene expression profiles. Furthermore, our leukemias seemed to most closely resemble Hardy fraction B and C. A more sophisticated principal component analysis revealed that each stage of B-cell development has an identifiable, unique transcriptional program that differs from all other stages of B-cell development. We also observed that in comparison to normal B-cells, B-progenitor leukemias harbor a significantly different gene expression signature. However, when we perform principal component analysis on normal B-cells and "lock" the analysis, and then add the leukemias onto this analysis we see that wild-type *BCR-ABL1* leukemias are most similar to Hardy fractions B and C. Interestingly, when one *Pax5* allele is lost, all the leukemias now are most similar to Hardy fraction B suggesting that *Pax5* haploinsufficiency contributes to a block in B-cell development. Furthermore, gene set enrichment analysis revealed that a geneset comprised of 335 genes that are commonly expressed in murine and human embryonic stem cells is enriched in our murine *BCR-ABL1* leukemias suggesting that our leukemias harbor a stem cell like gene expression profile. Interestingly, when we compare each Hardy fraction versus all other normal B-cells we see that the core embryonic stem cell-like geneset is only enriched in Hardy fraction B suggesting that Hardy fraction B cells harbor a latent stem cell like gene expression signature. Interestingly, *Pax5* is first expressed in Hardy fraction B and is absolutely necessary for B-cell commitment and development.[8,18] In the absence of Pax5, B-cells are arrested at Hardy stage B and in the context of appropriate growth factors can dedifferentiate into many other hematopoietic lineages.[18] Taken together, our data suggests that Hardy fraction B cells have a latent embryonic stem cell-like gene expression signature and that loss of Pax5 during leukemogenesis provides a block in B-cell development at Hardy fraction B providing the developing leukemia with a more stem cell-like environment. It is very likely that Hardy fraction B cells are the last progenitor stage in the B-cell lineage and wild-type *Pax5* expression is absolutely critical to commit B-cells to differentiate. *Pax5* is in this context a haploinsufficient tumor suppressor. *Pax5* haploinsufficiency not only cooperates with *BCR-ABL1* but also with other oncogenic lesions, like ENU, suggesting that its role as a haploinsufficient tumor suppressor is not limited to collaborating with the very strong *BCR-ABL1* oncogene.

Taken together, these data suggest that the lesions that have been identified by SNP analysis in human ALLs are cooperating oncogenic lesions and further examination of the pathways involved may yield important insights into future therapeutic avenues worth exploration in the global effort to cure ALL.

CHAPTER 4
BCR-ABL1 LYMPHOBLASTIC LEUKEMIA IS CHARACTERIZED BY THE DELETION OF IKAROS*

4.1 INTRODUCTION

Acute lymphoblastic leukemia (ALL) comprises a heterogeneous group of disorders characterized by recurring chromosomal abnormalities including translocations, trisomies and deletions. An ALL subtype with especially poor prognosis is characterized by the presence of the Philadelphia chromosome arising from the t(9;22)(q34;q11.2) translocation, which encodes the constitutively activated BCR-ABL1 tyrosine kinase. *BCR-ABL1* positive ALL constitutes 5% of paediatric B-progenitor ALL and approximately 40% of adult ALL.[21,22] Expression of BCR-ABL1 is also the pathologic lesion underlying chronic myelogenous leukemia (CML).[20] Data from murine studies demonstrates that expression of BCR-ABL1 in hematopoietic stem cells can alone induce a CML-like myeloproliferative disease, but cooperating oncogenic lesions are required for the generation of a blastic leukemia.[75,83] Although the p210 and p190 *BCR-ABL1* fusions are most commonly found in CML and paediatric *BCR-ABL1* ALL respectively, either fusion may be found in adult *BCR-ABL1* ALL.[84] Importantly, a number of genetic lesions including additional cytogenetic aberrations and mutations in tumour suppressor genes have been described in CML cases progressing to blast crisis.[85] However, the specific lesions responsible for the generation of *BCR-ABL1*ALL and blastic transformation of CML remain incompletely understood.[85] To identify cooperating oncogenic lesions in ALL, we recently performed a genome-wide analysis of paediatric ALL.[76] This analysis identified an average of 6.8 genomic copy number alterations (CNA) in 9 *BCR-ABL1* ALL cases, including deletions in genes that play a regulatory role in normal B cell development.

4.2 MATERIALS AND METHODS

4.2.1 Patients and Samples

Two hundred eighty two patients with acute lymphoblastic leukemia (ALL) treated at St. Jude Children's Research Hospital, 22 adult *BCR-ABL1* ALL patients treated at the University of Chicago, and 49 samples obtained from 23 adult patients with chronic myeloid leukaemia (CML) treated at the Institute of Medical and Veterinary Science, Adelaide. The CML cohort included 24 chronic phase, 7 accelerated phase and 15 blast crisis samples, and three samples obtained at complete cytogenetic response. All blast crisis samples were flow sorted to at least 90% blast purity prior to DNA extraction

---

* Adapted with permission. Mullighan C, Miller C, Radtke I, Phillips L, Dalton J, Ma J, White D, Hughes T, Le Beau M, Pui C, Relling M, Shurtleff S, Downing J. *BCR-ABL1* lymphoblastic leukaemia is characterized by the deletion of *Ikaros*. *Nature* **453**, 110-114 (2008).[19]

using FACSVantage SE (with DiVa option) flow cytometers (BD Biosciences, San Jose, CA) and fluorescein isothiocyanate labeled CD45, allophycocyanin labeled CD33 and phycoerythrin labeled CD19 and CD13 antibodies (BD Biosciences). Germline tissue was obtained by also sorting the non-blast population in 7 cases. Informed consent for the use of leukemic cells for research was obtained from patients, parents or guardians in accordance with the Declaration of Helsinki, and study approval was obtained from the SJCRH institutional review board.

4.2.2 Cell Lines Examined by SNP Array

Thirty-six acute myeloid and lymphoid leukemia cell lines were genotyped using the Affymetrix Mapping 250k Sty and Nsp arrays. These were the ALL cell lines 380 (*MYC-IGH* and *BCL2-IGH* B-precursor), 697 (*TCF3-PBX1*), AT1 (*ETV6-RUNX1*), BV173 (CML in lymphoid blast crisis), CCRF-CEM (*TAL-SIL*), Jurkat (T-ALL), Kasumi-2 (*TCF3-PBX1*), MHH-CALL-2 (hyperdiploid B-precursor ALL), MHH-CALL-3 (*TCF3-PBX1*), MOLT3 (T-ALL), MOLT4 (T-ALL), NALM-6 (B-precursor ALL), OP1 (*BCR-ABL1*), Reh (*ETV6-RUNX1*), RS4;11 (*MLL-AF4*), SD1 (*BCR-ABL1*), SUP-B15 (*BCR-ABL1*), TOM-1 (*BCR-ABL1*), U-937 (*PICALM-AF10*), UOCB1 (*TCF3-HLF*), YT (NK leukaemia); and the AML cell lines CMK (FAB M7), HL-60 (FAB M2), K-562 (CML in myeloid blast crisis), Kasumi-1 (*RUNX1-RUNX1T1*), KG-1 (myelocytic leukaemia), ME-1 (*CBFB-MYH11*), ML-2 (*MLL-AF6*), M-07e (FAB M7), Mono Mac 6 (*MLL-AF9*), MV4-11 (*MLL-AF4*), NB4 (*PML-RARA*), NOMO-1 (*MLL-AF9*), PL21 (FAB M3), SKNO-1 (*RUNX1-RUNX1T1*) and THP-1 (FAB M5). Cell lines were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany; the American Type Culture Collection, Manassas, VA, from local institutional repositories, or were gifts from Olaf Heidenreich (SKNO-1) and Dario Campana (OP1). Cells were culture in accordance with previously published recommendations.[86] The paediatric *BCR-ABL1* B-precursor ALL cell line OP1[87] was cultured in RPMI-1640 containing 100 units/ml penicillin, 100 μg/ml streptomycin, 2 mM glutamine and 10% fetal bovine serum. DNA was extracted from $5x10^6$ cells obtained during log phase growth after washing in PBS using the QIamp DNA blood mini kit (Qiagen, Valencia, CA).

4.2.3 Single Nucleotide Polymorphism Microarray Analysis

Collection and processing of diagnostic and remission bone marrow and peripheral blood samples for Affymetrix single nucleotide polymorphism microarray analysis has been previously reported in detail.[76] Affymetrix 250K Sty and Nsp arrays were performed on all samples. 50k Hind 240 and 50k Xba 240 arrays were performed for 252 ALL samples. SNP array CEL and SNP call TXT files (generated by Affymetrix GTYPE 4.0 using the DM algorithm) have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession numbers GSE9109-9113.

4.2.4 Fluorescence *In Situ* Hybridization

Fluorescence *in situ* hybridization for *IKZF1* deletion was performed using diagnostic bone marrow or peripheral blood leukaemic cells in Carnoy's fixative as previously described.[76] BAC clones CTD-2382L6 and CTC-791O3 (for *IKZF1*, Open Biosystems, Huntsville, AL) were labeled with fluorescein isothiocyanate, and control 7q31 probes RP11-460K21 (Children's Hospital Oakland Research Institute, Oakland, CA) and CTB-133K23 (Open Biosystems), were labeled with rhodamine. At least 100 interphase nuclei were scored per case.

4.2.5 *IKZF1* PCR, Cloning, Quantitative PCR and Genomic Sequencing

RNA was extracted and reverse transcribed using random hexamer primers and Superscript III (Invitrogen, Carlsbad, CA) as previously described.[76] *IKZF1* transcripts were amplified from cDNA using the Advantage 2 PCR enzyme (Clontech, Mountain View) as previously described[76] using primers that anneal in exon 0 and 7 of *IKZF1*. PCR products were purified, and sequenced directly and after cloning into pGEM-T-Easy (Promega, Madison, WI). Genomic quantitative PCR for exons 1-7 of *IKZF1*, and real-time PCR to quantify expression of Ik6 were performed as previously described.[76] All primers and probes are listed in Supplementary Table 3. Genomic sequencing of *IKZF1* exons 0-7 in all ALL and CML samples was performed as previously described.[76]

4.2.6 Western Blotting

Whole cell lysates of 3-6x10^6 leukemic cells were prepared and blotted as previously described[76] using N- and C-terminus specific rabbit polyclonal Ikaros antibodies (Santa Cruz Biotechnology, Santa Cruz, CA).

4.2.7 Methylation Analysis

Methylation status of the *IKZF1* promoter CpG island (chr7:50121508-50121714) was performed using MALDI-TOF mass spectrometry of PCR-amplified, bisulfite modified genomic DNA extracted from leukemic cells as previously described.[49,76]

4.2.8 Statistical Analysis

Associations between ALL subtype and *IKZF1* deletion frequency were calculated using the exact likelihood ratio test. Differences in Ik6 expression between *IKZF1* Δ3-6 and non-Δ3-6 cases was assessed using the exact Wilcoxon-Mann-Whitney test. All *P* values reported are two-sided. Analyses were performed using StatXact v8.0.0 (Cytel, Cambridge, MA).

4.3 RESULTS

4.3.1 *Ikaros* is Commonly Deleted in *BCR-ABL1* ALL

To extend this analysis and identify lesions that distinguish CML from *BCR-ABL1* ALL, we have now examined DNA from leukemic samples from 304 pediatric and adult ALLs (254 B-progenitor, 50 T-lineage), including 21 pediatric and 22 adult *BCR-ABL1* ALL, and 23 adult CML cases. Samples were analyzed using the 250k Sty and Nsp Affymetrix SNP arrays (and also the 100K arrays for most cases). This identified a mean of 8.79 somatic CNA per *BCR-ABL1* ALL case (range 1-26), with 1.44 gains (range 0-13) and 7.33 losses (range 0-25). No significant differences were noted in the frequency of CNAs between paediatric and adult *BCR-ABL1* ALL cases. The most frequent somatic CNA was deletion of *IKZF1*, which encodes the transcription factor Ikaros (Table 4-1). *IKZF1* was deleted in 36 (83.7%) of 43 *BCR-ABL1* ALL cases, including 76.2% of paediatric and 90.9% of adult *BCR-ABL1* ALL cases. *CDKN2A* was deleted in 53.5% of *BCR-ABL1* ALL cases, most of which (87.5%) also had deletions of *IKZF1* (Table 4-1). Conversely, of the *BCR-ABL1* ALL cases with *IKZF1* deletions, 41.6% lacked CDKN2A alterations. Deletion of *PAX5* occurred in 51% of *BCR-ABL* ALL cases, again with the majority also having a deletion of *IKZF1* (95%) (Table 4-1). No other defining CNAs were identified in the rare *BCR-ABL1* ALL cases that lacked a deletion of *IKZF1*.

Ikaros is a member of a family of zinc finger nuclear proteins that is required for normal lymphoid development.[88-91] Ikaros has a central DNA-binding domain consisting of four zinc fingers, and a homo- and heterodimerization domain consisting of the two C-terminal zinc fingers[92] (Figure 4-1A). Alternative splicing generates multiple Ikaros isoforms, several of which lack the N-terminal zinc fingers required for DNA binding; however, the physiological relevance of these isoforms in normal hematopoiesis remains unclear.[88-90,93] The *IKZF1* deletions identified in *BCR-ABL1* ALL were predominantly mono-allelic and were limited to the gene in 25 cases, conclusively identifying *IKZF1* as the genetic target (Figure 4-1B). In 19 cases the deletions were confined to a subset of internal *IKZF1* exons, most commonly exons 3-6 (Δ3-6; N=15). Importantly, the Δ3-6 deletion is predicted to encode an Ikaros isoform that lacks the DNA-binding domain but retains the C-terminal zinc fingers. The *IKZF1* deletions were confirmed by FISH and genomic quantitative PCR, and were in the predominant leukemic clone. Detailed analysis failed to reveal any evidence of either *IKZF1* point mutations or inactivation of its promoter by CpG methylation in primary ALL samples.

4.3.2 *Ikaros* Deletions Result in the Expression of a Dominant Negative Isoform Ik6

The expression of aberrant, dominant negative Ikaros isoforms in B- and T-lineage ALL has been previously reported by several groups,[94-101] although alternative splicing has been reported to be the underlying mechanism.[102] Importantly, the Δ3-6 isoform of Ikaros has been shown to function as a dominant negative inhibitor of the

Table 4-1 Frequency of recurring DNA copy number abnormalities in ALL

ALL subtype (N)	*IKZF1*	*CDKN2A*	*PAX5*	*C20orf94*	*RB1*	*MEF2C*	*EBF1*	*BTG1*	*DLEU*	*FHIT*	*ETV6*
B-progenitor (254)											
BCR-ABL1 (43)	36	23	22	10	8	6	6	6	4	4	3
- childhood (21)	16	13	12	3	4	4	3	2	3	2	1
- adult (22)	20	10	10	7	4	2	3	4	1	2	2
Hypodiploid (10)	5	10	10	0	0	0	1	1	0	1	2
Other B-ALL (75)	15	25	22	4	1	0	2	5	1	3	10
High hyperdiploid (39)	2	8	4	1	3	0	0	0	5	0	3
MLL-rearranged (22)	1	4	4	0	2	0	0	0	3	0	2
TCF3-PBX1 (17)	0	6	7	0	2	0	0	0	2	0	0
ETV6-RUNX1 (48)	0	14	16	6	2	0	5	7	4	6	33
T-lineage (50)	2	36	5	1	6	1	3	0	3	0	4
Total (304)	61	126	90	22	24	7	17	19	22	14	57
P	$6.6x10^{-27}$	$7.4x10^{-10}$	$1.4x10^{-9}$	$7.0x10^{-8}$	$1.1x10^{-6}$	0.0004	0.0247	$1.5x10^{-7}$	$2.6x10^{-6}$	0.0076	$9.1x10^{-15}$

The prevalence of recurring genomic abnormalities in *BCR-ABL1* B-progenitor ALL identified by SNP array analysis is shown for each ALL subtype. The exact likelihood ratio *P* value for variation in the frequency of each lesion across ALL subtypes is shown. The *DLEU* region at 13q14 incorporates the miRNA genes MIRN16-1 and MIRN15A.

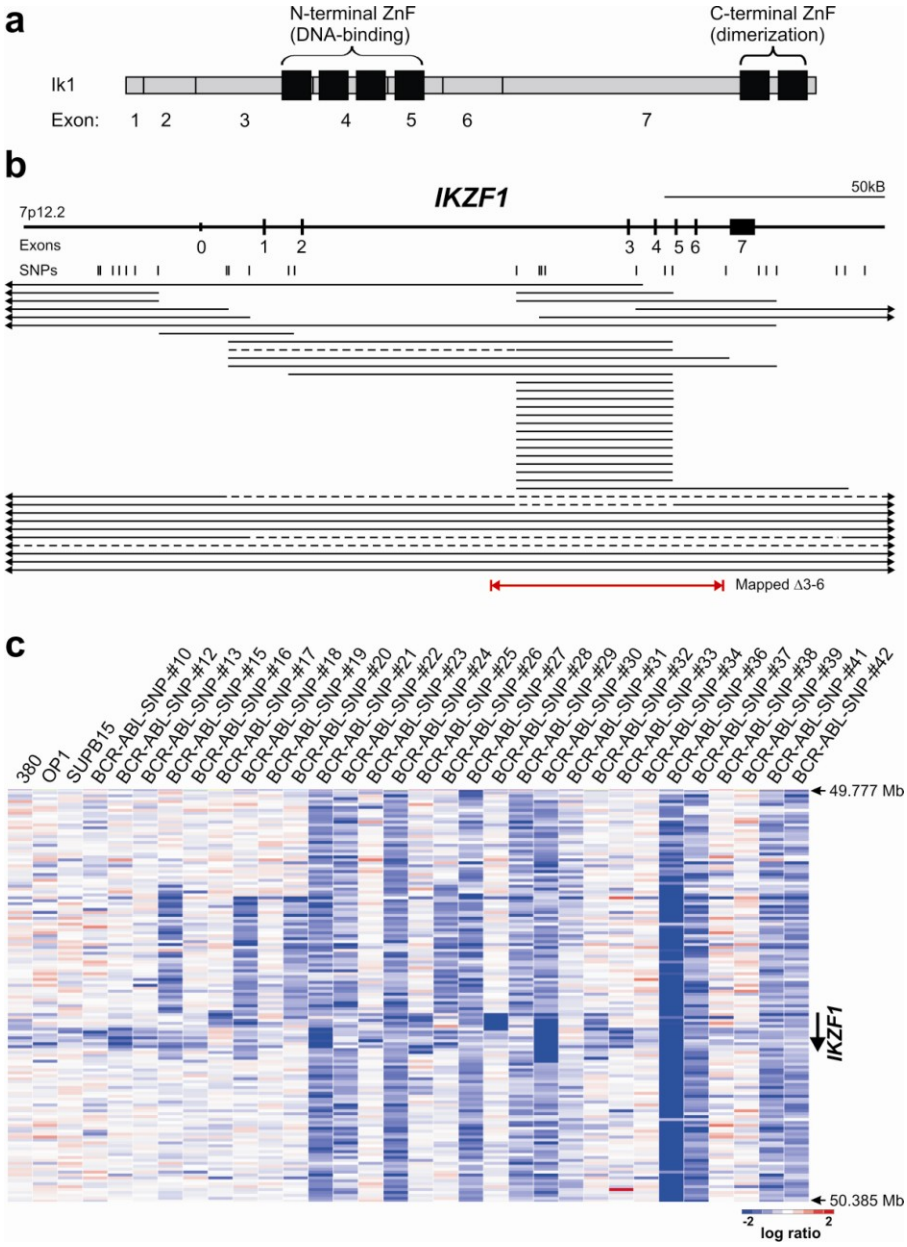


Figure 4-1 *IKZF1* deletions in BCR-*ABL1* ALL

A, Domain structure of IKZF1. Exons 3-5 encode four N-terminal zinc fingers (black boxes) responsible for DNA binding. The C-terminal zinc fingers encoded by exon 7 are essential for homo- and heterodimerization. **B**, Genomic organization of *IKZF1* and location of each of the 36 deletions observed in *BCR-ABL1* B-progenitor ALL. Each line depicts the deletion(s) observed in each case. In five cases, two discontiguous deletions were observed. Hemizygous deletions are solid lines and homozygous deletions dashed. Arrows indicated deletions extending beyond the limits of the figure. The exact boundaries of the deletions were defined by genomic qPCR, and for *IKZF1* Δ3-6, by long-range genomic PCR (red arrow). **C**, dChipSNP raw $\log_2$ ratio copy number data depicting *IKZF1* deletions for 29 *BCR-ABL1* cases and 3 B-progenitor ALL cell lines.

transcriptional activity of Ikaros and related family members.[92] Moreover, mice homozygous for either an *Ikzf1* null mutation[103] or a dominant negative *Ikzf1* mutation[67] exhibit profound defects in lymphoid development, and mice heterozygous for a dominant negative *Ikzf1* mutation develop clonal T cell expansions and lymphoproliferative diseases,[69] demonstrating that alteration in the level of *Ikzf1* expression is oncogenic.

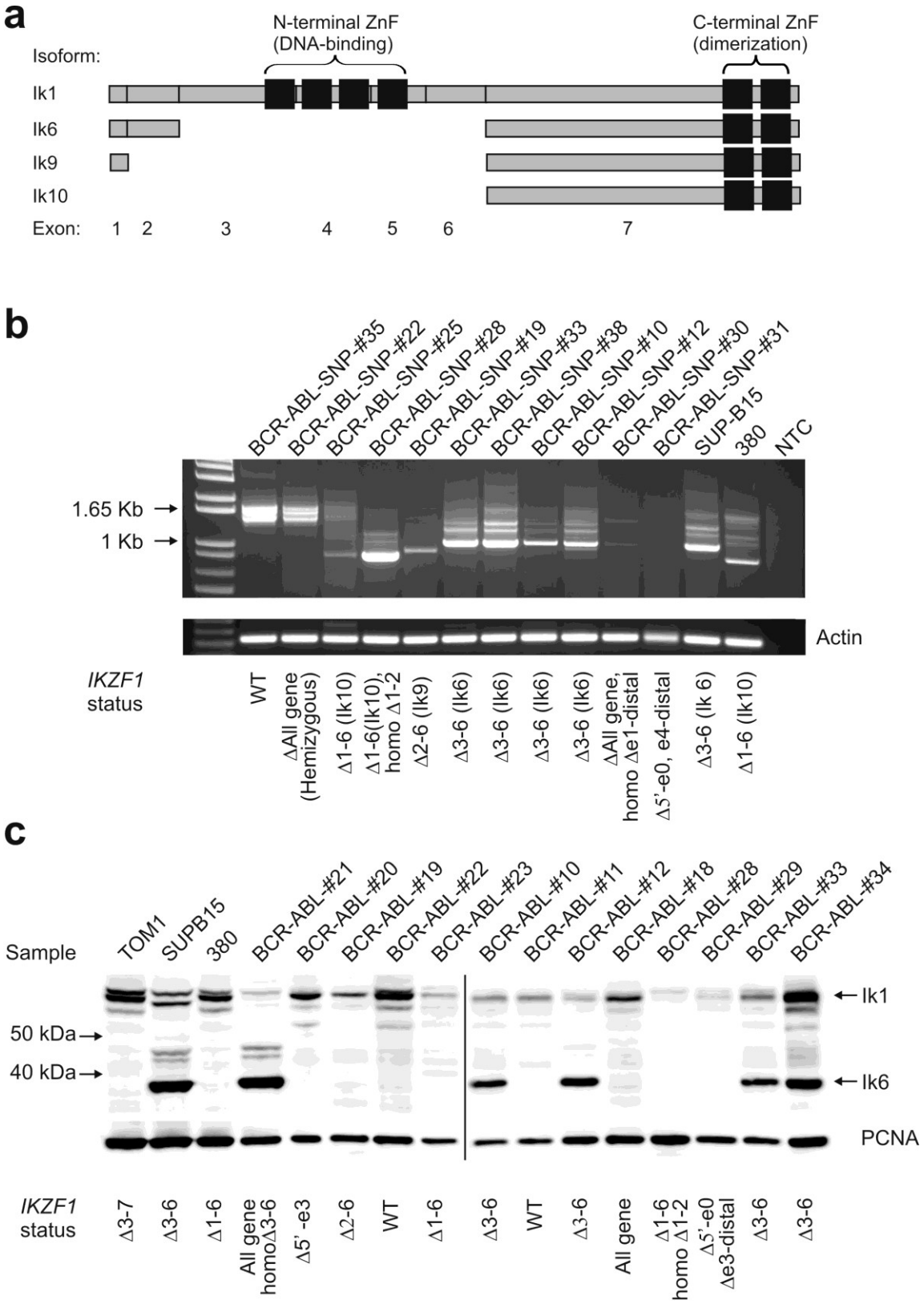
The high frequency of focal deletions in *IKZF1* in *BCR-ABL1* ALL suggests that expression of alternative *IKZF1* transcripts may be the result of specific genetic lesions, and not alternative splicing of an intact gene. To further explore this possibility, we performed RT-PCR analysis for *IKZF1* transcripts in 159 cases (Figure 4-2). This demonstrated that expression of the Ik6 transcript, which lacks exons 3-6, was exclusively observed in cases harboring the *IKZF1* Δ3-6 deletion (Figure 4-2B). Furthermore, we detected two novel Ikaros isoforms exclusively in cases with larger deletions; Ik9 in a case with deletion of exons 2-6, and Ik10 in three cases with deletion of exons 1-6 (Figure 4-2A-B). For each isoform, Ik6, 9 and 10, there was concordance between the transcripts detected by RT-PCR and the extent of deletion defined by SNP array and genomic PCR analysis (Figure 4-2B). Moreover, analysis of 22 *IKZF1* Δ3-6 and 29 non-Δ3-6 cases with a quantitative PCR assay specific for the Ik6 transcript confirmed that Ik6 expression was restricted to cases with the Δ3-6 deletion (P=6.41x10^{-15}). In addition, the Ik6 protein isoform was only detectable by western blotting in cases with a Δ3-6 *IKZF1* deletion (Figure 4-2C). We also did not observe expression of Ik6 following the enforced expression of *BCR-ABL1* in *Arf* null or wild type murine hematopoietic precursors. Together, these data indicate that the expression of non-DNA binding Ikaros isoforms is due to *IKZF1* genomic abnormalities, and not aberrant post-transcriptional splicing induced by BCR-ABL1, as has been suggested.[102]

4.3.3 Deletion of Ikaros is an Important Event in the Transformation of CML to Lymphoid Blast Crisis

To identify CNAs in CML, we performed SNP array analysis on 23 CML cases. In addition to chronic phase CML (CP-CML), we also examined matched accelerated phase (AP-CML, N=7) and blast crisis (BC-CML, N=15, 12 myeloid and 3 lymphoid) samples. This identified only 0.47 CNAs per CP-CML case (range 0-8), suggesting that *BCR-ABL1* is sufficient to induce CML, but alone does not result in substantial genomic instability. Importantly, no recurrent lesions were identified. In contrast, there was a mean of 7.8 CNAs per BC-CML case (range 0-28), with *IKZF1* deletions in four BC samples, including two of the three cases with lymphoid blast crisis (Figure 4-3A). Two of the *IKZF1* deletions involved the entire gene (CML-#4-BC and #22-BC), one Δ3-6 (CML-#1-BC, which was associated with Ik6 expression by RT-PCR) and one Δ3-7 (CML-#7-BC). CML-#7-BC also had an *IKZF1* nonsense mutation in the C-terminal zinc finger domain of exon 7 in the non-deleted allele (c.1520C>A, p.Ser507X, Figure 4-3B). One BC sample had a *CDKN2A* deletion, and four cases had CNAs involving *PAX5* (two deletions, one internal amplification, one trisomy 9), with two of these also having *IKZF1* deletion. CNAs were identified in two AP-CML samples.

Figure 4-2 Ikaros isoforms in ALL blasts

A, Domain structure of the *IKZF1* isoforms detected by RT-PCR, examples of which are shown in panel B. **B**, RT-PCR for *IKZF1* transcripts (using exon 0 and 7 specific primers) in representative cases with various *IKZF1* genomic abnormalities. Each case expressing an aberrant isoform had a corresponding *IKZF1* genomic deletion. *IKZF1* Δ3-6 was also detected in the *BCR-ABL1* ALL cell lines SUP-B15 and OP1, and Δ1-6 in the ALL cell line 380. **C**, Western blotting for Ikaros using a C-terminus specific polyclonal antibody. Ik6 was only detectable in cases with *IKZF1* Δ3-6. The Δ1-6 and Δ2-6 deletions do not produce a detectable protein. In three cases with multiple focal hemizygous deletions involving different regions of *IKZF1* (BCR-ABL-SNP-#26, -#29, and -#31), no wild type Ikaros was detectable by RT-PCR or western blotting, indicating that the deletions involve both copies of *IKZF1* in each case.


a
N-terminal ZnF
(DNA-binding)
C-terminal ZnF
(dimerization)
Isoform:
Ik1
Ik6
Ik9
Ik10
Exon: 1 2 3 4 5 6 7
b
BCR-ABL-SNP-#35
BCR-ABL-SNP-#22
BCR-ABL-SNP-#25
BCR-ABL-SNP-#28
BCR-ABL-SNP-#19
BCR-ABL-SNP-#33
BCR-ABL-SNP-#38
BCR-ABL-SNP-#10
BCR-ABL-SNP-#12
BCR-ABL-SNP-#30
BCR-ABL-SNP-#31
SUP-B15
380
NTC
1.65 Kb
1 Kb
Actin
IKZF1 status
WT
ΔAll gene (Hemizygous)
Δ1-6 (Ik10)
Δ1-6(Ik10), homo Δ1-2
Δ2-6 (Ik9)
Δ3-6 (Ik6)
Δ3-6 (Ik6)
Δ3-6 (Ik6)
Δ3-6 (Ik6)
ΔAll gene, homo Δe1-distal
Δ5'-e0, e4-distal
Δ3-6 (Ik 6)
Δ1-6 (Ik10)
c
Sample
TOM1
SUPB15
380
BCR-ABL-#21
BCR-ABL-#20
BCR-ABL-#19
BCR-ABL-#22
BCR-ABL-#23
BCR-ABL-#10
BCR-ABL-#11
BCR-ABL-#12
BCR-ABL-#18
BCR-ABL-#28
BCR-ABL-#29
BCR-ABL-#33
BCR-ABL-#34
50 kDa
40 kDa
Ik1
Ik6
PCNA
IKZF1 status
Δ3-7
Δ3-6
Δ1-6
All gene homoΔ3-6
Δ5' -e3
Δ2-6
WT
Δ1-6
Δ3-6
WT
Δ3-6
All gene
Δ1-6 homo Δ1-2
Δ5'-e0 Δe3-distal
Δ3-6
Δ3-6

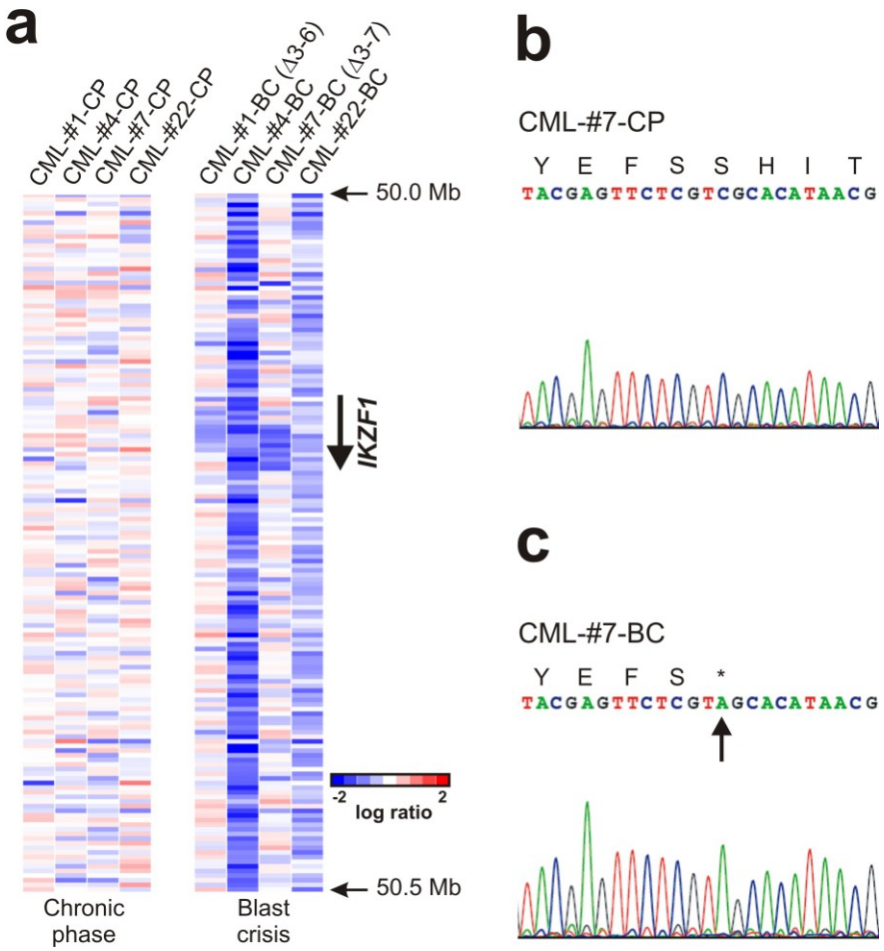


Figure 4-3 *IKZF1* deletions in blast crisis CML

A, dChipSNP log_2ratio copy number heatmaps of four CML cases showing acquisition of *IKZF1* deletions at progression to blast crisis. **B-C**, Pherograms of *IKZF1* exon 7 sequencing demonstrating acquisition of the c.1520C>A, p.Ser507X mutation at blast crisis in case CML-#7. As this case has a concomitant hemizygous *IKZF1* deletion involving exon 7, the mutation appears homozygous.

These data demonstrate an increased burden of genomic aberrations during progression of CML, with *IKZF1* mutation a frequent event in the transformation of CML to lymphoid blast crisis.

4.3.4 Ikaros Deletions are Likely a Result of Aberrant RAG Activity

To explore the mechanism responsible for the identified *IKZF1* deletion, we sequenced the *IKZF1* Δ3-6 genomic breakpoints. The deletions were restricted to highly localized sequences in introns 2 and 6. Moreover, heptamer recombination signal sequences (RSSs) recognized by the RAG enzymes during V(D)J recombination[104] were located immediately internal to the deletion breakpoints, and a variable number of additional nucleotides were present between the consensus intron 2 and 6 sequences, suggestive of the action of terminal deoxynucleotidyl transferase (TdT). Together, these data suggest that the *IKZF1* Δ3-6 deletion arises due to aberrant RAG-mediated recombination.

4.4 DISCUSSION

In summary, we have identified a high frequency of CNAs in *BCR-ABL1* ALL and BC-CML, but not in CP-CML. We observed a near obligate deletion of *IKZF1* in *BCR-ABL1* ALL, with 83.7% of paediatric and adult cases containing deletions that lead to a reduction in dose and/or the expression of an altered Ikaros isoform. By contrast, deletion of *IKZF1* was not detected in CP-CML, but was identified as an acquired lesion in 2 of 3 lymphoid BC-CML samples. These data, together with the low frequency of *IKZF1* deletions in other pediatric B-progenitor ALL cases, and the lack of focal *IKZF1* aberrations in recently reported genomic analysis of non-hematopoietic tumors[105] suggest that alterations in Ikaros directly contribute to the pathogenesis of *BCR-ABL1* ALL. How reduced activity of Ikaros, and possibly that of other family members through the expression of dominant negative Ikaros isoforms, collaborates with *BCR-ABL1* to induce lymphoblastic leukemia remains to be determined. Importantly, mice with attenuated Ikaros expression exhibit a partial block of B lymphoid maturation at the pro-B cell stage,[106] suggesting that Ikaros loss may contribute to the arrested B lymphoid maturation in *BCR-ABL1* ALL. However, the high co-occurrence of *PAX5* deletions in many cases suggests that *IKZF1* deletion contributes to transformation in additional ways. The frequent co-deletion of *CDKN2A* (encoding INK4A/ARF) with *IKZF1* in *BCR-ABL1* ALL is a notable finding. This suggests that attenuated Ikaros activity may either collaborate with disruption of INK4A/ARF-mediated tumor suppression, or act through alternative uncharacterized tumor suppressor pathways in ALL. Furthermore, the identification of aberrant RAG-mediated recombination as the mechanism underlying deletions of *IKZF1* suggests that the cellular target of this transforming event is downstream of the hematopoietic stem cell. Dissecting the contribution of altered Ikaros activity to *BCR-ABL1* leukaemogenesis should not only provide valuable mechanistic insights, but will also help to determine if the presence of this genetic lesion can be used to gain a therapeutic advantage against this aggressive leukemia.

CHAPTER 5
OVERALL SUMMARY AND FUTURE DIRECTIONS

Significant progress has been made in the treatment of ALL with cure rates approaching 90%. Despite the significant advances in the treatment of children with ALL, 10% of children do not survive and many children endure significant complications associated with current therapy. As a result, it is critical that the full complement of genetic mutations within ALL be identified so that therapy can be adjusted and improved in order to cure every child and to minimize therapy related complications as much as possible.

ALL is comprised of several distinct subtypes and in the majority of subtypes, the initiator or driver lesion is known. In order to identify additional cooperating oncogenic mutations we have performed genome-wide SNP analysis on nearly 250 ALL cases. The SNP analysis revealed that genes that are responsible for B-cell development and differentiation are mutated in 40% of B-progenitor ALLs. These mutations include deletions, amplifications, translocations, and point mutations. The most common target of mutation was the *PAX5* gene which was found to be mutated in 32% of cases. As discussed in chapter 2, there were several *PAX5* mutations identified, the majority of which resulted in either hemizygous loss of the gene or in the production of a hypomorphic allele suggesting that *PAX5* haploinsufficiency may be a cooperating lesion during leukemogenesis. This result raises several questions for future study that have not been addressed in this dissertation. We can clearly appreciate that *PAX5* mutations are common in ALL but what are the effects that *PAX5* mutations have on the normal function of the PAX5 protein? Chapter 3 discusses work that has been done elucidating the finding that *Pax5* haploinsufficiency cooperates with *BCR-ABL1* during leukemogenesis. However, there are additional mechanisms of *PAX5* mutation in addition to hemizygous loss, including point mutation, internal truncation, and translocation. It will be important to understand the effects that these *PAX5* mutations have on normal PAX5 function.

In order to determine the effects that these mutations have on normal PAX5 function we will need to introduce the various Pax5 mutants into *Pax5*$^{-/-}$ B-cells. The transduction experiments can be performed *in vitro* and the cells can then be evaluated for evidence of differentiation by flow cytometry, Southern or PCR analysis for immunoglobulin gene rearrangements, and bone marrow transplant assays for differentiation and tumorigenic potential. The next logical progression of experiments would be to generate knock-in animals expressing one or more of the *Pax5* mutations. By knocking in a Pax5 point mutation into the *Pax5* locus we can ask several important questions. First, is the mutation conducive to life, that is, does the presence of the mutation cause the death of the animal? Second, what is the spectrum of B-cell development within the bone marrow compartment of the animal? Are there any off-site effects of the *Pax5* mutation; does the mutation affect other tissues? A knock in animal model will also be very important for future work because expression of the *Pax5* mutation is controlled by the *Pax5* locus and not a viral LTR which results in abnormally

high expression of the associated gene. Thus, is *Pax5* mutant expression at normal levels that is sensitive to normal cellular signaling cues, sufficient for normal B-cell development or is it insufficient or perhaps even oncogenic? This series of experiments is critical because it is very possible that expression of hypomorphic allele at high levels is enough to confer normal Pax5 effects, whereas, expression of the same hypomorphic allele at normal physiologic levels may result in a block in B-cell development.

The generation of knock in animals that harbor Ikaros mutations will also be an important series of experiments to perform. We have shown in chapter 4 that deletions of the *IKZF1* (Ikaros) gene are almost a required event in Ph+ leukemia. Not only do most *BCR-ABL1* leukemias harbor deletions of the Ikaros gene, but we have shown in a recent study that deletion of Ikaros is a negative prognostic indicator in ALL.[107] Because many of the Ikaros deletions result in internally truncated mutants that are expected to behave as dominant negative inhibitors of normal Ikaros function, it will be important to generate an Ikaros mutant knock in animal model. Again, by knocking the mutant Ikaros allele into the normal germline Ikaros locus we will be able to better appreciate the changes associated with the mutant allele. We will ask questions like what are the consequences of Ikaros mutant; how does it affect normal hematopoiesis? If the Ikaros mutants are in fact dominant negative inhibitors of normal Ikaros function then we will be able to asses those effects by performing DNA binding assays, transactivation assays, and Western blot analyses to determine the effects that the mutant has alone and the effects that it has on normal Ikaros function. Additional studies regarding the leukemogenic effects of the Ikaros mutant knock-in, as well as, cooperation with *BCR-ABL1* or other oncogenes will be of great interest.

Importantly, many pediatric ALL cases harbor mutations of both *PAX5* and Ikaros suggesting that multiple lesions within the B-cell development and differentiation pathway may cooperate during leukemogenesis. Once the generation of both the *PAX5* and Ikaros knock-in animals is completed then these models can then be crossed and analyzed for effects on the hematopoietic system. It will be important to assess the ability of these mutations to cooperate to impair B-cell development both in the presence of an oncogene like *BCR-ABL1* and in the absence of any oncogene. The results of these experiments will help in our understanding of exactly how the loss of *PAX5* or Ikaros really contributes to the development of ALL.

In chapter 3 we discussed some of the insights we have gained into how the loss *Pax5* may contribute to the development of ALL. We discussed the previous findings that during normal B-cell development, as defined by Hardy fraction analysis, *Pax5* expression in turned on in fraction B and remains expressed in mature circulating B-lymphocytes. If one allele of *Pax5* is lost then the gene expression profiles of all Hardy fractions from B distally change. Even more impressive is the loss of both *Pax5* alleles which completely arrest B-cell development at Hardy fraction B. We have expanded the previously published data on the transcriptional networks that function during B-cell development. Our gene expression profiling of B-cell development has revealed a significant increase in the number and scope of genes that are regulated by Pax5 as defined by changes in the gene expression profiles of *Pax5* haploinsufficient and null B-

cells. Another important future experiment will be to elucidate which Pax5 responsive genes are primary targets by performing ChIP-sequencing analysis. We have also shown using principal component analysis and gene set enrichment analysis that Hardy fraction B is enriched for a set of genes commonly expressed in embryonic stem cells. It is no surprise then, that in pre-B leukemias, the stage of developmental arrest is Hardy fraction B, where there is expression of a set of stem cell genes. The enrichment of this core ESC-like geneset was also seen in our murine leukemias further supporting the link between Hardy fraction B arrest and leukemogenic potential. I would speculate that the loss of *Pax5* in the context of *BCR-ABL1* expression confers a block in B-cell development sufficient to prevent the leukemia from differentiating beyond Hardy fraction B and, as a result, arrests the leukemia at a stage of development that retains a stem cell-like phenotype.

Interestingly, leukemias that started as $Pax5^{+/-}$ $Arf^{+/-}$ still retained expression of the wild-type *Arf* allele at the protein level suggesting that the two pathways regulated by Pax5 and Arf somehow cooperate during leukemogenesis. *Arf* has long been proposed to be a prototypical two-hit tumor suppressor. And there is some evidence that suggests, in the context of *BCR-ABL1*, that if one *Arf* allele is lost then both *Arf* alleles will be lost during leukemogenesis[75]. Our data would suggest that in the context of *Pax5* haploinsufficiency, *Arf* haploinsufficiency cooperates and the wild-type *Arf* allele is intact suggesting that loss of both *Arf* alleles is not necessary unless Pax5 somehow regulates the *Arf* locus. If Pax5 were found to regulate *Arf* expression then this would explain a possible mechanism by which *Pax5* haploinsufficiency cooperates with *BCR-ABL1* during leukemogenesis; by acting as a function hit to the *Arf*/*Mdm2*/*p53* tumor suppressor pathway. Future studies aimed at elucidating the possible role that Pax5 has in regulating the *Arf*/*Mdm2*/*p53* pathway will provide much needed insight.

In order to elucidate the role that *Pax5* haploinsufficiency plays in regulating the Arf tumor suppressor pathway we will perform a series of *in vitro* studies aimed at assessing Arf protein levels in BCR-ABL1 positive B-cells of varying *Pax5* genotype. Specifically, we will transduce total bone marrow from wild-type, $Pax5^{+/-}$, $Arf^{+/-}$ animals with ecotropic retrovirus harboring the *BCR-ABL1* oncogene using the cationic polymer polybrene. Transduced bone marrow cells will then be cultured for seven days during which time BCR-ABL1 expressing B-cell cultures will develop. Upon maturation of the B-cell cultures, five million cells will be harvested for western blot and quantitative RT-PCR analysis. The relative level of Arf protein and *Arf* transcript will be assessed and compared from wild-type and $Pax5^{+/-}$ cultures, allowing us to determine the effect that hemizygous loss of *Pax5* has on *Arf* expression at the transcript and protein level.

Should we find that *Pax5* haploinsufficiency does affect the level of *Arf* expression *in vitro* then further experiments will be performed *in vivo*. We will harvest B-cells from wild-type, $Pax5^{+/-}$, $Pax5^{-/-}$, $Arf^{+/-}$ and $Pax5^{+/-}$ $Arf^{+/-}$ animals and use protein and RNA extracted from these cells in the same series of analyses in order to determine the effects that *Pax5* dose has on *Arf* expression. This series of experiments will allow us to determine the effects that Pax5 has on Arf expression during normal development in the absence of *BCR-ABL1* signaling. Additionally, using a Pax5 antibody directed at the

DNA binding domain, we will perform chromatin imunoprecipitation (ChIP) experiments in order to determine whether the changes in Arf expression is due to direct Pax5 binding to the *Arf* promoter or due to down-stream signaling effects. This series of experiments will be performed by harvesting B-cells from wild-type, $Pax5^{+/-}$, $Pax5^{-/-}$, $Arf^{+/-}$ and $Pax5^{+/-}$ $Arf^{+/-}$ animals and subjecting them to a cross-linking reagent in order to actively cross-link transcription factors like Pax5 to their DNA targets. Cross-linked cells are then sonicated in order to shear the chromatin into smaller fragments and then the cross-links are reversed, DNA is extracted, and quantitative genomic PCR is performed using assays designed for the *Arf* promoter as well as a control assay designed for the known Pax5 targets *CD19* and *Blnk*. Analysis of the relative enrichment of Arf promoter to either the *CD19* or *Blnk* promoters will allow us to determine if Pax5 directly binds to the Arf promoter.

In summary, our work to elucidate cooperating oncogenic lesions in pediatric ALL has been very fruitful. Further work validating that these secondary oncogenic lesions cooperate has shown that the lesions we have identified are real cooperating lesions. Future work aimed at describing the mechanism by which these lesions cooperate will be of vital importance if we are to progress toward somehow targeting these dysregulated pathways therapeutically in an effort to bring Danny Thomas' profound mission statement to pass "That no child should die in the dawn of life".

LIST OF REFERENCES

1. Miller, R.W., Young, J.L., Jr. & Novakovic, B. Childhood cancer. *Cancer* **75**, 395-405 (1995).
2. Gaynon, P.S.*, et al.* Children's Cancer Group trials in childhood acute lymphoblastic leukemia: 1983-1995. *Leukemia* **14**, 2223-2233 (2000).
3. Gurney, J.G., Severson, R.K., Davis, S. & Robison, L.L. Incidence of cancer in children in the United States. Sex-, race-, and 1-year age-specific rates by histologic type. *Cancer* **75**, 2186-2195 (1995).
4. Young, J.L., Jr., Ries, L.G., Silverberg, E., Horm, J.W. & Miller, R.W. Cancer incidence, survival, and mortality for children younger than age 15 years. *Cancer* **58**, 598-602 (1986).
5. Jaffe, E.S., Harris, N.L., Stein, H. & Vardiman, J.W. World Health Organization Classification of Tumours. Tumours of Haematopoietic and Lymphoid Tissues. *IARC press* (2001).
6. Greaves, M.F. & Wiemels, J. Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer* **3**, 639-649 (2003).
7. Mullighan, C.G.*, et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764 (2007).
8. Fuxa, M. & Busslinger, M. Reporter gene insertions reveal a strictly B lymphoid-specific expression pattern of Pax5 in support of its B cell identity function. *J Immunol* **178**, 8222-8228 (2007).
9. Fitzsimmons, D.*, et al.* Pax-5 (BSAP) recruits Ets proto-oncogene family proteins to form functional ternary complexes on a B-cell-specific promoter. *Genes & Development* **10**, 2198-2211 (1996).
10. Kozmik, Z., Wang, S., Dorfler, P., Adams, B. & Busslinger, M. The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Molecular and Cellular Biology* **12**, 2662-2672 (1992).
11. Nutt, S.L., Morrison, A.M., Dorfler, P., Rolink, A. & Busslinger, M. Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. *Embo J* **17**, 2319-2333 (1998).
12. Nutt, S.L., Urbanek, P., Rolink, A. & Busslinger, M. Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes & Development* **11**, 476-491 (1997).
13. Roessler, S.*, et al.* Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5. *Molecular and Cellular Biology* **27**, 579-594 (2007).
14. Schebesta, M., Pfeffer, P.L. & Busslinger, M. Control of pre-BCR signaling by Pax5-dependent activation of the BLNK gene. *Immunity* **17**, 473-485 (2002).
15. Delogu, A.*, et al.* Gene repression by pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity* **24**, 269-281 (2006).
16. Urbanek, P., Wang, Z.Q., Fetka, I., Wagner, E.F. & Busslinger, M. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79**, 901-912 (1994).

17. Pui, C.H., Relling, M.V. & Downing, J.R. Acute lymphoblastic leukemia. *The New England Journal of Medicine* **350**, 1535-1548 (2004).
18. Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449**, 473-477 (2007).
19. Mullighan, C.G.*, et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110-114 (2008).
20. Goldman, J.M. & Melo, J.V. Chronic myeloid leukemia--advances in biology and new approaches to treatment. *The New England Journal of Medicine* **349**, 1451-1464 (2003).
21. Gleissner, B.*, et al.* Leading prognostic relevance of the BCR-ABL translocation in adult acute B-lineage lymphoblastic leukemia: a prospective study of the German Multicenter Trial Group and confirmed polymerase chain reaction analysis. *Blood* **99**, 1536-1543 (2002).
22. Ribeiro, R.C.*, et al.* Clinical and biologic hallmarks of the Philadelphia chromosome in childhood acute lymphoblastic leukemia. *Blood* **70**, 948-953 (1987).
23. Bartram, C.R.*, et al.* Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia. *Nature* **306**, 277-280 (1983).
24. Shtivelman, E., Lifshitz, B., Gale, R.P. & Canaani, E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* **315**, 550-554 (1985).
25. Chan, L.C.*, et al.* A novel abl protein expressed in Philadelphia chromosome positive acute lymphoblastic leukaemia. *Nature* **325**, 635-637 (1987).
26. Clark, S.S., McLaughlin, J., Crist, W.M., Champlin, R. & Witte, O.N. Unique forms of the abl tyrosine kinase distinguish Ph1-positive CML from Ph1-positive ALL. *Science (New York, N.Y* **235**, 85-88 (1987).
27. Groffen, J.*, et al.* Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. *Cell* **36**, 93-99 (1984).
28. Witte, O.N., Dasgupta, A. & Baltimore, D. Abelson murine leukaemia virus protein is phosphorylated in vitro to form phosphotyrosine. *Nature* **283**, 826-831 (1980).
29. Bloomfield, C.D.*, et al.* Chromosomal abnormalities identify high-risk and low-risk patients with acute lymphoblastic leukemia. *Blood* **67**, 415-420 (1986).
30. Preti, H.A.*, et al.* Philadelphia-chromosome-positive adult acute lymphocytic leukemia: characteristics, treatment results, and prognosis in 41 patients. *The American Journal of Medicine* **97**, 60-65 (1994).
31. Westbrook, C.A.*, et al.* Clinical significance of the BCR-ABL fusion gene in adult acute lymphoblastic leukemia: a Cancer and Leukemia Group B Study (8762). *Blood* **80**, 2983-2990 (1992).
32. Peters, U.R.*, et al.* Aberrant FHIT mRNA transcripts are present in malignant and normal haematopoiesis, but absence of FHIT protein is restricted to leukaemia. *Oncogene* **18**, 79-85 (1999).
33. Jackson, A.*, et al.* Deletion of 6q16-q21 in human lymphoid malignancies: a mapping and deletion analysis. *Cancer Res* **60**, 2775-2779 (2000).

34. Okuda, T., *et al.* Frequent deletion of p16INK4a/MTS1 and p15INK4b/MTS2 in pediatric acute lymphoblastic leukemia. *Blood* **85**, 2321-2330 (1995).
35. Raynaud, S., *et al.* 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**, 2891-2899 (1996).
36. Raimondi, S.C., *et al.* Acute lymphoblastic leukemias with deletion of 11q23 or a novel inversion (11)(p13q23) lack MLL gene rearrangements and have favorable clinical features. *Blood* **86**, 1881-1886 (1995).
37. Heerema, N.A., *et al.* Abnormalities of chromosome bands 13q12 to 13q14 in childhood acute lymphoblastic leukemia. *J Clin Oncol* **18**, 3837-3844 (2000).
38. Cave, H., *et al.* Deletion of chromosomal region 13q14.3 in childhood acute lymphoblastic leukemia. *Leukemia* **15**, 371-376 (2001).
39. McCarroll, S.A., *et al.* Common deletion polymorphisms in the human genome. *Nat Genet* **38**, 86-92 (2006).
40. Rozen, S. & Skaletsky, H.J. Primer3 on the WWW for general users and for biologist programmers. in *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds. Krawetz, S. & Misener, S.) 365-386 (Humana Press, Totowa, NJ, 2000).
41. Garvie, C.W., Hagman, J. & Wolberger, C. Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**, 1267-1276 (2001).
42. Xu, H.E., *et al.* Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes & Development* **13**, 1263-1275 (1999).
43. Strehl, S., Konig, M., Dworzak, M.N., Kalwak, K. & Haas, O.A. PAX5/ETV6 fusion defines cytogenetic entity dic(9;12)(p13;p13). *Leukemia* **17**, 1121-1123 (2003).
44. Persons, D.A., *et al.* Retroviral-mediated transfer of the green fluorescent protein gene into murine hematopoietic cells facilitates scoring and selection of transduced progenitors in vitro and identification of genetically modified cells in vivo. *Blood* **90**, 1777-1786 (1997).
45. Campbell, R.E., *et al.* A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7877-7882 (2002).
46. Czerny, T. & Busslinger, M. DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Molecular and Cellular Biology* **15**, 2858-2871 (1995).
47. Andrews, N.C. & Faller, D.V. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res* **19**, 2499 (1991).
48. Maier, H., Colbert, J., Fitzsimmons, D., Clark, D.R. & Hagman, J. Activation of the early B-cell-specific mb-1 (Ig-alpha) gene by Pax-5 is dependent on an unmethylated Ets binding site. *Molecular and Cellular Biology* **23**, 1946-1960 (2003).
49. Ehrich, M., *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proceedings of the*

National Academy of Sciences of the United States of America **102**, 15785-15790 (2005).
50. Stanssens, P.*, et al.* High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res* **14**, 126-133 (2004).
51. Rahman, M.*, et al.* A repressor element in the 5'-untranslated region of human Pax5 exon 1A. *Gene* **263**, 59-66 (2001).
52. Mahmoud, M.S. & Kawano, M.M. Cloning and analysis of the human Pax-5 gene promoter. *Biochem Biophys Res Commun* **228**, 159-164 (1996).
53. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550 (2005).
54. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263-267 (1995).
55. Nutt, S.L., Heavey, B., Rolink, A.G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556-562 (1999).
56. Busslinger, M. Transcriptional control of early B cell development. *Annu Rev Immunol* **22**, 55-79 (2004).
57. Ying, H., Healy, J.I., Goodnow, C.C. & Parnes, J.R. Regulation of mouse CD72 gene expression during B lymphocyte development. *J Immunol* **161**, 4760-4767 (1998).
58. Cazzaniga, G.*, et al.* The paired box domain gene PAX5 is fused to ETV6/TEL in an acute lymphoblastic leukemia case. *Cancer Res* **61**, 4666-4670 (2001).
59. Bohlander, S.K. ETV6: a versatile player in leukemogenesis. *Semin Cancer Biol* **15**, 162-174 (2005).
60. Wlodarska, I.*, et al.* FOXP1, a gene highly expressed in a subset of diffuse large B-cell lymphoma, is recurrently targeted by genomic aberrations. *Leukemia* **19**, 1299-1305 (2005).
61. Bond, H.M.*, et al.* Early hematopoietic zinc finger protein (EHZF), the human homolog to mouse Evi3, is highly expressed in primitive human hematopoietic cells. *Blood* **103**, 2062-2070 (2004).
62. Hu, H.*, et al.* Foxp1 is an essential transcriptional regulator of B cell development. *Nat Immunol* **7**, 819-826 (2006).
63. Warming, S.*, et al.* Evi3, a common retroviral integration site in murine B-cell lymphoma, encodes an EBFAZ-related Kruppel-like zinc finger protein. *Blood* **101**, 1934-1940 (2003).
64. Ross, M.E.*, et al.* Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102**, 2951-2959 (2003).
65. Martoriati, A.*, et al.* dapk1, encoding an activator of a p19ARF-p53-mediated apoptotic checkpoint, is a transcription target of p53. *Oncogene* **24**, 1461-1466 (2005).
66. Conte, N.*, et al.* Carcinogenesis and translational controls: TACC1 is down-regulated in human cancers and associates with mRNA regulators. *Oncogene* **21**, 5619-5630 (2002).

67. Georgopoulos, K.*, et al.* The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**, 143-156 (1994).
68. O'Riordan, M. & Grosschedl, R. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity* **11**, 21-31 (1999).
69. Winandy, S., Wu, P. & Georgopoulos, K. A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. *Cell* **83**, 289-299 (1995).
70. Nutt, S.L.*, et al.* Independent regulation of the two Pax5 alleles during B-cell development. *Nat Genet* **21**, 390-395 (1999).
71. Robson, E.J., He, S.J. & Eccles, M.R. A PANorama of PAX genes in cancer and development. *Nat Rev Cancer* **6**, 52-62 (2006).
72. Busslinger, M., Klix, N., Pfeffer, P., Graninger, P.G. & Kozmik, Z. Deregulation of PAX-5 by translocation of the Emu enhancer of the IgH locus adjacent to two alternative PAX-5 promoters in a diffuse large-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6129-6134 (1996).
73. Pridans, C.*, et al.* Identification of Pax5 target genes in early B cell differentiation. *J Immunol* **180**, 1719-1728 (2008).
74. Schebesta, A.*, et al.* Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function. *Immunity* **27**, 49-63 (2007).
75. Williams, R.T., Roussel, M.F. & Sherr, C.J. Arf gene loss enhances oncogenicity and limits imatinib response in mouse models of Bcr-Abl-induced acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6688-6693 (2006).
76. Mullighan, C.G.*, et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764 (2007).
77. Ross, M.E.*, et al.* Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* **104**, 3679-3687 (2004).
78. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
79. Gentleman, R.C.*, et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
80. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* **57**, 289-300 (1995).
81. Andersson, A.*, et al.* Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 19069-19074 (2005).
82. Wong, D.J.*, et al.* Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333-344 (2008).
83. Daley, G.Q., Van Etten, R.A. & Baltimore, D. Blast crisis in a murine model of chronic myelogenous leukemia. *Proc Natl Acad Sci U S A* **88**, 11335-11338 (1991).
84. Melo, J.V. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood* **88**, 2375-2384 (1996).

85. Melo, J.V. & Barnes, D.J. Chronic myeloid leukaemia as a model of disease evolution in human cancer. *Nat Rev Cancer* **7**, 441-453 (2007).
86. Drexler, H.G. *The Leukemia-Lymphoma Cell Line Facts Book*, (Academic Press, London, 2001).
87. Manabe, A.*, et al.* Interleukin-4 induces programmed cell death (apoptosis) in cases of high-risk acute lymphoblastic leukemia. *Blood* **83**, 1731-1737 (1994).
88. Hahm, K.*, et al.* The lymphoid transcription factor LyF-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. *Mol Cell Biol* **14**, 7111-7123 (1994).
89. Molnar, A. & Georgopoulos, K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Molecular and Cellular Biology* **14**, 8292-8303 (1994).
90. Molnar, A.*, et al.* The Ikaros gene encodes a family of lymphocyte-restricted zinc finger DNA binding proteins, highly conserved in human and mouse. *J Immunol* **156**, 585-592 (1996).
91. Rebollo, A. & Schmitt, C. Ikaros, Aiolos and Helios: transcription regulators and lymphoid malignancies. *Immunol Cell Biol* **81**, 171-175 (2003).
92. Sun, L., Liu, A. & Georgopoulos, K. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *Embo J* **15**, 5358-5369 (1996).
93. Klug, C.A.*, et al.* Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. *Proc Natl Acad Sci U S A* **95**, 657-662 (1998).
94. Sun, L.*, et al.* Expression of dominant-negative Ikaros isoforms in T-cell acute lymphoblastic leukemia. *Clin Cancer Res* **5**, 2112-2120 (1999).
95. Sun, L.*, et al.* Expression of aberrantly spliced oncogenic ikaros isoforms in childhood acute lymphoblastic leukemia. *J Clin Oncol* **17**, 3753-3766 (1999).
96. Sun, L.*, et al.* Expression of dominant-negative and mutant isoforms of the antileukemic transcription factor Ikaros in infant acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **96**, 680-685 (1999).
97. Nakase, K.*, et al.* Dominant negative isoform of the Ikaros gene in patients with adult B-cell acute lymphoblastic leukemia. *Cancer Res* **60**, 4062-4065 (2000).
98. Olivero, S.*, et al.* Detection of different Ikaros isoforms in human leukaemias using real-time quantitative polymerase chain reaction. *Br J Haematol* **110**, 826-830 (2000).
99. Nishii, K.*, et al.* Expression of B cell-associated transcription factors in B-cell precursor acute lymphoblastic leukemia cells: association with PU.1 expression, phenotype, and immunogenotype. *Int J Hematol* **71**, 372-378 (2000).
100. Takanashi, M.*, et al.* Expression of the Ikaros gene family in childhood acute lymphoblastic leukaemia. *Br J Haematol* **117**, 525-530 (2002).
101. Tonnelle, C.*, et al.* Overexpression of dominant-negative Ikaros 6 protein is restricted to a subset of B common adult acute lymphoblastic leukemias that express high levels of the CD34 antigen. *Hematol J* **4**, 104-109 (2003).
102. Klein, F.*, et al.* BCR-ABL1 induces aberrant splicing of IKAROS and lineage infidelity in pre-B lymphoblastic leukemia cells. *Oncogene* **25**, 1118-1124 (2006).

103. Wang, J.H., *et al.* Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. *Immunity* **5**, 537-549 (1996).
104. Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. & Schatz, D.G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol* **18**, 495-527 (2000).
105. Weir, B.A., *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893-898 (2007).
106. Kirstetter, P., Thomas, M., Dierich, A., Kastner, P. & Chan, S. Ikaros is critical for B cell differentiation and function. *Eur J Immunol* **32**, 720-730 (2002).
107. Mullighan, C.G., *et al.* Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *The New England Journal of Medicine* **360**, 470-480 (2009).

VITA

Christopher Miller was born at Warner Robbins Air Force Base in Robbins, Georgia on June 24, 1975. Only a few months old, he moved to Zweibruken, Germany where he lived for three years before moving back to the United States to Berkeley, California where he spent the majority of his youth. After completing two years of undergraduate education at Solano Community College in Fairfield, California, he moved to Fresno, California where he spent a year in discipleship training at Masters Commission Fresno. The following year, Christopher, moved to Cleveland, Tennessee where he completed his undergraduate coursework at Lee University in 2001. In 2002 he entered medical school at the University of Tennessee College of Medicine in Memphis, Tennessee. After two years he opted to enter the PhD training program at the University of Tennessee and carried out his dissertation work under the supervision of Dr. James R. Downing at St. Jude Children's Research Hospital. Upon completion of his dissertation in 2009, he will reenter medical school and is expected to complete his medical training in 2011.