

7-1-2011

c-myb alternative splicing: a novel biomarker in leukemia

Ye Zhou

Follow this and additional works at: https://digitalrepository.unm.edu/biom_etds

Recommended Citation

Zhou, Ye. "c-myb alternative splicing: a novel biomarker in leukemia." (2011). https://digitalrepository.unm.edu/biom_etds/41

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Biomedical Sciences ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Ye Zhou
Candidate

Biomedical Science Graduate Program
Department

This dissertation is approved, and it is acceptable in quality
and form for publication:

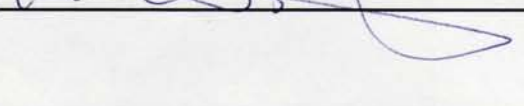
Approved by the Dissertation Committee:


_____, Chairperson









**C-MYB ALTERNATIVE SPLICING:
A NOVEL BIOMARKER IN LEUKEMIA**

by

YE ZHOU

B.S., PHARMACEUTICAL SCIENCE,
CHINA PHARMACEUTICAL UNIVERSITY, 2002

M.S., PHARMACOKINETICS
CHINA PHARMACEUTICAL UNIVERSITY, 2005

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY
Biomedical Science

The University of New Mexico

Albuquerque, New Mexico

July, 2011

ACKNOWLEDGEMENTS

First I would like to thank my mentor, Dr. Scott Ness, for giving me the opportunity to pursue my PhD studies in his laboratory. He has provided the perfect combination of reason, creativity and enthusiasm and has taught me much about the process of science. His high expectations have challenged and inspired me in my work and I appreciate him for his confidence in me.

My project involves developing biotechnological tools and I have benefited greatly from the special expertise of one of my committee members, Dr. Jeremy Edwards, and his group. I thank them for providing me unconditional help and advice.

In addition, I would like to thank all of my committee members, Dr. Michelle Ozbun and Dr. Helen Hathaway for all of their time and helpful conversations. I greatly thank my colleagues in the Ness lab, both past and present, for their friendship and helpful discussion over the past 5 years. I also appreciate the UNM Flow Cytometry Resource Center and the KECK - UNM Genomics facility. Without them, much of my work could not be done.

Finally, a warm thanks to my family: my husband and my parents for their encouragement and support and for helping me maintain perspective.

***c-myb* Alternative Splicing: A Novel Biomarker in Leukemia**

By

Ye E. Zhou

M.S Pharmacokinetics

Ph.D Biomedical Science

Abstract

c-myb encodes a transcription factor that is essential for hematopoiesis and for normal development of other tissues. There is ample evidence showing the activated alleles of *c-myb* gene can induce leukemias in several animal species. However, there was no clear evidence showing that *c-myb* could be oncogenic in human cancers until recent studies revealed that the human *c-myb* locus is subjected to chromosomal rearrangement in some pediatric T-ALL and adenoid cystic carcinoma (ACC) patient samples. The *c-myb* gene contains 15 constant exons and 6 alternative exons and can undergo extensive alternative splicing. These alternatively spliced variants encode variant Myb proteins sharing the same DNA binding domain but distinct C-terminal domains. Evidence suggests that despite their identical DNA binding domains, these c-Myb variant proteins have unique transcriptional activities: the differences in the proteins allow the proteins to regulate different sets of target genes. Most importantly, some of the variants have C-terminal deletions that resemble the deletions in v-Myb, the oncogenic variant of c-Myb that can induce acute leukemias in chickens and

mice. Previous studies showed that *c-myb* alternative splicing is tightly regulated during hematopoiesis. The profiles of *c-myb* alternative splicing in some leukemias are different from the profiles in normal bone marrow samples. We hypothesized that *c-myb* is a unique oncogene that is activated by alternative splicing in leukemias, resulting in the production of variant forms of c-Myb protein that contribute to hematopoietic cell transformation and the development of leukemia. To evaluate the clinical significance of *c-myb* alternative splicing and find out its correlation with leukemia, profiling of *c-myb* alternative splicing in a cohort of leukemia patient samples is needed. We developed a single-molecule exotyping method to characterize the alternative splicing of *c-myb* RNAs in a small cohort of pediatric pre-B-ALL samples and correlated the level or the pattern of *c-myb* splice variants with survival time of these patients. We detected that *c-myb* alternative splicing is more complex and more prevalent in leukemia patients than in normal cells. Furthermore, some splice variants encoding Myb proteins with truncations in the C-terminal domain are correlated with poor survival. We further explored the possibility of adapting high-throughput low-cost Next-Generation Sequencing methods to analyze *c-myb* alternative splicing at a large scale. We conclude that de-regulated *c-myb* alternative splicing could be an important contributor to the development of leukemia and that the analysis of *c-myb* splicing could provide a unique and novel approach of classifying leukemias, which is distinct from analyses performed with other technologies, such as microarray assays of gene expression.

TABLE OF CONTENTS

CHAPTER 1: MYB PROTEINS: ANGELS AND DEMONS IN NORMAL AND TRANSFORMED CELLS	1
ABSTRACT	1
INTRODUCTION	1
Significance of c-Myb in human disease.....	1
Features that make c-Myb unique	1
Questions to be addressed.....	3
THE TWO FACES OF MYB: A REGULATOR AND AN ONCOGENE.....	4
Myb as a critical regulator in normal cells	4
Myb unveiled becomes an oncogene.....	7
N-terminal deletions of c-Myb	9
C-terminal deletions affect intra- and intermolecular interactions	11
Wnt signaling and regulation of c-Myb stability	14
Structure and activities of the Myb DNA binding (SANT) domain	15
The related proteins, A-Myb and B-Myb, are not oncogenic.....	17
Mutations change the transcriptional activities of c-Myb.....	20
DOES MYB REGULATE DIFFERENTIATION, PROLIFERATION OR BOTH?	22
Reversible regulation of differentiation by v-Myb	22
Myb target genes and transforming activities are lineage- and differentiation-specific.....	24
Myb is a key regulator of stem cell fate.....	26
Myb target genes may change during differentiation and the cell cycle.....	28

MECHANISMS AFFECTING C-MYB ACTIVITY AND SPECIFICITY	30
Activation of the <i>c-myb</i> gene in tumors	30
Alternative RNA splicing as a novel mechanism for unleashing c-Myb oncogenicity	32
Regulation of <i>c-myb</i> expression through microRNAs	34
PERSPECTIVE	36
ACKNOWLEDGEMENTS	38
FIGURE LEGENDS	39
Figure 1. Myb genomic and protein structures	39
Figure 2. Wnt signaling affects c-Myb stability and activity.	40
Figure 3. Alternative splicing as a novel means of activating c-Myb.	40
 CHAPTER 2: ABERRANT C-MYB ALTERNATIVE SPLICING IN LEUKEMIAS	 45
AUTHOR CONTRIBUTION.....	46
SINGLE MOLECULE ANALYSIS OF C-MYB ALTERNATIVE SPLICING REVEALS NOVEL BIOMARKERS IN PRECURSOR B-ALL	47
Authors:	47
Affiliation:	47
Corresponding Author:.....	47
ABSTRACT	48
INTRODUCTION	50
RESULTS.....	52
Alternative splicing of <i>c-myb</i> transcripts is complex and combinatorial.....	52

Detection of <i>c-myb</i> alternative splicing variants using a polony-based assay	54
Comparison of the polony and QPCR assays.....	57
Identification of alternatively spliced variants of <i>c-myb</i> in pediatric B-ALL ...	59
Alternative splicing of <i>c-myb</i> as a potential prognostic marker	61
Analysis of the expression of <i>c-myb</i> splice variants relationship with the expression of oncogenic genes in pediatric B-ALL patient samples	62
DISCUSSION	63
MATERIALS AND METHODS.....	67
Cells, tissue culture and patient samples.....	67
RNA expression and structure assays.....	68
Polony amplification.....	68
Denaturation and Hybridization.....	69
Image Acquisition and Data Analysis.....	70
Sensitivity and linearity of the polony assay.....	71
ACKNOWLEDGEMENTS	71
FIGURE LEGENDS	71
Figure 1. Alternative splicing in the <i>c-myb</i> gene	71
Figure 2. Sensitivity and linearity of the polony assay	72
Figure 3. Comparison of exon-specific (QPCR) to single molecule (polony) assays.....	73
Figure 4. Alternative splicing of <i>c-myb</i> transcripts in normal and leukemic cells.....	74

Figure 5. Survival plots for patients grouped by <i>c-myb</i> variant expression levels.....	75
Figure 6. Relationship between the 9S/10 splice variant's expression level (measured by polony assay) and the Myb target genes' expression levels .	75
SUPPORTING INFORMATION LEGENDS	75
Figure S1. Conservation of the <i>c-myb</i> 9S/10 splice variant.	75
Figure S2. Polony exontyping of <i>c-myb</i> transcripts.	76
TABLES	85
Table 1. Expression of <i>c-myb</i> splice variants in different cell lines	85
SUPPORTING INFORMATION TABLES.....	86
Table S1. Primers used for QRT-PCR.....	86
Table S2. All primers used to perform polony amplification and exon-profiling	87
Table S3. The expression level of <i>c-myb</i> splice variants detected by polony assay in 13 pediatric pre-B ALL sample.....	88
 CHAPTER 3: NEXT-GENERATION MULTIPLEX SEQUENCING OF C-MYB ALTERNATIVE SPLICING IN A COHORT OF PEDIATRIC B-ALL SAMPLES: IMPLICATIONS FOR CLINICAL APPLICATIONS	 92
INTRODUCTION	92
RESULTS.....	95
BEAMing (beads, emulsion, amplification and magnetics), a high throughput method to detect <i>c-myb</i> alternative RNA splicing	95
Whole molecule <i>c-myb</i> exontyping by next generation sequencing.....	99

I. Generating individual short DNA fragment libraries	99
II. Dual Primer Emulsion PCR (DPePCR).....	102
III. Sequencing By Polonator	104
DISCUSSION	106
MATERIALS AND METHODS.....	112
I. Bead-based Polony assay	112
cDNA sample preparation.....	112
Preparation of DNA Primer Coupling Beads	113
PCR Reaction Mix Preparation and Formulation	113
Emulsification and Amplification	114
Breaking the Emulsion and Recovery of Beads.....	116
Enrichment of the Beads.....	117
Fluorescent labeled Probe Hybridization	118
Bead Immobilization.....	119
Exontyping on the polony sequencing system	120
Image Acquisition	120
II. Whole molecule <i>c-myb</i> exontyping by Next-Generation Sequencing.....	121
Sample preparation and library construction.....	121
Library Construction QC	125
Coupling of dual-biotinylated primer to the streptavidin beads.....	126
Dual Primer Emulsion PCR.....	127
Polonator Sequencing.....	128
Dual Primer ePCR Master Mixer.....	130

Silicone Oil Phase.....	130
Breaking DPePCR emulsions	131
Capping with dideoxynucleotide reaction.....	131
FIGURE LEGENDS	132
Figure 1. Molecular Barcoding Approach.....	132
Figure 2. Process schematic of bead based polony assay.	132
Figure 3. Process schematic of <i>c-myb</i> cDNA library construction.....	133
Figure 4. Oligos and intermediates in Acul inside adaptor ligation.	134
Figure 5. Agarose gels of selected library construction intermediates.	134
Figure 6. Dual Primer Emulsion PCR approach.....	135
Figure 7. The final DNA library for DPePCR.	136
Figure 8. Polonator sequencing.....	136
Figure 9. Distribution of sequence coverage for each exon.....	137
TABLES	149
Table 1. Primers for polony sequencing (bead-based) assay	149
Table 2. The splice variants in 4 pediatric pre-B-ALL patient samples detected by the bead based polony assay (ND: Not Detectable).....	150
Table 3. Primers for constructing sequencing libraries, emulsion PCR and sequencing by NGSM.....	153
CHAPTER 4: DISCUSSION	157
DO C-MYB SPLICE VARIANTS HAVE UNIQUE TRANSCRIPTIONAL ACTIVITIES THAT CONTRIBUTE TO ONCOGENESIS?	158
DO MYB PARTNERS DEFINE THEIR ACTIVITIES?.....	160

COULD C-MYB ALTERNATIVE SPLICING AFFECT STEM CELL FATE?.....	162
IMPLICATION FOR THE FUTURE	164
APPENDICES.....	166
APPENDIX I: CHARACTERIZATION OF VARIANT MYB PROTEINS IN NALM-6 CELLS	166
MATERIALS AND METHODS.....	167
Myb lentiviral Production.....	167
Making Stable Cell Lines	167
Immunoprecipitation and Western Blot	167
Chromatin Immunoprecipitation	168
Microarray analysis.....	169
RESULTS.....	169
FLAG-tagged Myb cell lines.....	169
Chromatin-IP on the stable cell lines.....	170
Gene-expression profiles in NALM-6 cells with variant Myb proteins expression	173
DISCUSSION	174
FIGURE LEGENDS	176
Figure 1. The expression of variant Myb proteins in stable cell lines	176
Figure 2. FLAG-tagged variant Myb proteins bind to different gene promoters.....	176
TABLES	179
Table 1: Promoter Primers for SYBR Green	179

Table 2: The gene expression changes induced by variant Myb proteins.. 180

APPENDIX II: DEVELOP CELL LINES ECTOPICALLY EXPRESSING TAP-TAGGED VARIANT MYB PROTEINS TO STUDY PROTEIN-PROTEIN

INTERACTION	181
MATERIALS AND METHODS.....	182
Generation of NTAP-MYB.....	182
Myb Lentiviral Production.....	182
Cell Culture	183
Transfection assay.....	183
Making Stable Cell Lines	183
TAP Purification and Western Blot.....	184
RESULTS.....	185
TAP-tag does not affect the transcriptional activities of c-Myb protein.....	185
Purification of c-Myb Complexes	186
IMPLICATIONS.....	187
FIGURE LEGENDS	188
Figure 3. Schematic of the tandem affinity purification of NTAP-c-Myb protein complex.	188
Figure 4. Activation of a c-Myb-responsive reporter gene by TAP-c-Myb and normal c-Myb proteins.	189
Figure 5. Tandem affinity purification of TAP-c-Myb protein.	189
REFERENCE.....	194

Chapter 1: Myb proteins: angels and demons in normal and transformed cells

Abstract

A key regulator of proliferation, differentiation and cell fate, the c-Myb transcription factor regulates the expression of hundreds of genes and is in turn regulated by numerous pathways and protein interactions. However, the most unique feature of c-Myb is that it can be converted into an oncogenic transforming protein through a few mutations that completely change its activity and specificity. The c-Myb protein is a myriad of interactions and activities rolled up in a protein that controls proliferation and differentiation in many different cell types. Here we discuss the background and recent progress that have led to a better understanding of this complex protein, and outline the questions that have yet to be answered.

Introduction

Significance of c-Myb in human disease

Features that make c-Myb unique

Several things make *c-myb* a unique type of oncogene and an unusual transcription factor. Unlike some oncogenes whose normal functions are closely linked to the regulation of the cell cycle or to the mitogenic response (e.g. SRC,

MYC, FOS), the normal role of *c-myb* is most closely associated with the regulation of differentiation, especially in immature hematopoietic cells^{21,22}. Transformation by the *v-myb* oncogenes is associated with a block in differentiation^{23,24} and disrupting the function of *v-myb* in transformed cells causes them to stop proliferating and to differentiate²⁵. This raises the question of whether c-Myb regulates both differentiation and proliferation, or whether changes in proliferation occur as a consequence of Myb-regulated changes in differentiation. For example, the more immature transformed cells may respond more vigorously to cytokines or growth factors, resulting in increased proliferation. Alternatively, c-Myb could be involved in the regulation of both proliferation and differentiation. Another unique feature of c-Myb, which may be the key to understanding how it functions as an oncogene, is that its activity is malleable: it regulates different genes in different situations and its activity can be dramatically altered by relatively minor mutations, even single amino acid changes^{26,27}. Thus, c-Myb appears to have many different functions and many different roles in various tissue types and situations, and mutations or mechanisms that disrupt these context-specific effects may increase oncogenic activity.

One of the most unique features of *c-myb* is the extremely complex nature of its regulation. The expression of the *c-myb* gene is regulated at several levels, including a sophisticated control of transcriptional elongation that occurs in the first exon and that may be regulated by the transcription factor NFkappaB or by estrogen receptor, depending on the cell type^{5,28,29}. The *c-myb* gene exhibits

quite complicated alternative RNA splicing that can produce dozens of different transcripts encoding different versions of c-Myb protein with unique transcriptional activities ¹⁹. This makes the *c-myb* gene multifaceted and capable of encoding subtly different transcription factors in different situations or cell types. At the post-transcriptional level the *c-myb* mRNAs are tightly controlled by a system of microRNAs, mostly that bind to the very long 3'-UTR region ³⁰⁻³⁴. However, the most sophisticated and least understood regulation occurs after translation. The c-Myb protein is subject to at least 10 and probably many more post-translational modifications, all of which have the ability to change the activity or stability of the protein or to affect which proteins c-Myb is able to interact with and therefore which genes it can regulate. Thus, the *c-myb* gene and the c-Myb protein lie at the intersection of many different regulatory pathways feeding into such diverse mechanisms as the control of transcription, alternative RNA splicing, translation and protein activity and stability.

Questions to be addressed

It has been over 30 years since *v-myb* was first discovered in a virus that caused leukemia in chickens. The diverse studies and approaches undertaken by numerous laboratories have uncovered a myriad of complexities related to the expression and regulation of c-Myb, but several important questions remain unanswered. It is still not clear how c-Myb regulates both differentiation and proliferation, two processes that seem contradictory. Although c-Myb is a DNA-binding transcription factor ²⁶, it is still not clear which Myb-regulated target genes

are the most important ones for the oncogenic activity of v-Myb, or whether Myb proteins regulate different genes in tumors than in normal cells. Although the turnover of c-Myb protein is tightly regulated, whether higher intracellular levels of c-Myb protein lead to a different outcome in terms of gene expression patterns, changes in differentiation or increased proliferation remains to be determined. The goal of this review is to address these and related questions in order to shed light on the most important and most relevant discoveries, and to set a framework for future studies on c-Myb and its role in normal cells and tumors.

The Two Faces of Myb: A Regulator and An Oncogene

Myb as a critical regulator in normal cells

Like many transcription factors, c-Myb has a highly conserved DNA binding domain (Figure 1A) ³⁵, recognizes and binds to specific sequences in DNA ²⁶ and activates the expression of specific target genes ^{26,36,37}. The unique feature of c-Myb is that a few mutations can convert it into a transforming protein that regulates a completely different spectrum of target genes ³⁶. Trying to understand the normal functions of c-Myb and how they become corrupted in the oncogenic v-Myb has been the motivation behind many of the studies involving the Myb proteins.

Because the viruses expressing v-Myb caused leukemias, much of the work involving c-Myb has focused on its functions in hematopoietic cells. Early studies showed that anti-sense oligos targeting c-Myb could block in vitro hematopoietic

cell differentiation ²¹ and that homozygous disruption of the murine *c-myb* gene causes a dramatic failure of definitive erythro- and myelopoiesis ²². Since then, c-Myb has been intensively investigated in the context of the hematopoietic system where it has been shown to regulate progenitor cell expansion and differentiation of a number of lineages. A variety of mutant alleles and conditional knockout mouse models have facilitated the study of c-Myb in hematopoiesis. For example, inducible *c-myb* knockout systems have shown that precise expression levels of c-Myb are required at distinct differentiation steps of each hematopoietic cell lineage ³⁸. Tissue-specific deletion of c-Myb in lymphoid progenitors has demonstrated the importance of c-Myb expression during B-cell ³⁹ and T-cell development ^{40,41}. In addition, low levels of *c-myb* expression lead to a myeloproliferative phenotype resembling a stem cell disorder ⁴², suggesting that c-Myb is required for proper differentiation and proliferation of hematopoietic stem cells.

Presumably, c-Myb is required in these diverse cell types because it regulates specific genes. Indeed, c-Myb appears to regulate some genes, such as c-Kit ⁴³⁻⁴⁶ in many cell types while other genes, such as Bcl-2 ^{47,48}, may only be targets in specific lineages. Many other c-Myb targets have been identified in hematopoietic cells, including the first identified target *mim-1* ²⁶, the stem cell antigen CD34 ⁴⁹, the T-cell receptor delta ⁵⁰, the cell cycle regulatory genes CDC2 ⁵¹ and CCNB1 ⁵² and the MYC oncogene ⁵³. Although many of these target genes play important roles in hematopoiesis or even in oncogenesis, none of the genes that have been identified have provided more than a partial

explanation for what the role of c-Myb is in hematopoietic or other cell types, nor have they led to a good explanation for why c-Myb should have oncogenic activity.

Recent studies have also demonstrated an important need for c-Myb expression in the normal development of some epithelial and other tissues. In situ hybridization studies have shown that *c-myb* is expressed at high levels in hematopoietic cells in developing embryos, but also in neural retina and in respiratory epithelial cells ⁵⁴. In the colon, c-Myb is required for colonic crypt homeostasis, including the integrity, normal differentiation, and steady-state proliferation of colon epithelial stem cells and progenitors ^{48,55,56}. Interestingly, c-Myb expression has been found to be important for neural progenitor cell proliferation and for maintenance of neural stem cell niche ⁵⁷, and an involvement of c-Myb in melanocytes links it to the functions of neural crest cells ⁴⁴. Clearly, the *c-myb* gene is widely expressed, and the c-Myb protein plays a role in many cell types, not just hematopoietic cells. As conditional knockout systems are expanded to other cell lineages, it seems likely that the number of tissues that require c-Myb for normal development will increase.

One complication of determining the role of the c-Myb protein in normal tissues is the presence of two related transcription factors, A-Myb (MYBL1) and B-Myb (MYBL2), which are often co-expressed with c-Myb. The three Myb proteins have nearly identical DNA binding domains ⁵⁸ and bind the same DNA sequences in vitro and in reporter plasmids ⁵⁹. However, they are biologically distinct, since mice lacking A-Myb or B-Myb have completely different

phenotypes than those lacking c-Myb^{22,60,61}. The genes encoding the three Myb proteins are all induced by activating the estrogen receptor in MCF-7 breast cancer cells⁶², making this system a good one for studying the activities of all three proteins. Microarray studies have shown that the three Myb proteins induce almost totally non-overlapping sets of target genes when ectopically over-expressed in MCF-7 cells. This suggests that the similar DNA binding domains are not enough to determine the specificities of the Myb proteins. Instead, the large transcriptional trans-activation and protein-protein interaction domains (Figure 1A) must play a role by helping to steer each of the transcription factors to specific target genes. Thus, it seems likely that the products of the three genes have completely different functions, although they likely evolved from a single progenitor Myb gene similar to the one in *Drosophila*⁶³, which is involved in chromosome condensation and genome stability^{64,65}.

Myb unveiled becomes an oncogene

The AMV and E26 viruses expressing different versions of v-Myb are both strongly oncogenic, capable of transforming immature avian hematopoietic cells in vitro and inducing leukemias in animals^{24,27,66,67}. In addition, both forms of v-Myb can induce leukemias in mice when expressed from mouse retroviruses^{68,69}. However, both v-Myb proteins are truncated and mutated derivatives of wild type c-Myb (Figure 1A), lacking the N-terminus and C-terminus of c-Myb while retaining the highly conserved DNA binding domain and more than half of the C-terminal domain, including the region required for transcriptional transactivation

⁷⁰. In contrast to v-Myb, full-length c-Myb is only capable of weak transformation in vitro ^{71,72} and does not induce leukemias in animals, not even in transgenic mice that over-express c-Myb in all tissues ⁷³. However, C-terminal deletion mutants of c-Myb are much more potent oncogenes and do induce leukemias in mice ^{74,75}. In addition, leukemias induced by retroviral insertions into the *c-myb* gene express either N- or C-terminal truncations of c-Myb ^{76,77}, suggesting that truncation of one end or the other of c-Myb is necessary for conversion into a potent oncogenic form. These results lead to the conclusion that c-Myb is subject to auto-inhibition or repression, which can be relieved by deletion of one end or the other, or both.

Researchers studying the differences between v-Myb and c-Myb have focused on the idea that the two proteins have quantitatively different activities: that they have essentially the same activities but the v-Myb protein is de-repressed, lacking the negative controls that keep c-Myb in check. However, this viewpoint now appears to be incorrect, since several types of evidence point to a qualitative difference in the activities of c-Myb and v-Myb. The first hint should have come from the study of the first identified Myb target gene, *mim-1*, which is activated by c-Myb and the v-Myb protein from E26 virus, but not the v-Myb encoded by AMV ^{26,27}. Instead, researchers focused on the fact that all the Myb proteins can bind the same DNA sequences and activate the same reporter gene plasmids as evidence that they must regulate the same genes. It is now clear that the results from reporter genes are misleading and fail to distinguish between the different Myb proteins. However, microarray assays clearly show

that the different Myb proteins activate completely different sets of target genes^{36,37,78}, suggesting that they are qualitatively and functionally distinct. These conclusions have important implications for our understanding of how Myb proteins work and what controls their specificities. If the DNA binding domains of the Myb proteins are not sufficient to direct them to target genes, then protein-protein interactions become much more important and explain why the different Myb proteins have such different activities. However, protein-protein interactions are subject to regulation by many mechanisms, including post-translational modifications, suggesting that not only the activities, but also the specificities of Myb proteins are highly regulated.

N-terminal deletions of c-Myb

N-terminal truncation appears sufficient to unleash at least some of the transformation potential of c-Myb. Retroviral insertional mutagenesis that results in a truncation of amino acids 1-20 of c-Myb has been found in non-bursal B cell lymphomas caused by ALV (avian leukosis virus) in chickens⁷⁹. This deletion occurs in recombinant retrovirus-induced rapid-onset tumors, including B-cell lymphomas, sarcomas, and adenocarcinomas in animals⁸⁰, suggesting that the N-terminal deletion is important for increasing the oncogenic activity of c-Myb. The precise role of the N-terminus of c-Myb is unclear, but it has been reported that deletion of the N-terminus can abrogate the ability of c-Myb protein to cooperate with other transcription factors (Figure 1), such as Ets-1⁸¹, which raises the possibility that oncogenic activation of c-Myb is linked to the loss of

cooperation between Myb and other cofactors. Truncation of the N-terminus and the first Myb repeat can cause a decreased affinity for DNA binding and affects interactions between c-Myb and Cyclin D1^{82,83}. Moreover, mutant forms of c-Myb with different N-terminal deletions display differences in transcriptional transactivation activities and the ability to interact with C/EBPbeta⁸⁴. All the N-terminal deletions lack a conserved casein kinase II (CKII) phosphorylation site that may influence DNA binding activity⁸⁵. However, mutagenesis of the CKII site does not influence the oncogenic activity of c-Myb⁸², suggesting that its importance may be limited to specific target genes or cell types.

There is provocative evidence suggesting that modifications of the N-terminus of c-Myb protein may occur in vertebrates through unusual transcriptional mechanisms. The first type of evidence is the identification of a second promoter in the *c-myb* gene, located in the normal intron 1, which would encode an mRNA lacking the normal exon 1 and a protein with a small N-terminal deletion lacking the CKII phosphorylation sites⁸⁶. The implication is that alternative use of the second promoter in some cell types or in some circumstances would produce a subset of c-Myb proteins lacking the N-terminus and whatever negative regulatory controls it confers. The second type of evidence suggests that some *c-myb* transcripts arise via a mechanism of trans-splicing, starting at a completely different promoter located on a different chromosome from the rest of the *c-myb* gene, but resulting in the production of a chimeric transcript encoding a protein that lacks the normal N-terminus of c-Myb^{87,88}. This line of evidence originated with the cloning and characterization of one

of the first cDNAs for *c-myb*, which contained sequences from two different chromosomes⁸⁹. Although the evidence for trans-splicing is compelling, it is not clear how important that mechanism is in the biology of *c-myb* or whether proteins encoded by the trans-spliced chimeric mRNAs have any role in controlling cell fate or in regulating specific sets of genes. However, given the complexity of *c-myb* gene regulation, it would be foolish to rule out trans-splicing just because it is unexpected without proof that it is not important.

C-terminal deletions affect intra- and intermolecular interactions

All the transforming alleles of c-Myb, including both forms of v-Myb, have C-terminal deletions, suggesting that elements at or near the C-terminus play an important negative regulatory role, perhaps by keeping the activity of the full length protein in check. The fact that overexpression of full-length c-Myb cannot efficiently transform cells in vitro or induce tumors in animals suggests that negative regulation of c-Myb is inherent in its structure. In support of this idea, the EVES motif (513aa-563aa) near the C-terminus of c-Myb (Figure 1A) has been shown to interact with the N-terminal DNA binding domain^{90,91}. This intramolecular interaction could form the basis for autoinhibition, explaining why c-Myb fails to transform cells, even if grossly over-expressed. The EVES motif overlaps a site of in vivo phosphorylation by p42 Mitogen-Activated Protein Kinase (p42MAPK)⁹² and is close to several sites of sumoylation and ubiquitinylation. Thus, activation of signal transduction cascades, leading to changes in p42MAPK activity and c-Myb phosphorylation could affect the

intramolecular interactions and transcriptional activity and/or stability of c-Myb^{93,94}. Although the potential for these regulatory mechanisms has been demonstrated, a direct link between signaling pathways, p42MAPK activity and changes in the transcriptional or oncogenic activity of c-Myb has yet to be demonstrated.

Recognition of the EVES motif in c-Myb led to the identification of a related domain in the transcriptional co-activator p100, also known as Tudor-SN and recently renamed SND1⁹⁵. The p100/SND1 protein appears to have multiple functions: it is a transcriptional co-activator^{96,97}, a component of the RISC complex involved in microRNA metabolism^{98,99} and a key component of RNA editing complexes^{100,101}. For c-Myb, the implication is that intramolecular interactions within the c-Myb protein compete for the intermolecular interactions with p100/SND1⁹⁰, which could also affect the interactions between c-Myb and Pim-1¹⁰², an oncogenic kinase that phosphorylates the DNA binding domain of c-Myb¹⁰³. C-terminal truncations of c-Myb could disrupt the intramolecular interaction, allowing other domains of c-Myb to recruit co-activators or co-repressors, leading to changes in transcriptional activity. Three co-repressors (Ski, N-CoR, and mSin3A) that bind to the DNA-binding domain of c-Myb form a complex with TIF1beta that binds to the C-terminus¹⁰⁴. Deletion of the C-terminus decreases the interactions with these co-repressor complexes and also weakens the co-repressor-induced negative regulation of Myb activity (Figure 1A). Meanwhile, the deletion of the C-terminus could also unmask binding sites for other proteins in the DNA-binding domain or N-terminal domain, such as

CBP/p300¹⁰⁵. These regulatory mechanisms are likely dependent on enzymes that bind c-Myb and catalyze changes in protein conformation. Examples include the peptidyl proline isomerases Cyp40, which binds c-Myb and induces a conformational change that affects DNA binding activity¹⁰⁶ and Pin1, which binds c-Myb in a phosphorylation-dependent manner⁹⁴. All these results paint an increasingly complex picture of the interdependence of each domain of c-Myb and the interplay of multiple cofactors in modulating its activity, and suggest that multiple different regulatory pathways interface with c-Myb to control its activity.

Deletion of the C-terminus of c-Myb plays a major role in changing the stability of the protein. Full-length c-Myb has a short half-life in most situations¹⁰⁷, which is greatly extended in proteins with C-terminal truncations. Some aspects of c-Myb activity are closely tied to the level of its expression^{38,42}, so controlling the stability of the protein could play a critical role in regulating c-Myb activity and cell fate in some situations. The regulation of c-Myb degradation is linked to the presence of a PEST sequence¹⁰⁸, which targets the protein for ubiquitinylation and degradation. The PEST sequence overlaps with the EVES motif and the p42MAPK phosphorylation site, suggesting that the mechanisms that target c-Myb to the proteasome are highly regulated by upstream signaling pathways.

Wnt signaling and regulation of c-Myb stability

The carboxyl terminal domain of c-Myb that is deleted in v-Myb plays a major role in negative regulation of c-Myb activity. Retrovirus-induced truncation of the C-terminus of c-Myb leads to a substantial increase in transactivation^{70,109,110} and transformation activity¹⁰⁹. Several types of post-translational modifications occur in the C-terminal domain of c-Myb, affecting stability, subcellular location¹¹¹ or interactions with other proteins. For instance, activation of the Wnt-1 signaling pathway leads to phosphorylation of c-Myb at multiple sites, followed by ubiquitinylation, targeting to the proteasome and degradation¹¹². The phosphorylation sites in the C-terminal domain are more critical for the Wnt-1-induced degradation since the interaction of c-Myb with E3 ubiquitin ligases, such as Fbxw7¹¹³ can be enhanced by these modifications. The oncogenic v-Myb protein lacks the phosphorylation and ubiquitinylation sites and is relatively resistant to this degradation pathway (Figure 2), which may partially explain the differential transforming activity of v-Myb vs c-Myb¹¹⁴.

Although the number of proteins known to interact with c-Myb increases each year, there is still relatively little information about how any of these interactions affect the activity of c-Myb or whether the interactions are significant for oncogenic activities. These limitations are mostly due to problems with the available assays. A major problem is that most investigators assess the activity of c-Myb only through transfection assays using plasmid-based reporter genes. However, it has been clear since the identification of the very first c-Myb target gene, *mim-1*, that reporter genes do not distinguish between the activities of

different Myb proteins or between normal and oncogenic alleles of c-Myb ^{26,27}.

Microarray studies have shown that c-Myb and v-Myb activate completely different, almost completely non-overlapping sets of target genes in human cells ^{36,78}. It will most likely be necessary to study the effects of the Myb-binding proteins in the context of real target genes in order to determine how they affect c-Myb activity. A related problem is that it is not yet clear which Myb target genes are linked to oncogenic activity. Some target genes, such as cyclin genes (CCNA1, CCNE1, CCNB1), other proto-oncogenes (c-MYC and c-KIT) or survival genes (BCL-2) seem likely to be important for transformation or oncogenesis. However, no target genes have yet been linked directly to the transformation activity of c-Myb or its oncogenic derivatives.

Structure and activities of the Myb DNA binding (SANT) domain

The 48 KDa v-Myb protein encoded by AMV, the 135 KDa Gag-Myb-Ets protein encoded by the E26 virus, and the 75 KDa c-Myb protein are all nuclear and bind to the same DNA sequence (PyAACG/TG) ¹¹⁵. All three share a highly conserved DNA binding domain composed of two approximately 50 amino acid long “Myb repeats”, each of which includes three tryptophan residues separated by a characteristic spacing ¹¹⁶. This motif is present in three tandem repeats (R1, R2, and R3) in c-Myb, the first of which is deleted in the AMV and E26 oncoproteins ^{117,118}. The first repeat has been implicated in stabilizing the binding to DNA ¹¹⁹ but it is not required for DNA binding ¹²⁰ and deletion of the first repeat increases the oncogenic activity of c-Myb ¹¹⁰. Thus, the core DNA binding

domain includes only the second and third Myb repeats (residues 90-192) ¹²⁰.

Four amino acid substitutions are present within the DNA-binding domain of AMV v-Myb. These substitutions do not appear to alter the sequence specificity of DNA binding ¹¹⁵, but do affect the ability of the protein to transform different cell types, to regulate specific genes and to be regulated by other cellular proteins ^{27,106,121}.

Although it is named for its ability to bind DNA, the most conserved part of c-Myb ⁵⁸ is also important for a large number of protein-protein interactions. As discussed above, the DNA binding domain is the site of intramolecular interactions with the C-terminal EVES domain ⁹⁰ and for intermolecular interactions with the structurally related p100/SND1 co-activator protein ^{90,102}. The DNA binding domain is also the site of interactions with and phosphorylation by protein kinases including Protein Kinase A ¹²² and Pim-1 ¹⁰³. The tetratricopeptide repeat domain of Cyp40 interacts with the DNA binding domain of c-Myb ¹⁰⁶, as do the cell cycle regulator Cyclin D1 ⁸³ and the transcriptional co-repressors Ski, N-CoR and mSin3A ¹⁰⁴. The solution structure of the Myb DNA binding domain ¹²⁰ shows that both the structure and the outside surface of the domain – which interacts with other proteins – is perfectly conserved amongst vertebrate Myb proteins. Interestingly, several of the mutations in the oncogenic v-Myb protein change surface residues in the DNA binding domain, suggesting that they enhance the oncogenic activity by affecting interactions with regulators ¹²² or co-activators that change the specificity of the protein.

In addition to recognizing specific promoters and binding to numerous proteins, the DNA-binding domain of c-Myb is also involved in chromatin remodeling and oncogenic mutations can abolish this function^{38,123}. This suggests that the DNA binding domain may play somewhat of a catalytic role, initiating changes in chromatin structure or histone modifications at regulated promoters. Recently the Myb repeat found in the Myb DNA binding domain was relabeled as a SANT domain, variants of which are found in several chromatin remodeling enzymes¹²³⁻¹²⁵. SANT domains are the critical DNA-and chromatin recognition domains in regulators such as Swi3p, Ada2p, Rsc8p and TRF2 and are responsible for recognition of histone tails. Thus, the DNA binding/SANT domain of c-Myb may play an important role in epigenetic regulation of target genes or in the initiation of chromatin remodeling in genes that become activated or repressed.

The related proteins, A-Myb and B-Myb, are not oncogenic

In addition to c-Myb, all vertebrates also express two closely related proteins with similar overall structures and only a few differences to the surface residues of their DNA binding domains⁵⁸. They both have been shown to be transcription factors^{126,127}, but neither has been shown to be an oncogene or to have transforming activity. B-Myb (MYBL2) is expressed in all proliferating cells and is required for cell proliferation in which it plays a role in cell cycle progression¹²⁸. B-Myb and c-Myb have some similar effects in hematopoietic cells¹²⁹, but they are regulated differently and are not interchangeable¹³⁰.

Overexpressed or amplified B-myb has been reported in different types of human malignancies ^{131,132}, but since its expression is linked to proliferation, overexpression of B-Myb may just signal that the tumor cells are dividing. It is suspected that B-Myb plays a role in human cancer, but there is still no direct evidence of its causative role. Although B-Myb recognizes the same DNA sequence as c-Myb (PyAACG/T/G), it regulates completely different sets of genes ³⁷. B-Myb has been identified as a component of LINC (LIN complex), a protein complex involved in the regulation of genes that are expressed in the G2/M phase of the cell cycle ¹³³. LINC is related to similar complexes dREAM and DRM from *Drosophila* and *C. elegans* ¹³³, suggesting that this function of B-Myb is evolutionarily conserved and primordial. Thus, B-Myb appears to be the vertebrate equivalent of the original Myb protein, only one of which is found in lower organisms. Compared to B-Myb, c-Myb appears to be a specialized version that has acquired oncogenic potential.

The other Myb family member, A-Myb (MYBL1), is the least well studied and remains the most enigmatic. In lymphoid cells, A-Myb expression is normally restricted to the proliferating B-cell centroblasts ¹³⁴ and is not correlated with that of c-Myb and B-Myb ¹²⁹. Human A-Myb is a strong transcriptional activator ¹³⁵ and is therefore more similar to the oncogenic v-Myb than to c-Myb. Transgenic mice with overexpressed A-Myb develop hyperplasia of the spleen and lymph nodes with over-expanded B lymphocyte populations ¹³⁶, suggesting that A-Myb may contribute to hyperplasia by increasing the rate of B cell proliferation. A-Myb is overexpressed or deregulated in several subtypes of human B-cell neoplasia

¹³⁷ and there is evidence that A-Myb regulates important target genes such as MYC ¹³⁸. However, there is no direct evidence that A-Myb is oncogenic or that it transforms cells in vitro or in animals. Although many studies of A-Myb activity have focused on its role in hematopoietic cells, it is widely expressed and is often co-expressed with c-Myb and B-Myb. For example, the expression of all three Myb proteins is stimulated by estrogen in MCF-7 breast cancer cells ⁶².

Many types of proliferating cells co-express the A-Myb, B-Myb and c-Myb proteins, and since all three proteins can bind the same DNA sequences and activate the same plasmid-based reporter genes, there is a chance that the three proteins regulate the same genes, perhaps in different phases of the cell cycle. Alternatively, they might have different activities, for example one might activate a promoter that the other represses. However, microarray studies showed that the three Myb proteins activated essentially completely different sets of target genes in human cells ³⁷. Furthermore, swapping the DNA binding domains between A-Myb and c-Myb did not alter their specificities, indicating that the other parts of the proteins, not the conserved DNA binding domains, were most important for determining which genes got regulated by the different Myb proteins. The best explanation is that the DNA binding domain is required for finding Myb binding sites, but the protein-protein interactions mediated by other domains are required for cooperative interactions necessary to stabilize the formation of transcription complexes. Thus, the other parts of the Myb protein determine which promoters get favored, and essentially determine the specificities and which genes get regulated. This has important implications for

how the Myb proteins are regulated, since post-translational modifications and mutations can affect protein-protein interactions and therefore determine which target genes get activated in different situations.

Mutations change the transcriptional activities of c-Myb

Many studies have investigated the regions of the c-Myb and v-Myb proteins involved in transcriptional activation, mostly using standard plasmid-based reporter genes that give information about activity, but not specificity. Deletion studies originally identified the transcriptional activation domain (residues 241-325) of c-Myb as a conserved region downstream from the DNA binding domain^{117,118}. However, other studies have shown that this small central domain is not sufficient for transcriptional activation in the context of the native v-Myb protein⁶⁷. Rather, a series of adjacent subdomains are required for transcriptional activation by v-Myb, including a heptad leucine repeat or “leucine zipper” (LZ) and the conserved FAETL motif^{139,140}. Substitution of specific leucines in the LZ with proline residues can activate the protein for both transcriptional regulation and for oncogenic transformation¹⁴¹, suggesting that the LZ is involved in negative regulation of c-Myb. In contrast, mutations in this region decrease the activities of v-Myb^{67,140}. The results suggest that the central, transactivation domain of c-Myb and v-Myb is a series of conserved motifs responsible for mediating interactions with co-activators such as p300/CBP^{142,143,144,145} as well as other proteins whose functions are less well defined, such as p160¹⁴⁶. It is likely that some of these interactions are more generic, affecting the

transcriptional activity of c-Myb measured in reporter gene assays, while others are more specific and may only become evident when tested in a more biologically relevant context or with specific sets of target genes.

As described above, the C-terminal domain of c-Myb plays an important role in controlling which target genes are regulated in different situations. Deletions and domain swap experiments have demonstrated this for panels of human target genes³⁶, and this is the best explanation for the finding that A-Myb and c-Myb regulate such different sets of genes, despite having nearly identical DNA binding domains^{37,59}. The C-terminal domain of c-Myb, which includes the transcriptional activation and LZ regions described above, contains numerous sites of post-translational modifications, including sites of acetylation^{147,148}, phosphorylation^{92,112,122,149-151}, sumoylation^{93,111,152-154} and ubiquitinylation^{107,113,155} and is the site of action of the proline isomerases Cyp40 and Pin1 that have dramatic effects on c-Myb activity. It seems likely that the activity and specificity of c-Myb is regulated by post-translational modifications, some of which could change in real time, such as during the cell cycle. Thus, c-Myb is likely to regulate different sets of target genes in different phases of the cell cycle. However, such time-dependent changes may not be detectable using the reporter gene assays or even the endogenous gene activation assays that have been used for most studies involving Myb proteins.

Overall, this section has described the activities of c-Myb and v-Myb, and how mutations can convert one into the other. The results from many laboratories and numerous different types of experiments suggest that the activities of c-Myb

are context-specific and affected by numerous post-translational modifications. It seems likely that the mutations that distinguish v-Myb mimic a specific set of modifications in c-Myb, for example by freezing v-Myb into a conformation that stimulates proliferation. Alternatively, the v-Myb mutations may completely change the activity of c-Myb, converting it into something more sinister. If the latter is true, it would mean that v-Myb and c-Myb are as different in activities as A-Myb, B-Myb and c-Myb, and that they are, for all intents and purposes, completely different transcription factors. This is in contrast to the notion that v-Myb is merely a constitutively activated or de-repressed version of c-Myb. The possibility that the normal and necessary activities of c-Myb could be corrupted by mutations into something qualitatively different is one of the defining features of the Myb proteins, and sets it apart from other oncogenes and other transcription factors.

Does Myb Regulate Differentiation, Proliferation or Both?

Reversible regulation of differentiation by v-Myb

In transformation assays, normal hematopoietic cells form small colonies in semi-solid medium, but eventually stop dividing and differentiate into mature myeloid, erythroid or lymphoid cells. When transformed by oncogenic v-Myb, the cells become fixed at an immature stage of differentiation and continue proliferating. As a result, they form large colonies of immature dividing “blast-like” cells that can be isolated and expanded for weeks in tissue culture. However,

transformation by v-Myb does not constitute a block in differentiation, but an acquired phenotype induced by the oncogenic transcription factor. This is evident from studies performed using a mutant v-Myb that is temperature-sensitive for DNA binding and for transformation^{25,156} and which transforms cells reversibly. Thus, the ts (temperature sensitive)-v-Myb transformed cells remain immature until shifted to the non-permissive temperature, which inactivates v-Myb and allows the cells to differentiate and stop proliferating. If shifted back to the lower temperature the cells de-differentiate, become immature again and resume proliferating¹⁵⁶. These results suggest that the active v-Myb induces a change in the transcription profile, which leads to proliferation and the adoption of an immature phenotype. When v-Myb is inactivated the cells resume their pre-programmed differentiation pathway and become mature, non-proliferating cells. Thus, v-Myb affects both differentiation and proliferation simultaneously. In contrast, cells transformed by the v-Myc oncogene differentiate almost completely, but retain their ability to proliferate, forming large colonies of mature monoblast-like cells^{157,158}. The v-Myc oncogene stimulates proliferation without blocking or affecting differentiation. If v-Myc is introduced into cells that are transformed by the ts-v-Myb oncogene, the doubly-transformed cells become temperature-dependent for differentiation, since the ts-v-Myb controls their differentiation state, but continue proliferating at both temperatures, due to the activity of v-Myc¹⁵⁸. This highlights the differences between v-Myb and v-Myc, but also shows the unique nature of v-Myb and its complex functions, which affect both differentiation and proliferation.

Myb target genes and transforming activities are lineage- and differentiation-specific

The c-Myb protein is expressed in many cell types, including all proliferating hematopoietic cells and most types of proliferating epithelial cells ^{7,8,54,159}.

However, evidence of transformation by the v-Myb proteins is restricted to hematopoietic ¹⁵⁷ and some neural crest cells ⁴⁴. Since Myb proteins are transcription factors, the oncogenic activity is presumably linked to the activation or repression of specific target genes. However, Myb proteins must cooperate with other transcription factors in order to regulate cell type-specific genes. The best example of this is activation of the *mim-1* gene, the first identified target gene for c-Myb ²⁶, which is activated by the combinatorial effects of c-Myb plus C/EBPbeta (also called NF-M) that binds adjacent to c-Myb on the *mim-1* promoter ^{160,161}. Thus, c-Myb can only activate the *mim-1* gene in cells that already express C/EBPbeta. However, co-transfection of c-Myb plus C/EBPbeta into other cell types, such as fibroblasts, can lead to activation of the endogenous *mim-1* gene ¹⁶¹.

Interestingly, the two different v-Myb oncogenes encoded by the AMV and E26 leukemia viruses transform related, but quite distinct types of cells. Both types of transformed cells are immature myeloid cells and require the cytokine cMGF for growth and survival ¹⁶² although AMV-transformed cells produce some cMGF and are therefore less cytokine-dependent, due to autocrine stimulation ¹²¹. Only the E26 transformed cells express the *mim-1* target gene, which is

activated by the E26 form of v-Myb and by c-Myb, but not by AMV v-Myb. The latter protein has three point mutations in its DNA binding domain that disrupt a protein interaction surface and prevent it from activating the *mim-1* gene²⁷. As shown in Figure 1B, the two v-Myb proteins have quite different structures. The AMV protein is truncated at both ends compared to c-Myb, and has 11 point mutations that contribute to its oncogenic potential^{82,121} and affect which target genes it regulates^{36,78}. The E26 protein is fused to the retroviral Gag protein at the N-terminal end and a portion of another transcription factor, Ets1, at the C-terminal end. The E26 virus transforms more immature Myeloid Erythroid Progenitor (MEP) cells¹⁶³. Interestingly, the ability to transform these immature cells in vitro requires both the Myb and Ets proteins, but still works when they are expressed *in trans* as separate proteins¹⁶⁴. However, the induction of leukemias in animals requires that the proteins be fused¹⁶⁵, suggesting that the fusion protein has unique activities, required for leukemia induction, that are not provided by the separate transcription factors. Indeed, there are even specific target genes that are only activated by the Myb-Ets fusion protein, and not by the two proteins expressed separately¹⁶⁶.

If transformation by v-Myb or mutants of c-Myb requires the activation of specific target genes, then combinatorial effects could explain why v-Myb only transforms hematopoietic cells and not other cell types. A number of c-Myb target genes have been identified that may be linked to transformation or leukemogenesis, including c-kit^{43, 45, 167, 168}, CD34+^{169, 170}, GATA-1¹⁶⁸ and Flt-3^{168,171}. All of these genes are expressed in hematopoietic stem/progenitor cells

but not other cell types and could explain the restriction of v-Myb oncogenic activity to hematopoietic cells. The ability of Myb proteins to activate these genes likely depends on combinatorial interactions with other transcription factors, cofactors, or accessory proteins that influence Myb protein specificity and the genes that are regulated in each cell type ¹⁷². This again highlights the importance of protein-protein interactions in gene regulation by Myb proteins. Mutations in c-Myb that disrupt interactions with C/EBPbeta affect the ability to activate the *mim-1* gene ²⁷ and other target genes ³⁶, suggesting that protein-protein interactions are important for Myb protein specificity and for transcriptional activity. This opens up new possibilities for identifying possible drugs targeting specific interactions between Myb proteins and other cooperating factors linked to transformation, that might inhibit the activities of “bad” oncogenic Myb proteins in transformed cells, while leaving “good” c-Myb functional in essential normal cells.

Myb is a key regulator of stem cell fate

Several types of evidence indicate that c-Myb is a key regulator of proliferation and differentiation in stem cells, and that it plays a central role in the determination of cell fate. In the hematopoietic system, the earliest evidence for an important role of c-Myb came from experiments using antisense oligonucleotides that showed the importance of c-Myb in normal hematopoiesis ^{21,173} and from the mouse knockouts showing that c-*myb* gene expression is required for definitive hematopoiesis in mammals ²². More recent studies have

used conditional knockout systems to demonstrate that c-Myb is also critical for adult hematopoietic stem cells ¹⁷⁴. These studies showed that c-Myb affects all the hematopoietic lineages, suggesting either that it functioned at the earliest stem cell stage or was required in all the lineage-specific progenitors, or both. Another type of early evidence came from transformation studies, which showed that even single amino acid changes in v-Myb led to the induction of completely different phenotypes in the transformed cells ²⁷, suggesting that the oncogenic Myb protein was playing a decisive role by shifting or determining the extent and direction of differentiation along different pathways.

The importance of c-Myb in hematopoietic stem cells has been demonstrated in several different systems. The *c-myb* mRNA is highly expressed in pluripotent hematopoietic stem cells ¹⁷⁵ and the products of c-Myb target genes c-Kit ¹⁷⁶ and CD34 ^{169,170,177} are used as cell surface antigens to isolate hematopoietic stem cells and immature progenitors. In mice, two different mutagenesis studies designed to find genes involved in stem cell regulation identified mutations in *c-myb* ^{178,179}, suggesting that even minor changes in the c-Myb protein can have dramatic effects on the balance between differentiation and proliferation in stem cells. However, the absolute levels of c-Myb are also important, since mice engineered to express low levels of c-Myb displayed myeloproliferative defects resembling stem cell disorders ⁴². Very similar studies have been performed using the zebrafish model, where the *c-myb* gene is a key marker of hematopoietic stem cells ¹⁸⁰ and imaging studies have been used to follow the development of *c-myb* positive hematopoietic stem cells ¹⁸¹. Thus, the

importance of c-Myb expression for hematopoietic stem cell proliferation and differentiation has been conserved throughout vertebrate evolution.

Although the role of c-Myb in hematopoietic stem cells has been best documented, several studies have shown that c-Myb plays an important role in stem cells from other tissues, including epithelial stem cells in colonic crypts⁵⁶ and neural stem cells in adult brain⁵⁷. Recent progress in the stem cell field has led to procedures for generating induced Pluripotent Stem (iPS) cells from adult fibroblasts, keratinocytes and other tissues^{182,183}. One of the important inducer genes used for the generation of iPS cells, c-Myc, is a known target gene for c-Myb⁵³ and bioinformatics approaches have identified c-Myb as a likely regulator of pluripotency and “stemness” in iPS cells¹⁸⁴, but so far no direct tests of the role of c-Myb, or the effects of oncogenic variants such as v-Myb, on the formation of iPS cells have been reported.

Myb target genes may change during differentiation and the cell cycle

Several c-Myb target genes important for stem cells or early progenitors have been identified, including c-Kit¹⁷⁶ and CD34^{169,170,177}, suggesting that c-Myb is required for the expression of genes that are important in stem cells and the earliest progenitors. However, c-Myb is also required for the expression of *mim-1*²⁶, myeloperoxidase¹⁸⁵ and other genes that are expressed in mature cells¹⁸⁶. Ironically, the mature cells express little or no c-Myb, which is primarily expressed in the immature proliferating cells. This suggests that c-Myb may be

involved in the initial activation of genes that are kept active by other transcription factors in the mature cells. Although the mechanisms remain poorly defined, it is clear that c-Myb regulates different genes in the immature and more mature cells, another example of its context-specific regulation activity. As described in a previous section, c-Myb regulates genes in combinations with other transcription factors, and its targets change as the presence or absence of the other factors change during differentiation.

Recently, c-Myb was shown to regulate the expression of Cyclin B1, a cell cycle regulatory gene involved in G2/M transition ⁵². These results are important because they suggest a mechanism for c-Myb regulation of the cell cycle and proliferation. The c-Myb protein has also been shown to form complexes with Cyclin D1 ⁸³ and with the cyclin-dependent kinases CDK4 and CDK6 ¹⁸⁷, which also affect c-Myb transcriptional activity via a mechanism that involves p27Kip1 ¹⁸⁷. These results raise the possibility that c-Myb activity and specificity could be regulated during the cell cycle. For example, c-Myb might regulate different sets of target genes in different cell cycle compartments. Cell cycle regulation could be controlled by phosphorylation or other post-translational modifications, or by the presence or absence of specific co-regulators, such as E2F or Rb. There is ample evidence that the related transcription factor B-Myb is regulated by association with Rb and other cell cycle-regulated proteins ^{133,188} and that B-Myb transcriptional activity can be regulated by Cyclin-dependent kinase phosphorylation ^{189,190}. However, direct evidence that c-Myb activity is regulated during the cell cycle is still unavailable.

Mechanisms Affecting c-Myb Activity and Specificity

Activation of the *c-myb* gene in tumors

The control of *c-myb* transcription is complex and can be affected by several different mechanisms. Overexpression or rearrangement of the *c-myb* gene has been reported in patients with head and neck carcinomas, breast cancer, melanoma and leukemias^{14-16,191}, implicating activated alleles of *c-myb* in the development of human tumors. However, the *c-myb* gene promoter is GC rich, resembling a constitutively expressed promoter from a housekeeping gene¹⁹², suggesting that expression of the gene is controlled through unconventional mechanisms. Early studies in murine hematopoietic cells demonstrated a direct correlation between the relative abundance of *c-myb* mRNA and the level of transcriptional arrest in intron 1 (Figure 1A)^{29,193}. Attenuation of transcription was linked to the rapid down-regulation of *c-myb* mRNA that occurs during the induced differentiation of human colorectal cancer (CRC) cell lines⁹ and sequence analysis of primary CRC tumor samples revealed frequent mutations in a poly-T tract disrupting a proposed stem-loop motif in intron 1. The mutations reduced transcriptional attenuation and allowed greater gene expression, suggesting that disruption of the attenuation mechanism plays an important role in the elevated levels of *c-myb* mRNA in CRCs¹⁰. In human breast tumors, *c-myb* was found to be over-expressed in estrogen receptor positive samples and run-on transcription assays showed that *c-myb* transcriptional elongation was

directly regulated by estrogen/ER signaling ⁵. These results show that attenuation or control of elongation is an important mechanism for regulating the levels of *c-myb* in normal and tumor cells.

Elevated levels of *c-myb* mRNA can also be caused by genomic alterations, such as reciprocal translocations or genomic amplification. In human T-cell acute leukemia (T-ALL), recurrent chromosomal translocations t(6;7)(q23;q34) involving the TCRB and *c-myb* loci were identified (Figure 1B) ¹⁶. These translocations were reciprocal and balanced, and led to the juxtaposition of the *c-myb* proto-oncogene near the TCRB regulatory sequence, which suggested deregulated expression. Chromosomal translocations involving the *c-myb* gene were also reported in other types of human tumors: the t(6;9)(q22-23;p23-24) translocation in adenoid cystic carcinomas (ACC) of the breast and head and neck consistently resulted in fused MYB-NFIB transcripts predominantly consisting of MYB exon 14 linked to the last coding exon of NFIB (Figure 1B) ⁶. This leads to loss of *c-myb* exon 15, which encodes the 3'-UTR, where several highly conserved target sites for microRNAs (miR15a/16 and miR-150) are located (Figure 1B). Thus, the translocation appears to deregulate *c-myb* by removing the microRNA binding sites.

The human *c-myb* locus was recently found to be flanked by 257-bp Alu repeats that can mediate *c-myb* gene tandem duplication by homologous recombination between related elements on sister chromatids ¹⁷. This Alu-mediated *c-myb* tandem duplication could be one of the reasons for genomic duplication of *c-myb* gene frequently reported in human T-ALL ^{16,18}, MYST3-

linked AMLs ¹⁹⁴, BRCA1-mutated breast cancer ⁴ and some other types of cancer. Intriguingly, comparative studies on healthy individuals and patients with T-ALL showed that this tandem duplication also occurs spontaneously during normal thymocyte development and is clonally selected during the molecular pathogenesis of human T-ALL. So rearrangement or duplication of the *c-myb* gene may occur frequently, but is apparently not sufficient to induce leukemogenesis or transformation without other cooperating events.

Alternative RNA splicing as a novel mechanism for unleashing c-Myb oncogenicity

Alternative RNA splicing provides a mechanism for encoding multiple different transcripts from a single gene and represents an important molecular mechanism of gene regulation in physiological processes such as developmental programming as well as in disease. Recent studies using next-generation sequencing technologies have shown that up to 94% of human genes undergo alternative RNA splicing ¹⁹⁵, suggesting that this mechanism is important for adding to the complexity of expressed transcripts. In cancers, splicing is significantly altered. Many cancer-associated splice variants arise from genes with an established role in oncogenesis, and their functions can be oncogenic. The *c-myb* gene contains 15 classical exons and 6 alternative exons (Figure 1B). It undergoes extensive alternative splicing producing dozens of different splice variants that encode many different isoforms of c-Myb protein ¹⁹. Alternative splicing in *c-myb* RNAs has been detected in both normal and tumor cells and in

several species¹⁹⁶⁻²⁰⁰. Recent studies revealed an array of *c-myb* variant transcripts, expressed in highly regulated, lineage-specific patterns, that are formed through the use of alternative exons 8A, 9A, 9B, 10A, 13A and 14A (Figure 3A)¹⁹. All of these identified splice variants encode Myb proteins sharing the same DNA binding domain but having unique transcriptional activation and C-terminal domains. Several lines of evidence have shown that relatively minor changes in the transcriptional activation domain and/or negative regulatory domain of c-Myb can dramatically affect its transcriptional activity, raising the possibility that the variant c-Myb proteins encoded by these splice variants may display unique transcriptional activities. Indeed, the variant c-Myb proteins encoded as a result of alternative RNA splicing have quantitative and qualitative differences in their transcriptional activities¹⁹.

Perhaps the best characterized of the alternatively spliced *c-myb* transcripts includes alternative exon 9B (previously named 9A) that encodes the p89 variant of c-Myb with a 121-amino-acid in-frame insertion in the transcriptional activation domain (Figure 3A)²⁰¹⁻²⁰³. Overexpression of this alternatively spliced isoform in avian hematopoietic cells resulted in increased transactivation and transformation activity²⁰⁴. The murine homologue of this variant enhanced cell survival in hematopoietic cell assays in which wild-type c-Myb accelerated cell death²⁰⁵. The 9B variant is conserved throughout vertebrates, suggesting that it plays an important role. However, a targeted knockout of the 9B/p89 variant had no apparent phenotype²⁰⁶, so it is possible that the variant has specialized

functions or that the loss of the variant might be compensated for by other forms of c-Myb.

Recent studies showed that the patterns and relative levels of *c-myb* splice variants changed dramatically in a lineage-specific manner during the in vitro differentiation of CD34+ cells ¹⁹, suggesting that the alternative splicing of *c-myb* transcripts is highly regulated in hematopoietic cell differentiation and that the ratios or the relative levels of alternatively spliced c-Myb variants play a role in lineage-specific functions of hematopoietic cells. Some alternative exons contain stop codons or cause a translation frame-shift, so the encoded variant proteins have C-terminal deletions, similar to the oncogenic v-Myb (Figure 3A) ¹⁹. Since the alternative splicing of *c-myb* transcripts produces a family of transcription factors, it is possible that aberrant or increased alternative splicing in tumors could lead to the production of truncated variants of c-Myb with transforming or leukemogenic activities. This would represent a novel mechanism of activation of an oncogene via alternative RNA splicing. Unfortunately, there is very little information about how alternative splicing is regulated, although some splicing factors have been shown to have oncogene like activities when over-expressed

207 .

Regulation of *c-myb* expression through microRNAs

MicroRNAs (miRs) are a large and growing class of ~22 nucleotide-long non-coding RNAs which function as repressors, negatively regulating the

expression of their target genes, in all known animal and plant genomes^{208 209}. In many cases, miRNAs bind to the 3'-UTRs of the mRNAs they regulate²⁰⁹. The *c-myb* gene encodes an mRNA with an extensive 3'-UTR that contains binding sites for several different microRNAs, and there are several examples of microRNAs that appear to regulate the turnover or translation of *c-myb* mRNA. For example, the microRNA miR-150, which binds the *c-myb* 3'-UTR, was found to be down regulated in granulocytes that over-expressed *c-myb* and that were isolated from patients with primary myelofibrosis (PMF)²¹⁰. The implication is that aberrant expression of miR-150 in the PMF patients led to over-expression of *c-myb* mRNA and subsequent defects in granulocyte differentiation. Similar studies have uncovered an important control circuit in which miR-150 plays a critical role in the regulation of *c-myb* during the development of B-cells³³ and other hematopoietic lineages^{32,211}. The link between miR-150 and *c-myb* has even been conserved in zebrafish³¹. In addition to miR-150, the *c-myb* 3'UTR also contains binding sites for several other microRNAs^{30,34,212}. These results indicate that deregulation of c-Myb expression by aberrant expression of microRNAs could be another mechanism of activating c-Myb in tumors.

Like the c-Myb protein, which is regulated by many different mechanisms, the *c-myb* gene is controlled at all possible levels, from regulation of the promoter, to control of transcriptional elongation, to alternative RNA splicing, to microRNAs that control the fate of the mRNAs. Each of these is a potential regulatory circuit that can respond to various extracellular cues or activated signaling pathways. Thus, the regulation of *c-myb* expression is a complicated

and rich area of study, and offers many opportunities for potential intervention in the design of novel therapeutic approaches.

Perspective

One of the most surprising and provocative results that has come out of studies of c-Myb and v-Myb is the finding that the activities of Myb proteins are variable and context-dependent and controlled by their interactions with other proteins. Thus, c-Myb regulates different target genes in different cell types, in immature vs. mature cells and perhaps even during different parts of the cell cycle. On the one hand, this realization seems obvious, since c-Myb is expressed in so many different cell types. However, there were expectations that c-Myb would be a “master regulator” of hematopoiesis or that it would have one or a few key targets that would explain its ability to become an oncogenic transforming protein. On the contrary, c-Myb has turned out to regulate hundreds of target genes, and to regulate different sets of genes in different cell types^{36,37,78,186}. The oncogenic activity of v-Myb may be linked to its ability to regulate different genes than c-Myb, or to do so at different times or in different circumstances than c-Myb. And oncogenic activity may be a combined effect of activating dozens or hundreds of genes, rather than one key target. The other big surprise about c-Myb is the large number of proteins it interacts with, and the large number of post-translational modifications that affect it. Dozens of Myb-binding proteins have been identified, including cell cycle regulators, transcription factors, transcriptional co-activators and modifying enzymes. Although some of these

interactions have been shown to affect the activity of c-Myb, it is not at all clear how all these potential interactions relate to one another. For example, do these interactions happen simultaneously in the same cell? Do they happen on the same c-Myb protein or do they define subpopulations of c-Myb with distinct protein partners and activities? It seems clear that the study of c-Myb needs to progress from the discovery phase and into a more systematic study of how the protein is regulated in a time- and context-dependent manner.

Despite more than 25 years of research, the question of what makes v-Myb a transforming protein remains unanswered. This is the central question, and the reason that studying c-Myb is significant and unique. We have learned that the v-Myb protein is lacking the C-terminal domains that cause c-Myb to be rapidly degraded and that promote its intramolecular auto-inhibition, that the v-Myb mutations affect protein-protein interactions and that v-Myb and c-Myb regulate different sets of genes, but we do not understand why v-Myb is an oncogene, what genes it regulates in transformed or leukemic cells or how those genes contribute to oncogenesis. Answering these questions may require using new technologies, like genome-wide chromatin immunoprecipitation assays, in order to identify all the v-Myb targets in transformed cells. A related question is whether mutants or splice variants of c-Myb that are expressed in tumors can mimic the activity of v-Myb, which appears to be so different from that of c-Myb. It will be necessary to understand the transforming activity of v-Myb in order to answer these questions about c-Myb and derivatives of it.

Although the transforming potential of c-Myb is still not understood, it has become clear that the regulation of c-Myb is extremely complex and involves many different pathways and regulators. The regulation of c-Myb activity involves changes in *c-myb* gene expression, changes in alternative splicing, regulation by microRNAs and then a myriad of post-translational modifications and protein interactions. It seems that c-Myb plays the role of integrator, at the intersection of many different signaling and regulatory pathways. With so many variables affecting its expression and activity, it should be no surprise that the activities of c-Myb are also complex. Hopefully the application of new technologies will illuminate the role of Myb proteins in transformation and oncogenesis, provide an understanding of how post-translational modifications and protein-protein interactions control c-Myb and help explain its important roles in determining cell fate and controlling differentiation and proliferation.

Acknowledgements

This work was supported by USPHS/NIH grants CA058443 and CA105257 (to SAN). The authors thank A.M. Quintana and J.P. O'Rourke for helpful discussions and comments on the manuscript.

Figure Legends

Figure 1. Myb genomic and protein structures.

(A) A diagram of the c-Myb protein structure shows the major functional domains including the highly conserved DNA binding domain (DBD) near the N-terminus, transcriptional trans-activation domain in the middle, and the C-terminal negative regulation domain. Shading indicates the most highly conserved portions of the protein and known sites of post-translational modifications are marked with asterisks. Structural details are labeled below the diagram. The DNA binding domain is comprised of Myb repeats R1, R2 and R3 and is also the binding site for a number of proteins including Ets-1, Cyp40, p100, c-Ski, N-CoR and mSin3A, as indicated. The central transactivation domain (TAD) is the interaction site for p300/CBP. The negative regulatory domain (NRD) extends from the LZ/FAETL region to the EVES motif and includes the binding sites for p160, Pin1 and TIF1beta. Arrows indicate potential intramolecular interactions. The lower section shows the structures of the AMV v-Myb and E26 Gag-Myb-Ets proteins. Compared to c-Myb, both have N- and C-terminal truncations. AMV v-Myb also has 11 point mutations (empty circles) that change its activity. (B) The *c-myb* gene contains 15 core exons (black boxes above the line) and 6 alternative exons (under the line). Important features of the gene are labeled (from left to right): The two promoters (bent arrows) upstream of exons 1 or 2, the translation attenuation site in intron 1 (star), a frequent site of chromosomal translocation in tumors (//) and the 3' untranslated region encoded by exon 15 containing microRNA binding sites.

Figure 2. Wnt signaling affects c-Myb stability and activity.

c-Myb protein is phosphorylated and degraded by Wnt-1 signal via the pathway involving TAK1, HIPK2 and NLK. Wnt-1 causes the nuclear entry of TAK1, which then activates HIPK2 and NLK. NLK directly binds to c-Myb DNA binding domain with HIPK2, resulting in the phosphorylation of c-Myb at multiple sites (indicated by P), which enhances the interaction of c-Myb with Fbw7, a ubiquitin ligase, and as well as the interaction with the SCF complex, leading to ubiquitinylation of c-Myb and degradation by the proteasome. Mutations in the DNA binding domain of v-Myb decrease the affinity for HIPK2, and the C-terminal truncation removes NLK phosphorylation and ubiquitinylation sites, helping v-Myb to evade this regulatory mechanism.

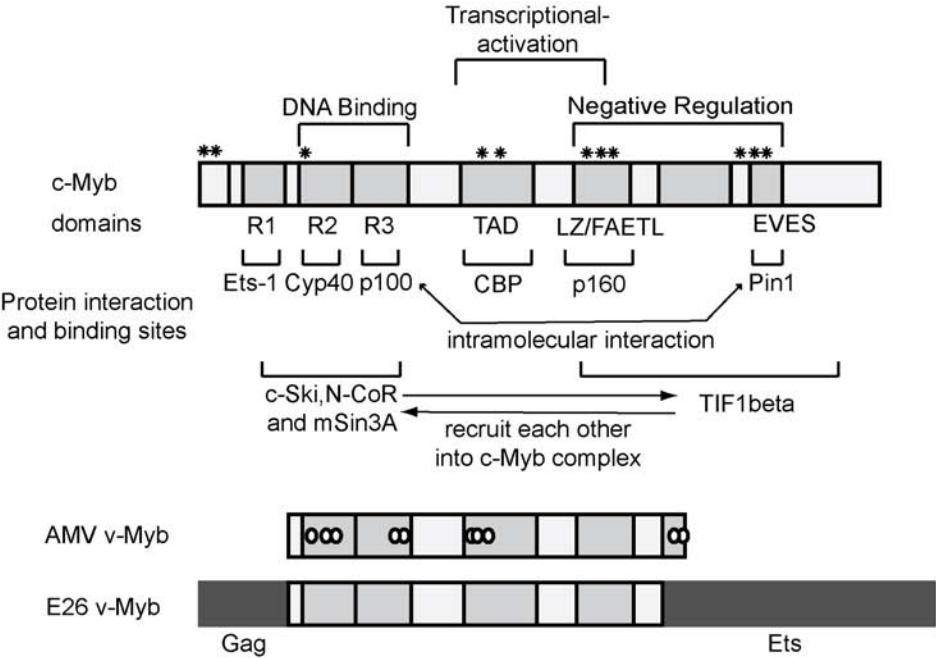
Figure 3. Alternative splicing as a novel means of activating c-Myb.

(A) The diagrams depict the protein structures of wild type c-Myb, alternatively spliced isoforms 9B, 8A, 9A, 10 and 9S/10 and AMV v-Myb. The DNA binding domain (DBD), transactivation domain (TAD) and negative regulatory domain (NRD) are labeled at the top. The numbers on the right indicate amino acid residues from c-Myb plus novel amino acids encoded by the alternative exons. (B) Different versions of Myb protein have different protein structures, permitting them to interact and cooperate with distinctive sets of

transcription factors or cofactors to form stable transcription complexes at different promoters.

Figure 1

A



B

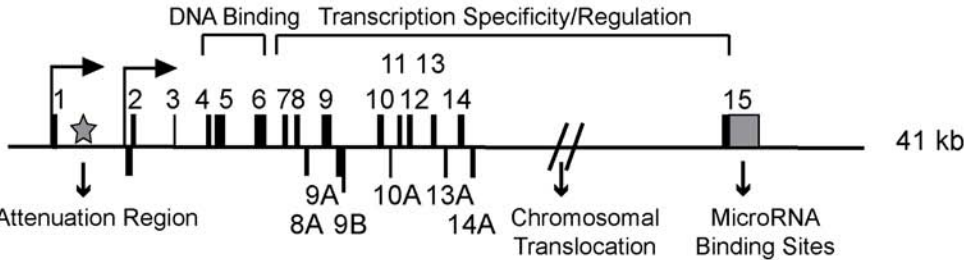


Figure 2

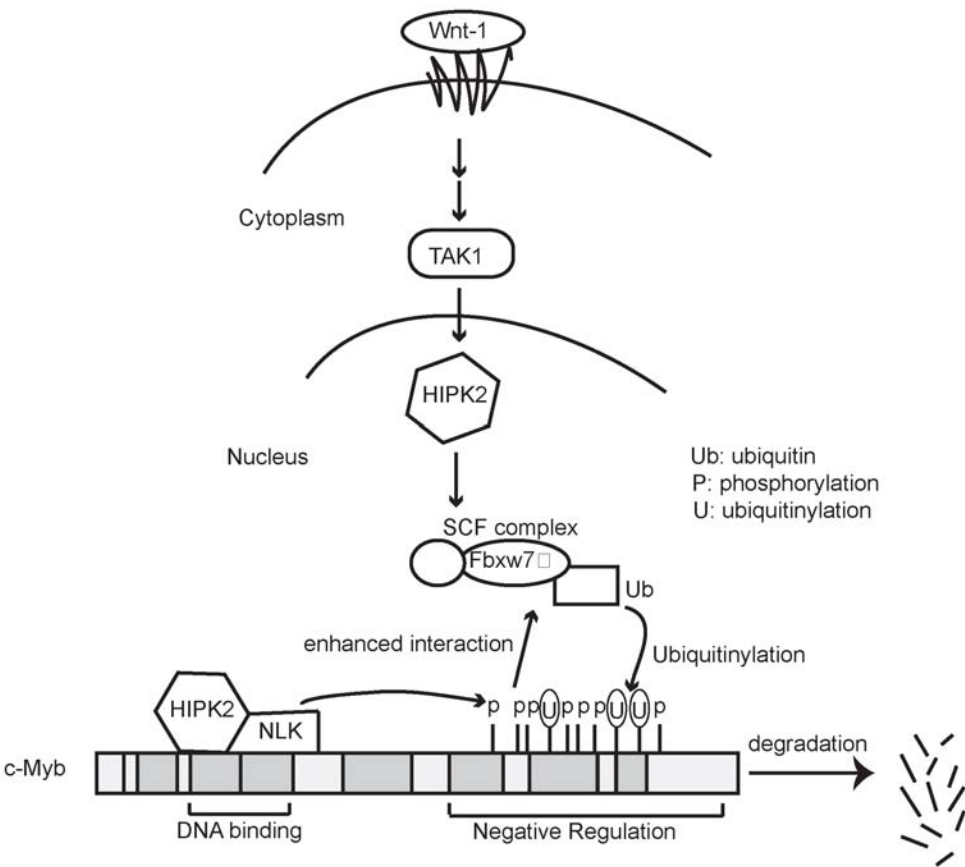
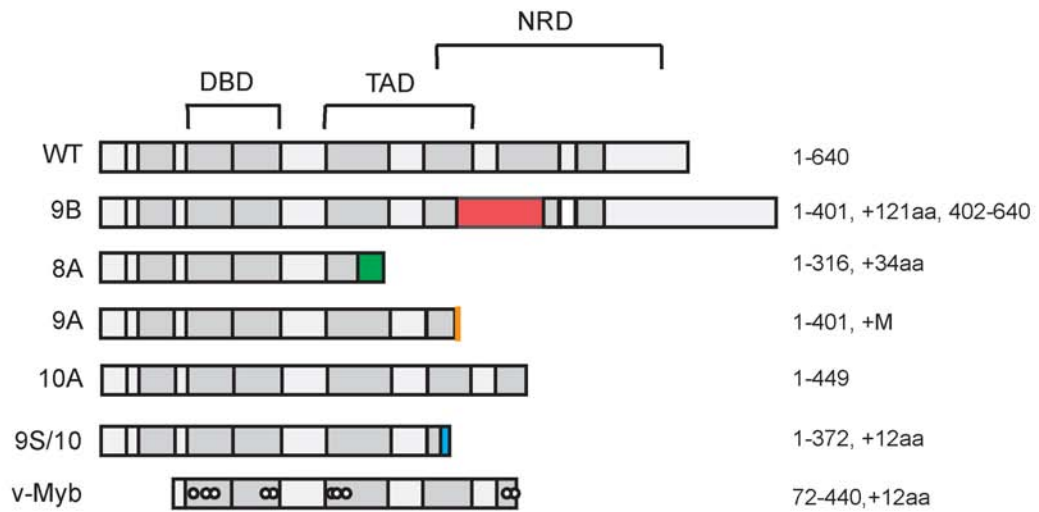
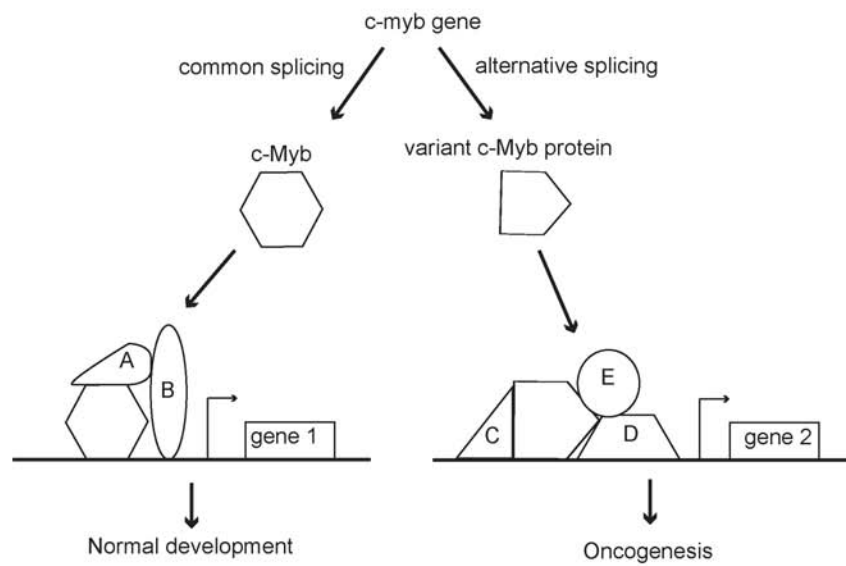


Figure 3

A



B



Chapter 2: Aberrant *c-myb* alternative splicing in leukemias

Alterations of the C-terminal domains, through post-translational modifications, mutations or deletions, have been implicated by many studies to impact the activities of c-Myb. Such mutations can direct c-Myb to different promoters and transform it to a potent oncoprotein. As discussed, the *c-myb* gene undergoes extensive alternative splicing, leading to the expression of variant forms of c-Myb protein. The variant c-Myb proteins contain the same DNA binding domain, but have changes in the C-terminal domains. With the identification of additional c-Myb target genes and the availability of assays that can detect differences in the genes targeted by different versions of c-Myb, it has become clear that many of these variant Myb proteins have distinct transcriptional activities and activate different sets of genes than the “wild type” c-Myb. Alternative splicing of *c-myb* RNAs appears to be tightly regulated in hematopoietic cell differentiation. It is plausible that alternative RNA splicing provides a mechanism for producing specific variants of c-Myb with unique functions in hematopoiesis. By shifting the pattern of *c-myb* alternative splicing, hematopoietic progenitor cells could produce a different repertoire of c-Myb proteins and express a different subset of c-Myb target genes than more differentiated cells. Thus, if the alternative splicing process skews, oncogenic forms of c-Myb could be produced, which could potentially affect the normal hematopoiesis and induce leukemias.

We hypothesized that aberrant *c-myb* alternative splicing could correlate with the formation or the outcome of leukemias. We developed a polony assay to characterize *c-myb* alternative splicing in several cell types, including a small cohort of leukemia samples. We also tested whether some of the *c-myb* splice variants that are generally expressed more in leukemias correlate with the activation of the genes that are potentially targeted by Myb proteins and play important roles in leukemias. Our data demonstrate that the levels of *c-myb* splice variants are more prevalent in leukemia patients and imply that some splice variants that encode Myb proteins with truncated C-terminal domains could be oncogenic and contribute to leukemogenesis.

Author Contribution

The following paper discusses our observations in a small cohort of pediatric pre-B-ALL with 4 independent authors Ye E. Zhou (YEZ), John P' O'Rourke (JPO), Jeremy S. Edwards (JSE) and Scott A. Ness (SAN). YEZ performed the research. JPO provided the preliminary data on *c-myb* splice variants in different cell types detected by qRT-PCR. JSE provided expertise and materials for the polony assay. SAN mentored YEZ, provided funding and supervised all the research. SAN also approved the manuscript for publication.

Single Molecule Analysis of *c-myb* Alternative Splicing Reveals Novel Biomarkers in Precursor B-ALL

Authors:

Ye E. Zhou, John P. O'Rourke, Jeremy S. Edwards and Scott A. Ness

Affiliation:

Department of Molecular Genetics and Microbiology
University of New Mexico Health Sciences Center
Albuquerque, New Mexico 87131-0001 USA

Corresponding Author:

Dr. Scott A. Ness
Department of Molecular Genetics and Microbiology, MSC08 4660
1 University of New Mexico, Albuquerque, NM 87131-0001 USA
Tel: (505) 272-9883
Email: ness@unm.edu

Abstract

The c-Myb transcription factor, a key regulator of hematopoiesis, has an N-terminal DNA binding domain and a large C-terminal domain responsible for transcriptional activation, negative regulation and protein-protein interactions that control its target gene specificity. Overexpression and rearrangement of the *c-myb* gene has been reported in patients with leukemias and other types of cancers, implicating activated alleles of *c-myb* in the development of human tumors. Alternative RNA splicing-produced variant c-Myb proteins with different C-terminal domains and unique transcriptional activities have been detected in both normal and tumor cells and in several species; however, no link was established between the aberrant splicing of *c-myb* and leukemias. Here, by performing a detailed, single molecule-based polony assay, we monitored *c-myb* splicing variants in both leukemia patient samples and normal donors. We found that *c-myb* alternative RNA splicing was elevated and much more complex in a small cohort of pediatric precursor B-ALL samples than in CD34+ hematopoietic progenitor cells from normal donors. The results revealed that leukemia samples express more than 60 different *c-myb* splice variants, most of which have multiple alternative splicing events and were not detectable by conventional microarray or PCR approaches. For example, the single molecule assay detected 21 and 22 splice variants containing the 9B and 9S exons, respectively, most of which encoded unexpected variant forms of c-Myb protein. Furthermore, the detailed analysis identified some splice variants whose expression correlated with poor survival in a small cohort of precursor B-ALL samples. Our findings

indicate that single molecule assays can reveal complexities in *c-myb* alternative splicing that have potential as novel biomarkers and could help explain the role of c-Myb variants in the development of human leukemia.

Introduction

The *c-myb* gene encodes a transcription factor that is absolutely required for normal hematopoiesis²² including the differentiation of stem cells into committed progenitors¹⁷⁹, the normal development of myeloid and erythroid lineages^{22,38} and for B-cell and T-cell differentiation^{40,213}. The c-Myb protein has a highly conserved DNA binding domain near its N-terminus and a large C-terminal domain required for transcriptional activation as well as negative regulation^{74,78,90}. In addition, mutations or modifications of c-Myb can change its activity, converting the normal regulator into a potent oncoprotein that transforms immature myeloid cells in tissue culture and induces leukemias in animals (for review see²¹⁴). Activated alleles of the *c-myb* gene have been identified in numerous human malignancies, including several types of leukemias as well as colon, breast and head and neck tumors^{6,16,18}. Thus, c-Myb activities are conflicting, affecting both differentiation and proliferation and from normal regulator to transforming protein. These duplicities may be due to the fact that relatively minor changes can completely change the activity and specificity of c-Myb. For example, individual point mutations change its ability to regulate specific target genes^{26,27}. Furthermore, microarray studies have shown that c-Myb regulates totally different sets of target genes than v-Myb, a truncated, mutated and oncogenic allele of c-Myb encoded by Avian Myeloblastosis Virus³⁶. Most of the differences in the activity of v-Myb are due to changes in the large C-terminal domain, which appears to control target gene specificity by affecting protein-protein interactions⁷⁸. The implication is that the activity of c-Myb is

highly variable and subject to change through mechanisms that alter its C-terminal domain, such as post-translational modifications, mutations or deletions. Such modifications can target c-Myb to different promoters and convert it from a normal regulator into a potent transforming protein.

One of the mechanisms affecting the C-terminal domains of c-Myb is alternative RNA splicing. The human *c-myb* gene is located on chromosome 6q22-q23 and spans 15 exons plus 6 alternative exons (Figure. 1A). Interestingly, all of the alternative exons are in the region of the gene encoding the C-terminal domains that affect target gene specificity. Alternatively spliced *c-myb* transcripts have been detected in both humans and animals and in numerous tissues^{19,200,204,215,216} and some *c-myb* splice variants with altered C-terminal domains have increased transforming activities^{204,205,217}. Our own studies showed that alternative splicing of *c-myb* is tightly regulated during the differentiation of primary human hematopoietic progenitor cells and that levels of alternatively spliced *c-myb* transcripts are elevated in leukemia samples compared to normal bone marrow cells¹⁹, suggesting that c-myb alternative splicing could be a whole new mechanism of unmasking its oncogenicity and play a role in leukemogenesis.

These initial observations prompted us to further investigate *c-myb* alternative splicing profiles in leukemia patient samples and to evaluate its clinical relevance in terms of prognosis. The conventional way to monitor alternative splicing variants is by using real-time PCR, microarray, or short-gun sequencing. However, any exons combinatorial information regarding individual

c-myb transcript is lost in these methods. In order to circumvent this problem, we applied a polony assay that allows for quantitative profiling complex alternative splicing at the single-molecule level ²¹⁸ to analyze the *c-myb* splicing patterns in 13 pediatric pre-B ALL patient samples and correlate the expression levels of the splice variants with the survival of these leukemia patients. We also tested whether c-Myb splice variants that are more highly expressed in leukemias are able to associate with the level of some oncogenes that are normally overexpressed in leukemia. Our results showed that alternative splicing of *c-myb* RNAs is complex and that leukemias produce elevated levels of c-Myb splice variants that can be potentially associated with clinical outcome, implicating deregulation of c-myb splicing could contribute to leukemogenesis.

Results

Alternative splicing of *c-myb* transcripts is complex and combinatorial

The transforming activity of c-Myb has often been linked to deletions or mutations in its C-terminal domains ²¹⁹, which can lead to the regulation of different sets of target genes ³⁶. Our previous studies showed that alternative splicing of c-myb transcripts is highly regulated in hematopoietic cell differentiation and in leukemias, and could lead to the production of variant c-Myb proteins with different C-terminal domains ¹⁹. We set out to determine whether the shift in the patterns of *c-myb* alternative splicing could be a novel mechanism for activating c-Myb oncogenicity in leukemias. However, we were faced with a

complication due to the complexity and combinatorial nature of *c-myb* alternative RNA splicing. As shown in Figure 1A, the *c-myb* gene spans 41 kb on chromosome 6q and contains 15 exons that compose the wild type transcript, plus 6 alternative exons (8A, 9A, 9B, 10A, 13A and 14A). The complexity is increased further by exons that have multiple splice donor sites. For example, as shown in Figure 1B, exon 9 can also be spliced as an 85 nt shorter “9S” version. The normal exon 9 and the short 9S form can be joined to exons 9A, 9B or 10 (labeled a, b, c and d, e, f, respectively), so just this small portion of the gene produces 6 different splice variants. All six variants encode different versions of the c-Myb protein, with identical N-terminal DNA binding domains (shaded black in Figure 1C), but different C-terminal domains that affect transcriptional activity and that control which target genes get regulated^{19,36,37,78}. Analysis of chicken and mouse bone marrow has shown that the use of the 9S splice site is conserved in other vertebrate species, suggesting that it is an integral and important part of the *c-myb* gene (Supporting information Figure S1). For simplicity, we designate the 9S exon spliced to exon 10 as the 9S/10 splice variant, and so on.

Our previous studies¹⁹ had also identified a large number of transcripts containing more than one alternative splice event, e.g. a transcript including exon 8A that also has exon 9S spliced to exon 10. These combinatorial transcripts are problematic, because the usual methods of measuring alternative splicing do not assess them correctly. For example, a real-time PCR (QPCR) assay measuring the 9S/10 splice junction would not distinguish whether exon 8A was also present

in the same transcript, but the presence or absence of exon 8A produces markedly different protein products (Figure 1D). Almost all *c-myb* alternative splicing occurs in the exon 6-11 region, which is too long to be analyzed using currently available next-generation DNA sequencing technologies. Therefore, in order to determine what c-Myb protein products were being produced in different samples, we had to find a quantitative methodology that would determine the levels of expression as well as the structures of the *c-myb* transcripts.

Detection of *c-myb* alternative splicing variants using a polony-based assay

To produce a quantitative and complete picture of *c-myb* alternative splicing, we adopted a single molecule analysis procedure utilizing polonies, or PCR colonies²¹⁸. In the polony procedure (Figure 2A), mRNA samples are converted to cDNA by reverse transcription, the mixture of individual cDNA template molecules are then seeded into and immobilized in a thin polyacrylamide gel containing primers and other reagents necessary for PCR amplification. In situ PCR is conducted directly in the gel, which prevents the PCR products from diffusing away, so that each template produces a “PCR colony”, or polony, each of which arose from the same template cDNA. The structures of the polonies are then interrogated through sequential rounds of hybridization with fluorescently labeled, exon-specific probes. To test the sensitivity of this assay on detecting a rare splice variant, we generated polonies from a region of the *c-myb* transcript spanning exons 6 to 11, where the majority of alternative splicing occurs. We used a mixture of wild type *c-myb* cDNA plus an alternatively spliced cDNA

containing exon 8A mixed in at different molar ratios (e.g. 5000:1). After in-gel amplification to produce the polonies, multiple rounds of hybridization were performed sequentially with fluorescently labeled probes specific for different exons or splice junctions (Supplemental Table S2). The microscope slide was scanned after each round of hybridization to record which polonies were positive for each probe. Since each polony was generated from a single template cDNA molecule and every polony had a fixed address on the polony slide, the sequential hybridization steps produced a series of images that were converted into an exon map for each polony (Supplemental Figure S2).

The Polony assay provides a distinct advantage over QPCR assays, which only measure one exon or one splice junction at a time. The polony approach can determine the structure of the entire cDNA molecule that served as the polony template. This is critical for the analysis of *c-myb* alternative splicing, since different splice variants produce different proteins with potentially unique activities. For example, all splice variants that include exon 8A, which includes a premature stop codon, produce the same protein, even if they also have additional alternative exons downstream (e.g. 9S, 9B or 10A). The polony assay can distinguish these complex splice variants to more accurately quantify which versions of c-Myb protein would be produced. The QPCR approach cannot distinguish a transcript that has only exon 10A from one that has both exons 8A and 10A, so it does not accurately predict which proteins will be produced.

As shown in Figure 2B, the polony assay was highly sensitive and able to detect one transcript containing alternative exon 8A in a background of more

than 5,000 wild type *c-myb* transcripts. Thus, the assay is well suited for detecting rare splice variants in a background of wild type transcripts. The linearity of the polony assay was examined by seeding different amounts of wild type *c-myb* plasmid templates. As shown in Figure 2C, the polony assay was linear over a wide range of concentrations ($r^2=0.99$). Thus, the polony approach provides a sensitive and linear methodology for analyzing the structures and relative expression levels of alternatively spliced *c-myb* transcripts, and it gives more structural information than conventional QPCR approaches.

To test the polony technique using more relevant samples, we examined the expression of *c-myb* transcripts in 2 human hematopoietic cell lines (K562 cells, Jurkat cells) and 2 types of primary cells (peripheral blood leukocytes, PBLs, and CD34+ progenitor cells) that were previously analyzed extensively for *c-myb* splice variants by QPCR and shotgun cloning and sequencing¹⁹. The four cell types expressed wild type (WT) plus 10 additional splice variants of *c-myb*, some of which have not been previously described (Table 1). The patterns of *c-myb* splice variants identified by the polony assay were complex, demonstrating the effectiveness of this assay to detect multiple splice events in a single transcript. For example, the 9B/10A splice variant, an alternatively spliced *c-myb* transcript containing both alternative exons 9B and 10A, was detected in both K562 (0.53%) and Jurkat (1.07%) leukemic cell lines, but not in two primary human hematopoietic cells. The 9S/10/10A splice variant, which includes the short (9S) version of exon 9 spliced to exon 10, plus an alternative exon 10A, was detectable in Jurkat (0.36%) and CD34+ cells (0.09%) but not in the other cell

types. Approximately 10%-15% of the *c-myb* transcripts in these samples were alternatively spliced variants and the polony assay detected levels of alternative exons 8A, 9A, 9B and 10A that were similar to the results obtained previously using either QPCR or direct shotgun sequencing techniques ¹⁹. We conclude that the polony assay provides quantitative information about *c-myb* alternative splicing that rivals or exceeds the results obtained using QPCR, and that it offers a valuable and powerful alternative for quantifying transcripts with complex structures.

Comparison of the polony and QPCR assays

The results described above demonstrated the usefulness of the polony assay for following patterns of alternative RNA splicing and suggested that the polony assay could provide different and more extensive information than QPCR. We next applied both QPCR and the polony assay to monitor *c-myb* expression and alternative splicing patterns in a small cohort (n=13) of pediatric precursor B-ALL patient samples. RNA samples from each patient were converted to cDNA, which was used for QPCR assays or used to seed the polony assays, which were interrogated by probes for various *c-myb* exons. Up to 1500 polonies per slide, or 3000 polonies in two replicate measurements, were queried for each patient sample. As shown in Figure 3A, the QPCR and polony assay results for the total amount of *c-myb* transcripts in the patient samples correlated quite well ($r^2=0.946$), suggesting that both assays were similarly efficient and were equally linear. This confirms the results described above showing that both the QPCR

and polony approaches are very sensitive and provide quantitative results, at least for measuring the total levels of *c-myb* expression.

However, the differences between the two methods were more dramatic when they were used to measure the expression of individual *c-myb* splice variants. As shown in Figure 3B, the QPCR and polony assays agreed only modestly ($r^2=0.815$) when they were used to assay the levels of the 9B splice variant. The situation was much worse for the 9S/10 splice variant (Figure 3C), where the results from the assays failed to correlate ($r^2=0.243$). Thus, although the two methods gave highly similar results for the total expression of *c-myb*, they gave quite different results for some splice variants, and the variation appeared to be sample specific, producing much more scatter in the leukemia sample plots, rather than some sort of systematic difference.

The differences in the QPCR and polony assay results for individual splice variants can be explained when the details of the polony results are scrutinized (Supplemental Table S3). While QPCR measures only the relative abundance of each exon, such as 9B, or each splice junction, such as 9S/10, the polony assay measures the structure of each template and the data are converted into predicted protein structures. As shown in Supplemental Table S3, the polony assay detected a total of 21 different splice variants that contained exon 9B, only one of which is predicted to produce the 9B splice variant protein, which includes residues 1-401 of c-Myb, followed by a novel 121 amino acid in-frame insert followed by residues 402-640 (abbreviated: 1-401, +121aa, 402-640). Similarly, the polony assay detected 22 different splice variants that included the short 9S

exon, only two of which produce the 9S/10 protein (1-372, +2aa). Thus, the QPCR assay overestimates the number of splice variants that will produce the 9B and 9S/10 protein products because it fails to take into account the other changes in the transcript. Indeed, if the polony assay data are compiled in order to simulate the QPCR assay results by pooling all the enumerated splice variants that contain the 9S/10 splice junction, the results of the two assays again correlate quite well ($r^2 > 0.88$, Figure 3D). Thus, the polony assay detects all the different splice variants containing specific exons and provides a more detailed glimpse of the spectrum of proteins produced as a consequence of alternative RNA splicing.

Identification of alternatively spliced variants of *c-myb* in pediatric B-ALL

Using the polony results (Supplemental Table S3), we compared the total levels of wild type and alternatively spliced transcripts in the leukemia samples, which revealed that *c-myb* alternative splicing was highly increased in the leukemias compared to normal CD34+ cells. In 3 normal donor samples (progenitor CD34+ cells), the combined expression level of *c-myb* splice variants was approximately 5%-15% of the total *c-myb* transcripts (Figure 4A). However, the splice variants were much more abundant in the 13 pediatric leukemia patient samples, accounting for 25% to 60% of the total *c-myb* transcripts, suggesting that alternative splicing of *c-myb* transcripts occurs much more frequently in leukemias than in normal hematopoietic cells.

The polony assay detected 61 different splice variants of *c-myb* in the leukemia samples (Supplemental Table S3). They can be categorized as: (a)

exon deletions, in which one or more exons are skipped (for example, Del8; Del9; Del10; Del8/Del9; Del8/9/Del10); (b) partial exon deletions by the activation of cryptic exonic splice sites (for example, 85 nucleotides of partial exon 9 deletion (9S)); (c) inclusion of one or more alternative exons (for example, 8A; 9A; 9B; 10A); (d) combinations of these splicing events (for example, 9S/9A; Del8/9S/9B; Del8/9/9B/10A; 9B/10A; 8A/10A). Some of these variants are predicted to encode the same variant c-Myb proteins, since the introduction of a premature termination codon is encoded by the upstream alternative exon(s) (8A, 9A, and 10A). Truncation of the c-terminus of c-Myb activates its leukemogenic activities^{74,109}, so these variant proteins could be expected to have increased transforming activities. The c-terminus of c-Myb protein also contains the major site of ubiquitination, and truncated variants are more stable¹⁰⁷, so the shorter proteins produced by the splice variants could also be expected to be more stable than wild type c-Myb protein.

The expression levels of the 13 most abundant splice variants in the normal CD34+ and leukemia samples (each accounting for more than 1% of the total *c-myb* transcripts) are shown in Figure 4B. The results revealed highly variable levels of each of these splice variants amongst the patient samples. For example, the levels of the 9S/10 transcript varied 6-fold in these 13 patients (from 1% to 6% of the total *c-myb* transcripts), the level of the 9B form ranged from 2% to 16% of the total transcripts, and the 10A transcript varied from 5% to 15% of the total. On average, the splice variants were far more abundant in patient samples than in normal CD34+ cells. In addition, the splicing repertoires of *c-myb*

were distinct in patient samples compared to healthy CD34+ cells and each patient displayed a unique pattern of splice variants: some patients expressed more than 25 *c-myb* isoforms, while others only expressed 12 splice variants (Supplemental Table S3). Taken together, these data confirm that overall alternative splicing of the *c-myb* gene was significantly elevated in leukemia samples compared to normal CD34+ cells, suggesting that alternative splicing of *c-myb* could be linked to prognosis or patient outcome.

Alternative splicing of *c-myb* as a potential prognostic marker

The higher rates of alternative splicing in leukemias suggested that expression of particular *c-myb* splice variants could be a potential prognostic signature. Analysis of the alternative splicing data (Supplemental Table S3) showed that only wild type c-Myb and 5 splice variant proteins: Del8, 8A, 9S/10, 9A and 10A were expressed in all the leukemia samples. By analyzing the levels of expression in leukemias and normal samples (Figure 4B) we identified three splice variants: Del8, 9A and 9S/10 that were consistently expressed more highly in the leukemias than in the normal CD34+ progenitor cells. Next, we used the polony exon profiling data to group the 13 patients based on whether they expressed above or below the median level of each of the most commonly expressed splice variants. The survival data for these patients were then analyzed to see if the groupings, based on splice variant expression, correlated with better or worse outcome. As shown in Figure 5A, the patients displaying the highest and lowest levels of total *c-myb* displayed similar survival curves, indicating that the absolute level of *c-myb* expression did not affect outcome in

this context. Similar results were obtained for the majority of the splice variants analyzed (not shown). However, the groups of patients with high and low levels of expression of the 9S/10 splice variant, which is generated through the use of an alternative splice donor site in exon 9 and which encodes a truncated version of c-Myb protein (Figure 1C), were quite striking. As shown in Figure 5B, the patients with below median expression of variant 9S/10 all survived at least 8 years (solid line), while the patients with above median expression of 9S/10 had significantly lower survival ($p < 0.05$). In contrast, high levels of the *c-myb* 9B variant, which has been implicated in altered regulation of apoptosis in some cell types²⁰⁵, was associated with increased survival ($p = 0.265$) (Figure 5C). Although these results are from a small cohort of patients, they suggest that increased levels of the 9S/10 splice variant could be a negative prognostic factor for pediatric precursor B-ALL and that analysis of *c-myb* alternative splicing could prove to be a useful prognostic marker for leukemias such as precursor B-ALL.

Analysis of the expression of *c-myb* splice variants relationship with the expression of oncogenic genes in pediatric B-ALL patient samples

The 9S/10 splice variant of *c-myb* is formed through the use of an alternative splice donor site located 85 nucleotides upstream of the normal splice site, which connects a shortened exon 9 to exon 10. This variant encodes a truncated c-Myb protein with a unique C-terminus whose structure is reminiscent of the v-Myb protein encoded by Avian Myeloblastosis Virus. Compared to c-Myb, the v-Myb protein is truncated at the N- and C- termini and is modified by several point mutations that unleash its oncogenic potential and its ability to

transform myelomonocytic cells ¹²¹. Several c-Myb target genes are involved in regulating cell proliferation, for example cell cycle regulator CCNA1 ³⁶ and transcription factor c-MYC ²²⁰; as well as genes important in metastasis, such as CXCR4 ³⁶. Some of these genes have been linked to aggressive childhood pre-B ALL and high frequencies of relapse ²²¹⁻²²⁴. We tested whether the aberrant expression of the 9S/10 splice variant could be correlated to the expression levels of these target genes in pediatric pre B-ALL leukemia patients. Quantitative RT-PCR was used to compare the expression levels of MYC, CCNA1, and CXCR4 to the level of splice variant 9S/10 measured by polony assay (Figure 6). However, the data showed no clear correlation between the expression of the 9S/10 splice variant and the expression of these Myb target genes, suggesting that there is no simple, direct relationship between the level of splice variant 9S/10 and overexpression of these genes.

Discussion

The *c-myb* gene is organized in way that the vast majority of alternative RNA splicing occurs in exons encoding the C-terminal domains of c-Myb protein, generating a family of transcription factors with the same N-terminal DNA binding domains but differences in the C-terminal domains affecting transcriptional activity, negative regulation and target gene selection ^{19,36,78}. Here, we have adapted a polony assay to detect the complex and combinatorial alternative splicing events in *c-myb* transcripts and to quantitatively measure the changes in the patterns of *c-myb* alternative splicing in different cell lines and in leukemia patient samples. We found that *c-myb* alternative splicing is greatly increased in

precursor B-ALL samples, the pattern of *c-myb* alternative splicing is more complex in leukemia patients than in normal CD34+ cells, and the pattern is distinct in each patient sample, suggesting that activation of alternative RNA splicing is a mechanism that could contribute to leukemogenesis. By analyzing a small cohort of patient samples, we identified a splice variant, 9S/10, whose expression correlated with poor survival. Although confirmation will require detailed analyses of larger patient cohorts, the results are intriguing and suggest that patterns of *c-myb* alternative splicing have the potential to provide an independent means of classifying patients into different treatment groups, similar to gene expression profiling²²⁵. Interestingly, the latter approach compares total levels of mRNAs, but our analyses (Figure 5) showed that total *c-myb* levels did not correlate with patient survival, while levels of some individual splice variants (e.g. 9S/10) did. This may explain why microarray assays did not identify *c-myb* as a component of the classifier distinguishing high risk patient groups, even though c-Myb is a key regulator of proliferation and differentiation in hematopoietic and other cells²¹⁴.

A key difference in our study compared to previously reported analyses of alternative splicing is our use of a single molecule polony assay to analyze the various *c-myb* transcripts. Unlike microarrays or QPCR, which only measure the relative levels of expression of specific exons or splice junctions, the polony assay provides structural information over the entire length of the template cDNA. Thus, the polony assay can identify transcripts that contain both an upstream variation, such as inclusion of exon 8A, and also a downstream

variation, such as inclusion of exon 9B. The other methods can detect each of these events but cannot determine whether both splice variations are present in the same, or in different RNA molecules. In this example, since exon 8A introduces a premature stop codon, mRNAs that include exon 8A either alone or in combination with exon 9B produce the same truncated protein. Thus, the polony assay is better able to predict which types of variant c-Myb proteins will be produced in a given sample. At present, even next-generation sequencing methods are limited to reading less than 1 kb of DNA sequence per run, so they are insufficient to determine the structure of the entire region of the *c-myb* transcript that becomes altered by alternative RNA splicing. Until next-generation methods are developed that can determine the structures of entire cDNAs, single molecule approaches such as the polony assay will remain the best alternative for following complex changes in alternative splicing.

An important question concerns why the alternative splicing patterns differ in leukemias compared to normal cells, and why they vary amongst different patient samples. There is little information about the regulation of alternative splicing in leukemias, although the levels of specific splicing regulatory factors are altered in tumors, can affect the alternative splicing of some genes, including *c-myb*, and can act as oncogenes in some situations²⁰⁷. Several types of regulatory pathways, including transcriptional and post-transcriptional regulation and signaling cascades leading to post-translational modifications can affect the activities of factors controlling alternative RNA splicing²²⁶. Many of these pathways can be modified in leukemias, which could lead to changes in splicing

patterns. In addition, histone modifications can affect alternative RNA splicing²²⁷, implicating important leukemia-associated oncogenes that affect histone modifications in the control of alternative splicing^{228,229}. Thus, the regulation of alternative splicing is a relatively unstudied area that could be affected by many of the known oncogene pathways and that could play an important role in tumor development or progression.

Perhaps the most profound result from our study is that leukemias express such a large fraction, in some samples up to 60%, of *c-myb* transcripts that are alternatively spliced products (Figure 4). Thus, a large fraction, or perhaps even a majority of the c-Myb proteins in these samples are likely to be variant forms, rather than the expected wild type c-Myb. Interestingly, the two leukemia cell lines we analyzed, K562 and Jurkat cells, were much more similar to normal cells than to the leukemia samples in this respect, since they only expressed 10-15% of alternatively spliced *c-myb* transcripts (Table 1). Neither of these cell lines was derived from a precursor B-ALL, so the high level of alternatively spliced products we observed in the leukemias could be specific to precursor B-ALL. Alternatively, the cell lines may be a poor model system for studying the expression of *c-myb* variants. Analysis of additional leukemia types will be necessary to distinguish these possibilities.

Overall, our study explored the possibility of quantitatively monitoring *c-myb* alternative splicing in different samples by polony assay. The increased *c-myb* alternative splicing events in leukemia patient samples and the survival analyses suggested a tight link between the regulation of alternative splicing, the

production of variant c-Myb proteins and their possible roles in leukemogenesis. Further investigation of more patients using the polony assay or related techniques will be needed to shed light on the clinical value of following *c-myb* alternative splicing in leukemia samples, which could potentially be used as a unique type of biomarker.

Materials and Methods

Cells, tissue culture and patient samples

Human erythroleukemia K562 cells (CCL-243, ATCC, Manassas, VA) and human Jurkat T-cells (TIB 152, ATCC, Manassas, VA) were maintained in RPMI1640 medium (GIBCO, Carlsbad, CA) with 10% Fetal Bovine Serum (PAA, Morningside QLD Australia). Buffy coat blood samples were purchased from United Blood Services (Albuquerque, NM) and peripheral blood leukocytes were purified by Ficoll (Amersham, Piscataway, NJ) density centrifugation and cultured in Iscove's Modified Dulbecco's Medium (IMDM from GIBCO) with 10 U each of interleukin-2 (IL-2; PeproTech, Rocky Hill, NJ) and phytohaemagglutinin per ml for 4 days prior to RNA isolation. Cytokine-mobilized CD34+ cells (purchased from the Fred Hutchison Cancer Research Center Large-Scale Cell Processing Core) were cultured in IMDM supplemented with BITS serum substitute, IL-3 (20 ng/ml), IL-6 (20 ng/ml), Stem Cell Factor (100 ng/ml), and FLT-3 ligand (100 ng/ml) (all from Stem Cell Technology, Vancouver, Canada). Pre-B-ALL Patient

bone marrow RNA samples were obtained from the Children's Oncology Group Cell Bank.

RNA expression and structure assays

Total RNA was isolated using RNeasy mini kits (Qiagen, Valencia, CA), cDNA was synthesized using a first-strand cDNA synthesis kit (Invitrogen, Carlsbad, CA), SYBR green-based real-time PCR used a Maxima qPCR kit (Fermentas, EU) and Taqman probe-based QRT-PCR used Taqman Universal PCR Master Mix (Applied Biosystems/Life Technologies, Foster City, CA), according to the manufacturers protocols. PCR reactions were performed in quadruplicate using primers described in Supporting Information Table S1. The results of relative gene expression assays were normalized to the level of GAPDH, and the data were analyzed using the comparative threshold cycle method²³⁴.

Polony amplification

A variety of cell lines and cell types, and 13 pediatric precursor B-ALL samples (provided by COG) were used for the *c-myb* alternative splicing patterns studies. After reverse transcription (as described above), the polony amplification was performed as previously described (²¹⁸). Briefly, the three-step procedure involves application of the gel matrix on a slide, infusion of PCR reagents and in situ PCR in a Slide Cycling '16/16' Dual PCR Block (MJ Research, MA). Gel mix [10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.01% gelatin, 1.5 mM MgCl₂, 6% acrylamide, 0.35% DATD, 0.035% Bis-acrylamide, 1 μM acrydite modified primer

(Table S2), 0.1% Tween-20 and 0.2% BSA] was freshly prepared, ammonium persulfate (APS) and TEMED were added to a final concentration of 0.087% each and 15 μ m thick gels were poured on glass microscope slides that had been partially coated with Teflon masks (Erie Scientific), which served as spacers between the glass surfaces and glass coverslips (22 mm x 30 mm, Fisher Scientific). The gels were allowed to polymerize in a dedicated pre-PCR hood for 15 min. The slides were washed 20 min in ddH₂O, dried under the hood for about 25 min and 23 μ l of polony amplification mix [10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.01% gelatin, 0.2% BSA, 0.1% Tween-20, 1.33 μ l primer (Table S2), 200 μ M dNTP, 0.335 U Jumpstart Taq and desired amount of template] was poured on the slides and covered with a coverslip. The slides were covered with mineral oil utilizing adhesive incubation chambers (Secure Seals SA 500, GRACE Bio-Labs), then cycled in a PTC-200 twin tower thermocycler as follows: denaturation (3 min at 94 °C), amplification for 48-52 cycles (30 sec at 94 °C, 45 sec at 62 °C, 3 min at 72 °C), extension (6 min at 72 °C), chill to 4 °C. After cycling, mineral oil was washed off by rinsing the slides in hexane.

Denaturation and Hybridization

After polony amplification, the unattached strand of DNA was removed by incubating the slides (after removing the glass cover slip) in pre-warmed denaturing buffer (70% formamide, 1 x SSC) at 70 °C for 15 min. The slides were subsequently washed in ddH₂O for 3 min, 2 x 4 min each in wash buffer 1E (10 mM Tris-HCL pH 7.5, 50 mM KCL, 2 mM EDTA, 0.01% Triton X-100) and 90 μ l of annealing mix [0.5 μ M of each hybridization probe (4 probes at a time), 6 x

SSPE, 0.01% Triton X-100] (all of the probes were listed in Table S2) was added over the gel and sealed with a frame seal chamber (HybriWell HBW2240, GRACE Bio-Labs). The slides were heated at 94 °C for 6 min, followed by 58 °C for 15 min. Free probes were removed by washing the slides 2 x 4 min in wash buffer 1E. Following the first cycle, subsequent denaturation steps were performed at 65 °C for sequential exon hybridization. Denaturation in the denaturing buffer will result in removal of the fluorophore labeled exon probes and therefore is done before the annealing of a new exon probes.

Image Acquisition and Data Analysis

All images of gels were acquired on a ScanArray 5000 instrument (Perkin Elmer) at 10 µm resolution using four lasers (635 nm, 532 nm, 488 nm, 594 nm). Gels were scanned after hybridization with labeled probes, as well as after probe stripping (to assess background signal). The myb-exon probed polony slide images were processed with Matlab to generate a composite. Colocalization of polonies allowed us to detect the exon context of the original molecule. A grid was overlaid on the images and a systematic random approach was performed to select fields for analysis. A total of 250–1500 polonies per case were counted in this fashion to provide estimates of expression. For each case, there were at least two slides created and counted. All counts were performed without knowledge of diagnosis.

Sensitivity and linearity of the polony assay

PCR products containing unique alternative exons were cloned in frame in a human c-Myb expression vector (pCDNA3; Invitrogen). To assess the alternative splice variant detection sensitivity of the polony assay, we performed the polony assay on different molar ratio of mixtures of wild-type c-*myb* transcript plasmid and the plasmid containing unique alternative exon 8A. To evaluate the quantitative ability of the polony assay, we serially diluted the wild-type plasmid from 5×10^{-2} ng/ μ l to 6×10^{-4} ng/ μ l, and used 1 μ l from each dilution to perform the polony assay.

Acknowledgements

We thank Jun Zhu (Duke University) for his helpful advice on improving the polony assay, Meenakshi Devidas (University of Florida) for performing the survival curve analysis and Anita Quintana and Julie Torres for helpful discussions. The authors report no conflicts of interest. Some of the experiments used the Keck-UNM Genomics Resource and the Shared Flow Cytometry Resource in the University of New Mexico Cancer Center.

Figure Legends

Figure 1. Alternative splicing in the c-*myb* gene

(A) Structure of the human c-*myb* gene. The diagram shows the exon structure of the human c-*myb* gene, which spans more than 41,000 nucleotides. The arrow marks the major transcription initiation site and exons 1-15, which are included in the wild type version of c-Myb, are indicated above the line. The

highly conserved DNA binding domain of c-Myb is encoded by exons 4-6, and the C-terminal domains that control transcriptional activity and specificity are encoded by exons 7-15, as indicated. The latter region also includes the 6 alternative exons, shown below the line. (B) Alternative splicing in exons 9-10. A cryptic splice donor site in exon 9 generates a short (9S) version lacking 85 nt (shaded). Both the long and short forms of exon 9 can be spliced to alternative exons 9A or 9B or to exon 10. (C) Diagrams of wild type c-Myb and the 9A, 9B, 9S/9A, 9S/9B and 9S/10 splice variants. The structure of the v-Myb protein encoded by Avian Myeloblastosis Virus is shown for comparison. The highly conserved Myb DNA binding domain is shaded black. The variant proteins lack the normal C-terminus of c-Myb and have unique regions, indicated by shaded boxes. The numbers on the right summarize the structures of the proteins with the included amino acids (aa) and the changes relative to wild type c-Myb. (D) Structures of the proteins encoded by splice variants 8A, 8A/9S/10 and 9S/10. These variants are described in the text. Labels and numbering are as in panel (C).

Figure 2. Sensitivity and linearity of the polony assay

(A) Outline of polony assay. Gene-specific primers are used to convert mRNA to cDNA, which is then seeded into a thin acrylamide gel containing primers, enzymes and other components necessary for in situ PCR, which results in the formation of “PCR colonies”, or polonies. The polonies are fixed in the gel, but can be interrogated by sequential hybridization with specific

fluorescently labeled oligonucleotides. (B) Sensitivity of polony assay. Plasmids containing cDNAs for wild type *c-myb* or the 8A splice variant were mixed in a ratio of 5,000:1. After in situ PCR, the slide was hybridized sequentially with fluorescently labeled probes specific for exon 8, exon 8A or exon 11. The full polony slide is shown at left and the enlargement at right shows a single detected exon 8A-containing splice variant (white). In the false color image the polonies containing only exons 8 and 11 appear Cyan in color, the polony containing all three exons appears white. (C) Linearity of polony assay. Different amounts of wild-type *c-myb* plasmid were used as template in the polony assay and the number of polonies detected with the exon 11 probe was plotted against amount of plasmid template used in the assay. The assay was very linear, as shown by the correlation coefficient r^2 , which was 0.99.

Figure 3. Comparison of exon-specific (QPCR) to single molecule (polony) assays

Scatter plots compare the levels of different *c-myb* splice variants detected by bulk QPCR or single molecule polony-based assays for (A) total *c-myb* transcripts or (B-C) the 9B or 9S/10 splice variants, respectively. The QPCR and polony assays agree nicely ($r^2=0.946$) for total *c-myb* transcripts, but the results for the 9B variant agree only modestly ($r^2<0.82$) and the results for the 9S/10 variant did not correlate well ($r^2<0.25$). (D) Compiling the polony data by combining all the detected transcripts that contain the 9S/10 exon produces data that correlate much better ($r^2>0.88$) with the QPCR data. Thus the differences in

the two assays are due to the exclusion of variants in the polony results that are predicted to produce different protein products, which are detectable in the polony assay but not by QPCR.

Figure 4. Alternative splicing of *c-myb* transcripts in normal and leukemic cells

(A) Rates of *c-myb* alternative splicing in normal cells and leukemias.

Polony assays were used to measure the levels of normal and alternatively spliced *c-myb* transcripts in normal CD34+ progenitors from healthy donors (n=3) or in pediatric precursor B-ALL patient samples (n=13). The black and gray sections of each bar indicate the fractions of alternatively spliced or wild type transcripts, respectively. (B) Patterns of *c-myb* alternative splicing in pediatric precursor B-ALL samples. Dots indicate levels of wt or splice variant expression, as a fraction of total *c-myb* transcripts detected, as determined by polony assay for pediatric precursor B-ALL samples (n=13, filled circles) or CD34+ cells from normal donors (n=3, empty diamonds). Each symbol represents the average of duplicate measurements and the horizontal bars represent the median values for the leukemia samples.

Figure 5. Survival plots for patients grouped by *c-myb* variant expression levels

Precursor B-ALL patients (n=13) were grouped according to their expression above (dotted line) or below (solid line) the mean expression level for (A) wild type *c-myb* or (B) the 9S/10 variant or (C) the 9B variant. The plots show overall survival curves, with p-values as indicated.

Figure 6. Relationship between the 9S/10 splice variant's expression level (measured by polony assay) and the Myb target genes' expression levels

Patient RNA samples were analyzed by QRT-PCR to assess the expression level of the CCNA1, MYC, and CXCR4 genes, relative to the GAPDH control gene and the patient sample that has the lowest expression. Scatter plots correlate the levels of 9S/10 splice variant in each patient detected by polony assay with the levels of the CCNA1 (A), MYC (B), and CXCR4 (C) genes measured by QRT-PCR, respectively. R is the observed correlation coefficient.

Supporting Information Legends

Figure S1. Conservation of the *c-myb* 9S/10 splice variant.

(A) Comparison of the cryptic splice site in exon 9 from human, mouse and chicken and the predicted C-terminal amino acids in the corresponding c-Myb 9S/10 variant proteins. The 5' splice site (gt) is bolded. All the predicted proteins

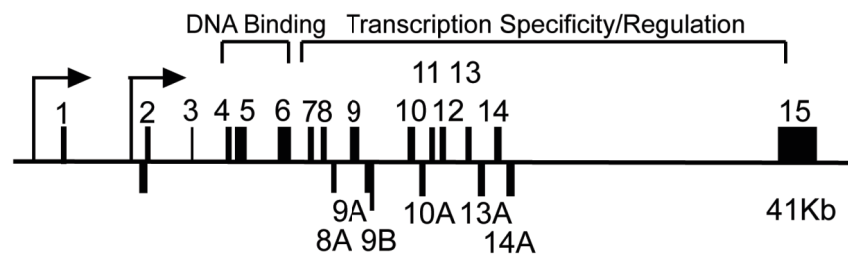
terminate with two serine residues (in bold). B) Expression of the *c-myb* 9S/10 variant in primary chicken and mouse bone marrow cells. RNA samples isolated from the chicken cell line HD3, primary chicken bone marrow or primary mouse bone marrow were subjected to RT-PCR using a forward primer spanning the unique 9S/10 splice junction and a reverse primer found in exon 10. Nucleotide sequencing of the PCR product confirmed that it was the expected splice junction product from the 9S/10 splice variants.

Figure S2. Polony exotyping of *c-myb* transcripts.

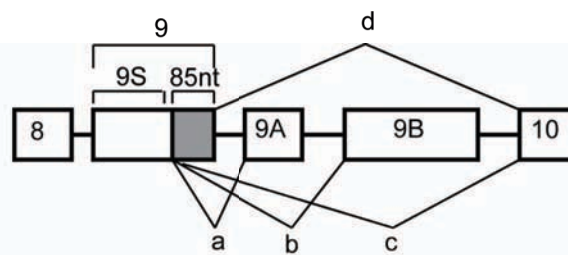
(A) Scheme of *c-myb* alternative splicing. 11 exons including constant and alternative ones were analyzed in this experiment. The same slide was sequentially hybridized with 11 different fluorophore-labeled exon probes. The probed exons are shown in different colors. Exon8: red; Exon 8A: green; Exon 10A: blue. All four possible alternative splicing isoforms are indicated. (B) Individual and merged images hybridized with exon 8, exon 8A and exon 10A are shown. Different combinations were identified as indicated. White polonies are transcripts with 8A+10A+8, purple are transcripts with 8+10A, and yellow are transcripts with 8A+8.

Figure 1

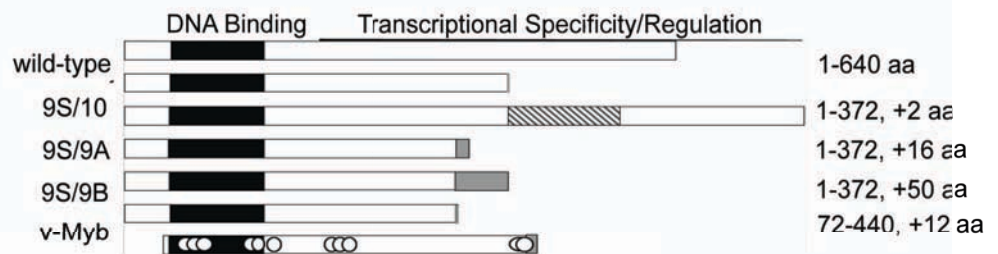
A



B



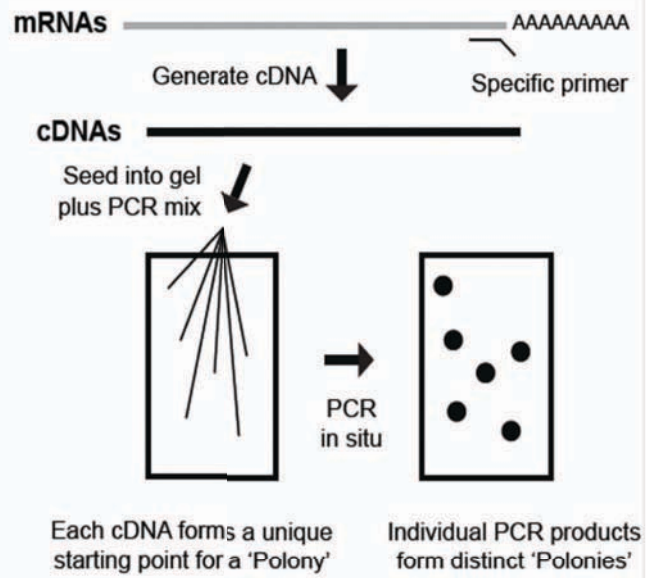
C



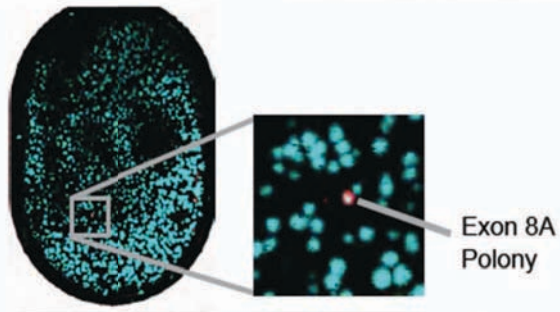
D



Figure 2
A



B



C

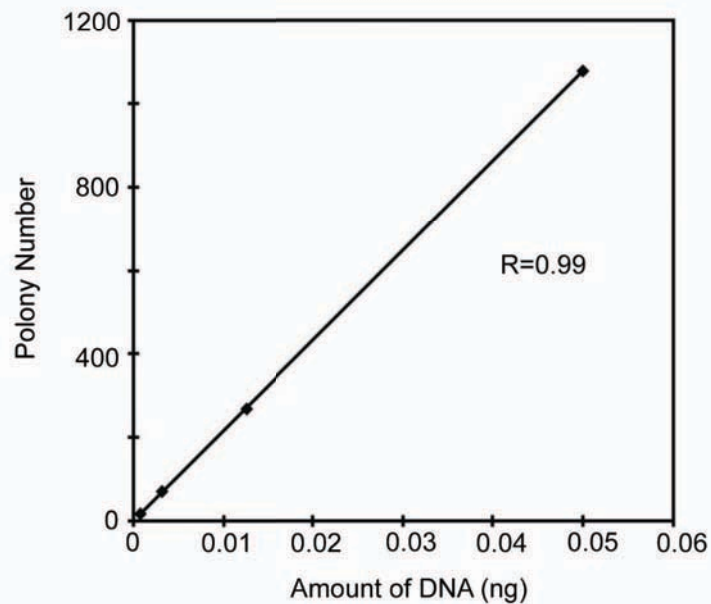


Figure 3

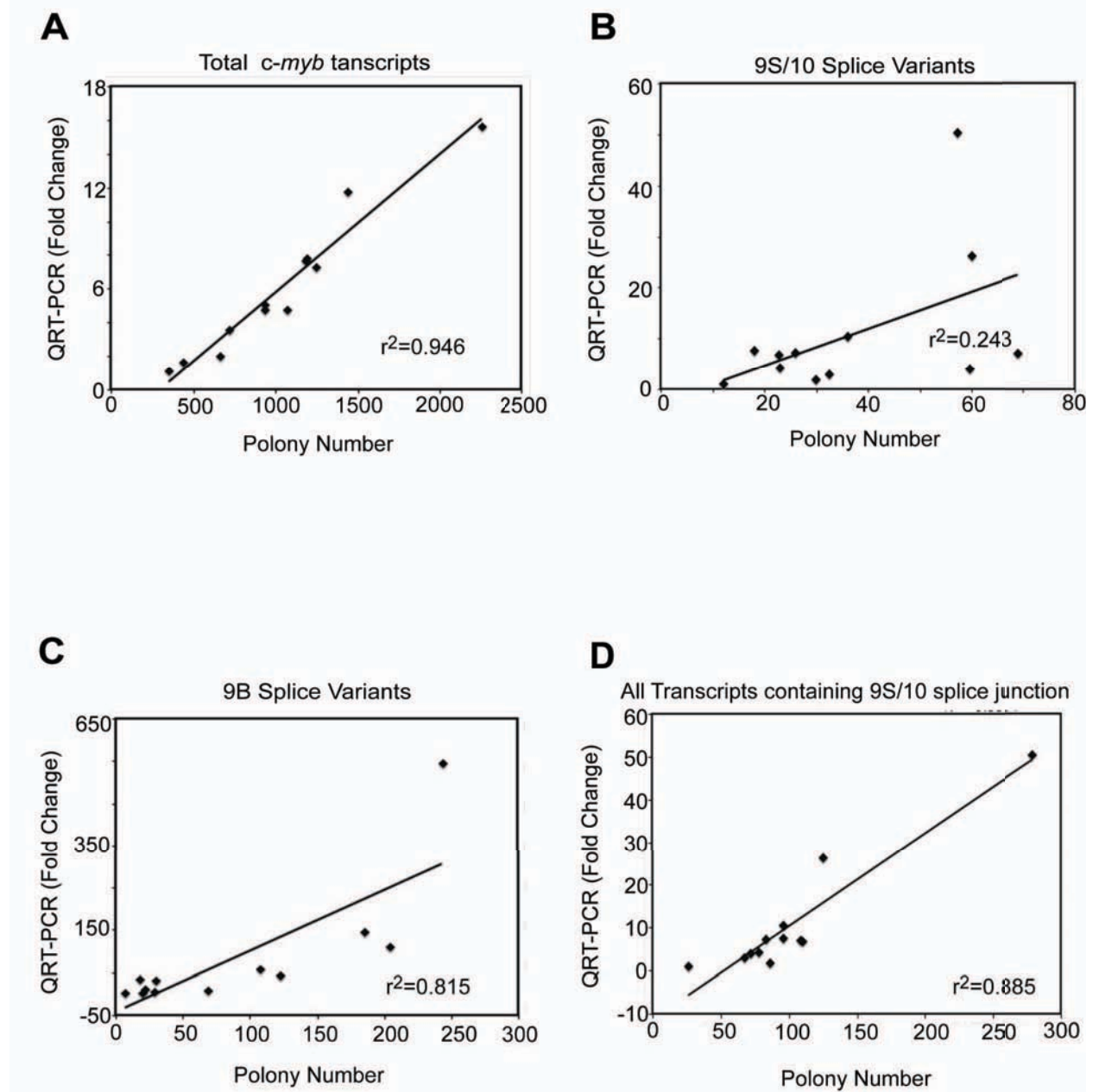
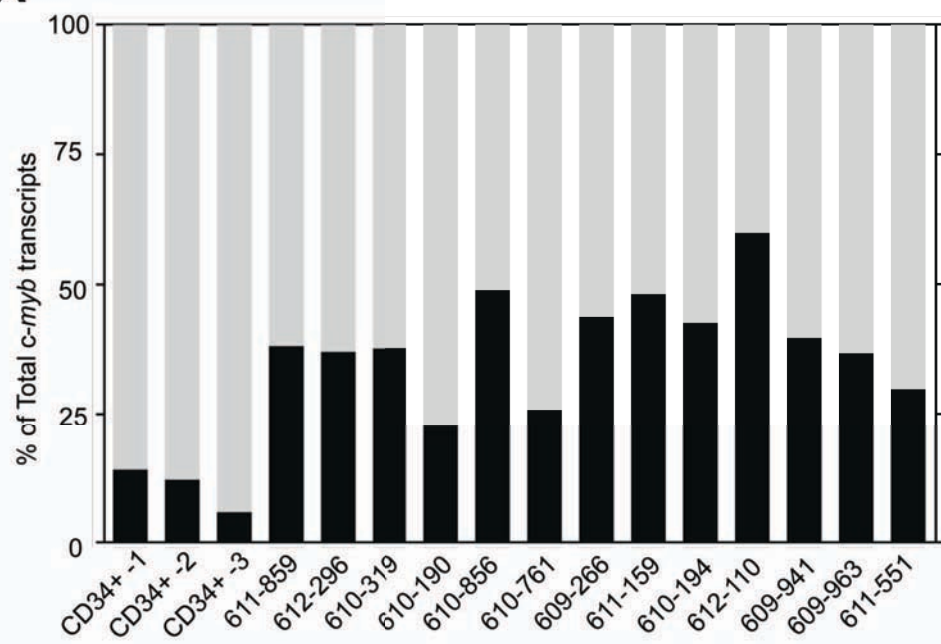


Figure 4

A



B

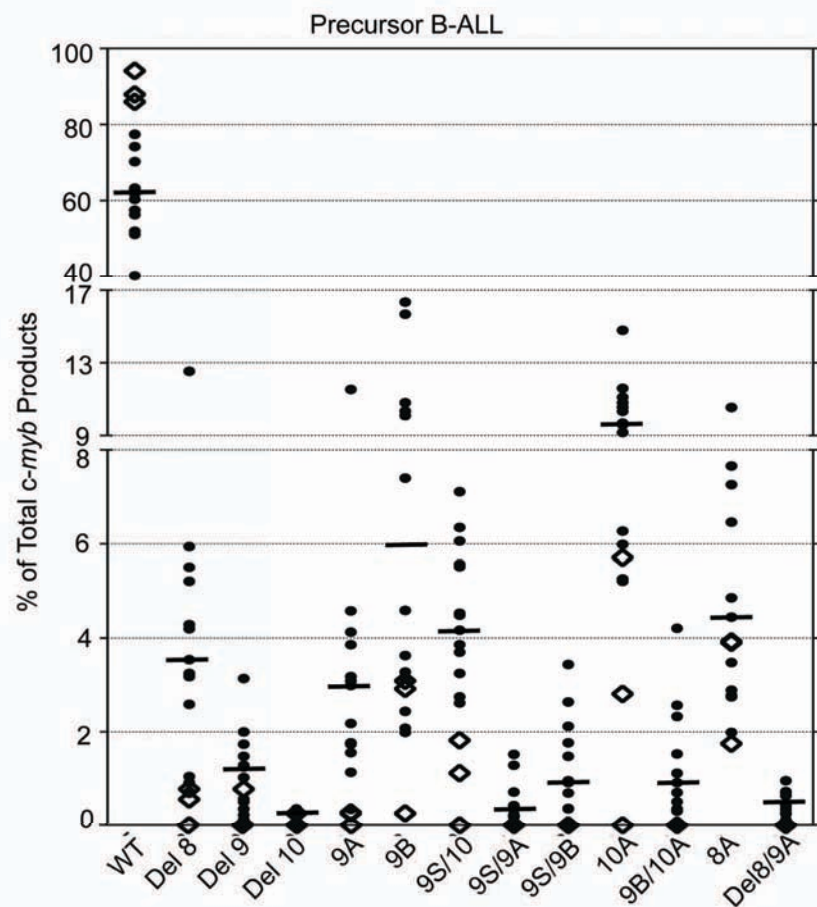
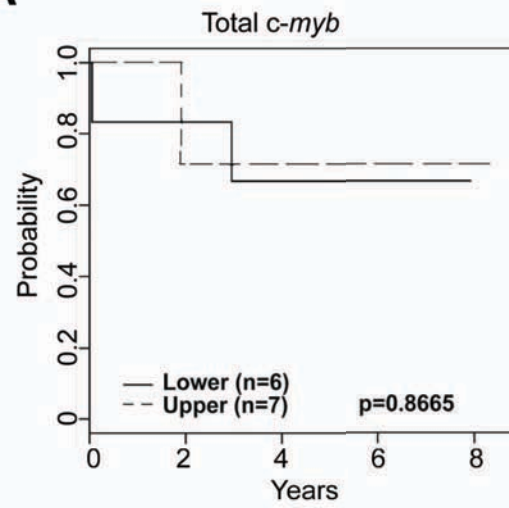
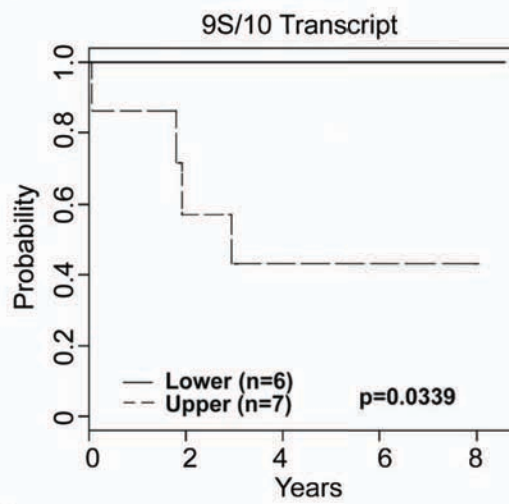


Figure 5

A



B



C

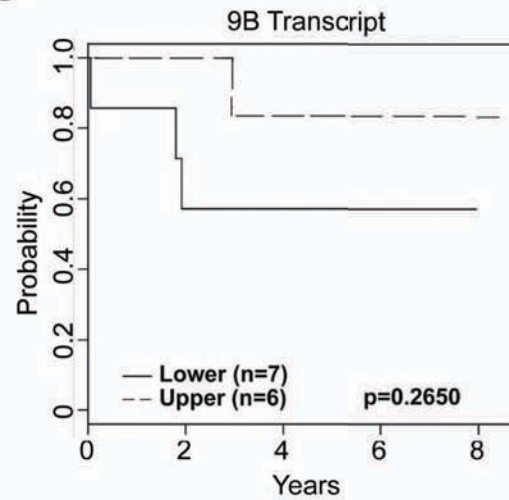
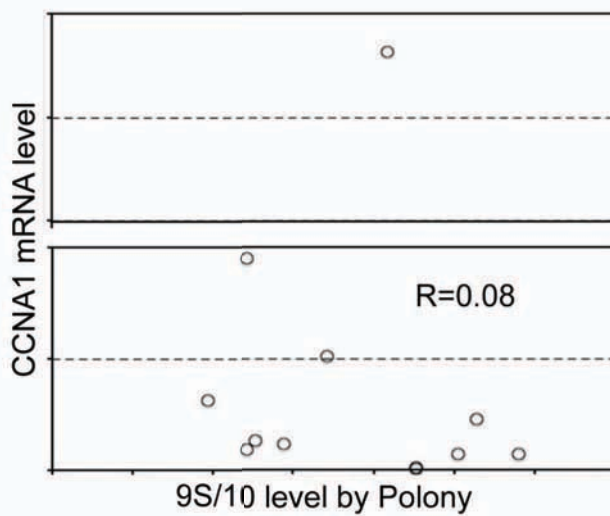
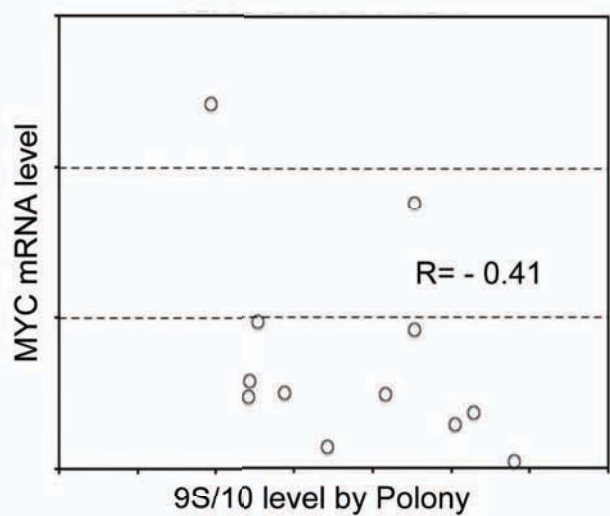


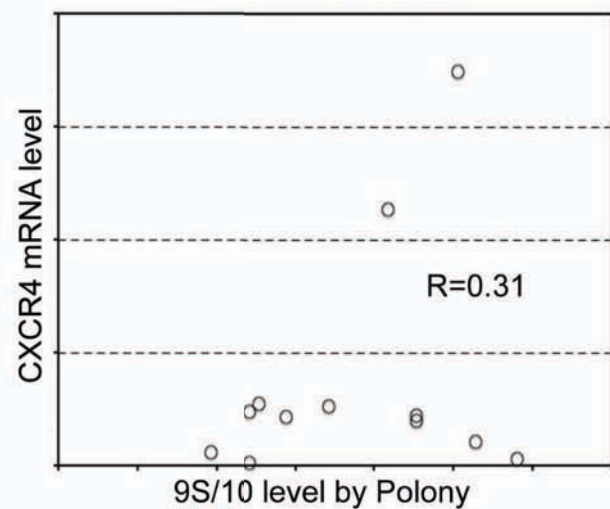
Figure 6
A



B



C



Supplemental Figures

Figure S1

A

Species	Exon 9S Splice Sequence	c-Myb 9S/10 C-Terminus
Human	GCAAG gt gcatgac	PEESASP <u>SS</u>
Mouse	GCAAG gt gcatgac	PEESASP <u>SS</u>
Chicken	GCA C G gt gcatgac	PEESASP <u>RS</u>

B

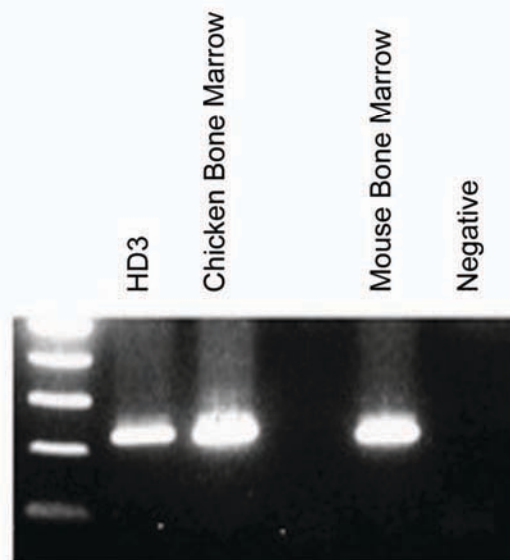
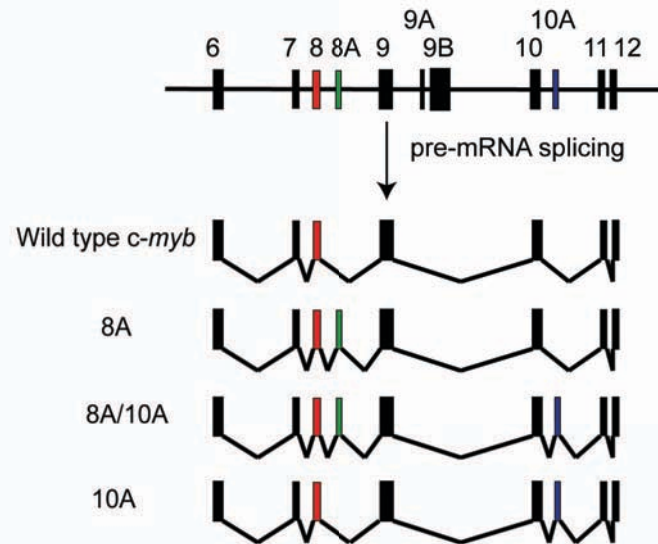
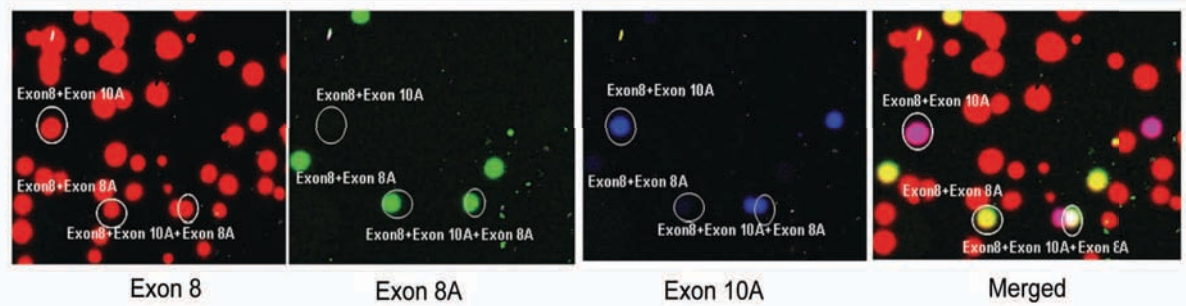


Figure S2

A



B



Tables

Table 1. Expression of c-myb splice variants in different cell lines

Description	Accession No.	Exontyping	K562	Jurkat	PBL	CD34+
Del9	AY787448	6-7-8-10-11	ND	1.78%	ND	0.26%
Del8	AY787447	6-7-9-10-11	1.33%	ND	1.05%	0.45%
9S/10	AY787470	6-7-8-9S-10-11	2.13%	0.36%	1.83%	0.89%
Del10	NA	6-7-8-9-11	0.27%	ND	ND	0.09%
Del8/9A	NA	6-7-9-9A-10-11	0.27%	ND	ND	ND
9A	AY787464	6-7-8-9-9A-10-11	0.54%	0.36%	0.52%	0.18%
9B	AY787467	6-7-8-9-9B-10-11	6.91%	5.69%	2.09%	2.09%
9B/10A	NA	6-7-8-9-9B-10-10A-11	0.53%	1.07%	ND	ND
9S/10/10A	AY787471	6-7-8-9S-10-10A-11	ND	0.36%	ND	0.09%
10A	AY787450	6-7-8-9-10-10A-11	2.66%	4.63%	6.54%	2.84%
8A	AY787454	6-7-8-8A-9-10-11	1.06%	1.43%	3.40%	3.76%
WT	AY787475	6-7-8-9-10-11	84.57%	84.34%	84.55%	89.35%

Table 1 Legend

Expression of *c-myb* alternative splicing isoforms measured by polony assay and expressed as percentage of the total detected *c-myb* transcripts. ND: not detected. NA: not applicable, since these products were only detected using the polony assay, their sequences have not been determined and no accession numbers are available.

Supporting Information Tables

Table S1. Primers used for QRT-PCR.

SYBR Green Detection

Name		Sequence
<i>c-myb</i> 9S/10	Forward	5'-CCTCGCCAGCAAGTTCTTA
	Reverse	5'-GGTGGAAGTTAAAGAAGGCATTT
<i>c-myb</i> 9B/10	Forward	5'-TTCAACTCCCAAGCGTTCC
	Reverse	5'-AAAGAAGGCATTTCCAAGTCTG
CCNE1	Forward	5'-CCATCCTTCTCCACCAAAGA
	Reverse	5'-CTGGAGGTGGCTGGTGTACT
c-MYC	Forward	5'-CAGCTGCTTAGACGCTGGAT
	Reverse	5'-ACCGAGTCGTAGTCGAGGTC
GAPDH	Forward	5'-GCTGTGTGGCAAAGTCCAAG
	Reverse	5'-GGTCAGGCTCCTGGAAGATA

Taqman Expression Assay

Primer Catalogue Number

CXCR4 HS00607978_s1

c-myb Hs00193527_m1

Table S2. All primers used to perform polony amplification and exon-profiling

Name	Sequence (5'-3')
Gene specific reverse primer	AGAAGACGAGTGTGGTGACC
Myb_Forward	Acrydite/GGAAGGTCGAACAGGAAGGTTATCTG
Myb_Reverse	CTAGATGAGAGGGTGTCTGAGGTAGCA
Myb_E6R	Alexa532/TAGTCGTTGTTAACAGTGGGCTGG
Myb_E7R	Alexa488/GGTATGGAACATGACTGGAGAC
Myb_E8R	Alexa647/TTTATTCGCTTTTCCTTCTCAGGG
Myb_E8AR	Alexa594/CTGCCCAGTCGCTAAAGATGTC
Myb_E9R	Alexa532/CATGAGGTCTGGTGTGGTCGGCAA
Myb_E9AR	Alexa488/TGAGCAGTTGCTCTATTATACTGC
Myb_E9BR	Alexa594/CTACAGTCTCCAGTGGCTAAGG
Myb_E10R	Alexa532/CAGTCAATTTGTGACCAATGAG
Myb_E10AR	Alexa594/ATTCCTGGGGATGTCAGAAGAG
Myb_E11R	Alexa488/GAATGGTGTAGGAGTTCTTGGA
Myb_E(9-9S)R	Alexa647/TTAACATTATCCAGAATGGTGCC

Table S3. The expression level of c-myb splice variants detected by polony assay in 13 pediatric pre-B ALL sample

	Isoform	Protein Size	Exontyping	611-859	612-296	610-319	610-190	610-856	610-761	609-266	611-159	610-194	612-110	609-941	609-963	611-551
1	Del8/Del9	1-281, 402-640	6-7-10-11	0.14%	ND	ND	ND	ND	ND	ND	ND	0.24%	ND	ND	0.47%	ND
2	Del8/Del10	1-281, 371-402	6-7-9-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.16%	ND
3	Del8	1-281, 317-640	6-7-9-10-11	4.20%	12.54%	3.18%	5.21%	3.54%	1.05%	4.29%	5.51%	2.59%	3.25%	0.92%	5.94%	0.73%
4	Del9	1-316,402-640	6-7-8-10-11	0.36%	0.57%	1.03%	1.74%	1.29%	3.14%	0.51%	1.48%	ND	0.12%	0.23%	1.25%	2.00%
5	Del10	1-402	6-7-8-9-11	0.21%	0.28%	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.16%	0.36%
6	WT	1-640	6-7-8-9-10-11	61.85%	62.96%	62.30%	77.43%	51.13%	74.22%	56.31%	51.91%	57.46%	40.14%	60.32%	63.28%	70.24%
7	Del8/9S	1-281, 317-372, +2aa	6-7-9S-10-11	0.14%	1.71%	ND	0.35%	0.96%	0.35%	0.25%	0.64%	0.47%	ND	0.23%	0.78%	ND
8	Del8/9S		6-7-9S-10-10A-11	0.07%	ND	ND	ND	2.57%	ND	ND	ND	ND	ND	ND	ND	ND
9	9S/10	1-372, +2aa	6-7-8-9S-10-11	2.42%	3.42%	2.43%	4.17%	1.93%	4.53%	5.81%	2.54%	5.05%	2.88%	5.28%	4.53%	6.17%
10	9S/10		6-7-8-9S-10-10A-11	1.07%	0.28%	0.19%	ND	1.93%	1.05%	0.25%	0.21%	0.47%	0.36%	1.83%	ND	0.18%
11	Del8/9A	1-281, 317-402, +M	6-7-9-9A-10-11	0.36%	ND	ND	ND	ND	ND	0.25%	ND	0.35%	0.36%	ND	ND	0.18%
12	Del8/9A		6-7-8-9A-9B-10-11	0.07%	N/A	0.09%	ND	ND	ND	0.25%	ND	0.12%	0.24%	ND	ND	ND
13	Del8/9A		6-7-9-9A-10-10A-11	0.07%	ND	ND	ND	ND	ND	ND	ND	ND	0.12%	ND	ND	ND
14	Del8/9A		6-7-9-9A-9B-10-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.12%	ND	ND	ND
15	9S/9A	1-372, +16aa	6-7-8-9S-9A-10-11	0.28%	ND	0.19%	ND	0.32%	0.35%	0.25%	ND	0.24%	0.60%	ND	ND	0.36%
16	9S/9A		6-7-8-9S-9A-9B-10-11	ND	ND	ND	ND	0.32%	ND	ND	ND	ND	ND	ND	ND	ND
17	9S/9A		6-7-8-9S-9A-10-10A-11	ND	ND	ND	ND	0.64%	ND	0.25%	ND	0.12%	ND	0.23%	ND	ND
18	9S/9A		6-7-8-9S-9A-9B-10A-11	0.14%	ND	ND	ND	ND	ND	1.01%	ND	ND	0.12%	ND	ND	ND
19	9A	1-401, +M	6-7-8-9-9A-10-11	1.57%	0.85%	2.43%	0.35%	2.25%	1.39%	1.26%	1.27%	2.35%	5.77%	2.75%	1.56%	1.09%

	Isoform	Protein Size	Exontyping	611-859	612-296	610-319	610-190	610-856	610-761	609-266	611-159	610-194	612-110	609-941	609-963	611-551
20	9A	1-401, +M	6-7-8-9-9A-9B-10-11	0.57%	ND	0.65%	ND	ND	ND	0.51%	1.27%	1.65%	3.13%	0.23%	ND	0.54%
21	9A		6-7-8-9-9A-10-10A-11	0.64%	0.28%	ND	ND	0.64%	0.35%	ND	0.42%	0.12%	1.92%	0.92%	ND	0.54%
22	9A		6-7-8-9-9A-9B-10-10A-11	0.21%	ND	ND	ND	0.96%	ND	ND	0.21%	0.47%	0.72%	0.23%	ND	ND
23	Del8/8A	1-281, +34aa	6-7-8A-9S-10-11	ND	ND	0.28%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
24	Del8/8A		6-7-8A-9-10-11	0.21%	0.28%	0.19%	ND	ND	ND	ND	ND	ND	ND	0.23%	0.47%	ND
25	Del8/8A		6-7-8A-9-10-10A-11	0.36%	ND	ND	ND	0.64%	ND	ND	0.21%	0.24%	ND	0.23%	ND	0.73%
26	Del8/8A		6-7-8A-9-9A-10-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.12%	ND	ND	ND
27	Del8/8A		6-7-8A-9S-9B-10-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	0.12%	ND	ND	ND	ND
28	Del8/8A		6-7-9-9A-10-10A-11	0.07%	ND	ND	ND	0.32%	ND	0.25%	0.21%	ND	ND	0.23%	ND	ND
29	8A	1-316, +34aa	6-7-8-8A-9S-10-11	0.21%	ND	0.19%	ND	ND	ND	0.25%	ND	ND	ND	0.92%	ND	ND
30	8A		6-7-8-8A-9-11	0.07%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
31	8A		6-7-8-8A-9-10-11	5.69%	0.28%	3.55%	0.35%	ND	ND	1.77%	2.33%	0.12%	0.12%	6.65%	ND	ND
32	8A		6-7-8-8A-9S-9A-10-11	0.07%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
33	8A		6-7-8-8A-9-9A-10-11	0.14%	ND	0.09%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
34	8A		6-7-8-8A-9S-9B-10-11	0.07%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
35	8A		6-7-8-8A-9-9B-10-11	0.14%	ND	0.09%	ND	ND	ND	0.25%	0.21%	ND	ND	ND	ND	ND
36	8A		6-7-8-8A-9-9A-9B-10-11	0.07%	ND	0.19%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
37	8A		6-7-8A-10-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	6.72%	ND
38	8A		6-7-8-8A-10-10A-11	0.07%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
39	8A		6-7-8-8A-9S-10-10A-11	0.14%	0.28%	ND	ND	0.32%	0.35%	ND	ND	0.35%	0.12%	0.23%	ND	0.18%

	Isoform	Protein Size	Exontyping	611-859	612-296	610-319	610-190	610-856	610-761	609-266	611-159	610-194	612-110	609-941	609-963	611-551
40	8A	1-316, +34aa	6-7-8-8A-9-10A-11	0.07%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
41	8A		6-7-8-8A-9-9A-10-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	0.24%	0.24%	ND	ND	ND
42	8A		6-7-8-8A-9S-9B-10-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	0.24%	0.24%	ND	ND	ND
43	8A		6-7-8-8A-9S-9A-10-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.18%
44	8A		6-7-8-8A-9-10-10A-11	1.21%	1.42%	0.56%	2.43%	1.61%	3.14%	0.51%	0.21%	4.82%	2.16%	2.75%	0.94%	6.72%
45	8A		6-7-8-8A-9S-9A-9B-10-11	ND	ND	0.19%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
46	8A		6-7-8-8A-9-9B-10-10A-11	0.07%	ND	ND	ND	0.96%	ND	ND	ND	0.71%	1.56%	ND	ND	0.18%
47	Del8/9S/9B	1-281, 317-372, +50aa	6-7-9S-9B-10-11	ND	ND	ND	ND	0.32%	ND	1.01%	ND	0.24%	ND	ND	ND	ND
48	Del8/9S/9B		6-7-9S-9B-10-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.12%	ND	0.16%	ND
49	Del9/10A	1-316, 402-449	6-7-8-10-10A-11	ND	ND	ND	ND	ND	0.35%	ND	0.21%	ND	ND	ND	ND	ND
50	Del9/10A		6-7-8-10-10A-11	ND	ND	0.09%	ND	0.32%	ND	ND	ND	ND	ND	ND	ND	ND
51	9S/9B	1-372, +50aa	6-7-8-9S-9B-10-11	0.21%	ND	0.37%	ND	0.32%	0.70%	1.52%	1.27%	2.00%	1.80%	0.69%	3.28%	ND
52	9S/9B		6-7-8-9S-9B-10-10A-11	0.14%	ND	ND	ND	0.64%	ND	0.25%	0.21%	0.12%	0.84%	0.23%	0.16%	ND
53	9B/Del10	1-401, +121aa, 402	6-7-8-9-9B-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.12%	ND	ND	ND
54	9B	1-401, +121aa, 402-640	6-7-8-9-9B-10-11	3.13%	1.99%	10.10%	2.08%	7.40%	2.44%	10.35%	15.68%	10.81%	16.35%	4.59%	ND	3.63%
55	Del8/10A	1-281, 317-449	6-7-9-10-10A-11	0.71%	1.14%	0.19%	ND	0.96%	0.35%	1.52%	0.64%	0.24%	0.60%	ND	0.63%	0.36%
56	Del9/Del10/10A	1-316, +13aa	6-7-8-10A-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.12%	ND	0.63%	ND
57	Del10/10A	1-401, +13aa	6-7-8-9-10A-11	0.07%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
58	10A	1-449	6-7-8-9-10-10A-11	11.60%	11.11%	10.57%	5.21%	14.79%	6.27%	9.60%	10.81%	5.99%	10.34%	9.17%	9.69%	5.26%

59	9B/10A	1- 401,+121aa,402- 449	6-7-8-9-9B- 10-10A-11	0.71%	ND	1.12%	0.35%	2.57%	ND	0.51%	2.33%	1.53%	4.21%	0.92%	0.31%	0.36%
60	Del8/9S/9A	1-281,317- 372,+16aa	6-7-9S-9A- 10-11	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.16%	ND
61	Del8/9B	1-281,317- 401,+121aa,402- 640	6-7-9-9B-10- 11	0.36%	0.28%	ND	0.35%	0.32%	ND	1.01%	0.21%	0.59%	0.96%	ND	ND	ND

Chapter 3: Next-Generation multiplex sequencing of *c-myb* alternative splicing in a cohort of pediatric B-ALL samples: Implications for clinical applications

Introduction

As predicted from the early studies demonstrating that activation of the *c-myb* gene is leukemogenic in several animal species, it has now become clear that the *c-myb* gene is also deregulated in human leukemias. Recent studies revealed that a subset of T-ALL shows selective amplification and overexpression of *c-myb* mediated by somatic homologous recombination between Alu-elements flanking the *c-myb* locus¹⁶⁻¹⁸. A Myb-Tcrb fusion resulted from t(6,7)(q23;q34) translocations was also identified in a novel subgroup of T-ALL associated with very young age and a proliferation/mitosis gene expression signature¹⁶. However, these mutations on the *c-myb* gene are not associated with prognosis or other clinical parameters of the cancer patients. The *c-myb* gene can undergo extensive alternative splicing in different cell lines, cell types, some leukemias and other types of human cancers, producing a big repertoire of *c-myb* splice variants encoding multiple forms of variant c-Myb proteins¹⁹. Our previous findings (see details in Chapter 2) showed that *c-myb* alternative splicing is more prevalent in pre-B-ALL leukemia samples than in normal cells, and the level of some splice variants could

be tightly correlated with clinical outcome of these leukemia patients, suggesting that aberrant *c-myb* alternative splicing may be a mechanism of activating the *c-myb* gene and unleashing its oncogenicity and could potentially be useful as a novel biomarker in pre-B-ALL.

Our preliminary data are very promising, but are based on analyzing a very small cohort of leukemia samples. We need to confirm these results using a larger data set. To have 80% statistic power, we would need to analyze at least 100 pre-B-ALL patient samples. However, it is not feasible to use a traditional polony assay to analyze such a large set of samples. It is important to develop a high throughput assay. Here, we describe two methods that have multiplex and high-throughput capacity to monitor *c-myb* alternative splicing at a scale and ease heretofore unattainable. In the bead based polony assay, each *c-myb* cDNA molecule was amplified with an agarose bead within an emulsified microdroplet. After amplification, each agarose bead was bound by thousands of copies of DNA identical in sequence to the original molecule. Therefore, each bead in this assay was like the “PCR colony” in the traditional polony assay, representing each individual starting cDNA molecule. Exon-structure of each cDNA molecule could be simply assessed by hybridizing with exon-specific probes via automatic polony sequencing system. Simultaneous analysis of many samples could be easily accomplished by adding barcodes to patient samples in the process of cDNA synthesis, and the identity of each cDNA molecule (each agarose bead) could be interrogated by hybridizing with

barcode-specific probes. We applied this method to analyze 4 leukemia patient samples in parallel and the results suggest this method has high-throughput capacity and could efficiently exotype mRNA transcripts in multiple samples.

We have also addressed the possibility of adapting Next-Generation Sequencing methods (NGSMs) into our application by developing a novel method to construct the barcode-labeled mRNA libraries for each individual *c-myb* mRNA molecule in each patient. We further adopted dual primer emulsion PCR method, which combines the emulsion PCR and bridge amplification on beads, to amplify the short DNA fragment libraries for paired-end next generation sequencing. We named this novel approach “barcoded-mRNA-Seq”. This method can map out an entire cDNA molecule, which fills up an important niche in the applications of Next-Generation Sequencing Methods (NGSMs) for profiling alternative splicing of the genes. We applied “barcoded-mRNA-Seq” to sequence *c-myb* alternatively spliced variants in a large cohort of pediatric pre-B-ALL patient samples. We showed the utility, throughput, accuracy and robustness of this approach using different quality control methods. Overall, our work provides important clues to the molecular pathogenesis of pre-B-ALL this high malignant leukemia and opens up new avenues for improved diagnosis and identification of novel targets for pharmaceutical interventions.

Results

BEAMing (beads, emulsion, amplification and magnetics), a high throughput method to detect *c-myb* alternative RNA splicing

In general, new polony sequencing (from beads) has replaced the use of the original polony assay (in situ DNA amplification) in an acrylamide gel matrix. One of the key innovations was developed by Bert Vogelstein's lab at Johns Hopkins University, which was the emulsion PCR technology²³⁵, which is also used by 454 as well as by Applied Biosystems' SOLiD system. However, there still remain a few specific applications for the original polony approach. The most common reason for using the original polony method is to study long PCR amplicons. It is currently difficult to generate long amplicons (>1 kb) using the bead-based polony sequencing method. Due to the amplicon size limitations, the traditional polony assay is currently the best method for assaying large amplicons, such as characterizing the splicing patterns in long transcripts. In the previous chapter, I applied the in situ amplification polony assay to monitor *c-myb* alternative splicing in a small cohort of pediatric pre-B ALL patient samples. The results revealed that aberrant *c-myb* alternative splicing could be a novel biomarker in pre-B-ALL.

In order to further confirm the preliminary data and establish statistically significant links between the patterns or the levels of *c-myb* splice variants and clinical parameters, we need to analyze a large set of patient samples. Since the traditional polony assay could only analyze one

sample at a time and each sample needed to be hybridized with 3 rounds of exon probes manually, it was not feasible to apply that method to analyze a large set of patient samples. Emulsion PCR or BEAMing (Beads, Emulsion, Amplification and Magnetics) is a higher throughput adaptation of the polony assay suitable for analyzing a large set of samples. BEAMing is essentially a bead-based polony assay, which can amplify individual DNA molecules and simultaneously attach the PCR products to beads within emulsion droplets²³⁵. BEAMing permits the analysis of $>10^7$ beads simultaneously coated with DNA, which greatly increases the throughput of the polony-based assays. However, in traditional BEAMing, the size of the beads is 1 μm and the size of the emulsion microdroplets are about 10 times bigger than the beads, which highly restricts the size of the amplicon that could be amplified within the emulsion droplet²³⁶.

To address this problem and adapt this high-throughput assay in our analysis, we modified the traditional BEAMing and developed a novel bead based polony assay. In this bead based polony assay procedure, mRNA samples are converted to cDNA by reverse transcription, in which mRNA samples from different patients are labeled with unique barcodes (Figure 1). These labeled mRNA samples are then mixed and subjected to emulsion PCR. Emulsion PCR is conducted directly on the beads within the microdroplet. Ideally, each microdroplet only contains one bead with one cDNA molecule, so that the DNA products on each bead are amplified

from one single molecule and each bead with DNA products is similar to a “PCR colony” in traditional polony assay. However, unlike the traditional emulsion PCR, our approach uses 30 μm agarose beads instead of the traditionally used 1 μm magnetic beads and amplifies individual molecule in 100 μm (in diameter) emulsified microdroplets. With these modifications, 1.0-1.5 kb DNA amplicons could be easily amplified on the beads within the droplets, allowing us to map out the exon structure of the entire *c-myb* mRNA molecules. The structure and the identity of the DNA (from which patient) on individual bead could be interrogated by hybridizing with fluorescently labeled, exon-specific and barcode-specific probes. This entire exotyping process is very similar as the one described in traditional polony assay (see details in Chapter 2) but can be performed by an automated system.

To test the high-throughput and multiplex exotyping abilities of the agarose bead based polony assay, we examined the expression and the pattern of *c-myb* transcripts in 4 pediatric leukemia patient samples that have been previously investigated by traditional polony assay. In brief, *c-myb* cDNA samples were synthesized from 4 pediatric pre-B-ALL patient bone marrow RNA samples and were labeled with 4 different barcode sequences through reverse transcription. These cDNAs were then briefly amplified (8-10 cycles) with a forward primer complementary to exon 6 and a reverse primer specific to the universal T7 sequence T7 in the reverse primers used in the reverse transcription (Figure 1). After PCR, all

of the cDNA molecules were linearly amplified without PCR artifacts and the generated DNA molecules all contained the corresponding patient barcode sequences. The PCR products from the 4 leukemia patients were then mixed and subjected to the following emulsion PCR and exotyping.

Emulsion PCR, beads immobilization and mounting the flow cell onto the microscope stage were performed as described in “Materials and Methods” section (Figure 2A). To determine the structure of the DNA on the beads, 4 cycles of 3 to 4 fluorescent probes, which included 11 exon specific probes and 4 barcode specific probes (Table 1), were sequentially flowed through the polony sequencing system to hybridize with the beads. The images were obtained after each cycle of hybridization to record which beads were positive for each probe. The majority of the beads only showed one color after hybridizing with 4 specific barcode probes, indicating that the DNA products coupled to the bead were amplified from a single molecule (Figure 2B). The whole exotyping process was automated and was completed in 20 hours. After image acquisitions, a series of images generated from the sequential hybridization steps were translated into an exon map for each bead. Total 49 *c-myb* splice variants including the wild type *c-myb* transcript were detected in these 4 patients. Approximately 30-40% of the total *c-myb* transcripts in these samples were alternatively spliced variants. The patterns and the levels of *c-myb* splice variants identified by the bead based polony assay were very similar to the ones detected by the traditional polony assay (Table 2). For

example, the 9S/10 splice variant, the level of which has been correlated with poor leukemia prognosis, was detected in all 4 patients and the detected level in each patient was consistent with the previous results from the polony assay (Table 2). With this bead based polony assay, the profiles of *c-myb* splice variants in 4 patient samples could be obtained in one single experiment. We conclude that the bead based polony assay could be a powerful tool for multiplex profiling *c-myb* alternative splicing in a large set of patient samples.

Whole molecule *c-myb* exotyping by next generation sequencing

I. Generating individual short DNA fragment libraries

Compared to the original polony assay, the bead-based polony sequencing method can generate more readouts and works under more automated conditions. However, to perform the emulsion PCR with agarose beads, to recover and enrich the beads after emulsion PCR and immobilize them to the coverslips, is still difficult and time-consuming. Also, the cost and the capacity of the bead based polony sequencing cannot match the capabilities of newer next generation sequencing technologies. Since vast quantities of genetic data could be obtained affordably and with relative ease through these next generation sequencing methods, we were prompted to explore the feasibility of the use of these new tools in profiling *c-myb* alternative splicing in a large cohort of patient samples.

Next generation sequencing methods (NGSMs) have been widely used in different applications, such as ChIP-Seq^{237 238 239}, genome sequencing^{240 241}, and exon and transcriptome sequencing^{238,240,242,243}. The value of next generation sequencing technologies is in the high-throughput ability and low cost for sequencing a large number of samples²⁴⁴. At present, next generation sequencing platforms use slightly different technologies for sequencing, such as pyrosequencing, sequencing by synthesis or sequencing by ligation. However, most platforms adhere to a common library preparation procedure, with minor modifications. The procedure includes fragmenting the DNA (sonication or nebulization), followed by DNA repair and end polishing (blunt end or A overhang) and, finally, platform-specific adaptor ligation. Each short DNA fragment in this library is sequenced and reassembled to generate final results. However, these short DNA fragments in the library are derived from different DNA molecules, the final results can only represent the general DNA sequence information but not the sequence information of individual DNA molecule. While adequate for genome sequencing and exon profiling, this technique cannot map the exon combination of each individual mRNA transcript. To apply next generation sequencing methods to single molecule exotyping, we prepared a complex pool of sequencer-ready libraries for each individual mRNA transcript as described in “Materials and Methods” section. First, RNA samples were converted into cDNA. In the process of reverse transcription, every mRNA molecule from different patients was

labeled differently by using reverse primers containing 7 random nucleotides (N_7) and a 2 nucleotides patient barcode (BC) (there are a total of 8 different patient barcodes). After reverse transcription, these cDNA samples were amplified for 10-12 cycles, which ensures that all of the cDNA molecules were amplified simultaneously without bias and the PCR products arising from each cDNA molecule still contained the original barcode information. In this PCR reaction, an optimized amount of dUTP was included in the mix, which resulted in the generated PCR products containing randomly incorporated uracil moieties in the sequence. After PCR, the products were treated with UDG/APE1, which removed the uracil moieties and fragmented the DNA. This step is similar to the shearing step in the common way of making DNA libraries. The size of these fragments was from 200 bp to 2 kb, with the barcodes retained at one end (Figure 3 and Figure 5A).

The DNA fragments were end polished, A-tailed and then ligated to an adaptor (Acui-Inside Adaptor). After ligation, the DNA fragments that contained barcode information were selectively amplified with specific primers (Figure 3 and Figure 4). These PCR products were then circularized by intramolecular hybridization and ligation. The newly generated circular DNA molecules all contained the barcode information.

The circularized DNA molecules were digested with Acui to linearize them and were further ligated to the polonator-specific adaptors (FDV and RDV). The ligation products were amplified for 12-15 cycles with adaptor

specific primers (Figure 3). The resulting 180 bp PCR products were isolated and purified from the polyacrylamide gels (Figure 5E). These isolated products were the finalized DNA libraries and were the templates for the next generation sequencing reactions. Every DNA molecule in this library should contain a fragment of *c-myb* sequence and a barcode sequence identifying which patient and which original cDNA it was derived from. To test this, the final products were cloned in TA vectors and sequenced by Sanger sequencing method. The results showed that the template DNA fragments contained the FDV or RDV adaptor sequence on each end, the *myb* sequence and the barcode sequence, which flanked an inside DNA segment generated by circularization (Figure 7). The *c-myb* sequences span from exon 6 to exon 15 of the *c-myb* mRNA, and the barcode sequence showed which mRNA molecule the *c-myb* sequence originally derived from and which patient it was from (Figure 3 and Figure 7).

II. Dual Primer Emulsion PCR (DPePCR)

For the emulsion PCR step, we adapted the bridge amplification method used by Illumina. Instead of coupling both reverse and forward primers onto the solid phase, we coupled the primers to the magnetic beads (Dual-Primer emulsion PCR). Since the amplicons were confined to the droplets, the amplification efficiency was increased by including free primers in the aqueous phase (Figure 6A). With 120 PCR cycles, a single DNA fragment in an emulsion drop could be amplified effectively (Figure

6B and Figure 6C). After amplification, the dsDNA bridges were formed on the bead inside the emulsion droplets. These DNA bridges could be denatured, but the proximity of the complementary strands led them to immediately re-hybridize, which blocked annealing of the sequencing primers. To overcome this problem, during library construction, some enzyme restriction sites (MmeI, AclI or SnaBI) were included, so that the amplified products could be digested before sequencing. Here, we placed a restriction site SnaBI at the middle part of inside adaptor (Figure 6C). After digestion with SnaBI, the DNA bridges were cut into two DNA fragments that were both on the bead (Figure 6D) and then they were made single-stranded (Figure 6E). This provided the accessibility of both DNA strands to the sequence primers and the DNA products could be sequenced by standard sequencing by ligation (SBL).

The SBL sequencing strategy for Dual Primer ePCR (DPePCR) beads is similar to the one for standard ePCR beads. The only difference is that because there were two different paired-end fragments on the beads, both fragments could be sequenced independently from both the 3'-5' and 5'-3' directions using four different sequence primers (Figure 6E).

To validate the DNA bridges forming on the bead after DPePCR, the beads were treated with 0.1 M NaOH (no restriction enzyme digestion). After denaturing, a FAM-labeled oligo (complementary to the outside adaptor, FDV-FAM) was annealed to the bead-bound DNA fragment. The results indicated that the FAM-oligo could not hybridize to the DPePCR

products, which suggests the formation of double-stranded DPePCR product was in a bridged conformation (data not shown). Following restriction enzyme digestion, we were able to anneal the sequence primers and successfully sequenced the DNA on the beads from both ends of the fragment.

Theoretically, the DNA products on each bead arise from a single molecule, which is the critical requirement for all of the next generation sequencing platforms. To validate the clonal amplification, we performed SBL on the DPePCR beads in solution. When sequenced, the beads showed a single color during each SBL cycle, which indicated that the DNA molecules on the beads were clonal products of a single template molecule (Figure 6F).

III. Sequencing By Polonator

RNA from a total of 145 pediatric B-ALL patients was labeled with unique barcodes and grouped into 21 samples containing 4-8 differently labeled patient RNA samples. These 21 samples were converted into 21 libraries and further amplified on the beads for sequencing. The libraries were analyzed using sequencing by ligation method on the G007 Polonator (Westborough, MA, USA). The flow cell used in the Polonator contains 8 wells, which can simultaneously hold 8 samples. Prior to sequencing, the beads were treated with BS3 reagent, which provides covalent bond to the amino group on the amine-functionalized glass coverslip of the flow cell. The beads were then flowed in the individual

wells in which they were allowed to attach. Any unattached beads were removed by flushing the flow cell. A total of 4 sequence primers were sequentially flowed in (See Table 3), and each sequence primer could be used to sequence 6 to 7 nucleotides (Figure 8). The Polonator could capture and process four-color images of fluorescent beads in real time. After all of the SBL runs, the base calling program decoded the color of the beads to the corresponding nucleotide. For each cycle, each bead was dim in three of the four fluorescent colors, and bright in one, allowing an accurate base call to be made. The Polonator base caller output a separate read file for each well on the flow cell, with each read file containing ~ 20 million reads. For each bead, the optimal read length was approximately 26 bases, in which 13 bases were the DNA sequence of the *c-myb* gene and the other 13 bases (with 4 bases overlap) were the barcode sequence (Figure 7).

To be able to generate the final profile of *c-myb* alternative splicing in 145 patients, sophisticated bioinformatics tools are needed to analyze the raw data and reassemble the single molecule sequence. Preliminary data analysis has been run on the raw data files generated from 21 libraries. Briefly, all the beads with signals were first categorized into individual molecules from each patient using the 9 bp barcode. Then the bowtie program was used to align back the 13 bases read of *c-myb* sequence to the individual reference exon sequence of *c-myb* mRNA²⁴⁵. The beads with the reads that hit on the same exon were categorized into the same

exon group. After preliminary data processing, the reads of the beads for each library were categorized into different groups in which *c-myb* tags hit back to the same exon sequence and shared the same barcode sequence. Since a sequencing library was prepared and subjected to extensive amplification (PCR in the library preparation), the bias could be increased. We plotted the abundance of the tags of these reads found in each exon against the entire sequence of the exon (Figure 9). The plots showed that there were a few “hotspots” found but the majority of these tags were evenly distributed along the entire exon sequence, suggesting no obvious bias in the construction of the sequencing libraries.

Further analysis is needed to determine the combinatorial exon information of individual mRNA molecule and the abundance of the mRNA transcripts for each patient. Analysis of this large cohort of leukemia patient samples will allow us to generate statistically significant results, which will be important for evaluating whether *c-myb* alternative splicing could be used as a signature to either classify leukemia patient sample or a biomarker to predict the outcome of leukemia patients.

Discussion

Several lines of evidence show that *c-myb*, a key regulator of normal hematopoiesis, can be activated through multiple mechanisms and is a *bona fide* oncogene in human cancers. Our previous study showed that compared to normal hematopoietic cells, leukemia patients express a different and more diverse set of *c-myb* splice variants encoding a family

of c-Myb proteins with different C-terminal domains and different transcriptional activities. Importantly, the expression levels of some splice variants correlate with clinical outcome. The above findings indicate that aberrant alternative splicing of *c-myb* mRNA could be a novel and unexpected mechanism for expressing activated forms of c-Myb protein that participate in leukemogenesis and profiling *c-myb* splice variants could provide a unique and novel approach of classifying leukemias. However, current methods, such as QRT-PCR and in-situ polony assay, are not feasible to profile *c-myb* splice variants in large numbers of patient samples, which points to the need for new and improved methods of exotyping. Here, we have developed two bead-based methods, and both methods utilized emulsion PCR and were capable of simultaneous exotyping many RNA (cDNA) samples. One method was combined with automated polony sequencing stage to exotype *c-myb* splice variants by hybridization. The other method was coupled to the next-generation technology to sequence *c-myb* splice variants. We successfully applied both methods to analyze *c-myb* splicing patterns in several patient samples in parallel.

The polony assay is a method for in-situ amplifying a large number of distinct individual DNA molecules, which is ideal to analyze long amplicons and map the transcript splicing patterns at the single molecule level. Bead based emulsion PCR is a method to amplify individual DNA molecules on beads. Each bead coated with DNA can be considered as a

polony in the traditional polony assay. These beads cannot overlap, which makes the image analysis easier than traditional polony assays.

Furthermore, DNA on the beads is more concentrated than gel-base assays, increasing the signals to noise ratio. Overall, this approach allows us to simultaneously analyze $>10^7$ beads with DNA, which greatly increases the throughput of the polony-based assays. However, there is a critical limitation of current bead based technology; it is not possible to generate DNA coated beads with amplicons longer than a few hundred bases. In order to solve this problem, we replaced the traditional 1 μm magnetic beads with 30 μm agarose beads and generated microdroplets with the diameter around 100 μm , which were stabilized by the surrounding “microfines” during emulsifying. With these modifications, the larger surface area of the beads and the increased size of the compartments allowed the amplification of amplicons up to 1.5 kb and binding them to the beads via emulsion PCR. Compared to traditional polony assay, this bead-based method can be coupled with the automated polony sequencing system and the entire exotyping process can be run automatically. In addition to the automated system, which reduces a lot of manual labor and accelerates the exotyping process, multiple parallel analyses could further facilitate the throughput. Here, molecular barcoding of each patient and identifying them later through hybridization with specific probes were shown useful for this purpose. cDNA samples from 4 different patients were added with unique barcodes and mixed for the

bead-based polony assay. A total of 2×10^5 beads with DNA were queried and the profiles of *c-myb* alternative splicing in 4 patient samples were obtained simultaneously in one run.

Another way to increase the throughput would be to utilize next-generation sequencing methods. Next-generation sequencing methods (NGSMs) have been applied to a wide range of fields in biology and medicine. In addition, NGSMs have changed the scope and speed of standard sequencing methods by several orders of magnitude. In spite of the massive amounts of throughput generated in a single run, the main drawback to NGSMs is the relatively short read lengths. For profiling a single gene transcriptome, which means reading out the exon composition of an entire mRNA transcript up to several kilobases, the current NGSMs, even Roche 454 GS FLX with read length up to 600 bp, are not sufficient. In order to broaden the application scope of these NGSMs and adapt them in our research, we developed a novel protocol for the construction of sequencing libraries for each individual *c-myb* cDNA molecule. Next-generation sequencing libraries for amplification-based methods comprise short DNA (50 bp-300 bp) templates usually called “sequencing features”. In our prepared libraries, these sequencing features contain a short fragment from *c-myb* transcripts and the barcode sequence that indicates which cDNA molecule and which patient this *c-myb* fragment is originally from. By ligating to platform-specific adapters, these sequencing features could be amplified using the adapter sequence as priming sites or

hybridization targets. Here, we used Polonator G.007 sequencing to separately but in parallel sequence each of the millions of single clonally amplified targets. The generated read length from the Polonator is 26 bp. For relatively less complex *c-myb* gene transcriptome, short reads of the *c-myb* gene are sufficient to generate an alignment for individual *c-myb* mRNA transcript according to their barcode information. Our data showed that the generated short reads of *c-myb* are at high coverage of the *c-myb* gene, indicating there was no bias during library construction. Our results also suggest that these short reads of *c-myb* could be easily grouped, based on the reads of the barcode, into individual cDNA molecule for each patient, and these grouped reads could be mapped back to each exon of the *c-myb* gene. Hence, the exon combinatorial information of individual molecule could be mapped out and *c-myb* alternative splicing could be profiled for each patient.

Next-generation sequencing could generate volumes of data that are both a blessing and a curse. We faced the common and major problem that often happens using next generation sequencing; that is, the handling of massive amounts of sequencing data. It is necessary to develop effective bioinformatics tools to assist us in analyzing the raw data and generating the final profiles of *c-myb* alternative splicing in patients. With the data processed, the clinical value of *c-myb* alternative splicing will be evaluated by correlating alternative spliced isoforms with critical clinical parameters of these patients. Compared to the previous agarose bead

based polony assay, this NGSM provides a more efficient and cost-effective means to profile *c-myb* alternative splicing in multiple samples simultaneously. Adaptation of our sequencing library protocol may allow us to perform a comprehensive profiling of alternatively spliced mammalian transcriptomes.

Overall, our study has explored the possibility of multiplexed profiling *c-myb* alternative splicing in large amounts of samples by combining molecular barcodes, emulsion PCR with polony sequencing system/next generation sequencing platforms. The bead based polony assay breaks the limitation of conventional emulsion PCR that requires the analyzed amplicon shorter than 300 bp. The novel way of constructing libraries for next generation sequencing (“barcoded mRNA-Seq”) circumvents certain limitations of RNA sequencing, in which the analysis of the full sequence of individual mRNA molecule is impossible. In a single experiment, “barcoded mRNA-Seq” provides a wealth of qualitative as well as quantitative information that cannot be obtained by any other method. Indeed, “barcoded mRNA-Seq” is a promising tool in the analysis of a single gene transcriptome or whole transcriptomes in a disease across hundreds of individuals for clinical purpose. With this unique application, NGSMs could be used as a powerful tool in the studies that evaluate aberrantly expressed splice variants as a disease biomarker or the studies that discover new oncogenic forms of proteins as targets for developing novel therapeutic strategies.

Materials and Methods

I. Bead-based Polony assay

cDNA sample preparation

Patient bone marrow RNA samples were obtained from the Children's Oncology Group (COG) Acute Lymphocytic Leukemia (ALL) Cell Bank. *c-myb* cDNA was synthesized by using a first-strand cDNA synthesis kit (Invitrogen, Carlsbad, CA). In order to monitor *c-myb* alternative splicing in several patient samples in parallel, the reverse primers used in cDNA synthesis were comprised of the *c-myb* gene-specific reverse primer followed by a 20 bp barcode sequence and a universal T7 primer for the later amplification and emulsion PCR (Figure 1). Different patient RNA samples were reverse transcribed with unique reverse primers having their own barcodes (Table 1). The generated cDNAs were concentrated and purified through MiniElute columns (Qiagen, Valencia, CA) and were then subjected to PCR (8-10 cycles of 94 °C for 30 sec, 58 °C for 35 sec, 72 °C for 1min 45 sec) in a 50 µl reaction consisting of 5 units of Jumpstart (Sigma, St Louis, MO), 1x PCR buffer, 2.5 mM MgCl₂, 200 µM dNTPs, and 200 nM of reverse primer T7 Rev and forward primer SN-3 Forward. These products were the final templates for emulsion PCR.

Preparation of DNA Primer Coupling Beads

Packed beads from a 1 ml N-hydroxysuccinimide ester (NHS)-activated Sepharose HP affinity column (Amersham Biosciences, Piscataway, NJ) were activated as described in the product literature (Amersham Pharmacia Protocol # 71700600AP). In brief, 1 ml of 1 mM amine-labeled *c-myb* primer (5'-Amine-3 sequential 18-atom hexa-ethyleneglycol spacers CCTATCCCCTGTGTGCCTTG-3') (IDT Technologies, Coralville, IA, USA) in standard coupling buffer (0.2 M NaHCO₃, 0.5 M NaCl, pH 8.0), was injected into the column at a speed of 0.5 ml/min. The column was then sealed and incubated at room temperature for 30 min, allowing the amine labeled primers to couple to the beads. After coupling, the column was washed with buffer A (0.5 M ethanolamine, 0.5 M NaCl, pH 8.3) and buffer B (0.1 M NaOAc, 0.5 M NaCl, pH 4) alternatively 12 times each to remove the uncoupled primers and deactivate the uncoupled NHS group on the beads. The packed beads were removed from the column, resuspended in 2 ml Wash 1E buffer [10 mM Tris-Cl (FisherScientific), 50 mM KCl (Sigma, St.Louis, MO), 2 mM EDTA (USB Corporation, Cleveland, OH), 0.01% Triton X-100 (Dow Chemicals, Midland, MI)] and stored at 4 °C until needed.

PCR Reaction Mix Preparation and Formulation

To reduce the possibility of contamination, the PCR reaction mix was prepared in a UV-treated laminar flow hood. For each emulsion PCR reaction, 175 µl of reaction mix [1x Platinum HiFi Buffer (Invitrogen,

Carlsbad, CA), 0.25 mM dNTPs (Promega, Madison, WI), 3.5 mM MgSO₄ (Invitrogen, Carlsbad, CA), 0.1% BSA (NEB, Ipswich, MA), 0.01% Tween-20, 0.003 U/μl thermostable pyrophosphatase (NEB, Ipswich, MA), 0.05 μM forward primer SN-3 and 0.95 μM biotinylated-reverse primer T7 rev (Invitrogen, Carlsbad, CA) and 0.15 U/μl Platinum Hi-Fi Taq Polymerase (Invitrogen, Carlsbad, CA)] were prepared in a 1.5 ml tube. 15 μl of the reaction mix was removed and stored in an individual 200 μl PCR tube for use as negative control. Both the reaction mix and the negative controls were stored on ice until needed. Additionally, 240 μl of mock amplification mix [1x Platinum HiFi Buffer (Invitrogen, Carlsbad, CA), 2.5 mM MgSO₄ (Invitrogen, Carlsbad, CA), 0.1% BSA, 0.01% Tween-20] for every emulsion were prepared in a 1.5 ml tube, and similarly stored at room temperature until needed.

Emulsification and Amplification

The emulsification process creates a heat-stable water-in-oil emulsion with approximately 1,000 discrete PCR microreactors per μl, which serve as a matrix for single molecule, clonal amplification of the individual molecules in the target library. The reaction mixture and the *c-myb* forward primer-coupled beads for a single reaction were emulsified in the following manner: 160 μl of PCR reaction mix was mixed with 10 μl coupled agarose beads ($>10^6$), then 2 μl of desired amounts of cDNA samples was added into the PCR-bead mixture. 400 μl of emulsion oil [50 % (w/w) DC 5225C Formulation Aid (Dow Chemical Co., Midland, MI),

25% (w/w) DC 749 Fluid (Dow Chemical Co., Midland, MI), and 25% (w/w) AR20 Silicone Oil (Sigma, St Louis, MO)] were dispensed into a flat-topped 2 ml centrifuge low retention tube (Bioexpress, Lodi, CA). The 240 μ l of mock amplification mix was then added to 400 μ l of emulsion oil, the tube was capped and placed on a TissueLyser (Qiagen, Valencia, CA). The emulsion was homogenized for 5 min at 25 HZ to generate the extremely small emulsions, or “microfines”, that confer additional stability to the reaction.

After the microfines were formed, the combined beads and PCR reaction mix were briefly mixed and added. The mixture was then placed on the TissueLyser and homogenized for 4 min and 30 sec at 15 HZ. The lower homogenization speed created water droplets in the oil mix with an average diameter of 100 to 150 μ m, which is sufficiently large to amplify longer amplicons (>1 kb).

The total volume of the emulsion was approximately 800 μ l contained in one 2 ml flat-topped low retention tube. The emulsion was dispensed into 7-8 separate PCR tubes each containing roughly 100 μ l. The tubes were placed in an MJ thermocycler along with the 15 μ l negative control made previously. The following cycle times were used: 94 °C for 4 min, for Hotstart initiation; 60 cycles of 30 sec at 94 °C, 2 min at 58 °C, 1.5 min at 68 °C, for amplification; 13 cycles of 30 sec at 94 °C, 6 min at 63 °C, for hybridization extension. After completion of the PCR program, the

reactions were removed and the emulsions either broken immediately or the reactions stored at 4 °C before initiating the breaking process.

Breaking the Emulsion and Recovery of Beads

The emulsion of amplified material in each PCR tube was removed and combined in a 2 ml flat-topped low retention tube. 100 µl AR2 silicon oil (Sigma, St Louis, MO) was added to each PCR tube to remove the remaining emulsion solution into the same 2 ml tube. The tube was vortexed and then centrifuged at high speed for 2 min. The supernatant was removed and 1 ml of AR2 silicon oil was added to the tube. The tube was mixed on a vortex and spun in a benchtop centrifuge and the supernatant was removed. This silicon oil wash step was repeated 6 times. Then 1 ml of isopropyl alcohol (General Chemical Corp, Parsippany, NJ) was added to the tube and vortexed until mix well. The tube was then spun at high speed for 1 min and the supernatant was removed. This washing step with isopropyl alcohol was repeated 6 times as well. After isopropyl alcohol cleaning, 1 ml of 80% Ethanol/1x Annealing Buffer (80% Ethanol, 20 mM Tris-HCl, pH 7.5, 5 mM Magnesium Acetate) was used for washing, which was repeated 3 times. Then the beads were washed with 1 ml of 1x Annealing Buffer with 0.1% Tween (0.1% Tween-20, 20 mM Tris-HCl, pH 7.5, 5 mM Magnesium Acetate) 3 times. The beads were finally washed with 1 ml Wash 1E Buffer 3 times and ddH₂O 2 times. The final clean beads were resuspended in 200 µL EB buffer and stored at 4 °C until needed.

Enrichment of the Beads

Up to this point the beads were comprised of both beads with amplified DNA products, and null beads with no amplified product. The enrichment process was utilized to selectively capture beads with amplified DNA while rejecting the null beads. In brief, the clean agarose beads were washed with Bind and Wash Buffer (B+W Buffer) (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl) 3 times and resuspended in 150 μ l B+W Buffer. While washing the agarose beads, 30 μ l well-resuspended Dynabeads M-280 streptavidin beads (Invitrogen, Carlsbad, CA) was also washed with B+W Buffer 3 times and resuspended in 100 μ l B+W Buffer. The washed magnetic bead solution was added into the washed agarose beads. The bead mixture was mixed by gentle flicking and mixed at room temperature for 1 hr, allowing the streptavidin beads to bind to the biotinylated DNA strand on the agarose beads. The bead mixture was then spun at 2,000RPM for 3 min, after which the beads were gently flicked until the beads were resuspended. Then 300 μ l of B+W Buffer was added and incubated at room temperature for 5 min. Following the incubation, the bead mixture was placed onto a DynaMag-2 magnetic Particle Concentrator (MPC) (Invitrogen, Carlsbad, CA), and the beads were left undisturbed for 2 min to allow the beads to pellet against the magnet, after which the supernatant (containing the null agarose beads) was carefully removed and discarded. This step was repeated 4 times with 300 μ l B+W Buffer to ensure removal of all null agarose beads. To remove

the biotinylated DNA strand and the magnetic streptavidin beads from the selected agarose beads, the bead mixture was resuspended with 0.1 M NaOH and incubated at room temperature for 5 min. Then the bead mixture was vortexed and placed on MPC. The supernatant, containing the enriched agarose beads with the immobilized DNA strands, was transferred to a separate 1.7 ml low retention tube. The NaOH wash was repeated additional time and the supernatant was combined with the first supernatant. The beads attached on the MPC were discarded. The supernatant containing the enriched beads was placed on the MPC again, and the supernatant transferred to a fresh 1.5 ml tube, ensuring maximal removal of remaining magnetic streptavidin beads. The final enriched beads were washed with 500 μ l Wash 1E Buffer 2 times and resuspended in 200 μ l Wash 1E Buffer. The average yield for the enrichment process was 30% of the original beads added to the emulsion.

Fluorescent labeled Probe Hybridization

Before immobilizing the beads to the coverslip, the enriched beads were checked for the signals on the attached DNA. 30 μ l of the enriched beads were used to do hybridization with different fluorescent labeled probes, either specific to different exons or to the patient barcode sequences. The hybridization step was mainly carried out as follows: a mixture of different fluorescent labeled probes was added to the enriched bead solution to make the final concentration of each of these probes 1 μ M. The beads-probes mixture was mixed on a vortex and placed in a MJ

thermocycler (BioRad, Hercules, CA) for the following 3-stage hybridization program: 4 min at 94 °C, 10 min at 58 °C, and 10 min at 42 °C. Upon the completion of the hybridization program, the beads were washed with 100 µl EB buffer 2 times and resuspended in 20 µl EB Buffer. The images of these labeled beads were then recorded with a fluorescent microscope (Nikon, Melville, NY) (Figure 1).

Bead Immobilization

The large size of the agarose beads (30 µm) makes it really difficult to attach to the coverslip using traditional chemical conjugation of the hydroxyl- group on the DNA to the amine group on the coverslip. Therefore, the best way to immobilize the agarose beads to the coverslip is by using a polyacrylamide gel. 40 µm thick 6% polyacrylamide gels containing the agarose beads were prepared as described ²¹⁸, except that polymerization was initiated by addition 1.25 µl 5% TEMED (BioRad, Hercules, CA) and 1.875 µl 1% ammonium persulphate solution (EM Science, Gibbstown, NJ). In order to allow the beads to settle onto the surface of the gel, the coverslip was inverted during polymerization. The whole polymerization was performed in a vacuum chamber. The coverslip with the immobilized beads was placed into a flow cell and then mounted to the Polony sequencing system, consisting of a fluorescent microscope, automated stage, automated filter wheels and external shutters, PIFOC objective Z-control, 1 megapixel camera, autosampler, syringe pump, and flow cell with temperature controller.

Exontyping on the polony sequencing system

The polony sequencing system has been successfully used for bacterial genome sequencing (unpublished data by Jeremy Edwards group). The cyclical delivery of exontyping reagents into the flow cell and washing off the exontyping reaction byproducts from the cell was achieved using an automated fluidics system. The controller program was written in MATLAB, specifying the reagent name (wash buffer, exon probes, annealing buffer and so on), flow rate and duration of each step, and specific temperature for each step. The whole program ran as following: Wash the flow cell with 500 μ l Wash 1E 3 times; flow in 400 μ l 70% formamide and denature at 56 °C for 12 min; wash with 500 μ l H₂O at 30 °C; flush the flow cell with Wash 1E at 42 °C for 6 min; flow in 500 μ l 3x SSPE [diluted from 20x SSPE buffer (Ambion, Austin, TX)] with 0.01% Tween-20 with 4 different fluorescent labeled exon probes (2 μ M of each exon probe) or barcode probes, anneal at 90 °C for 12 min and 60 °C for 35 min; wash with 500 μ l H₂O once and 4 times of 500 μ l 0.1x SSPE (Ambion, Austin, TX) with 0.01% Tween-20 at 45 °C; the duration time of each wash within the hybridization step was 4 min. For each hybridization step, the total run time was about 120 min.

Image Acquisition

The bright field image for each flow cell was recorded preceding the entire exontyping run. After acquiring the bright field image, the flow cell was fixed with no movement during the exontyping process. Based on the

fixed position of the flow cell, the image analysis software could determine the location (in CCD pixel coordinates) of each bead within the flow cell. In operation, the entire coverslip was defined as an imaging area and was split into several frames (400 pixel X 400 pixel). For each frame, the attached CCD moved up and down to autofocus on the beads in this frame, set the image position for this frame and then recorded the image. This autofocus and recording process was repeated across the entire imaging area. Then, after hybridization, based on the previously recorded image position of each frame, the CCD went across the entire imaging area and recorded the fluorescent images of each frame. All of the images of the same frame were overlapped and processed in the same way as described in Chapter 2. Image analysis was also carried out in the same way as described in Chapter 2.

II. Whole molecule *c-myb* exotyping by Next-Generation

Sequencing

Sample preparation and library construction

Total RNA of patient samples was provided by the Children's Oncology Group (COG). *c-myb* cDNA was synthesized with a gene-specific reverse primer (Oligo Acui -7mer-2BC-RT) using a first-strand cDNA synthesis kit (Invitrogen, Carlsbad, CA). This gene-specific primer consists of a specific sequence complementary to the *c-myb* gene, 2 barcode nucleotides for labeling each patient, 7 random nucleotides for

labeling each cDNA molecule synthesized from the patient samples, and 6 nucleotides of Acul restriction sites. Approximately 8×10^{-4} pmol of cDNA from each patient was first subjected to PCR (12-15 cycles of 94 °C for 30 sec, 58 °C for 35 sec, 72 °C for 1 min 35 °C) in a 50 µl reaction consisting of 5 units of Jumpstart (Sigma, St Louis, MO), 1x PCR buffer, 2.5 mM $MgCl_2$, 200 µM dNTPs, 10 µM dUTP, and 200 nM of reverse primer Rev-Oligo Acul and forward primer SN-3-For. This process resulted in amplification of all splice variants of *c-myb* between exon 6 and exon 15 (~1.4 kb) with randomly incorporated dU moieties. The PCR products were treated at 37 °C for 60 min, 80 °C for 20 min with 5 units of a Uracil DNA Glycosylase (UDG) (NEB, Ipswich, MA) /APE1 (Enzymatics, Beverly, MA) cocktail to create random fragments with average size of about 700 bp. The products generated from 5 to 8 patient samples with different barcodes were mixed and purified with MiniElute column (Qiagen, Valencia, CA). Approximately 0.3 pmol of purified DNA was treated with 1.5 µl End-Repair enzyme mix (Epicentre Biotechnologies, Madison, WI) and incubated at 25 °C for 45 min and 75 °C for 15 min in a reaction consisting of 1x End-Repair buffer, 100 µM dNTP Mix, and 100 µM ATP, to create 5' phosphorylated blunt termini. The end-repaired DNA fragments were then 3'-A tailed with 7.5 units Bsu DNA polymerase (NEB, Ipswich, MA). These A-tailed, blunt-ended DNA fragments were ligated to 10-fold molar excess of synthetic inside adaptor (Acul-Inside Adaptor) for 30 min at 25 °C in a reaction containing 50 mM Tris-HCl (pH 7.8), 10 mM

MgCl₂, 1 mM ATP, and 1,000 units of T4 DNA ligase (Enzymatics, Beverly, MA). The Acul-Inside Adaptor has one terminus with 4 bases of 3'- protruding and the other terminus with 5'-T overhang and 3'-phosphate. In this T4 DNA ligation reaction, 5'-T overhang terminus of the inside adaptor was ligated to 3' A overhang DNA termini, producing an intermediate structure. And the ligated DNA products had both termini with 4 bases of 3'- protruding. The DNA was incubated for 30 min at 37 °C with 10 units of ExonIII (NEB, Ipswich, MA), heat inactivated at 75 °C for 20 min, then incubated with 10 units of Exon I (NEB, Ipswich, MA) at 37 °C. This process is to degrade unincorporated inside adaptors and unligated DNA, and only the ligated DNA with 3'-overhang (4 bases) on both termini, which is resistant to ExoIII, would survive.

Half of the ligated DNA was subjected to PCR (25-30 cycles of 95 °C for 30 sec, 58 °C for 30 sec, 72 °C for 1 min 30 sec) in a 400 µl reaction consisting of 20 units of PfuTurbo Cx (Stratagene, La Jolla, CA) 1x pfu Turbo Cx buffer, 200 µM dNTPs, and 200 nM Circularization-U primer sets. This process selectively amplified the DNA fragments containing the random 7 bases and barcode sequences on the left side and inside adaptor Acul-In on the other side, to produce PCR product incorporating dU moieties at specific locations within both termini. The PCR products were treated with 10 units of a UDG/APE1 cocktail to create both termini with complementary 3' overhangs. Approximately 1.2 pmol Minielute-purified DNA was diluted to a concentration of 3 nM in a reaction

consisting of 16.5 mM Tris-OAc (pH7.8), 33 mM KOAC, 5 mM MgOAc, and 1 mM ATP, heated to 55 °C for 10 min, and cooled to 16 °C for 10 min, to favor intramolecular hybridization (circularization). The reaction then was incubated at 16 °C for 1h with 5400 units of T4 DNA ligase (Enzymatics, Beverly, MA) to form monomeric dsDNA circles containing two Acul sites right next to the inside bridge formed by intramolecular hybridization. The circle was incubated at 37 °C for 30 min with 100 units of Exonuclease III/Exonuclease I to eliminate residual linear DNA.

Approximately 0.6 pmol of this purified circle was digested at 37 °C for 1h with 12 units of Acul (NEB, Ipswich, MA) according to the manufacturer's instruction to generate linear dsDNA structures containing the bridge flanked by two DNA fragments of either 14 bases or 16 bases, in which one is part of the *c-myb* gene and the other contains the 7 random bases and 2 bases of the patient barcode. The digested DNA was end-repaired with 1.5 µl End-Repair Enzyme Mix as above, incubated for 30 min at 37 °C in a reaction containing 1 mM dATP and 7.5 units of Bsu DNA polymerase (NEB, Ipswich, MA) to add 3' A overhang. The resulting 5' phosphorylated, 3' A overhang DNA was incubated for 1 h at 25 °C in a reaction containing 50 mM Tris-HCl, 10 mM MgCl₂, 5 mM dithiothreitol, 1 mM ATP and 1500 units of T4 DNA ligase, and two 10-fold molar excess of adaptors FDV and RDV, of which both were designed to ligate to the 3'A overhang of the DNA, thereby yielding a ligated DNA product with 4 bases of 3' protruding termini. The ligated DNA products were treated with

10 units of Exonuclease III, and then with 10 units of Exonuclease I to eliminate the unincorporated adaptors and the inefficiently ligated DNA (either unligated DNA or the DNA with only one end ligated to the adaptor). PCR was performed on the resulting DNA with primers FDV-BC-PCR and RDV-BC-PCR (12-15 cycles of 94 °C for 30 sec, 64 °C for 20 sec, 72 °C for 20 sec) in a 400 µl reaction consisting of 40 units of Jumpstart (Sigma, St Louis, MO), 1x Jumpstart buffer, 2.5 mM MgCl₂, 200 µM dNTPs and 200 nM each PCR primer, to specifically amplify the ligated DNA containing both FDV and RDV adaptors. After MiniElute purification, the PCR products were run on a 10% polyacrylamide gel (BioRad, Hercules, CA), and the DNA fragments with the length of approximately 180 bp, which contain both FDV and RDV adaptors, and the inside bridge formed by circularization, were extracted from the gel and recovered by phenol-chloroform precipitation.

Library Construction QC

To assess the efficiency of each step, intermediate constructs were assayed by PCR amplification with Jumpstart DNA polymerase (Sigma, St Louis, MO) and specific primer sets. To check the efficiency of the ligation to inside adaptor, PCR 1 was performed with the inside adaptor-specific primers (Figure 5B). To check the efficiency of the reaction of circularization, PCR 2 was performed on the circular DNA product after being ligated by T4 ligase or the generated DNA products without using T4 ligase (Figure 5C). To check whether the circle dsDNA was efficiently

digested by AclI, PCR 3 was performed on the linearized DNA with the inside bridge-specific primers but no PCR product was generated (Figure 5D).

To assess library construct structure and coverage bias, the final 180 bp library DNA was PCR-amplified (PCR 4) with Jumpstart Taq DNA polymerase (Sigma, St Louis, MO) and FDV- and RDV-adaptors specific PCR primers (Figure 5E). These PCR products were then cloned with pCR®2.1 TA cloning kit (Invitrogen, Carlsbad, CA), and plasmid mini-prep (Qiagen, Valencia, CA) was used to extract DNA from independent colonies. Sequence information was collected from these DNA products by DNA sequencing service (Health Science Center, UNM, NM) and blasted against the *c-myb* gene sequence (Accession Number: U22376).

Coupling of dual-biotinylated primer to the streptavidin beads

As previously described²⁴⁶, to bind the PCR primers to MyOne streptavidin beads (Invitrogen, Carlsbad, CA), 240 µl MyOne streptavidin beads solution were mixed with 240 µl B+W buffer by inversion. The beads were then placed on MPC to remove liquid by pipetting. The beads were washed twice with 480 µl B+W buffer, then resuspended in 480 µl B+W buffer. 2.4 µl FDV2-dual biotin primer (1 mM) and 2.4 µl RDV2-dual biotin primer (1 mM) were then added into beads solution. The mixture was mixed briefly and incubated at room temperature for 20 min with mixing. After incubation, the beads were washed three times with 600 µl

B+W Buffer. The final pellet beads were resuspended in 640 µl TE and stored at 4 °C.

Dual Primer Emulsion PCR

Emulsion PCR was performed as described²⁴⁶. 2 µL aliquots of constructed library (at the appropriate concentration to allow a single molecule per drop) was mixed with 265 µl Dual Primer emulsion PCR (DPePCR) master mix (see “Dual Primer ePCR Master Mix” table), 13 µl Taq B (Enzymatic, Beverly, MA), 0.1 µM FDV-BC-PCR and 0.1 µM RDV-BC-PCR and 12 µl Dual Primer beads. The mixture was added to 1 ml silicone oil phase (see details in “Silicone Oil Phase”) in a 2-ml tube and put on a TissueLyser II (Qiagen, Valencia, CA) to emulsify (13.5 Hz for 3 min). The DPePCR emulsion mixture was equally dispensed into PCR tubes and amplified in a MJ thermocycler (94 °C 3 min, 100 cycles of 90 °C 15 sec, 58 °C 2 min, 72 °C 45 sec, 20 cycles of 90 °C 15 sec, 60 °C 5 min, then 72 °C for 15 min). For the same library, to be able to fill up one lane on the flow cell, 10 of the above reactions are required. The emulsion beads were broken and cleaned using the protocol in the “Breaking Emulsion”. Finally, all of the beads were suspended in 1 ml TE containing 0.1% Triton X-100.

The DPePCR beads were digested with 5 µl SnaBI (5000 U/ml) (NEB, Ipswich, MA) in 1x NEBuffer 4 by incubating at 37 °C for 60 min. After SnaBI digestion, the beads were washed 2X with 1x NXS [10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0), 100 mM NaCl, and 1% Triton X-

100] and TE with 0.1% Triton. The beads were then treated with 0.1 M NaOH for 5 min to single strand DNA and washed 2 times with 1x NXS and TE with 0.1% Triton. The beads were finally resuspended in 25 μ l 1x PBS buffer (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na₂HPO₄• 2H₂O, 1.76 mM KH₂PO₄, pH 7.4) before loading onto the flow cell for polonator sequencing.

Polonator Sequencing

To test the sequencing before loading the beads onto the flow cell, the beads with DNA products were checked by in-solution sequencing by ligation method. For sequencing of the FDV strand in 3'-5' direction, the beads were annealed with 10 μ M sequence primer 5'phosphorylated-Sequence primer FDV-Minus at 75 °C for 5 min, 55 °C for 5 min and 25 °C for 10 min, and washed three times with Wash 1E buffer. Then 1 μ l T4 DNA ligase (600 unit/ μ l) and 200 μ l ligation buffer [2 μ l of each nonamer (300 μ M)(IDT, Coralville, IA) for one queried position and 200 μ l 1x DNA quick ligase buffer (Enzymatics, Beverly, MA)] were added to the beads and incubated at room temperature for 30 min. The beads were then wash with 2 times 1x NXS buffer and TE/0.1% Triton X-100, and resuspended in 50 μ l TE/0.1% Triton for observation on fluorescent microscope.

The DPePCR beads were interrogated from the FDV and RDV strand respectively using the same strategy from both the 3'-5' and 5'-3' directions. As describe in Polonator sequencing protocol, during the run, a series of sequencing primers (see details in Table 3) are flowed through

the cells, where they hybridize to the FDV strand or RDV strand. For 3'-5' direction sequencing, the sequence primers have 5'-phosphorylated modification. While for 5'-3' direction sequencing, the sequence primers have no modification, however, the beads need to be capped with dideoxynucleotide before hybridization and ligation (see details in "Capping with dideoxynucleotide reaction"). Once a sequence primer is hybridized, a mixture of fully degenerate nonanucleotides ('nonamers') and T4 DNA ligase (600 unit/ μ l) is flowed into the cell; each of the nonamer mixture's four components being labeled with one of four fluorophores (see Table 3), which correspond to the base type at the query position. The fluorophore-tagged nonamers selectively ligate onto the sequence primer, providing a fluorescent signal that identifies the corresponding base on the DNA library tag.

Once the probes are ligated, fluorescently labeling the beads, the polonator images the array in four colors. Each bead on the array will fluoresce in only one of the four images, indicating whether there is an A, C, G, or T at the position being queried. After imaging, the Polonator chemically strips the array the annealed primer-fluorescent probe complex, as well as residual enzyme, using 0.1 M NaOH.

After each cycle has been completed, and the primer-fluorescent probe complex has been stripped, the sequence primer is replaced, and a new mixture of fluorescently tagged nonamers is introduced, for which the query position is shifted one base further into the library DNA tag. Seven

bases are queried in this fashion, with the sequence performed from the 5' end of the FDV side 14 base pair DNA library tag, followed by six base reads with a different sequence primer from the 3' end of this same 14 base pair DNA library tag, for a total of 13 base pair reads for this FDV-side DNA library tag. This sequencing is then repeated for the 5' and 3' ends of the RDV side DNA library tag, resulting in another 13 base pair reads.

Dual Primer ePCR Master Mixer

Reagent	Final Concentration
10 X EXT Buffer (FINNZYMES, Lafayette, CO)	1.2 x
MgCl ₂ (25 mM)	7 mM
dNTP (10 mM)	1 mM
FDV-PCR primer	0.1 µM
RDV-PCR primer	0.1 µM
Taq B enzyme	0.1 U/µl
Triton X-100 (10% (w/v))	0.1%
BSA (10 mg/ml)	1 mg/ml
Glycerol (50% (w/v))	5%
H ₂ O	—

Silicone Oil Phase

The silicone oil phase is prepared by mixing 3.75 g AR 20 silicone oil (Sigma, St Louis, MO), 7.5 g 749 Fluid (INCI name: cyclopentasiloxane and trimethylsiloxysilicate; Dow Chemicals Co., Midland, MI) and 10 g

5225C Formulation Aid (INCI name: clyclopentasiloxane and PEG/PPG-18/18 dinethicone; Dow Chemicals Co., Midland, MI). The oil phase is mixed on a vortex at high speed and 400 µl of oil mixture was immediately dispensed equally into 2 ml low retention tube. Another 600 µl AR 21 silicone oil was added into this 400 µl mixture and mixed on a vortex to make the final oil phase for emulsion PCR.

Breaking DPePCR emulsions

After PCR, all of the emulsion solution in the PCR tubes was transferred into a low retention 2 ml tube, mixed and spun at 15,000 rpm in a benchtop centrifuge. The beads were washed 12 times with 1 mL isopropanol, and attached on MPC and the supernatant was removed. After the isopropanol washes, the beads were then washed with 1 ml 80% Ethanol/annealing buffer 4 times, then washed 4 times with 400 µl 1x NXS buffer (combine all the same samples into new tubes at first wash). The beads were finally washed with 2 times 800 µl TE buffer and 4 times with 800 µl TE buffer /0.1% Triton. After final wash, the beads were resuspended in 1 ml TE buffer/0.1% Triton and stored at 4 °C until needed.

Capping with dideoxynucleotide reaction

The beads were incubated with 15 µl (2.5 mM), 114 µl ddH₂O, 15 µl 10 x Tailing buffer (NEB, Ipswich, MA), 4 µl 1.25 mM (each) dideoxynucleotide mix (USB, Cleveland, OH), and 2 µl terminal

transferase (20,000 U/ml) (NEB, Ipswich, MA) at 37 °C for 60 min. The beads were washed 3 times with 500 µl 1x NXS buffer, 1x TE buffer with 0.1% Triton X-100, and 1x PBS.

Figure Legends

Figure 1. Molecular Barcoding Approach.

cDNA of each patient was tagged at the time of reverse transcription with the barcoded reverse primers. The sequence adjacent to the 5' end of the barcode sequence is a universal PCR primer (T7), which is shared by all the reverse primers. All of the PCR products from each individual patient contain the corresponding barcode information, which can be distinguished by hybridizing with the complementary fluorescently labeled barcode probes.

Figure 2. Process schematic of bead based polony assay.

(A) The general steps of bead based polony technique are shown. The agarose beads were conjugated to the forward primers. An aqueous mix containing all of the necessary components for PCR, primers-coupled beads, and template DNA was mixed with a silicone oil phase to create emulsified microdroplets. These microdroplets were temperature-cycled as in a conventional PCR. If a DNA template and a bead were present

together in a single compartment, the oligonucleotides bound to the bead acted as primers for amplification. After emulsion PCR, the microdroplets were broken and the DNA products coupled beads were purified and pooled together. By binding to the magnetic beads, the agarose beads with amplified DNA products were enriched. These enriched beads with coupled DNA were immobilized to a coverslip by casting a polyacrylamide gel. (B) The merged image hybridized with exon 11 (blue), barcode 1 (green), and barcode 2 (red) is shown. Exon 11 was used as a positive control, which should be present in every *c-myb* transcript. Different combinations were identified as indicated: White beads are transcripts with exon 11+ barcode 1 + barcode 2, indicating the DNA on the beads was not clonal products from one single molecule; purple are transcripts with exon 11+barcode 2; cyan are transcripts with exon 11+barcode 1; and blue are transcripts with exon 11 only, suggesting these beads contained probably the DNA products from the other two barcoded patient samples.

Figure 3. Process schematic of *c-myb* cDNA library construction.

see “Materials and Methods” for details.

Figure 4. Oligos and intermediates in Acul inside adaptor ligation.

Oligos and intermediates in Acul inside adaptor ligation. Adaptor (in grey shadow) is oriented as it would be in circle formation. 5', 3', and 5'-phosphate oligo termini are indicated as 5, 3, 5-P, respectively.

Phosphodiester linkages to insert sequences are indicated by >- for the top strand and <- for the bottom strand. Oligo names correspond to details listed in Table 3.

Figure 5. Agarose gels of selected library construction intermediates.

M1 (Marker) for the gels contains 2000 bp, 1500 bp, 1000 bp, 500 bp. M2 (Marker) for the gel contains 400 bp, 300 bp, 200 bp, 100 bp. (A) dUTP incorporation titration. Different ratios of dUTP:dNTP (1: 5, 1:20, 1:100) were used in the first PCR step. After amplification, the DNA products were fragmented with UDG/APE1 and compared to the products amplified with no dUTP (dUTP:dNTP=0) incorporation. (B) PCR 1. To check the efficiency of Acul inside adaptor ligation with fragmented DNA, circularization U forward primer and circularization U reverse primer were used to selectively amplify the adaptor ligated DNA product (Acul-DNA). (C) PCR 2. To check the intramolecular circularization of the DNA after ExoIII and I nucleases digestion, a negative control experiment was set up, in which T4 ligase was not included in the circularization reaction. Circularization U-For and circularization U-Rev primers were used to

amplify the DNA products from both reactions (W/ T4 and W/O T4 ligase).

(D) PCR 3. To check the efficiency of Acul digestion of linearizing the circle DNA, circularization U-For and circularization U-Rev primers were used for PCR amplification. The generated PCR products (W/Acul) were compared to the PCR products from the circle DNA prior to Acul digestion (W/O Acul). (E) PCR 4. After ligating to the outside adaptors FDV and RDV, the ligated DNA products were amplified with FDV-BC-PCR and RDV-BC-PCR primers. The PCR products (FDV-DNA-RDV) were compared to the PCR products (FDV/RDV) arose from the negative control ligation reaction in which there were only two adaptors but no DNA. The 180 bp DNA products on the agarose gel were the PCR products amplified from the final DNA libraries.

Figure 6. Dual Primer Emulsion PCR approach.

(A) Dual primer beads, diluted DNA fragment, ePCR reagents, and both free primers were mixed with silicone oil phase. Ideally, each microdroplet should only contain 1 bead, 1 DNA molecule and enough reagents for clonal amplification. (B) After 1 cycle, the first DNA bridge was formed inside the microdroplet and bound to the bead, which would be used as the template for later DNA amplification on the bead. (C) After 120 cycles, the beads were covered with numerous DNA bridges that were clonal products from a single molecule. (D) After breaking emulsion, the beads were digested with SnaBI enzyme. This digestion site was at the inside adaptor, which cut the DNA fragment into half. This digestion

left one segment (*c-myb* sequence) of the fragment DNA on one strand and the other segment (barcode sequence) on the other strand. (E) After digestion, the beads were treated with NaOH to single strand the dsDNA, which provides the accessibility of the DNA strand for sequencing. Specific sequence primers were hybridized at both ends of the forward and reverse strands (individually). (F) Merged images recorded after hybridizing with sequencing primers and ligating to the nonamers. After hybridizing with different sequencing primers, the SBL (sequencing by ligation) sequencing strategy was used, followed by recording the images.

Figure 7. The final DNA library for DPePCR.

About 85 bp DNA fragments were ligated to FDV and RDV outside adaptors. Since the ligation reaction is not oriented, the two adaptors could be ligated to either side of the DNA fragments. The size of the final DNA products in the library is about 180 bp, in which the inside DNA sequence formed by circularization was flanked by two segments containing 13 bp *c-myb* sequence and 9 bp barcode sequence, respectively.

Figure 8. Polonator sequencing.

After emulsion PCR, the DNA on the beads was digested with *SnaBI* and single-stranded. For 3'-5' direction sequencing, sequence primers

FDV-Minus and RDV-Minus were sequentially hybridized to the FDV or RDV adaptor sequence of the DNA strands coupled to the beads. A set of four fluorescently labeled nonamers competed for ligation to the sequencing sequence primers. Specificity of the nonanomer is achieved by interrogating the 1st base (-1 position) in the ligation reaction. Multiple cycles of denaturation, hybridization, and ligation are performed with the number of cycles determining the eventual read length for the 3'-5' direction. For 5'-3' direction sequencing, sequence primers FDV-Plus and RDV-Plus were individually hybridized to the sequence adjacent to SnaBI digestion site. A set of four 5'phosphorylated fluorescent nonamers competed to ligate to the plus sequencing primers. Specificity of the nonamer was achieved by interrogating the 1st base (+1 position) in the ligation reaction. The 5'-3 direction sequencing was performed the same way as the 3'-5' direction sequencing. The unknown sequences in the red box are the interrogated bases.

Figure 9. Distribution of sequence coverage for each exon.

After decoding the barcode information of each bead and categorizing the raw data into the groups with the same barcode, the *c-myb* tags were mapped back to different position at individual exon sequence. The frequency of the tags found in the same position was plotted against the entire sequence of the exon.

Figure 1

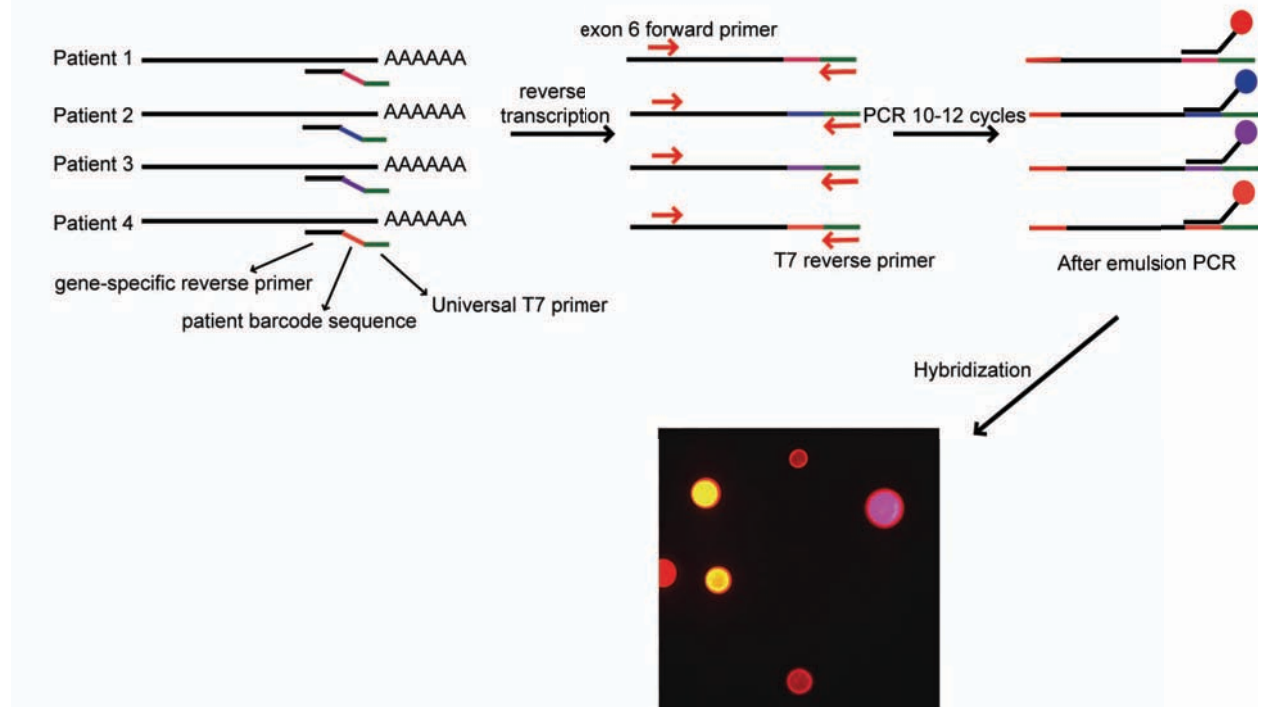
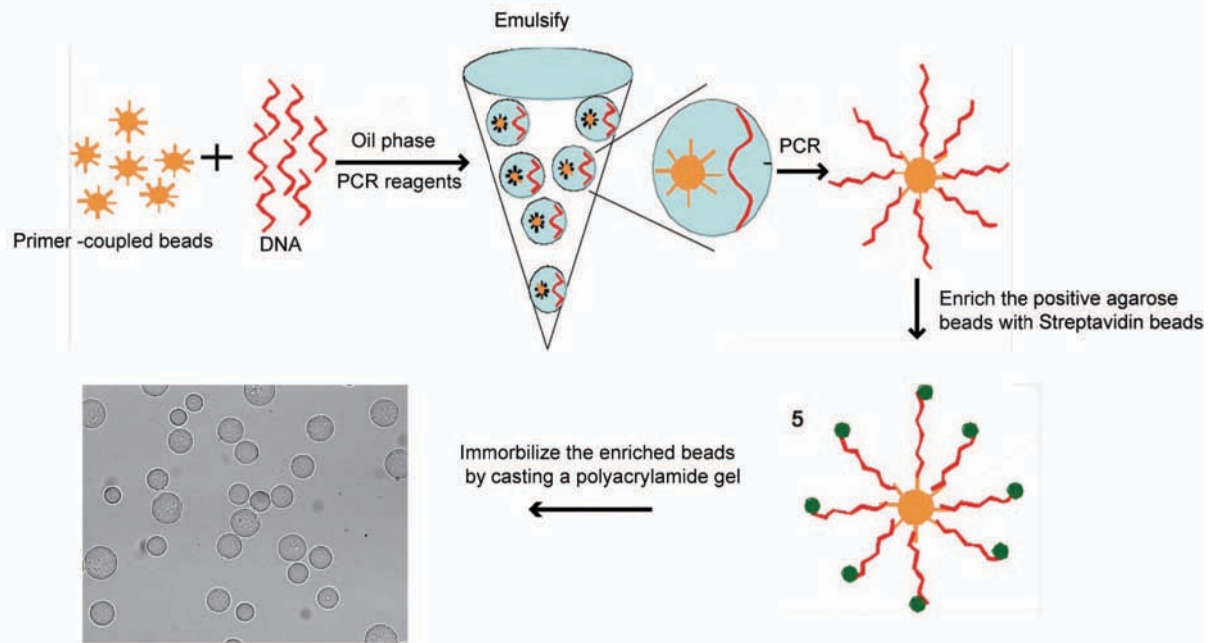
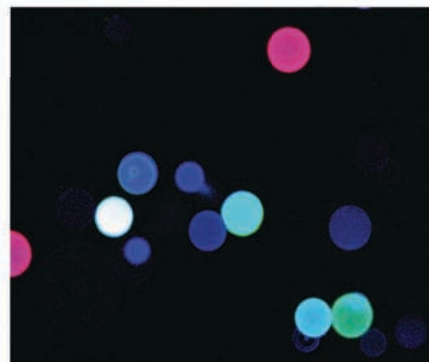


Figure 2

A



B



white: barcode 1+barcode 2+exon11
 Purple: barcode 2+exon11
 Cyan: barcode 1+exon11
 Blue: exon11

Figure 3

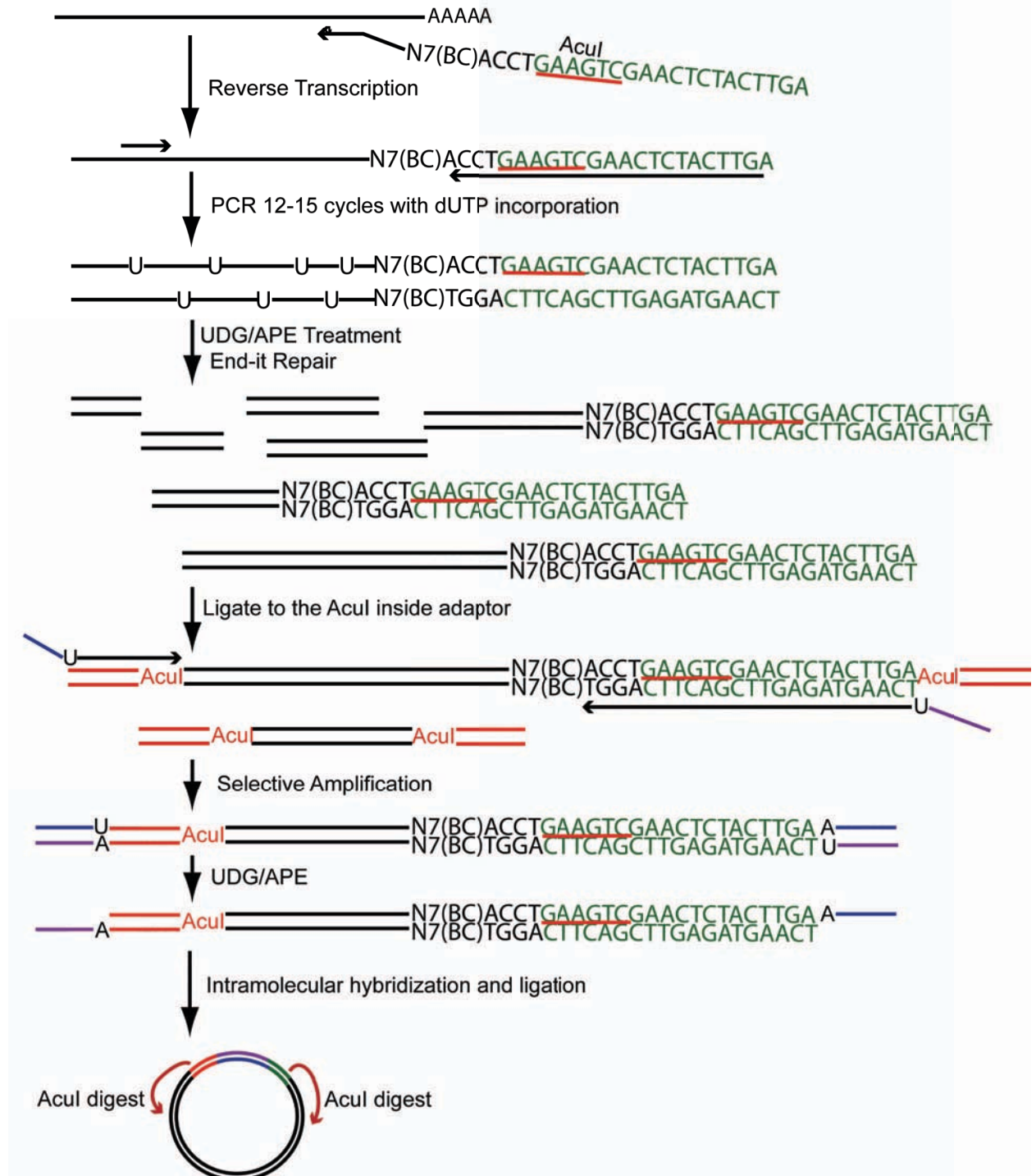


Figure 3

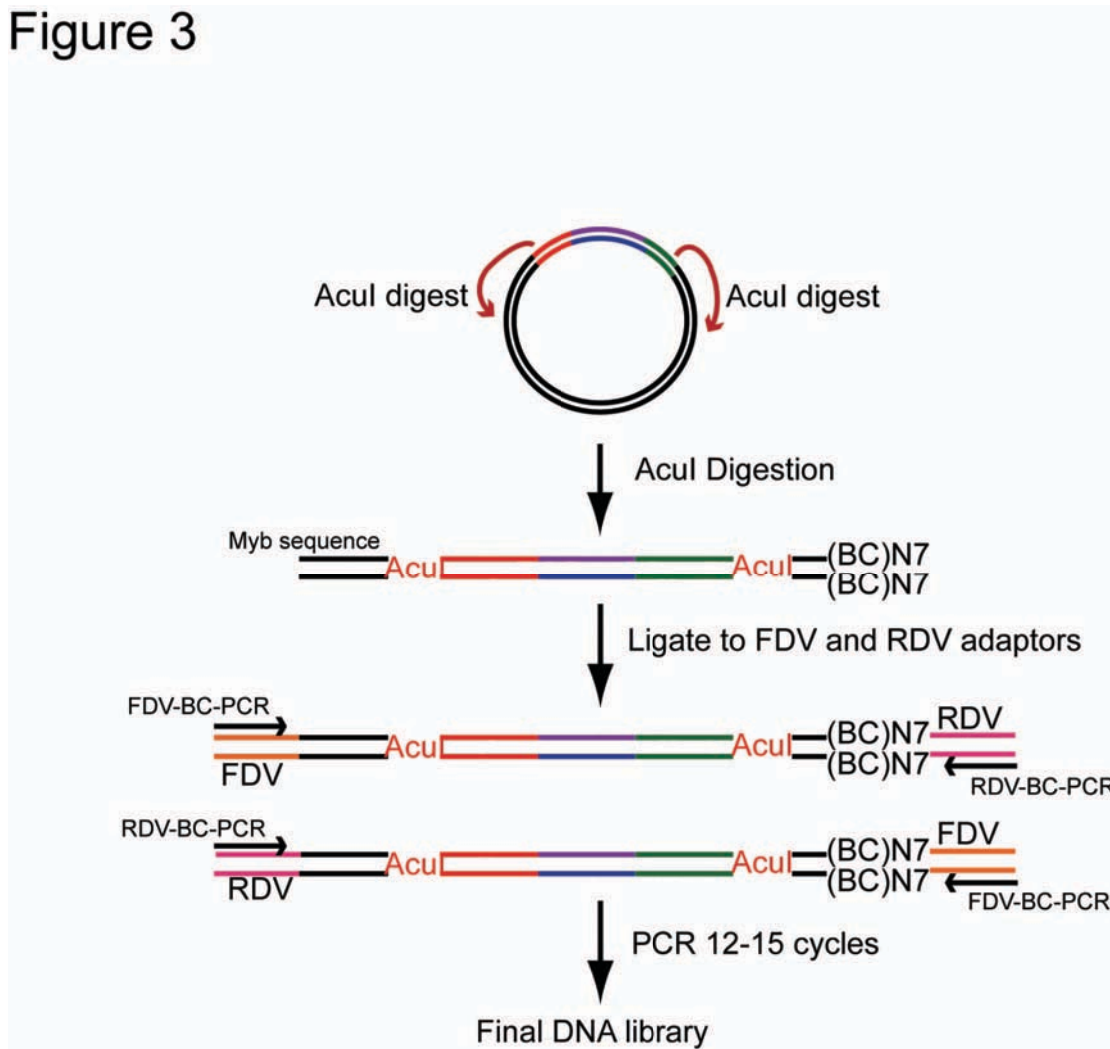


Figure 4

Acul Inside Adaptor

3-TGAAGTCTAGTACGAGATAGGAC-5
5-P-CTTCAGATCATGCTCTATCCTGCATT-3

Acul Inside adaptor ligated DNA product + PCR 1 primer sets

>-N7(BC)TGGACTTCAGCTTGAGATGAACTTTACTTCAGATCATGCTCTATCCTGCATT-3
<-N7(BC)ACCTGAAGTCGAACTCTACTTGAATGAAGTCTAGTACGAGATAGGAC-5
3-ACCTGCCGTCGAACTCTACTTGAUGCATAGCACGTA-5 Circularization U-Rev
Circularization U For 5-ACGTATCGTGCAUCAGGATAGAGCATGATCTGAA-3
5-CAGGATAGAGCATGATCTGAAGT>-
3-TTACGTCCTATCTCGTACTAGACTTCA<-

Acul inside adaptor ligated DNA PCR 1 product

>-N7(BC)TGGACTTCAGCTTGAGATGAACTACGTATCGTGAT-3
<-N7(BC)ACCTGAAGTCGAACTCTACTTGAUGCATAGCACGTA-5
5-ACGTATCGTGCAUCAGGATAGAGCATGATCTGAAGT>-
3-TGCATAGCACGTAGTCCTATCTCGTACTAGACTTCA<-

UDG/APE treated PCR 1 product

>-N7(BC)TGGACTTCAGCTTGAGATGAACTACGTATCGTGAT-3
<-N7(BC)ACCTGAAGTCGAACTCTACTTGA-5
5-CAGGATAGAGCATGATCTGAAGT>-
3-TGCATAGCACGTAGTCCTATCTCGTACTAGACTTCA<-

Intramolecular circularization and ligation product

>-N7(BC)TGGACTTCAGCTTGAGATGAACTACGTATCGTGATCAGGATAGAGCATGATCTGAAGT>-
<-N7(BC)ACCTGAAGTCGAACTCTACTTGATGCATAGCACGTAGTCCTATCTCGTACTAGACTTCA<-

Figure 5

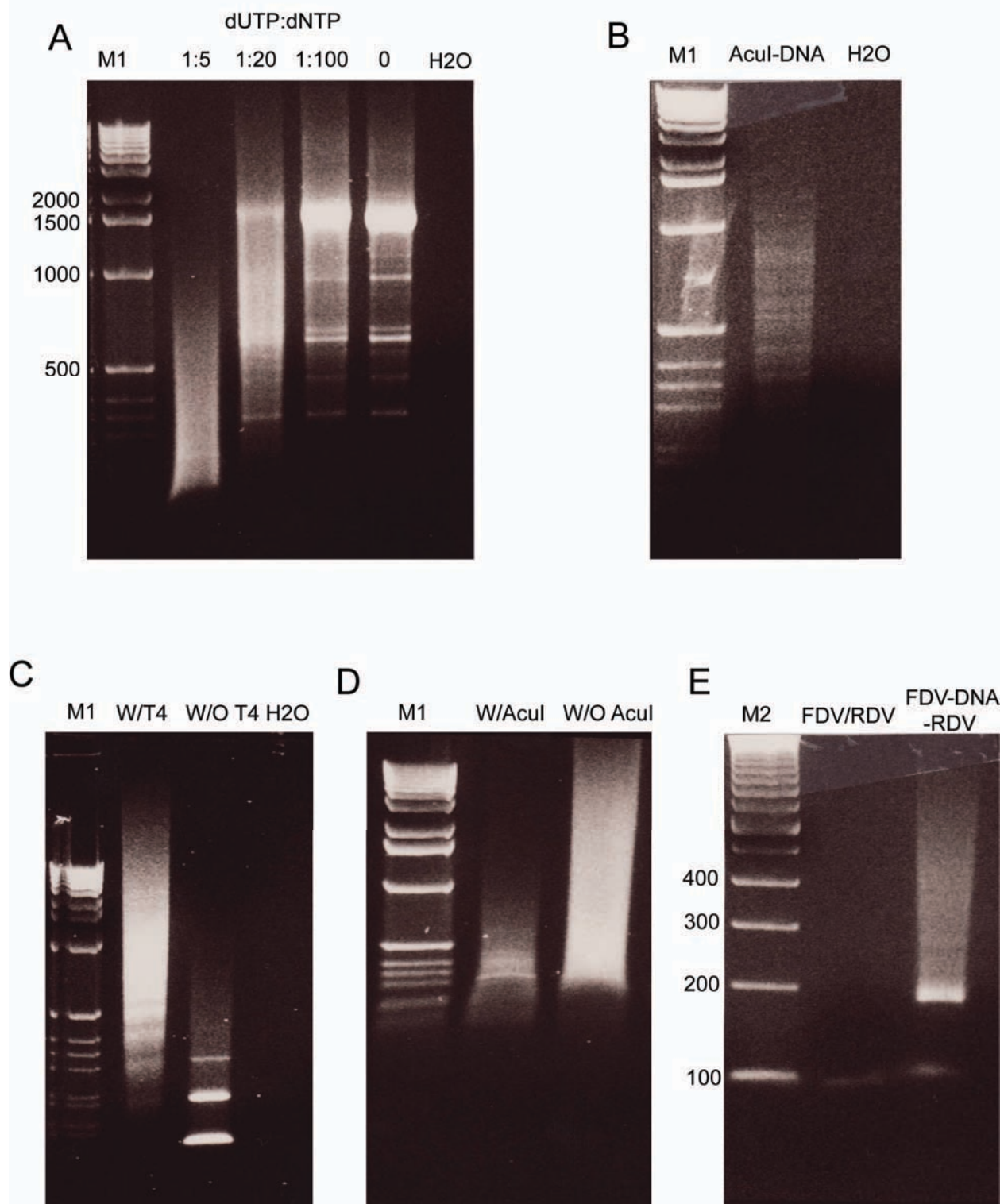


Figure 6

A

FDV RDV

dNTP
Free Primer
Taq
Mg²⁺

B

PCR 1 Cycle

C

PCR 120 Cycles

D

SnaBI digestion

E

Single strand and
anneal to anchor primers

F

Sequencing By Ligation

A
T
C
G

Cy5
Cy3
TxRed
FITC

Four colors overlap

Minus Anchor Primer
Plus Anchor Primer

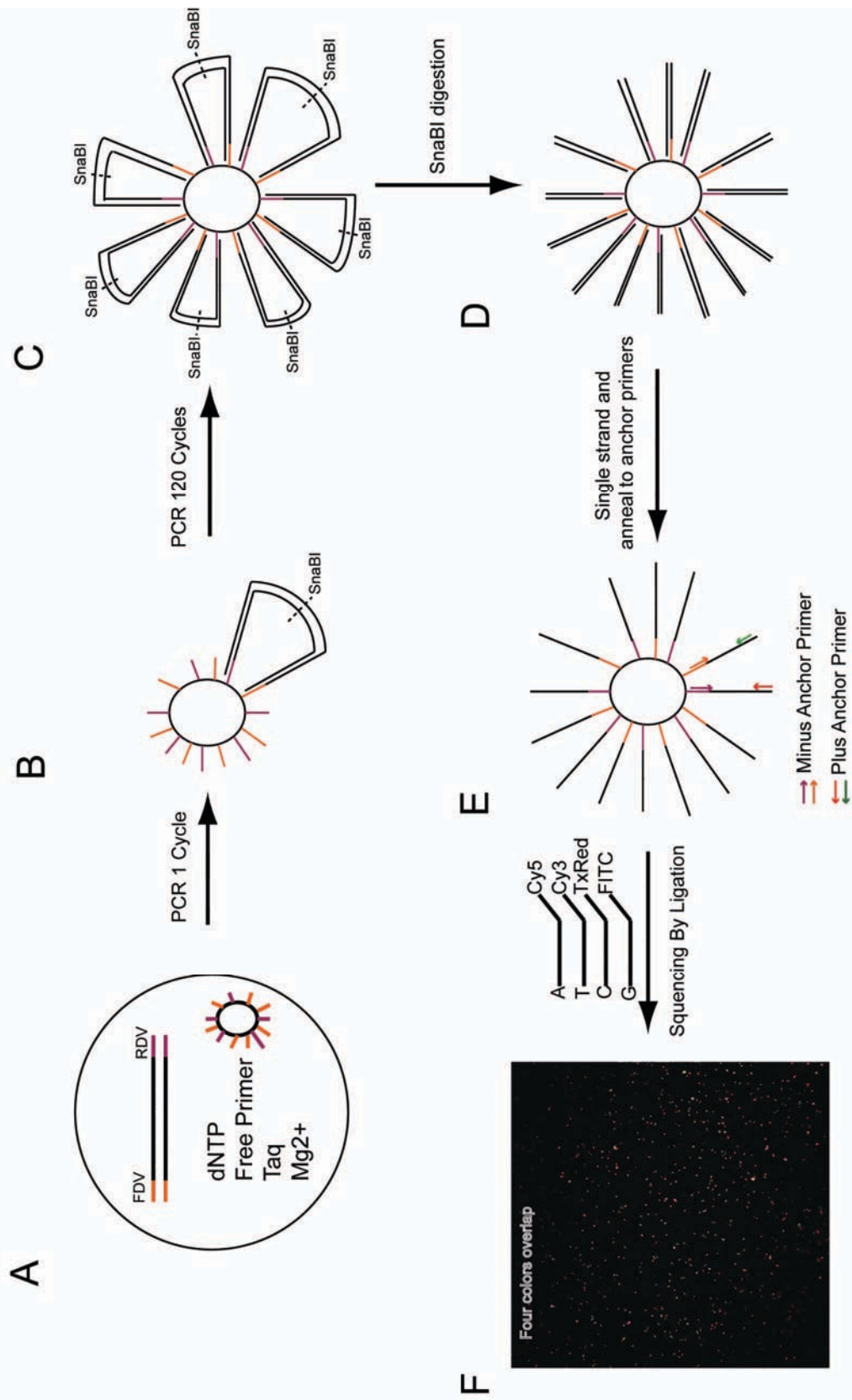


Figure 7

FDV-Adaptor

3'-TCAACCTCATACCTATAACCATATTTTCGTATCTCTCCTTTGCGCTCCG-5'
5'-Phos-GTTGGAGTATGGATATTGGTATAAAGCATAGAGAGGAAAGCGGAGGCTTAG-3'

RDV Adaptor

3'-TGAAGTCATCAGTACTCTATCCTTTACTTACTCCTTGGGCCCCGTCA-5'
5'-Phos-CTTCAGTAGTCATGAGATAGGAAATGAATGAGGAACCCGGGGCAGTCATT-3'

Final DNA library

Structure 1

Barcode sequence
5'-GCCTCCGCTTTCTCTCTATGCTTTATACCAATATCCATACTCCAACCTNNNNNNNNBBTGGA
3'-CGGAGGCGAAAGGAGAGATACGAAATATGGTTATAGGTATGAGGTTGAANNNNNNNNBBACCT

SnaBI
5'-CTTCAGCTTGAGATGAACTACGTATTCGTGCATCAGGATAGAGCATGATCTGAAG
3'-GAAGTCGAACTCTACTTGATGCATAGCACGTAGTCCTATCTCGTACTAGACTTC

5'-TNNNNNNNNNNNNNACTTCAGTAGTCATGAGATAGGTTTTGAATGAGGAACCCGGGGCAGT-3'
3'-ANNNNNNNNNNNNTGAAAGTCATCAGTACTCTATCCAAAACCTACTCCTTGGGCCCCGTCA -5'
c-myb sequence

Structure 2

c-myb sequence
5'-GCCTCCGCTTTCTCTCTATGCTTTATACCAATATCCATACTCCAACCTNNNNNNNNNNNNNNN
3'-CGGAGGCGAAAGGAGAGATACGAAATATGGTTATAGGTATGAGGTTGAANNNNNNNNNNNNNNN

SnaBI
5'-ACTTCAGATCATGCTCTATCCTGATGCACGATACGTAGTTTCATCTCAAGCTGAAG
3'-TGAAGTCTAGTACGAGATAGGACTACGTGCTATGCATCAAGTAGAGTTCGACTTC

5'-TCCABBNNNNNNNNACTTCAGTAGTCATGAGATAGGTTTTGAATGAGGAACCCGGGGCAGT-3'
3'-AGGTBBNNNNNNNTGAAAGTCATCAGTACTCTATCCAAAACCTACTCCTTGGGCCCCGTCA -5'
Barcode Sequence

Figure 8

The diagram illustrates the construction of a nanobiosensor. It shows the ligation of fluorescent nanomeres to DNA primers, followed by SnaBI digestion to create a hairpin structure. The final structure is a hairpin with a loop containing the target sequence.

Top Section: Ligation of nanomeres to FDV-Minus primer

5'-GCCTCCGCTTTCCTCTCTATGCTTTATACCAATATCCATACTCCAACCTTNNNNNNNB-3' (Anchor Primer FDV-Minus)

3'-TATGGTTATAGGTATGAGGTTGAA-P-5' (Fluorescent nanomeres)

Ligation

Middle Section: SnaBI Digestion and Ligation to FDV-Plus primer

5'-FDV-NNNNNNNB-3' (Anchor Primer FDV-Minus)

3'-ACCTGAAGTCGAACTCTACTTG-5' (Anchor Primer FDV-Plus)

SnaBI Digestion

Ligation

5'-phosphorylated-fluorescent nanomeres

Bottom Section: Ligation of nanomeres to RDV-Minus primer

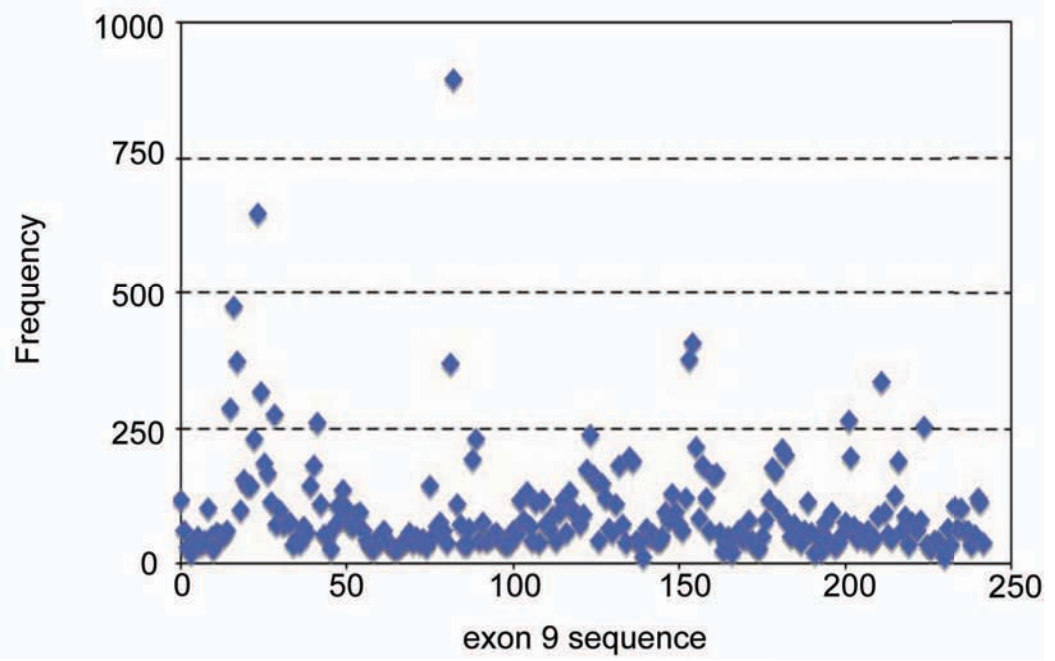
5'-P-ACCTCAGTAGTCATGAGATAGG-3' (Anchor Primer RDV-Minus)

3'-CATAGCACGTAGTCCTATCTCGTACTAGACTTCANNNNNNNN-RDV-5' (Anchor Primer RDV-Plus)

Ligation

5'-phosphorylated-fluorescent nanomeres

Figure 9



Tables

Table 1. Primers for polony sequencing (bead-based) assay

Primers for Polony Sequencing

	Sequence
Universal T7-Barcode1-RT	5'-GTAATACGACTCACTATAGGGCGAATTGGGTAAAGGCACGTATCATATCCCTCTTTCCACAGGATGCAGGTT-3'
Universal T7-Barcode2-RT	5'-GTAATACGACTCACTATAGGGCGAATTGGGTAAACATATCCTAATACAGGCGCCTTTCCACAGGATGCAGGTT-3'
Universal T7-Barcode3-RT	5'-GTAATACGACTCACTATAGGGCGAATTGGGTAGACGATCCTTATACTCGATGCTTTCCACAGGATGCAGGTT-3'
Universal T7-Barcode4-RT	5'-GTAATACGACTCACTATAGGGCGAATTGGGTAGAAGACAGTTATACCCATGCCTTTCCACAGGATGCAGGTT-3'
T7 Rev PCR Primer	5'-GTAATACGACTCACTATAGGGCGAATTGGGTA-3'
SN-3 Forward Primer	5'-GGAAGGTGGAACAGGAAGGTTATCTG-3'
T7 Rev PCR Primer-biotin	5'-Biotin-GTAATACGACTCACTATAGGGCGAATTGGGTA-3'

Probes for Exontyping

Sequence

Barcode 1 probe-cy3	5'-Alexa 532-TAAAGGCACGTATCATATCCCTC-3'
Barcode 2 probe-cy5	5'-Alexa 647-TAACATATCCTAATACAGGCGCC-3'
Barcode 3 probe-FITC	5'-FAM-TAGACGATCCTTATACTCGATGC-3'
Barcode 4 probe-TxRed	5'-Alexa 594-TAGAAGACAGTTATACCCATGCC-3'

The exon-specific probes are the same as the ones used in gel-based polony assay in Chapter 2.

Table 2. The splice variants in 4 pediatric pre-B-ALL patient samples detected by the bead based polony assay

(ND: Not Detectable)

	Isoform	Protein Size	Exontyping	610-194	610-856	610-319	611-551
1	Del8/Del9	1-281, 402-640	6-7-10-11	0.28%	ND	ND	ND
2	Del8/Del10	1-281, 371-402	6-7-9-11	ND	ND	ND	0.09%
3	Del8	1-281, 317-640	6-7-9-10-11	2.35%	3.89%	2.72%	0.68%
4	Del9	1-316,402-640	6-7-8-10-11	ND	1.29%	1.09%	1.56%
5	Del10	1-402	6-7-8-9-11	ND	ND	ND	0.19%
6	WT	1-640	6-7-8-9-10-11	58.63%	52.37%	64.50%	67.54%
7	Del8/9S	1-281, 317-372, +2aa	6-7-9S-10-11	0.47%	0.96%	ND	0.19%
8	Del8/9S		6-7-9S-10-10A-11	ND	2.57%	ND	0.09%
9	9S/10	1-372, +2aa	6-7-8-9S-10-11	4.17%	2.28%	1.93%	7.17%
10	9S/10		6-7-8-9S-10-10A-11	0.82%	1.60%	ND	0.09%
11	Del8/9A	1-281, 317-402, +M	6-7-9-9A-10-11	0.35%	0.09%	ND	ND
12	Del8/9A		6-7-9-9A-9B-10-11	0.12%	0.19%	0.09%	ND
13	Del8/9A		6-7-9-9A-10-10A-11	ND	ND	ND	ND
14	Del8/9A		6-7-9-9A-9B-10-10A-11	ND	ND	ND	ND
15	9S/9A	1-372, +16aa	6-7-8-9S-9A-10-11	0.24%	0.32%	0.19%	0.58%
16	9S/9A		6-7-8-9S-9A-9B-10-11	ND	0.32%	ND	0.19%
17	9S/9A		6-7-8-9S-9A-10-10A-11	0.12%	0.64%	ND	ND
18	9S/9A	1-372, +16aa	6-7-8-9S-9A-9B-10A-11	ND	ND	ND	ND
19	9A	1-401, +M	6-7-8-9-9A-10-11	2.54%	2.25%	2.48%	2.09%
20	9A		6-7-8-9-9A-9B-10-11	1.45%	0.09%	0.49%	0.36%
21	9A		6-7-8-9-9A-10-10A-11	0.12%	0.64%	ND	0.54%
22	9A		6-7-8-9-9A-9B-10-10A-11	0.47%	0.76%	ND	ND
23	Del8/8A	1-281, +34aa	6-7-8A-9S-10-11	ND	ND	0.29%	ND
24	Del8/8A		6-7-8A-9-10-11	ND	ND	0.19%	ND
25	Del8/8A		6-7-8A-9-10-10A-11	0.24%	0.64%	ND	0.56%
26	Del8/8A		6-7-8A-9-9A-10-10A-11	ND	ND	ND	0.19%

	Isoform	Protein Size	Exontyping	610-194	610-856	610-319	611-551
27	Del8/8A	1-281, +34aa	6-7-8A-9S-9B-10-10A-11	0.12%	ND	ND	ND
28	Del8/8A		6-7-9-9A-10-10A-11	ND	0.32%	ND	ND
29	8A	1-316, +34aa	6-7-8-8A-9S-10-11	ND	ND	0.19%	ND
30	8A		6-7-8-8A-9-11	ND	ND	0.13%	ND
31	8A		6-7-8-8A-9-10-11	0.12%	ND	3.59%	ND
32	8A		6-7-8-8A-9S-9A-10-11	ND	ND	ND	ND
33	8A	1-316, +34aa	6-7-8-8A-9-9A-10-11	ND	ND	0.09%	ND
34	8A		6-7-8-8A-9S-9B-10-11	ND	ND	ND	ND
35	8A		6-7-8-8A-9-9B-10-11	ND	ND	0.09%	0.19%
36	8A		6-7-8-8A-9-9A-9B-10-11	ND	ND	ND	0.24%
37	8A		6-7-8A-10-11	ND	ND	0.09%	ND
38	8A		6-7-8-8A-10-10A-11	ND	ND	ND	ND
39	8A		6-7-8-8A-9S-10-10A-11	0.35%	0.32%	ND	0.12%
40	8A		6-7-8-8A-9-10A-11	ND	ND	ND	ND
41	8A		6-7-8-8A-9-9A-10-10A-11	0.24%	ND	ND	ND
42	8A		6-7-8-8A-9S-9B-10-10A-11	0.24%	ND	ND	ND
43	8A		6-7-8-8A-9S-9A-10-10A-11	ND	ND	ND	ND
44	8A		6-7-8-8A-9-10-10A-11	4.82%	1.61%	0.56%	7.45%
45	8A		6-7-8-8A-9S-9A-9B-10-11	ND	ND	0.09%	ND
46	8A		6-7-8-8A-9-9B-10-10A-11	0.71%	0.96%	0.09%	0.09%
47	Del8/9S/9B	1-281, 317-372, +50aa	6-7-9S-9B-10-11	0.24%	0.32%	ND	0.09%
48	Del8/9S/9B		6-7-9S-9B-10-10A-11	ND	ND	ND	ND
49	Del9/10A	1-316, 402-449	6-7-8-10-10A-11	ND	ND	ND	ND
50	Del9/10A		6-7-8-10-10A-11	ND	0.32%	0.09%	ND
51	9S/9B	1-372, +50aa	6-7-8-9S-9B-10-11	1.57%	0.19%	0.41%	ND
52	9S/9B		6-7-8-9S-9B-10-10A-11	0.47%	0.64%	ND	ND
53	9B/Del10	1-401, +121aa, 402	6-7-8-9-9B-11	ND	ND	ND	ND
54	9B	1-401,+121aa, 402-640	6-7-8-9-9B-10-11	11.32%	7.50%	9.25%	4.24%
55	Del8/10A	1-281,317-449	6-7-9-10-10A-11	0.24%	0.96%	0.19%	0.36%
56	Del9/Del10/10A	1-316,+13aa	6-7-8-10A-11	ND	ND	ND	ND
57	Del10/10A	1-401,+13aa	6-7-8-9-10A-11	ND	ND	ND	ND
58	10A	1-449	6-7-8-9-10-10A-11	4.98%	13.87%	10.27%	5.26%

	Isoform	Protein Size	Exontyping	610-194	610-856	610-319	611-551
59	9B/10A	1-401,+121aa,402-449	6-7-8-9-9B-10-10A-11	1.33%	1.97%	0.90%	0.54%
60	Del8/9S/9A	1-281,317-372,+16aa	6-7-9S-9A-10-11	ND	ND	ND	ND
61	Del8/9B	1-281,317-401,+121aa,402-640	6-7-9-9B-10-11	0.89%	0.12%	ND	ND

Table 3. Primers for constructing sequencing libraries, emulsion PCR and sequencing by NGSM

Primers For Library Construction	Sequence
Oligo Acul-7mer-2BC-RT-1	5'-AGTTCATCTCAAGCTGAAGTCCA(CT)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Oligo Acul-7mer-2BC-RT-2	5'-AGTTCATCTCAAGCTGAAGTCCA(AG)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Oligo Acul-7mer-2BC-RT-3	5'-AGTTCATCTCAAGCTGAAGTCCA(AA)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Oligo Acul-7mer-2BC-RT-4	5'-AGTTCATCTCAAGCTGAAGTCCA(CC)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Oligo Acul-7mer-2BC-RT-5	5'-AGTTCATCTCAAGCTGAAGTCCA(GC)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Oligo Acul-7mer-2BC-RT-6	5'-AGTTCATCTCAAGCTGAAGTCCA(TT)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Oligo Acul-7mer-2BC-RT-7	5'-AGTTCATCTCAAGCTGAAGTCCA(CA)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Oligo Acul-7mer-2BC-RT-8	5'-AGTTCATCTCAAGCTGAAGTCCA(TA)NNNNNNNTGCACTTGGTCGTGCTCT-3'
Rev-Oligo Acul	5'-TCATCTCAAGCTGAAGTCCA-3'
SN-3-For	5'-GGAAGGTGCAACAGGAAGGTTATCTG-3'
Acul-Inside Adaptor-1	5'-CAGGATAGAGCATGATCTGAAGT-3'
Acul-Inside Adaptor-2	5'-Phos-CTTCAGATCATGCTCTATCCTGCATT-3'
Circularization U-For	5'-ACGTATCGTGCAUCAGGATAGAGCATGATCTGAA-3'
Circularization U-Rev	5'-ATGCACGATACGUAGTTCATCTCAAGCTGCCGTCCA-3'
FDV-adaptor-1	5'-GCCTCCGCTTTCTCTCTATGCTTTATACCAATATCCATACTCCA-3'
FDV-adaptor-2	5'-Phos-GTTGGAGTATGGATATTGGTATAAAGCATAGAGAGGAAAGCGGAGGCTTAC-3'
RDV-adaptor-1	5'-ACTGCCCCGGGTTCTCATTCTATCTCATGACTACTGAAGT-3'
RDV-adaptor-2	5'-Phos-CTTCAGTAGTCATGAGATAGGAAATGAATGAGGAACCCGGGGCAGTCATT-3'
FDV-BC-PCR	5'-GCCTCCGCTTTCTCCTATG-3'
RDV-BC-PCR	5'-ACTGCCCCGGGTTCTCATT-3'
Acul-Inside Adaptor	3'-TGAAGTCTAGTACGAGATAGGAC-5'
	5'-Phos-CTTCAGATCATGCTCTATCCTGCATT-3'
FDV-Adaptor	3'-TCAACCTCATACCTATAACCATATTTTCGTATCTCTCCTTTTCGCCTCCG-5'
	5'-Phos-GTTGGAGTATGGATATTGGTATAAAGCATAGAGAGGAAAGCGGAGGCTTAG-3'
RDV-Adaptor	3'-TGAAGTCATCAGTACTCTATCCTTTACTTACTCCTTGGGCCCCGTCA-5'
	5'-Phos-CTTCAGTAGTCATGAGATAGGAAATGAATGAGGAACCCGGGGCAGTCATT-3'

Primers for Emulsion PCR and Polonator Sequencing	Sequence
FDV-DualBiotin	5'-dual biotin-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAACCACTACGCCTCCGCTTTCCTCTCTATG-3'
RDV-DualBiotin	5'-dual biotin-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATATGTCAACTGCCCCGGGTTCTCATTC-3'
FDV-BC-PCR	5'-GCCTCCGCTTTCCTCCTATG-3'
RDV-BC-PCR	5'-ACTGCCCCGGGTTCTCATTC-3'
Sequence primer FDV-Minus	5'-/5Phos/-AAGTTGGAGTATGGATATTGGTAT-3'
Sequence primer RDV-Minus	5'-/5Phos/-ACTTCAGTAGTCATGAGATAGG-3'
Sequence primer FDV-Plus	5'-GTTTCATCTCAAGCTGAAGTCCA-3'
Sequence primer RDV-Plus	5'-GCATCAGGATAGAGCATGATCTGAAGT-3'
FDV-FAM	5'-FAM-AAGTTGGAGTATGGATATTGGTAT-3'

Nonamers	Sequence
-1 sequencing nonamers	5'Cy5-NNNNNNNNT 5'Cy3-NNNNNNNNA 5'TexasRed-NNNNNNNNNC 5'FAM-NNNNNNNNG
+1 sequencing nonamers	5'-/5Phos/-TNNNNNNNN/3Cy5/-3' 5'-/5Phos/-ANNNNNNNNN/3Cy3/-3' 5'-/5Phos/-CNNNNNNNN/3TxRed/-3' 5'-/5Phos/-GNNNNNNNN/3FAM/-3'
-2 sequencing nonamers	5'Cy5-NNNNNNNTN 5'Cy3-NNNNNNNAN 5'TexasRed-NNNNNNNNCN 5'FAM-NNNNNNNGN
+2 sequencing nonamers	5'-/5Phos/-NTNNNNNNNN/3Cy5/-3' 5'-/5Phos/-NANNNNNNNN/3Cy3/-3' 5'-/5Phos/-NCNNNNNNNN/3TxRed/-3' 5'-/5Phos/-NGNNNNNNNN/3FAM/-3'
-3 sequencing nonamers	5'Cy5-NNNNNNNTNN 5'Cy3-NNNNNNNANN 5'TexasRed-NNNNNNNCNN 5'FAM-NNNNNNNGNN
+3 sequencing nonamers	5'-/5Phos/-NNTNNNNNNN/3Cy5/-3' 5'-/5Phos/-NNANNNNNNN/3Cy3/-3' 5'-/5Phos/-NNCNNNNNNN/3TxRed/-3' 5'-/5Phos/-NNGNNNNNNN/3FAM/-3'
-4 sequencing nonamers	5'Cy5-NNNNNTNNN 5'Cy3-NNNNNANNN 5'TexasRed-NNNNNNCNNN 5'FAM-NNNNNGNNN
+4 sequencing nonamers	5'-/5Phos/-NNNTNNNNNN/3Cy5/-3' 5'-/5Phos/-NNNANNNNNN/3Cy3/-3' 5'-/5Phos/-NNNCNNNNNN/3TxRed/-3' 5'-/5Phos/-NNNGNNNNNN/3FAM/-3'
-5 sequencing nonamers	5'Cy5-NNNNNTNNNN 5'Cy3-NNNNNANNNN 5'TexasRed-NNNNNCNNNN 5'FAM-NNNNNGNNNN
+5 sequencing nonamers	5'-/5Phos/-NNNNTNNNNN/3Cy5/-3' 5'-/5Phos/-NNNNANNNNN/3Cy3/-3' 5'-/5Phos/-NNNNCNNNNN/3TxRed/-3' 5'-/5Phos/-NNNNGNNNNN/3FAM/-3'
-6 sequencing nonamers	5'Cy5-NNNNNTNNN 5'Cy3-NNNNNANNN 5'TexasRed-NNNNNNCNNN 5'FAM-NNNNNGNNN
+6 sequencing nonamers	5'-/5Phos/-NNNTNNNNNN/3Cy5/-3' 5'-/5Phos/-NNNANNNNNN/3Cy3/-3' 5'-/5Phos/-NNNCNNNNNN/3TxRed/-3' 5'-/5Phos/-NNNNGNNNNN/3FAM/-3'

-7 sequencing nonamers 5'Cy5-NNNNNNTNN
 5'Cy3-NNNNNNANN
 5'TexasRed-NNNNNNNCNN
 5'FAM-NNNNNNGNN

All nonamers were purified via high performance liquid chromatography (HPLC) and degenerated bases were hand-mixed to the 25:25:25:25 ratios.

Chapter 4: Discussion

Overall, our study demonstrates that *c-myb* alternative splicing is highly elevated in leukemia patient samples and the pattern of *c-myb* alternative splicing is more complex and distinct in leukemia patients than in normal samples. Our data also suggests that activation of *c-myb* alternative splicing could be a novel mechanism that contributes to leukemogenesis. We proposed the hypothesis that aberrant alternative splicing of *c-myb* transcripts in leukemias leads to the production of truncated variants of c-Myb with transforming or leukemogenic activities, which could be used as a biomarker in leukemias. Our data here support the hypothesis by showing that some splice variants with the protein structures resembling the oncoprotein v-Myb are more highly expressed in leukemia patients, which implies *c-myb* alternatively spliced variants could play an important role in tumor development or progression. Moreover, we performed survival analysis that identified some splice variants have a tight link to leukemia outcome demonstrating the *c-myb* alternative splicing could be a unique biomarker for classifying leukemia patients into different treatment groups, as gene expression-profiling does. To be able to identify the clinical value of following *c-myb* alternative splicing in leukemia samples, we further developed new techniques combining with next generation sequencing platforms to profile *c-myb* splice variants in leukemia patients at a large scale. This is the first time that next-generation methods are applied to determine the structures of entire cDNAs. The novel use of NGSs accelerates the speed of analyzing *c-myb* alternative splicing in a large number of patient samples simultaneously by

several orders of magnitude, which may become a promising tool facilitating medical research in analysis of the cause of leukemias, development of new drugs and diagnostics.

Do *c-myb* splice variants have unique transcriptional activities that contribute to oncogenesis?

Several recent studies^{36,37,78} have suggested that relatively minor changes in the transcriptional activation domain of c-Myb can dramatically affect its transcriptional activity. The alternative splicing of the *c-myb* gene products is a mechanism for producing variants of c-Myb with different transcriptional activation domains. Our laboratory has shown that variant c-Myb proteins encoded as a result of alternative splicing have quantitative and qualitative differences in their transcriptional activities¹⁹. However, very little is known about the biological roles of the variant c-Myb proteins. Since *c-myb* undergoes extensive alternative splicing producing vast amount of variant c-Myb proteins, it is not feasible to study all of these variant proteins. We need to focus on the ones that are highly related to disease. Our data identifies some *c-myb* splice variants that could have an important role in leukemia and reveals the expression level of these splice variants could affect the survival time of the patients. These data provide a foundation for characterizing the transcriptional specificities and biological functions of the aberrantly expressed splice variants in leukemias.

Variant c-Myb proteins share very similar or identical DNA binding domains, they recognize a degenerate consensus sequence motif (C/T)AAC(G/T)G^{26,247}, which is expected to occur approximately three million times in the human

genome. Thus, while the DNA binding domain is critical for its activity, the protein binding sites in the C-terminal domains probably play a crucial role in stabilizing multi-protein complexes and combinatorial and cooperative interactions that allow each type of c-Myb proteins to associate with specific sets of target gene promoters, or allow each c-Myb variant to associate with different sets of co-factors, resulting in different transcriptional activities. The C-terminal region of c-Myb also contains the sequences responsible for triggering its rapid ubiquitinated and proteolytic turnover^{112 113}, so truncated variants like 9S/10 might be much more stable than the full length protein and could accumulate to higher levels than expected, even though they are encoded by a relative small fraction of transcripts. Chromatin-ImmunoPrecipitation (ChIP) assays are one of the means to study the binding with target genes and follow the transcriptional specificities of variant Myb proteins¹⁸⁶. ChIP assays could determine the specific locations (promoters) that variant c-Myb proteins are associated with, indicating the target genes of these proteins. The uses of specific siRNA for splice variants, which could selectively knockdown variant Myb proteins, followed by transcriptome sequencing or microarray analysis could further assist with the identification of the genes that are regulated by variant Myb proteins. It is essential to develop antibodies that specifically recognize these variant proteins to study the difference in the target genes of c-Myb protein and its variants proteins. With genome wide ChIP assays and variant Myb protein knockdown experiments, a set of clinically relevant Myb target genes would be identified. By comparing to the target genes of oncoprotein v-Myb in transformed or leukemia

cells and a comparison of the expression levels of the target genes in normal cells versus in leukemia samples will offer an insight into the role of aberrantly expressed variant c-Myb proteins in leukemogenesis.

Do Myb partners define their activities?

Several lines of evidences show that C-terminal domains of Myb proteins affect intra- and intermolecular interactions with their protein co-factors and are subject to numerous post-translational modifications, which can regulate Myb proteins transcriptional activities (see review ¹⁷²). Our recent studies show that alternatively spliced *c-myb* transcripts are associated with polyribosomes, which indicates that they are translated to produce different forms of c-Myb proteins. Variant c-Myb proteins have different C-terminal domains, in which some variants with C-terminal deletion resembling v-Myb are identified by our research to correlate with human leukemias. These variant proteins are likely to interact with different co-factors and form different protein complex to regulate different gene sets.

The C-terminal domain of c-Myb protein is comprised of several sub-domains, such as the transactivation domain and negative regulatory domain, each of which contributes interdependently to c-Myb transcriptional activity (see review ²¹⁹). To identify proteins that interact with each domain, a generic tandem-affinity protein purification (TAP) method could be performed to isolate Myb protein complexes in different types of cells and use proteomics to identify the proteins that different variant Myb proteins bind to. Classical biochemical purification methods rely heavily on the biophysical properties of a given protein.

Variant Myb proteins are unique and the protein purification protocol for each protein needs to be designed individually, therefore, there is a critical need for developing widely applicable and efficient purification procedures for isolating protein complexes formed by variant Myb proteins.

Tandem affinity purification is a two-step affinity purification protocol for isolation of TAP-tagged proteins together with associated proteins ²⁴⁸. It has proven to be highly effective in the identification of protein complexes in yeast and some mammalian systems ^{249 248 250}. The TAP tag is comprised of two affinity tags, protein A and calmodulin-binding peptide (CBP), separated by a TEV protease cleavage site, which is fused to either the N- or C-terminus of the target protein. Extracts prepared from cells expressing the TAP-tagged protein are subjected to two successive purification steps. The first purification step involves binding of the protein A tag to an IgG column and eluting the bound material by cleaving with TEV protease. This elution method significantly reduces the contaminating proteins bound nonspecifically to the column. The elutes that contain protein complexes are further purified by binding with calmodulin-coated beads to remove TEV protease and contaminants. Highly purified protein complexes are eluted by adjusting the concentration of calcium ions with chelating reagents. The retrieved protein complexes are then subjected to proteomics to identify the proteins in the complexes.

To apply this generic purification approach to characterize the protein partners of variant Myb proteins, we could construct TAP-tagged variant Myb proteins, ectopically express them in cells and perform TAP purification. Many

proteins have been identified to interact with c-Myb and interfere with its activity, some of which are suggested to contribute to the oncogenic capacity of activated forms of Myb ²⁵¹. Therefore, if the interacting proteins characterized for different isoforms of c-Myb were identified as cell cycle regulators, kinases that might modify Myb proteins, and/or the proteins that are involved in chromatin remodeling, we will further study whether these different partners will regulate or affect the activity of variant c-Myb proteins.

Could c-myb alternative splicing affect stem cell fate?

c-Myb protein is clearly involved in the regulation of stem cell proliferation and differentiation in many cell types, including hematopoietic, epithelial and neural cells ^{179 56,57}. Recent developments in the stem cell field have shown that the combined ectopic expression of the transcription factors, KLF4, MYC, SOX2 and OCT4 in primary fibroblasts can induce a stem cell phenotype and the expression of characteristic stem cell specific genes such as NANOG, resulting in the production of induced pluripotent stem (iPS) cells ^{183,252,253}. Interestingly, the MYC gene is regulated by c-Myb in some cell types, suggesting that c-Myb could play a role in stem cell formation by regulating MYC ^{53,254,255}. Moreover, bioinformatics approaches have identified c-Myb as a likely regulator of pluripotency and “stemness” in iPS cells ¹⁸⁴.

Evidence of the influence of alternative splicing in stem cells is rapidly growing. Recent studies indicate the levels of some splice variants of stem cell-enriched genes (stemness genes) correlate with particular stages of differentiation. For example, the POU5F1 gene is an example of a central

stemness gene that is regulated by alternative splicing. This gene encodes transcription factor OCT4 that is highly expressed in stem cells and appears to be essential for reprogramming differentiated cells to an induced pluripotent state²⁵⁶. Finer discrimination among the POU5F1 gene products revealed differentially regulated splice variants²⁵⁷. The levels and the patterns of these variants are different in embryonic stem cells compared with more differentiated cells^{257 258} and they have distinct roles in self-renewal of stem cells²⁵⁹. Our laboratory has shown that alternative splicing of *c-myb* transcripts is highly regulated during the differentiation of primary human hematopoietic progenitor cells¹⁹. Furthermore, our data demonstrates that the alternative splicing occurs more frequently in leukemia and the profile of *c-myb* alternative splicing in each patient is unique. All of the above suggest that by shifting the pattern of *c-myb* alternative splicing, stem cells could produce a different repertoire of c-Myb proteins and express a different subset of Myb target genes than more differentiated cells. Aberrant alternative splicing could produce oncogenic forms of c-Myb that affect the “stemness” of the cells, making the normal hematopoiesis process go awry, resulting in leukemia.

We have different viral vectors that express different c-Myb proteins, including 9S/10, 9B and c-Myb, which can be used to infect different types of cells. We can utilize this viral system combining with other transcription factors that are required for inducing pluripotent stem cells to generate iPS cells. These iPS cells will be cultured in a standardized in-vitro culture system, and their abilities to differentiate along multiple hematopoietic lineages could be assessed.

Alternatively, we can use CD34+/CD133+ hematopoietic progenitor cells as an experimental model for similar purpose. CD34+ hematopoietic progenitor cells have the ability to differentiate along multiple lineages. After infecting the CD34+ cells with variant Myb proteins, these cells will be cultured in semi-solid methylcellulose cultures containing appropriate cytokines and growth factors to permit proliferation and differentiation along multiple pathways. By measuring the numbers and ratios of the formed colonies and analyzing the differentiation status of the cells, we could detect the differences in the activities of variant c-Myb proteins on affecting normal hematopoiesis. These experiments could characterize the activities of variant c-Myb proteins in hematopoietic stem cell differentiation and proliferation. These experiments could also enhance our knowledge of whether *c-myb* alternative splicing could be an important component to determine stem cell fate and if deregulation of this process could contribute to leukemogenesis.

Implication for the future

It is clear that alternative splicing provides a mechanism for producing specific variants of c-Myb with unique functions, which greatly increases the complexity of c-Myb activities in hematopoiesis and oncogenesis. Our studies investigating *c-myb* alternative splicing in leukemia patient samples established an association between alternatively spliced *c-myb* variants and clinical parameters in pediatric pre-B-ALL. The total expression level of the *c-myb* gene has never been indicated by microarray analysis as a component for classifying high-risk leukemia patients. Our results from the novel single molecule analysis

methods reveal aberrant *c-myb* alternative splicing as a novel biomarker in leukemia. Further studies testing the hypothesis that alternative splicing is a novel mechanism to unmask c-Myb oncogenicity will promote deep understanding of the normal and oncogenic functions of c-Myb. Future experiments will determine whether aberrantly expressed Myb variants could regulate hematopoiesis differently, whether these variant Myb proteins regulate target genes that are transforming relevant and whether the interacting protein complexes define their activities. Results from these experiments will expand our knowledge on how c-Myb becomes oncogenic and have a considerable effect on the novel molecularly targeted oncotherapeutics.

APPENDICES

Appendix I: Characterization of variant Myb proteins in NALM-6 cells

Identification of variant Myb proteins target genes is essential for the complete characterization of Myb protein activity. Myb proteins exert their activity through regulating their target genes. For example, c-Myb binds to the genes that are linked to cell proliferation and metastasis, including CXCR4, a gene encoding the receptor for the chemokine SDF-1, and the CCNA1, CCNB1 and CCNE1 genes, which encode the important cell cycle regulators Cyclin A1, Cyclin B1 and Cyclin E1, respectively. These target genes indicate the potential function of c-Myb in cell proliferation and cell survival. Therefore, identification of the genes that are directly bound by variant Myb proteins could assist exploring the underlying molecular mechanism how they contribute to leukemias. By far the best method to follow the activity and specificity of variant Myb proteins is ChIP assay¹⁸⁶. With this experiment, the chromatin-embedded gene promoters that are bound by transcription factor Myb proteins can be identified and compared to one another among variant forms of c-Myb. However, ChIP assay relies heavily on the biophysical properties of the interested protein, for each variant Myb protein, a specific antibody has to be developed and each protein is unique so that developing a general protocol for individual antibodies specific to each variant is not feasible. Therefore, we utilized a lentiviral system to ectopically express Flag-tagged Myb proteins in the cells and only used one Flag-antibody for all of the interested variant Myb proteins to perform ChIP assay^{220,260}.

Materials and Methods

Myb lentiviral Production

N-terminal FLAG-tagged human c-Myb, 9S/10, and 9B transcripts were cloned into the unique PacI site of the pHR IRES GFP lentiviral vector downstream of an EF1 alpha promoter. Plasmids were transfected into 293 FT cells (Invitrogen, Carlsbad CA) by calcium phosphate transfection along with the lentiviral packaging plasmid delta 8.9 and the pMD.G plasmid expressing the vesicular stomatitis virus glycoprotein. Supernatant was collected at 24 hr intervals for 48 hr. Ultrafiltration using an Ambion Ultracell 100 kDa NMWL filter unit (Millipore, Billerica MA) was used to concentrate viral supernatants.

Making Stable Cell Lines

NALM-6 cells were transduced individually with different lentiviruses that express different versions of Myb proteins, at MOI 1:100 in the presence of 8 µg/ml of polybrene (Sigma, St. Louis, MO) in two constitutive days. After 7 day post-transduction, the cells were sorted to enrich for GFP+ transduced cells. These GFP positive cells were then expanded by culturing in IMDM (Iscove's Modified Dulbecco's Medium) (GIBCO, Carlsbad, CA) supplemented with 10%(v/v) FBS, 1 mM L-glutamine, and 50 µM β-mercaptoethanol.

Immunoprecipitation and Western Blot

Cell pellets (2×10^7) were lysed with 1 ml RIPA buffer [1% (v/v) Nonidet P-40 (Sigma, St. Louis, MO), 0.1% SDS, 150 mM NaCl, 50 mM Tris-HCl pH 7.5] with protease inhibitor (1 µM each chymostatin, leupeptin, antipain, pepstatin-A;

1 mM each phenylmethylsulfonyl fluoride and benzamidine) for 30 min at 4 °C. After lysing, samples were spun at 13,000 rpm for 10 min at 4 °C to remove cell debris. About 10% of the supernatant were removed as input for the following western blot, and the rest was subjected to immunoprecipitation (IP). IP was performed with anti-FLAG M2 beads (Sigma, St. Louis, MO) (40 µl) or non-specific IgG beads for 2 hr. Samples were mixed by rotation at 4 °C for 1 hr. After binding, the beads were washed with 800 µl of RIPA buffer 3 times to remove all non-specific protein bindings. Then the proteins were eluted from the beads with 5 x SDS protein loading buffer. Samples were boiled for 10 min, beads were removed by centrifugation and samples were subjected to Western blot analysis with anti-c-Myb.

Chromatin Immunoprecipitation

Standard methods were used except chromatin was sheared in a 200 µl volume of 50 mM Tris pH 8.0 (Sigma, St. Louis, MO) by adding 36 units of Micrococcal nuclease (USB, Cleveland, Ohio) for 10 min at 37 °C^{220,260}. Adding EDTA to 10 mM stopped the reaction and cells were lysed in 1% v/v SDS. Equal aliquots were subjected to immunoprecipitation with anti-FLAG beads or anti-Myb monoclonal 1-1 antibodies (Millipore, Burlington, MA) or control non-specific IgG beads for 24 hr. Anti-Myb monoclonal 1-1 antibodies were used here to recognize the endogenous c-Myb protein. Primers used for QPCR reactions using SYBR green (Biorad, Hercules, CA) are listed in Table 1.

Microarray analysis

RNA was extracted with the RNeasy Mini kit (Qiagen, Valencia, CA) for each cell line (stably expressed GFP, FLAG-c-Myb, FLAG-9S/10, and FLAG-9B). RNA quality was judged using the Lab-on-a-Chip Bioanalyzer 2100 (Agilent, Palo Alto, CA), and total about 2 µg RNA for each cell line was subjected to microarray analysis. Hybridization probes were prepared and hybridized to U133A 2.0 GeneChip exactly as described by the manufacturer (Affymetrix, Santa Clara, CA), using the Superscript Choice System (Invitrogen, Carlsbad, CA) and the RNA transcript labeling kit (ENZO, Farmingdale, NY) for cRNA preparation. Microarray analysis was performed using the Affymetrix GeneArray Scanner G2500A, and the samples were scaled to a median fluorescence of 500 units using Microarray Suite 5.0 software (Affymetrix, Santa Clara, CA). Replicate samples were generated and analyzed on different days. The data was normalized to the average of the control samples and up- or down-regulated genes were identified using GeneSpring GX (Agilent Technologies, Santa Clara, CA).

Results

FLAG-tagged Myb cell lines

The development of Chromatin-IP assays for the detection of Myb proteins target genes provided a means to follow the activity and specificity of variant Myb proteins. However, performing ChIP assay on every variant Myb protein requires developing specific antibody and different antibody works under different

condition. In order to circumvent these possible problems, we used lentiviral vectors to generate stably transduced derivative of NALM-6 cells expressing FLAG epitope-tagged version of variant Myb proteins (GFP, 9S/10, 9B and c-Myb). For these cells, only one antibody (anti-Flag) is needed for all of the ChIP assays. Since our previous data on the significance of variant Myb proteins in leukemia were generated from pre-B ALL patients, we chose the pediatric pre-B-ALL cell line NALM-6. We first measured the expression level of the ectopically expressed Myb proteins in various stable cell lines. As shown in Figure 1A, FLAG-9B was expressed at levels comparable to the level of FLAG-c-Myb and FLAG-9S protein was expressed at a slightly higher level in these cells. Endogenous c-Myb proteins were also expressed in these cells, and the levels of the ectopically expressed variant Myb proteins were very close to the level of endogenous c-Myb protein. This may avoid the possible artifacts caused by high level of overexpression of these proteins, such as, non-specifically binding and regulating some genes. To further confirm that one single antibody anti-FLAG can be used for all of these FLAG-tagged variant Myb proteins in ChIP assay, immunoprecipitation experiments were carried out on these stable cell lines. The results showed that the anti-FLAG antibody could efficiently immunoprecipitated the epitope-tagged variant Myb proteins from these cells (Figure 1B).

Chromatin-IP on the stable cell lines

The results described above suggest that anti-FLAG antibody would be suitable for performing ChIP assay on all of the variant Myb proteins^{220,260}. In order to be able to tell whether ChIP assay works in these stable cell lines, it is

necessary to know some of the target genes that might be regulated by variant Myb proteins. For c-Myb protein, many target genes have been identified in the past, but little is known about the target genes regulated by variant Myb proteins. Furthermore, many studies suggest that the Myb proteins are subject to context-specific regulatory mechanisms, and in different cell types the Myb proteins form different transcription complex with different cellular components and activate specific sets of genes (see review ¹⁷²). While c-Myb play an important role in B-cell development ^{38,39,213,261}, little is known about the target genes regulated by c-Myb during B-cell differentiation and leukemogenesis. Our laboratory has identified genome-wide target genes regulated by c-Myb in MCF-7 cells by using ChIP-chip. Some target genes (MAT2A, CCNB1) are shown regulated by c-Myb in some types of hematopoietic cells, such as Jurkat, CD34+ cells. Therefore, by performing ChIP assay, we assessed the association of c-Myb, 9S/10 and 9B proteins with the promoter of MAT2A, CCNB1, and RAG2 (a target gene regulated by c-Myb protein in B-cell) ^{262,263} genes in variant Myb proteins expressed NALM-6 cell lines. Briefly, NALM-6 cells with ectopically expressed GFP, FLAG-c-Myb, FLAG-9S/10 and FLAG-9B were fixed with formaldehyde and the chromatin complexes were immunoprecipitated using control non-immune IgG, anti-flag beads or a c-Myb specific antibody (1-1) that recognizes a domain (after 440 aa in c-Myb protein) in c-terminus. This c-Myb specific antibody (1-1) could be used to pull down endogenous c-Myb protein formed chromatin complex, while the anti-FLAG antibody could pull down ectopically expressed variant Myb proteins formed chromatin complexes. The immunoprecipitated

chromatin complexes were then assayed for the enrichment of the MAT2A, CCNB1, RAG2 promoters, or control promoters, using quantitative real-time PCR (QPCR) assays (see details in Table 1).

Our results showed that anti-Myb antibody detected endogenous c-Myb at the MAT2A and RAG2 promoter (Figure 2) in all of the NALM stable cell lines, but no c-Myb could be detected at the CCNB1 promoter (data not shown) and control promoter (GAPDH), suggesting that c-Myb associated with the MAT2A and RAG2 promoter in NALM-6 cells. When using anti-FLAG beads, we first tested the ability of the anti-FLAG beads to enrich for the MAT2A and RAG2 promoters in NALM GFP cells with only endogenous c-Myb expression and NALM-6 FLAG-c-Myb cells with both endogenous and exogenous c-Myb expression. As shown in Figure 2A, no gene promoter was enriched for anti-FLAG beads in NALM-6 GFP cells, indicating there is no non-specific DNA binding to the anti-FLAG beads. While in FLAG-c-Myb cells, the anti-FLAG beads showed the similar enrichment on the MAT2A and RAG2 promoters (Figure 2B) compared to the anti-Myb antibodies, implying that the FLAG-tagged c-Myb protein has the same activity as the endogenous c-Myb protein on regulating genes.

In NALM-6 FLAG-9S/10 cells, the anti-c-Myb antibodies efficiently enriched for the MAT2A and RAG2 promoters, but the anti-FLAG beads did not enrich for either of the two gene promoters (Figure 2C). This result suggested that 9S/10 has different activity on the MAT2A and RAG2 genes from the activity of c-Myb. We further tested whether 9S/10 variant protein has different activity on the other

gene promoter (CCNB1) that was not enriched by c-Myb in NALM-6 cells and there was no detected enrichment on the CCNB1 promoter (data now shown). In NALM-6 FLAG-9B cells, the anti-FLAG beads enriched efficiently on the MAT2A promoter and the RAG2 promoter (Figure 2D) but not on the CCNB1 promoter. Comparing to endogenous c-Myb protein, FLAG-9B showed very similar activities on regulating these genes in NALM-6 cells. Based on all of these results, we conclude that c-Myb, 9S/10 and 9B variant proteins associate with different target gene promoters in NALM-6 cells and that ChIP assay on the cells stably expressing FLAG-tagged variant Myb proteins could be a powerful tool to detect the activity and specificity of Myb proteins.

Gene-expression profiles in NALM-6 cells with variant Myb proteins

expression

Our laboratory has used microarray method to compare the transcription specificities of Myb proteins (A-Myb, B-Myb, c-Myb and v-Myb) in different cell types, and identified sets of putative target genes of Myb proteins^{36,37,78}, though a lot of them still need to be further verified. Here, we applied the same idea on the various NALM-6 cells with the stable expression of different variant Myb proteins, and utilized a microarray approach to compare the specificities of c-Myb, 9S/10 and 9B proteins. Briefly, NALM-6 cells expressing variant Myb proteins and control GFP were seeded at 5×10^5 the night before. After 16 h, RNA was isolated and changes in gene expression were determined using Affymetrix microarrays. The entire experiment was conducted twice, and the data was analyzed to identify the genes that were reproducibly up or down-regulated

at least twofold as a result of expressing one of the variant Myb proteins. However, unexpectedly, there were very few genes (total 21 genes) affected by the expression of variant Myb proteins and the changes in gene expression for these genes were not significant (see details in Table 2). To partially validate the microarray results, we performed real-time PCR analysis to independently assess the changes in gene expression for several representative genes, such as MLL2 gene that was only induced by the expression of FLAG-c-Myb, PLAC8 gene that was only induced by the expression of 9B variant, and GNG11 gene that was more significantly repressed by the expression of 9S/10 variant. The results of the real-time PCR analysis (data not shown) were very similar as the microarray results. All of the above indicate that variant Myb proteins had non-distinct and almost similar effects on gene expression in these NALM-6 stable cell lines.

Discussion

These data demonstrated that by developing cells expressing FLAG-tagged variant Myb proteins, ChIP assays on different versions of Myb proteins could be greatly simplified by using only one anti-FLAG antibody, and it would be a useful tool to identify the target genes of variant Myb proteins. However, we were not able to identify the transcriptional specificities of each variant Myb protein in NALM-6 cells by microarray approach. It is possible that normally we compare the transcriptional specificities of Myb proteins by using recombinant adenovirus system, which can ectopically express proteins at very high level (much higher than the level of endogenous proteins). While we were using lentiviral system to

express variant Myb proteins in these NALM-6 cells, the expression levels of these variant proteins were very close to the level of endogenous c-Myb protein. Also, it is very hard to introduce exogenous genes into NALM-6 cells. We have tried used other ways such as adenovirus or electroporation to introduce these c-Myb transcript variants into NALM-6 cells, and they all failed to do so. When using lentiviral system, the cells that were positively transduced (GFP+) were only 3-10% of the total population after 7 days, and these cells were sorted out for making stable cell lines. In the whole process of making stable cell lines of NALM-6 cells, these cells went through a long term of massive selection and the initial changes in gene expression led by different activities of these variant Myb proteins probably could have been set off.

To circumvent these possible problems, we could choose different cell models, such as Jurkat cells. This cell line is much easier for introducing exogenous genes and the transcriptional activities of c-Myb protein in this cell line has been well-characterized by ChIP assay²⁶⁰. Or instead of growing the positively transduced NALM-6 cells into stable cell lines for a long term, we could just sort the positive cells after being transduced for 4 days, and directly perform ChIP assay and microarray on these sorted cells and do a short term experiment. Using these new model systems, the progress in discovering the transcriptional specificities of variant Myb proteins and identifying the different sets of target genes of these variant Myb proteins will be made at tremendous strides.

Figure Legends

Figure 1. The expression of variant Myb proteins in stable cell lines

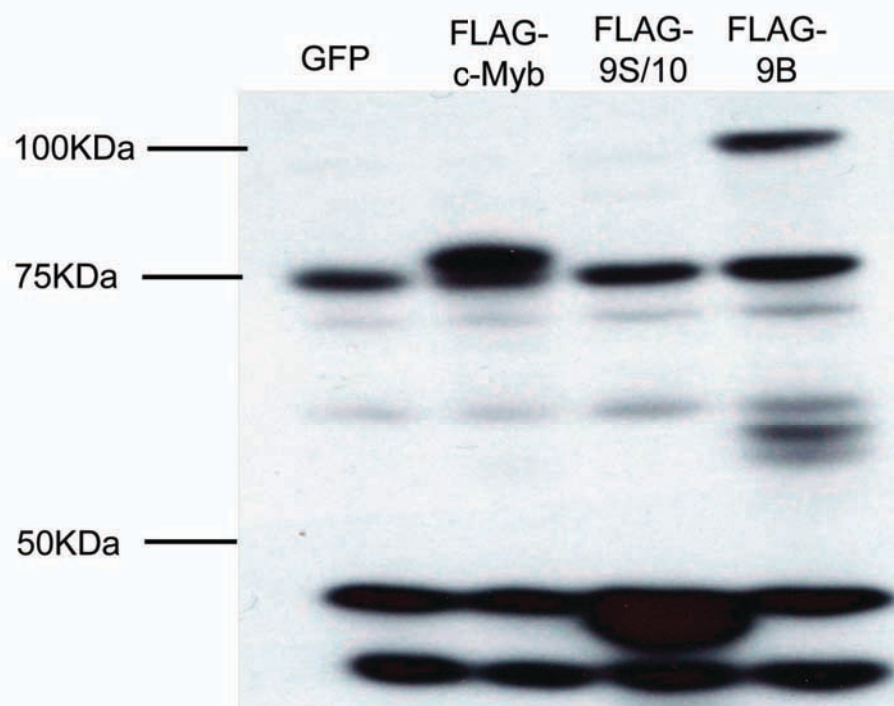
(A) Western blot of lentiviral transduced NALM-6 cells. Cell lysates were directly subject to Western Blot analysis. Each cell line expresses FLAG-tagged c-Myb, 9S/10 and 9B, respectively. Western Blot was analyzed with anti-c-Myb antibody. (B) Immunoprecipitation on lentiviral transduced NALM-6 cells. Lysates from cells were subjected to IP with anti-FLAG beads or control IgG beads and Western Blot was performed with anti-c-Myb antibody.

Figure 2. FLAG-tagged variant Myb proteins bind to different gene promoters.

ChIP assay using anti-FLAG beads and anti-c-Myb antibody 1-1 and chromatin from GFP control (A), FLAG-c-Myb (B), FLAG-9S/10 (C), and FLAG-9B (D) expressing NALM-6 cells. QPCR was performed with primers specific to the MAT2A and RAG2 promoters, and primers designed to recognize no-specific binding site within the GAPDH promoter. Black bar in each plot represents the ChIP results from anti-c-Myb 1-1 antibody, and the white bar represents the ChIP results from anti-FLAG beads. Error bars represent standard deviation from triplicate QPCR reactions. Each sample is normalized to IgG control.

Figure 1

A



B

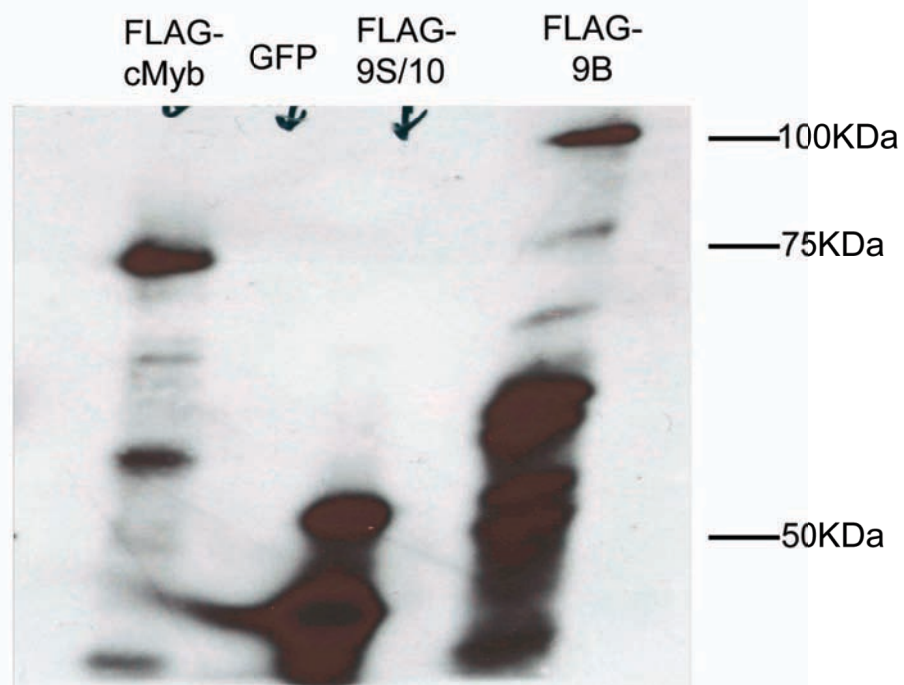
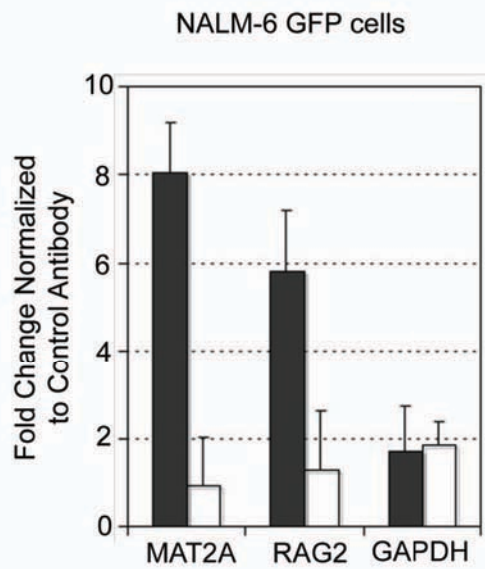
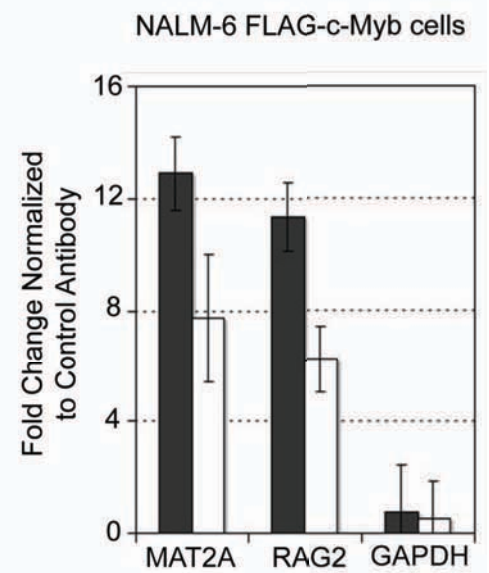


Figure 2

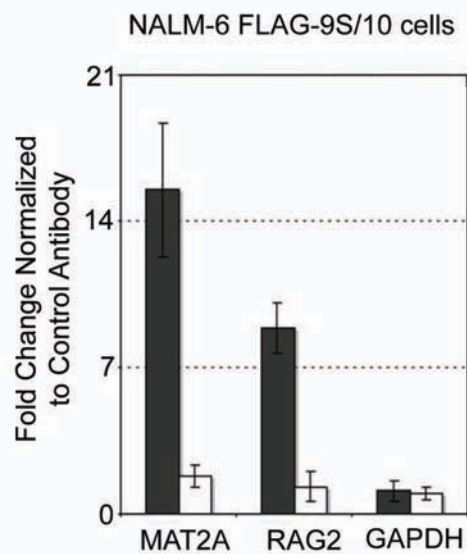
A



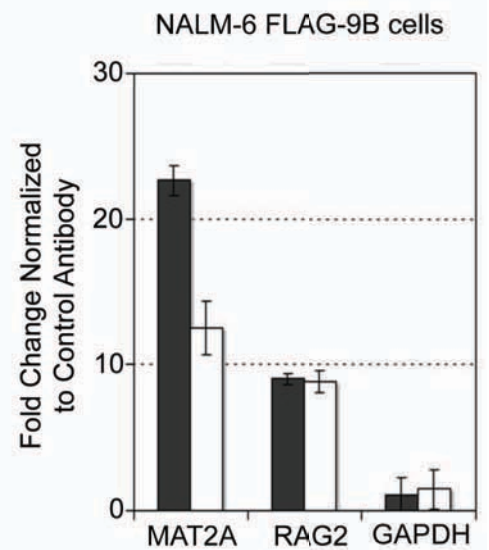
B



C



D



Tables

Table 1: Promoter Primers for SYBR Green

Gene Promoters	Sequence
MAT2A promoter For	ACCCACCAAACGGGACTCCATACTC
MAT2A promoter Rev	TCTCTTGACGGCGTGGAATGG
RAG2 promoter For	GTGAATTGTGTTGCCATTGTTGC
RAG2 promoter Rev	TGTGTGCCTACAGATGTTC
CCNB1 promoter For	TTGTGCCCCACCTTAAT
CCNB1 promoter Rev	CATGTTGATCTTCGCCTTATT
GAPDH promoter For	GTGCAGGAGTTCATGACCAAC
GAPDH promoter Rev	CACCTGGGCTAGAGTAC

Table 2: The gene expression changes induced by variant Myb proteins

Symbol	Unigene	Description	FLAG-9B	FLAG-9S	FLAG-Myb
GNG11	HS.83381	guanine nucleotide binding protein (G protein), gamma 11	0.65	0.27	0.61
RAP1A	Hs.190334	RAP1A, member of RAS oncogene family	0.28	0.09	0.05
CD99	Hs.654354	CD99 molecule	2.63	1.72	1
RBP1	Hs.529571	retinol binding protein 1, cellular	0.93	1.25	0.41
GZMA	Hs.90708	granzyme A(granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3)	0.79	0.69	3.38
KCNJ2	Hs.1547	postassium inwardly-rectifying channel, subfamily J, member 2)	0.48	0.58	1.43
PIP3-E	Hs.146100	phosphoinositide-binding protein PIP3-E	0.47	0.68	1
PLAC8	Hs.546392	placenta-specific 8	2.11	1.38	0.94
SORBS1	Hs.696027	sorbin and SH3 domain containing 1	0.96	0.76	2.17
SGCB	Hs.438953	sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)	0.79	0.98	0.38
MLL2	Hs.120228	myeloid/lymphoid or mixed-lineage leukemia 2	1.31	1.13	3.47
	Hs.655939				
	Hs.675711				
	Hs.679013				
SPP1	Hs.313	secreted phosphoprotein 1(osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	0.37	0.4	0.44
LEF1	Hs.555947	lymphoid enhancer-binding factor 1	1.99	1.54	1.57
CALD1	Hs.490203	caldesmon 1	0.44	0.52	0.67
		1-acylglycerol-3-phosphate-O-acyltransferase 2(lysophosphatidic acid acyltransferase,beta)			
AGPAT2	Hs.320151		0.53	0.61	0.66
SMOC2	Hs.487200	SPARC related modular calcium binding 2	0.48	0.77	0.52
DDIT4L	Hs.480378	DNA-damage-inducible transcript 4-like	0.28	0.41	0.5
		G-protein signaling modulator 1 (AGS3-like, c.elegans)			
GPSM1	Hs.239370		0.75	0.92	0.45
NPY	Hs.1832	neuropeptide Y	0.76	0.83	0.49
CASP1	Hs.2490	caspase 1, apoptosis-related cysteine peptidase(interleukin 1, beta, convertase)	0.81	0.67	0.43
	Hs.				
IRX1	424156	iroquois homeobox 1	0.71	0.83	0.45

Table 2 Legends. The gene expression data of 21 genes that were most significantly up or down-regulated in response to stably ectopically expressing FLAG-c-Myb, FLAG-9S or FLAG-9B in NALM-6 cells. Repressed genes are shown in green, and activated genes are in red over a range of 0.2-5 fold activation. Each gene is identified by its Affymetrix probe set identification number and by its common name, when available.

Appendix II: Develop cell lines ectopically expressing TAP-tagged variant Myb proteins to study protein-protein interaction

Our recent studies show that alternately spliced *c-myb* transcripts are associated with polyribosomes, which indicates that they are translated to produce different forms of c-Myb proteins ¹⁹. Variant proteins with different structures, especially with unique C-terminal regions, are likely to bind different co-factors that may confer different biological activities on them.

To identify the different proteins interacting with Myb proteins we used a new technique tandem affinity purification to isolate Myb protein-complex and use proteomics to identify the proteins that different versions of c-Myb bind to. Tandem affinity purification is a generic two-step affinity purification protocol for the isolation of TAP-tagged proteins together with associated proteins ²⁶⁴. It requires fusion of the TAP tag, either N- or C-terminally, to the target protein of interest. All of the steps in TAP purification are illustrated in Figure 3.

For our purpose, we generated a series of vectors expressing TAP-tagged variant Myb proteins, in which the TAP tag was fused to the N-termini of Myb proteins. By using a luciferase assay to test the functional ability of TAP-tagged c-Myb protein, we confirmed that the TAP tag did not change c-Myb protein functions. Further, we established two cell lines (Jurkat and MCF-7 cells) that can stably express TAP-tagged c-Myb and performed TAP purification on these cells. The proteins recovered from the TAP purification was subject to western blotting to check whether c-Myb formed protein complex could be efficiently purified.

Our preliminary data indicates that TAP-tag does not affect the normal function of Myb proteins, TAP-tagged Myb variant proteins could be stably expressed in different cell lines and TAP purification could be efficiently applied to the study on comparing the difference in the protein complexes formed with variant Myb proteins.

Materials and Methods

Generation of NTAP-MYB

TAP-tag sequence was amplified from pZome-1-N vector (CellZome, Germany) and then cloned in pcDNA3 vector (Invitrogen, Carlsbad CA) to generate pcDNA3-NTAP vector. The human c-Myb transcript sequence and its variants 9S/10 and 9B were then cloned into pcDNA3-NTAP vector, respectively. The lentiviral vectors with TAP-tagged Myb were constructed by cloning NTAP-Myb into pLenti6/V5-GW/LacZ vector (Invitrogen, Carlsbad CA), which would replace the entire sequence from attB1 to V5 epitope in the original vector.

Myb Lentiviral Production

N-terminal TAP-tagged human c-Myb, 9S/10, and 9B transcripts were cloned into the unique BamHI and XhoI site of the pLenti6/V5-GW/LacZ lentiviral vector downstream of a CMV promoter. Plasmids were transfected into 293 FT cells (Invitrogen, Carlsbad CA) by calcium phosphate transfection along with the lentiviral packaging plasmid delta 8.9 and the pMD.G plasmid expressing the vesicular stomatitis virus glycoprotein. Supernatant was collected at 24 hr

intervals for 48 hr. Ultrafiltration using an Ambion Ultracell 100 kDa NMWL filter unit (Millipore, Billerica MA) was used to concentrate viral supernatants.

Cell Culture

MCF-7 cells (ATCC Manassas, VA) and 293 cells (ATCC CRL-1573) were cultured at 37°/5% CO₂ in DMEM (Dulbecco's Modified Eagle Medium) (Invitrogen, Carlsbad, CA) supplemented with 10% (v/v) fetal bovine serum (Invitrogen, Carlsbad, CA). Human Jurkat T-cells (ATCC TIB152 Manassas, VA) were cultured at 37°/5% CO₂ in RPMI + Glutamax medium (Invitrogen, Carlsbad, CA) supplemented with 10% (v/v) fetal bovine serum (Invitrogen, Carlsbad, CA).

Transfection assay

For reporter gene activation assays, triplicate samples of 293 cells in 24-well dishes were transfected with 350 ng of the c-Myb-responsive EW5 reporter gene, along with 50 ng of the pcDNA3 c-Myb expression plasmid (or pcDNA3-NTAP-c-Myb expression plasmid or pcDNA3 plasmid) and 10 ng of a renilla luciferase plasmid as a control for transfection efficiency. This transfection assay were used Polyfect as instructed by manufacturer (QIAGEN, Valencia, CA). Luciferase levels were analyzed 2 days later with a dual luciferase kit (Promega, Madison, WI).

Making Stable Cell Lines

Cells were transduced individually with different lentiviruses that express different versions of Myb proteins, at MOI 1: 100 in the presence of 8 µg/ml of polybrene (Sigma, St. Louis, MO) in two constitutive days. After 7 day post-

transduction, the cells were sorted to by adding 8 -10 µg/ml blasticidin to drug-select for positively transduced cells. These blastcidin-resistant positive cells were then expanded by culturing in their respective media as described in “cell culture” with 8 - 10 µg/ml blasticidin.

TAP Purification and Western Blot

TAP purification was performed as described^{265,266}. In brief, approximately 2×10^7 cells were harvested. Cells were flash frozen and lysed with 5 ml lysis buffer (10% glycerol, 50 mM Hepes-KOH pH 8.0, 100 mM KCl, 2 mM EDTA, 0.1% NP-40, 2 mM DTT, 1X Sigma Protein inhibitor, 10 mM NaF, 0.25 mM NaOVO₃, 5 nM okadaic acid, 5 nM calyculin A and 50 mM beta-glycerolphosphate). After lyzing, the debris was spun down for 10 min at maximum speed at 4 °C. The clear supernatant was recovered and 100 µl was taken for input control. Then 30 µl IgG beads (Amersham, Piscataway, NJ) (60 µl of slurry) was added to the extract and rotated at 4 °C for 2 hr (after binding, the supernatant was aliquoted a portion for western blot analysis). The beads were then washed 3 times with 1ml lysis buffer and 3 times with TEV buffer (10 mM Hepes-KOH pH8.0, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA and 1mM DTT). Beads were resuspended in 120 µl of TEV buffer, 40 units of AcTEV protease (Invitrogen, Carlsbad, CA) and rotated for 4 hr at 4 °C. After digesting, the supernatant was transferred into the same amount of Calmodulin beads (30 µl) (here also aliquot a small portion) and the IgG beads were washed 2 times with 1 volume of Calmodulin binding buffer (10 mM beta-mercaptoethanol, 10 mM Hepes-KOH pH8.0, 150 mM NaCl, 1 mM MgOAC, 1 mM imidazole, 0.1% NP-40,

and 2 mM CaCl₂). These washing buffers were also transferred to Calmodulin beads (Amersham, Piscataway, NJ). 1/250 volume of 1 M CaCl₂ was also added into the beads. The calmodulin beads were rotated for 90 min at 4 °C (after binding, the supernatant here was also aliquoted for western blot analysis). The beads were washed 3 times with 1 ml Calmodulin binding buffer. The bound protein complex was then eluted from the calmodulin beads by calmodulin elution buffer (50 mM Amm.bicarb pH 8.0, 75 mM NaCl, 1 mM MgOAc₂, 1 mM imidazole, 50 mM EDTA and 2 mM CaCl₂). All of the protein portions that aliquoted from each step and the final eluted protein were subject to western blot.

Results

TAP-tag does not affect the transcriptional activities of c-Myb protein

Compare to other available tag protein, such as HIS, FLAG, or GST, TAP tag is a relatively bigger, about 21 KDa, protein tag. To make sure that fusing such big tag to N-terminal of variant Myb proteins would not affect their normal functions, we compared the activities of c-Myb protein and TAP-tagged c-Myb protein by using a well-established co-transfection assay incorporating a Myb responsive reporter gene. Myb-negative 293 cells were transfected with plasmids expressing normal and TAP-tagged c-Myb proteins, along with a Myb-responsive reporter plasmid containing five high-affinity Myb binding sites from the Myb-regulated *mim-1* gene promoter upstream of a minimal promoter^{26,27,36,267} and a luciferase reporter gene. As shown in Figure 4, the expression of c-Myb resulted in an approximately 14 folds increase in reporter gene activity compared to the

vector (pCDNA3) control. The TAP-tagged c-Myb protein resulted in a 12-fold activation in luciferase, which is similar to that of the normal c-Myb protein. The results suggest that TAP tag does not affect the normal activities of variant Myb proteins.

Purification of c-Myb Complexes

Tandem affinity purification of protein complexes has become an important tool in the field of proteomic research. Through performing TAP purification, novel protein-protein interactions could be identified. To apply the TAP method to the identification of different protein-partners of variant Myb proteins, we constructed TAP-c-Myb vectors and prepared lentiviruses, to infect human breast cancer MCF-7 cells and T lymphocyte Jurkat cells. Through blasticidin selection, the positive-infected cells were further cultured to make stable cell lines expressing TAP-c-Myb proteins. We compared the expression level of the TAP tagged c-Myb protein with the level of endogenous c-Myb protein in these stable cell lines. As shown in Figure 5A, endogenous c-Myb was not detectable in MCF-7 cells while the TAP-c-Myb proteins (100KDa) in both MCF-7 cells and Jurkat cells were expressed at the levels comparable with the expression level of endogenous c-Myb protein in Jurkat cells. To further test whether a tandem affinity tag purification method works in our system, we first chose to purify the c-Myb protein complex from MCF-7 cells expressing TAP-c-Myb. In brief, total 2×10^7 cells were seeded in 150-mm culture dishes at 5×10^5 cells/ml the night before and were harvested 14 hr later. The cells were harvested by flash freezing. Protein extracts were prepared from these cells and incubated with IgG-

Sepharose 6 Fast Flow beads. The IgG beads were recovered by brief centrifugation, and subsequently incubated with TEV protease 4 hr at 4 °C. A portion of the elutes obtained after TEV cleavage were analyzed by western blot. As shown in Figure 5B, compared to original cell lysates, after cleaving, the c-Myb protein with only Calmodulin binding domain was efficiently released. The rest elutes were further incubated with Calmodulin beads for 90 min at 4 °C. A portion of the unbound supernatant was removed for western blot analysis. After washing the beads, proteins bound to calmodulin beads were eluted by calmodulin elution buffer and subject to western blot. We found that most of the elutes after TEV cleavage were efficiently bound to Calmodulin beads, a few was left in the unbound supernatant. After second purification with Calmodulin beads, the 75-KDa c-Myb proteins were successfully eluted from the beads. Our results suggest that the two-step tandem purification procedure could efficiently purify TAP- Myb protein complexes from the cells expressing TAP tagged variant Myb proteins.

Implications

In many transcription systems the co-activators have turned out to be considerably more complex than initially envisioned. To fully understand the mechanism of action of any given transcription factor, knowledge of its interacting partners is essential. Therefore, it is important to develop an efficient purification method for isolating protein complexes. Affinity purification appears to be the most efficient and gentle separation technique for the retrieval of protein complexes, such as well-characterized small affinity tags FLAG and GST. TAP

purification procedure consists of two successive affinity purification steps, which can produce an extremely pure preparation of the protein of interest with associated proteins in yeast and mammalian systems^{264 265,268-270}. Here, we explored the possible application of the TAP purification approach in identifying the protein partners of variant Myb proteins. The expression of the TAP-tagged exogenous c-Myb protein in the cell lines were similar to endogenous protein levels, which is important for the identification of protein partners present in physiologically relevant Myb protein complexes. The efficiency of the TAP purification was demonstrated by our western blot results that significant amount of c-Myb was recovered even from small cell numbers. For future work, we could perform TAP purification on more starting numbers of cells to obtain enough proteins, which could be enough for separating on SDS-PAGE gel and visualizing by silver staining. Visible bands will be excised and subjected to proteomics. From these preliminary results, it is clear that the TAP purification could assist the future research on identifying the different protein-partners of variant Myb proteins and defining the possible mechanisms that differentiate variant Myb proteins specificities.

Figure Legends

Figure 3. Schematic of the tandem affinity purification of NTAP-c-Myb protein complex.

TAP tag comprised Protein A, a TEV protease cutting sequence, and calmodulin binding domain (CBD). NTAP-c-Myb formed protein complex were

first pull down by proteinA-IgG beads binding. After binding to the IgG beads, the complex was released by cleaving with AcTEV protease. Then the second affinity purification was performed by CBD binding to calmodulin beads. The final protein complex was eluted by elution buffer, and the protein partners of c-Myb were subject to proteomic analysis.

Figure 4. Activation of a c-Myb-responsive reporter gene by TAP-c-Myb and normal c-Myb proteins.

Cells were co-transfected with plasmids expressing the indicated Myb proteins and a Myb responsive reporter plasmid, and extracts were prepared and assayed for Myb-dependent luciferase activity after 2 days. Results are plotted as the average normalized luciferase activities from triplicate measurements in a representative experiment performed twice. Error bars represent standard deviation.

Figure 5. Tandem affinity purification of TAP-c-Myb protein.

(A) TAP-tagged c-Myb expressed in lentiviral transduced MCF-7 cells and Jurkat cells. Total cell lysates were subjected to western blot analysis. The blot was performed with anti-c-Myb antibody 1-1. (B) Tandem affinity purification on lentiviral transduced MCF-7 cells. A portion of the proteins from each step was analyzed by Western Blot. The blot was detected by blotting with anti-c-Myb antibody 1-1. Lane 1,3,5, and 7 are from MCF-7 LacZ control cell lines; Lane

2,4,6, and 8 are from MCF-7 TAP-c-Myb expressing cells. Lane 1, and 2 are total cell lysate prior to purification. Lane 3, and 4 are protein elutes after cleaving with TEV protease. Lane 5, and 6 are supernatants after proteins binding to calmoduline beads. Lane 7, and 8 are proteins eluted from calmodulin beads. Numbers at left are molecular masses (KDa).

Figure 3

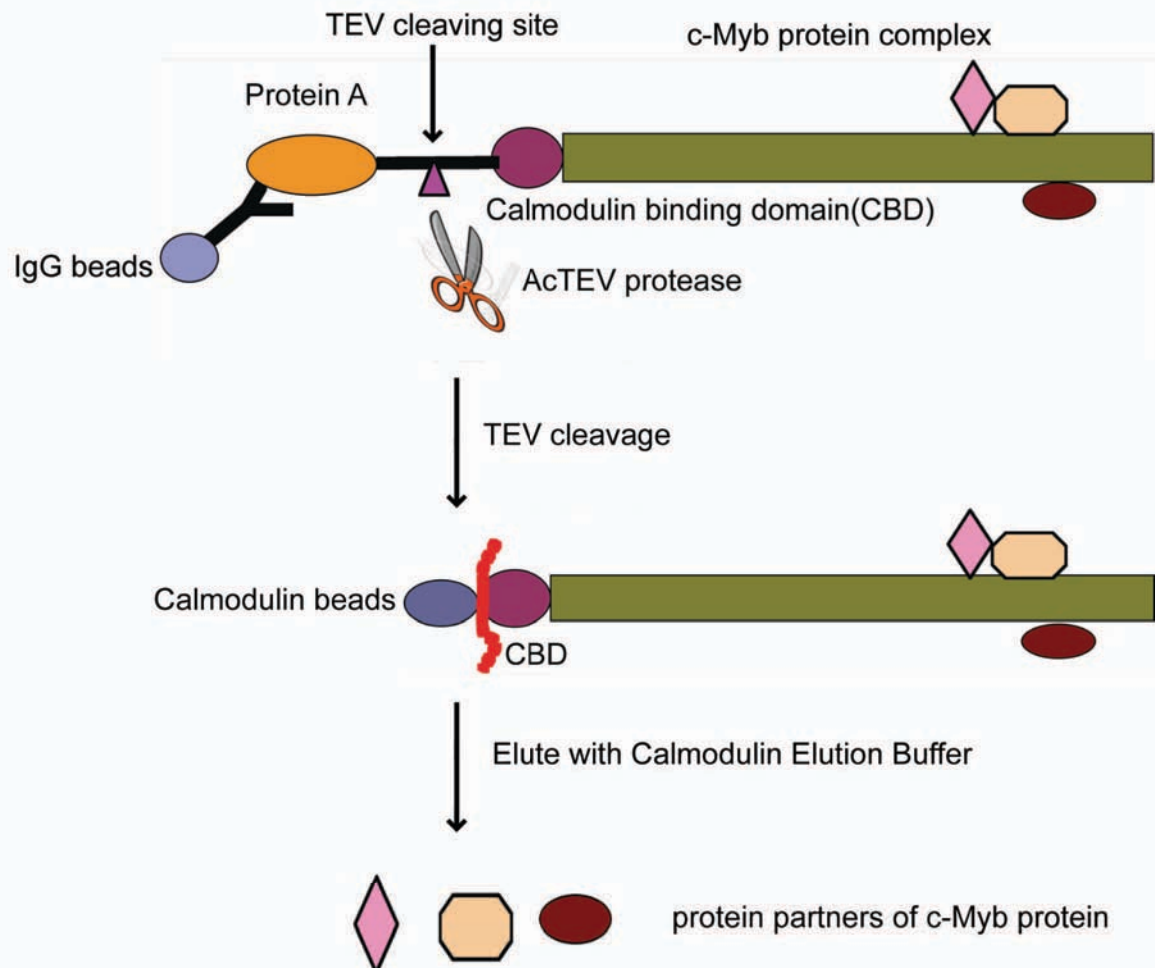


Figure 4

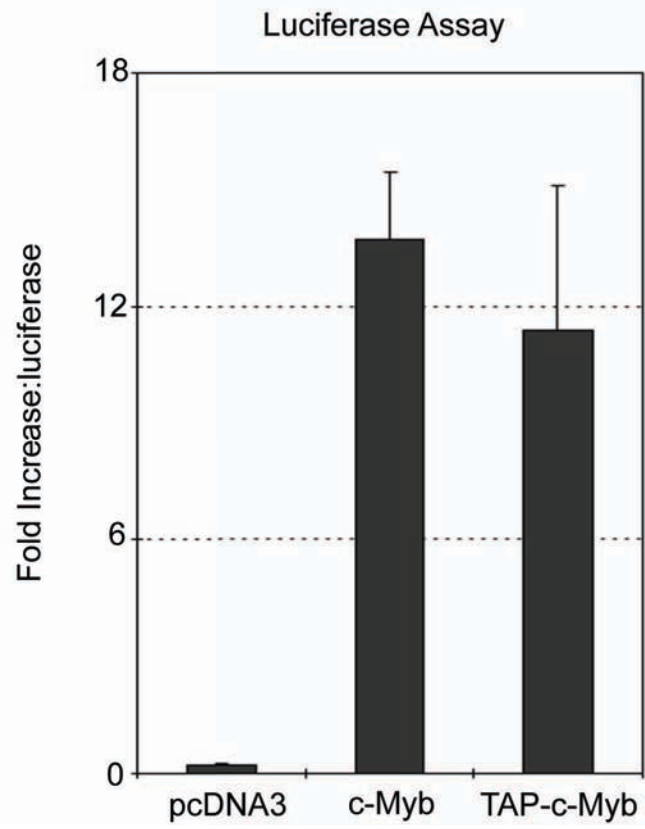
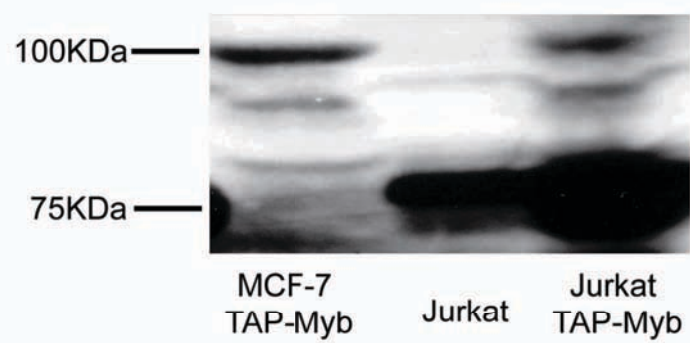
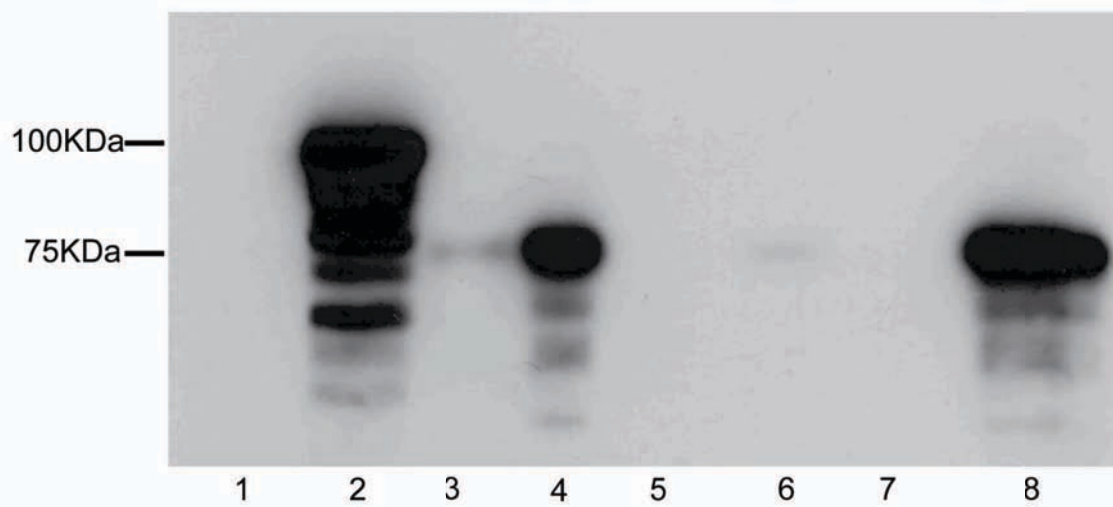


Figure 5

A



B



Reference

1. Klemphauer KH, Gonda TJ, Bishop JM. Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. *Cell*. 1982;31:453-463.
2. Gonda TJ, Bishop JM. Structure and transcription of the cellular homolog (c-myb) of the avian myeloblastosis virus transforming gene (v-myb). *J Virol*. 1983;46:212-220.
3. Pelicci PG, Lanfranccone L, Brathwaite MD, Wolman SR, Dalla-Favera R. Amplification of the c-myb oncogene in a case of human acute myelogenous leukemia. *Science*. 1984;224:1117-1121.
4. Kauraniemi P, Hedenfalk I, Persson K et al. MYB oncogene amplification in hereditary BRCA1 breast cancer. *Cancer Res*. 2000;60:5323-5328.
5. Drabsch Y, Hugo H, Zhang R et al. Mechanism of and requirement for estrogen-regulated MYB expression in estrogen-receptor-positive breast cancer cells. *Proc Natl Acad Sci U S A*. 2007;104:13762-13767.
6. Persson M, Andren Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proc Natl Acad Sci U S A*. 2009;106:18740-18744.
7. Torelli G, Venturelli D, Colo A et al. Expression of c-myb protooncogene and other cell cycle-related genes in normal and neoplastic human colonic mucosa. *Cancer Res*. 1987;47:5266-5269.

8. Ramsay RG, Thompson MA, Hayman JA, Reid G, Gonda TJ, Whitehead RH. Myb expression is higher in malignant human colonic carcinoma and premalignant adenomatous polyps than in normal mucosa. *Cell Growth Differ.* 1992;3:723-730.
9. Thompson MA, Flegg R, Westin EH, Ramsay RG. Microsatellite deletions in the c-myb transcriptional attenuator region associated with over-expression in colon tumour cell lines. *Oncogene.* 1997;14:1715-1723.
10. Hugo H, Cures A, Suraweera N et al. Mutations in the MYB intron I regulatory sequence increase transcription in colon cancers. *Genes Chromosomes Cancer.* 2006;45:1143-1154.
11. Wallrapp C, Muller-Pillasch F, Solinas-Toldo S et al. Characterization of a high copy number amplification at 6q24 in pancreatic cancer identifies c-myb as a candidate oncogene. *Cancer Res.* 1997;57:3135-3139.
12. Welter C, Henn W, Theisinger B, Fischer H, Zang KD, Blin N. The cellular myb oncogene is amplified, rearranged and activated in human glioblastoma cell lines. *Cancer Lett.* 1990;52:57-62.
13. Engelhard HH. Antisense Oligodeoxynucleotide Technology: Potential Use for the Treatment of Malignant Brain Tumors. *Cancer Control.* 1998;5:163-170.
14. Dasgupta P, Linnenbach AJ, Giaccia AJ, Stamato TD, Reddy EP. Molecular cloning of the breakpoint region on chromosome 6 in cutaneous malignant melanoma: evidence for deletion in the c-myb locus and translocation of a segment of chromosome 12. *Oncogene.* 1989;4:1201-1205.

15. Walker MJ, Silliman E, Dayton MA, Lang JC. The expression of C-myb in human metastatic melanoma cell lines and specimens. *Anticancer Res.* 1998;18:1129-1135.
16. Clappier E, Cuccuini W, Kalota A et al. The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. *Blood.* 2007;110:1251-1261.
17. O'Neil J, Tchinda J, Gutierrez A et al. Alu elements mediate MYB gene tandem duplication in human T-ALL. *J Exp Med.* 2007;204:3059-3066.
18. Lahortiga I, De Keersmaecker K, Van Vlierberghe P et al. Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. *Nat Genet.* 2007;39:593-595.
19. O'Rourke JP, Ness SA. Alternative RNA splicing produces multiple forms of c-Myb with unique transcriptional activities. *Mol Cell Biol.* 2008;28:2091-2101.
20. Poenitz N, Simon-Ackermann J, Gratchev A et al. Overexpression of c-myb in leukaemic and non-leukaemic variants of cutaneous T-cell lymphoma. *Dermatology.* 2005;211:84-92.
21. Gewirtz AM, Calabretta B. A c-myb antisense oligodeoxynucleotide inhibits normal human hematopoiesis in vitro. *Science.* 1988;242:1303-1306.
22. Mucenski ML, McLain K, Kier AB et al. A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell.* 1991;65:677-689.

23. Moscovici MG, Jurdic P, Samarut J, Gazzolo L, Mura CV, Moscovici C. Characterization of the hemopoietic target cells for the avian leukemia virus E26. *Virology*. 1983;129:65-78.
24. Radke K, Beug H, Kornfeld S, Graf T. Transformation of both erythroid and myeloid cells by E26, an avian leukemia virus that contains the myb gene. *Cell*. 1982;31:643-653.
25. Frykberg L, Metz T, Brady G et al. A point mutation in the DNA binding domain of the v-myb oncogene of E26 virus confers temperature sensitivity for transformation of myelomonocytic cells. *Oncogene Res*. 1988;3:313-322.
26. Ness SA, Marknell A, Graf T. The v-myb oncogene product binds to and activates the promyelocyte-specific mim-1 gene. *Cell*. 1989;59:1115-1125.
27. Introna M, Golay J, Frampton J, Nakano T, Ness SA, Graf T. Mutations in v-myb alter the differentiation of myelomonocytic cells transformed by the oncogene. *Cell*. 1990;63:1289-1297.
28. Suhasini M, Pilz RB. Transcriptional elongation of c-myb is regulated by NF-kappaB (p50/RelB). *Oncogene*. 1999;18:7360-7369.
29. Watson RJ. A transcriptional arrest mechanism involved in controlling constitutive levels of mouse c-myb mRNA. *Oncogene*. 1988;2:267-272.
30. Chung EY, Dews M, Cozma D et al. c-Myb oncoprotein is an essential target of the dleu2 tumor suppressor microRNA cluster. *Cancer Biol Ther*. 2008;7:1758-1764.

31. Lin YC, Kuo MW, Yu J et al. c-Myb is an evolutionary conserved miR-150 target and miR-150/c-Myb interaction is important for embryonic development. *Mol Biol Evol.* 2008;25:2189-2198.
32. Lu J, Guo S, Ebert BL et al. MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Dev Cell.* 2008;14:843-853.
33. Xiao C, Calado DP, Galler G et al. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. *Cell.* 2007;131:146-159.
34. Zhao H, Kalota A, Jin S, Gewirtz AM. The c-myb proto-oncogene and microRNA-15a comprise an active autoregulatory feedback loop in human hematopoietic cells. *Blood.* 2009;113:505-516.
35. Biedenkapp H, Borgmeyer U, Sippel AE, Klempnauer K-H. Viral myb oncogene encodes a sequence-specific DNA-binding activity. *Nature.* 1988;335:835-837.
36. Liu F, Lei W, O'Rourke JP, Ness SA. Oncogenic mutations cause dramatic, qualitative changes in the transcriptional activity of c-Myb. *Oncogene.* 2006;25:795-805.
37. Rushton JJ, Davis LM, Lei W, Mo X, Leutz A, Ness SA. Distinct changes in gene expression induced by A-Myb, B-Myb and c-Myb proteins. *Oncogene.* 2003;22:308-313.
38. Sakamoto H, Dai G, Tsujino K et al. Proper levels of c-Myb are discretely defined at distinct steps of hematopoietic cell development. *Blood.* 2006;108:896-903.

39. Greig KT, de Graaf CA, Murphy JM et al. Critical roles for c-Myb in lymphoid priming and early B-cell development. *Blood*. 2010;115:2796-2805.
40. Hu T, Simmons A, Yuan J, Bender TP, Alberola-Ila J. The transcription factor c-Myb primes CD4+CD8+ immature thymocytes for selection into the iNKT lineage. *Nat Immunol*. 2010;11:435-441.
41. Yuan J, Crittenden RB, Bender TP. c-Myb promotes the survival of CD4+CD8+ double-positive thymocytes through upregulation of Bcl-xL. *J Immunol*. 2010;184:2793-2804.
42. Garcia P, Clarke M, Vegiopoulos A et al. Reduced c-Myb activity compromises HSCs and leads to a myeloproliferation with a novel stem cell basis. *EMBO J*. 2009;28:1492-1504.
43. Ferrao P, Gonda TJ, Ashman LK. Expression of constitutively activated human c-Kit in Myb transformed early myeloid cells leads to factor independence, histiocytic differentiation, and tumorigenicity. *Blood*. 1997;90:4539-4552.
44. Karafiat V, Dvorakova M, Pajer P, Cermak V, Dvorak M. Melanocyte fate in neural crest is triggered by Myb proteins through activation of c-kit. *Cell Mol Life Sci*. 2007;64:2975-2984.
45. Ratajczak MZ, Perrotti D, Melotti P et al. Myb and ets proteins are candidate regulators of c-kit expression in human hematopoietic cells. *Blood*. 1998;91:1934-1946.
46. Vegiopoulos A, Garcia P, Emambokus N, Frampton J. Coordination of erythropoiesis by the transcription factor c-Myb. *Blood*. 2006;107:4703-4710.

47. Frampton J, Ramqvist T, Graf T. v-Myb of E26 leukemia virus up-regulates bcl-2 and suppresses apoptosis in myeloid cells. *Genes Dev.* 1996;10:2720-2731.
48. Thompson MA, Rosenthal MA, Ellis SL et al. c-Myb down-regulation is associated with human colon cell differentiation, apoptosis, and decreased Bcl-2 expression. *Cancer Res.* 1998;58:5168-5175.
49. He XY, Antao VP, Basila D, Marx JC, Davis BR. Isolation and molecular characterization of the human CD34 gene. *Blood.* 1992;79:2296-2302.
50. Hernandez-Munain C, Krangel MS. Regulation of the T-cell receptor delta enhancer by functional cooperation between c-Myb and core-binding factors. *Mol Cell Biol.* 1994;14:473-483.
51. Ku DH, Wen SC, Engelhard A et al. c-myb transactivates cdc2 expression via Myb binding sites in the 5'-flanking region of the human cdc2 gene. *J Biol Chem.* 1993;268:2255-2259.
52. Nakata Y, Shetzline S, Sakashita C et al. c-Myb contributes to G2/M cell cycle transition in human hematopoietic cells by direct regulation of cyclin B1 expression. *Mol Cell Biol.* 2007;27:2048-2058.
53. Nakagoshi H, Kanei-Ishii C, Sawazaki T, Mizuguchi G, Ishii S. Transcriptional activation of the c-myc gene by the c-myb and B-myb gene products. *Oncogene.* 1992;7:1233-1240.
54. Sitzmann J, Noben-Trauth K, Klempnauer KH. Expression of mouse c-myb during embryonic development. *Oncogene.* 1995;11:2273-2279.

55. Zorbas M, Sicurella C, Bertoncello I et al. c-Myb is critical for murine colon development. *Oncogene*. 1999;18:5821-5830.
56. Malaterre J, Carpinelli M, Ernst M et al. c-Myb is required for progenitor cell homeostasis in colonic crypts. *Proc Natl Acad Sci U S A*. 2007;104:3829-3834.
57. Malaterre J, Mantamadiotis T, Dworkin S et al. c-Myb is required for neural progenitor cell proliferation and maintenance of the neural stem cell niche in adult brain. *Stem Cells*. 2008;26:173-181.
58. Ness SA. The Myb oncoprotein: regulating a regulator. *Biochim Biophys Acta*. 1996;1288:F123-39.
59. Rushton JJ, Ness SA. The conserved DNA binding domain mediates similar regulatory interactions for A-Myb, B-Myb, and c-Myb transcription factors. *Blood Cells Mol Dis*. 2001;27:459-463.
60. Tanaka Y, Patestos NP, Maekawa T, Ishii S. B-myb is required for inner cell mass formation at an early stage of development. *J Biol Chem*. 1999;274:28067-28070.
61. Toscani A, Mettus RV, Coupland R et al. Arrest of spermatogenesis and defective breast development in mice lacking A-myb. *Nature*. 1997;386:713-717.
62. Hodges LC, Cook JD, Lobenhofer EK et al. Tamoxifen functions as a molecular agonist inducing cell cycle-associated genes in breast cancer cells. *Mol Cancer Res*. 2003;1:300-311.

63. Peters CW, Sippel AE, Vingron M, Klempnauer KH. Drosophila and vertebrate myb proteins share two conserved regions, one of which functions as a DNA-binding domain. EMBO J. 1987;6:3085-3090.
64. Manak JR, Mitiku N, Lipsick JS. Mutation of the Drosophila homologue of the Myb protooncogene causes genomic instability. Proc Natl Acad Sci U S A. 2002;99:7438-7443.
65. Manak JR, Wen H, Van T, Andrejka L, Lipsick JS. Loss of Drosophila Myb interrupts the progression of chromosome condensation. Nat Cell Biol. 2007;9:581-587.
66. Perbal B, Reinisch-Deschamps F, Kryceve-Martinerie C et al. Transforming potential of the v-myb oncogene from avian myeloblastosis virus: alterations in the oncogene product may reveal a new target specificity. Biochimie. 1986;68:969-980.
67. Ibanez CE, Lipsick JS. trans activation of gene expression by v-myb. Mol Cell Biol. 1990;10:2285-2293.
68. Aurigemma RE, Blair DG, Ruscetti SK. Transactivation of Erythroid Transcription Factor GATA-1 by a *myb-ets*-Containing Retrovirus. J Virol. 1992;66:3056-3061.
69. Davies J, Badiani P, Weston K. Cooperation of Myb and Myc proteins in T cell lymphomagenesis. Oncogene. 1999;18:3643-3647.
70. Dubendorff JW, Whittaker LJ, Eltman JT, Lipsick JS. Carboxy-terminal elements of c-Myb negatively regulate transcriptional activation in cis and in trans. Genes Dev. 1992;6:2524-2535.

71. Ferrao P, Macmillan EM, Ashman LK, Gonda TJ. Enforced expression of full length c-Myb leads to density-dependent transformation of murine haemopoietic cells. *Oncogene*. 1995;11:1631-1638.
72. Fu SL, Lipsick JS. Constitutive expression of full-length c-Myb transforms avian cells characteristic of both the monocytic and granulocytic lineages. *Cell Growth Differ*. 1997;8:35-45.
73. Furuta Y, Aizawa S, Suda Y et al. Degeneration of skeletal and cardiac muscles in c-myb transgenic mice. *Transgenic Res*. 1993;2:199-207.
74. Gonda TJ, Buckmaster C, Ramsay RG. Activation of c-myb by carboxy-terminal truncation: relationship to transformation of murine haemopoietic cells in vitro. *EMBO J*. 1989;8:1777-1783.
75. Gonda TJ, Cory S, Sobieszczuk P, Holtzman D, Adams JM. Generation of altered transcripts by retroviral insertion within the c-myb gene in two murine monocytic leukemias. *J Virol*. 1987;61:2754-2763.
76. Mukhopadhyaya R, Wolff L. New sites of proviral integration associated with murine promonocytic leukemias and evidence for alternate modes of c-myb activation. *J Virol*. 1992;66:6035-6044.
77. Nason-Burchenal K, Wolff L. Activation of c-myb is an early bone-marrow event in a murine model for acute promonocytic leukemia. *Proc Natl Acad Sci U S A*. 1993;90:1619-1623.
78. Lei W, Rushton JJ, Davis LM, Liu F, Ness SA. Positive and negative determinants of target gene specificity in Myb transcription factors. *J Biol Chem*. 2004;279:29519-29527.

79. Kanter MR, Smith RE, Hayward WS. Rapid induction of B-cell lymphomas: insertional activation of c-myb by avian leukosis virus. *J Virol.* 1988;62:1423-1432.
80. Jiang W, Kanter MR, Dunkel I, Ramsay RG, Beemon KL, Hayward WS. Minimal truncation of the c-myb gene product in rapid-onset B-cell lymphoma. *J Virol.* 1997;71:6526-6533.
81. Burk O, Klempnauer KH. Myb and Ets transcription factors cooperate at the myb-inducible promoter of the tom-1 gene. *Biochim Biophys Acta.* 1999;1446:243-252.
82. Dini PW, Lipsick JS. Oncogenic truncation of the first repeat of c-Myb decreases DNA binding in vitro and in vivo. *Mol Cell Biol.* 1993;13:7334-7348.
83. Ganter B, Fu S, Lipsick JS. D-type cyclins repress transcriptional activation by the v-Myb but not the c-Myb DNA-binding domain. *EMBO J.* 1998;17:255-268.
84. Oelgeschlager M, Kowenz-Leutz E, Schreek S, Leutz A, Luscher B. Tumorigenic N-terminal deletions of c-Myb modulate DNA binding, transactivation, and cooperativity with C/EBP. *Oncogene.* 2001;20:7420-7424.
85. Luscher B, Christenson E, Litchfield DW, Krebs EG, Eisenman RN. Myb DNA binding inhibited by phosphorylation at a site deleted during oncogenic activation. *Nature.* 1990;344:517-522.
86. Jacobs SM, Gorse KM, Westin EH. Identification of a second promoter in the human c-myb proto-oncogene. *Oncogene.* 1994;9:227-235.

87. Vellard M, Soret J, Viegas-Pequignot E et al. C-myb proto-oncogene: evidence for intermolecular recombination of coding sequences. *Oncogene*. 1991;6:505-514.
88. Vellard M, Sureau A, Soret J, Martinerie C, Perbal B. A potential splicing factor is encoded by the opposite strand of the trans-spliced c-myb exon. *Proc Natl Acad Sci U S A*. 1992;89:2511-2515.
89. Soret J, Vellard M, Martinerie C, Perbal B. Organization of 5'-proximal c-myb exons in chicken DNA. Implications for c-myb tissue-specific transcription. *FEBS Lett*. 1988;232:227-234.
90. Dash AB, Orrico FC, Ness SA. The EVES motif mediates both intermolecular and intramolecular regulation of c-Myb. *Genes Dev*. 1996;10:1858-1869.
91. Kiewitz A, Wolfes H. Mapping of protein-protein interactions between c-myb and its coactivator CBP by a new phage display technique. *FEBS Lett*. 1997;415:258-262.
92. Aziz N, Wu J, Dubendorff JW, Lipsick JS, Sturgill TW, Bender TP. c-Myb and v-Myb are differentially phosphorylated by p42mapk in vitro. *Oncogene*. 1993;8:2259-2265.
93. Bies J, Markus J, Wolff L. Covalent attachment of the SUMO-1 protein to the negative regulatory domain of the c-Myb transcription factor modifies its stability and transactivation capacity. *J Biol Chem*. 2002;277:8999-9009.

94. Pani E, Menigatti M, Schubert S et al. Pin1 interacts with c-Myb in a phosphorylation-dependent manner and regulates its transactivation activity. *Biochim Biophys Acta*. 2008;1783:1121-1128.
95. Lienard P, Riviere M, Van Vooren P, Szpirer C, Szpirer J. Assignment of SND1, the gene encoding coactivator p100, to human chromosome 7q31.3 and rat chromosome 4q23 by in situ hybridization. *Cytogenet Cell Genet*. 2000;90:253-254.
96. Broadhurst MK, Lee RS, Hawkins S, Wheeler TT. The p100 EBNA-2 coactivator: a highly conserved protein found in a range of exocrine and endocrine cells and tissues in cattle. *Biochim Biophys Acta*. 2005;1681:126-133.
97. Tong X, Drapkin R, Yalamanchili R, Mosialos G, Kieff E. The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. *Mol Cell Biol*. 1995;15:4735-4744.
98. Schwarz DS, Tomari Y, Zamore PD. The RNA-induced silencing complex is a Mg²⁺-dependent endonuclease. *Curr Biol*. 2004;14:787-791.
99. Caudy AA, Ketting RF, Hammond SM et al. A micrococcal nuclease homologue in RNAi effector complexes. *Nature*. 2003;425:411-414.
100. Yang W, Chendrimada TP, Wang Q et al. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol*. 2006;13:13-21.
101. Scadden AD. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat Struct Mol Biol*. 2005;12:489-496.

102. Levenson JD, Koskinen PJ, Orrico FC et al. Pim-1 kinase and p100 cooperate to enhance c-Myb activity. *Mol Cell*. 1998;2:417-425.
103. Winn LM, Lei W, Ness SA. Pim-1 phosphorylates the DNA binding domain of c-Myb. *Cell Cycle*. 2003;2:258-262.
104. Nomura T, Tanikawa J, Akimaru H et al. Oncogenic activation of c-Myb correlates with a loss of negative regulation by TIF1beta and Ski. *J Biol Chem*. 2004;279:16715-16726.
105. Pattabiraman DR, Sun J, Dowhan DH, Ishii S, Gonda TJ. Mutations in multiple domains of c-Myb disrupt interaction with CBP/p300 and abrogate myeloid transforming ability. *Mol Cancer Res*. 2009;7:1477-1486.
106. Levenson JD, Ness SA. Point mutations in v-Myb disrupt a cyclophilin-catalyzed negative regulatory mechanism. *Mol Cell*. 1998;1:203-211.
107. Bies J, Wolff L. Oncogenic activation of c-Myb by carboxyl-terminal truncation leads to decreased proteolysis by the ubiquitin-26S proteasome pathway. *Oncogene*. 1997;14:203-212.
108. Bies J, Nazarov V, Wolff L. Identification of protein instability determinants in the carboxy-terminal region of c-Myb removed as a result of retroviral integration in murine monocytic leukemias. *J Virol*. 1999;73:2038-2044.
109. Hu YL, Ramsay RG, Kanei-Ishii C, Ishii S, Gonda TJ. Transformation by carboxyl-deleted Myb reflects increased transactivating capacity and disruption of a negative regulatory domain. *Oncogene*. 1991;6:1549-1553.
110. Grasser FA, Graf T, Lipsick JS. Protein truncation is required for the activation of the c-myb proto-oncogene. *Mol Cell Biol*. 1991;11:3987-3996.

111. Molvaersmyr AK, Saether T, Gilfillan S et al. A SUMO-regulated activation function controls synergy of c-Myb through a repressor-activator switch leading to differential p300 recruitment. *Nucleic Acids Res.* 2010
112. Kanei-Ishii C, Ninomiya-Tsuji J, Tanikawa J et al. Wnt-1 signal induces phosphorylation and degradation of c-Myb protein via TAK1, HIPK2, and NLK. *Genes Dev.* 2004;18:816-829.
113. Kanei-Ishii C, Nomura T, Takagi T, Watanabe N, Nakayama KI, Ishii S. Fbxw7 acts as an E3 ubiquitin ligase that targets c-Myb for nemo-like kinase (NLK)-induced degradation. *J Biol Chem.* 2008;283:30540-30548.
114. Kanei-Ishii C, Nomura T, Tanikawa J, Ichikawa-Iwata E, Ishii S. Differential sensitivity of v-Myb and c-Myb to Wnt-1-induced protein degradation. *J Biol Chem.* 2004;279:44582-44589.
115. Weston K. Extension of the DNA binding consensus of the chicken c-Myb and v-Myb proteins. *Nucleic Acids Res.* 1992;20:3043-3049.
116. Ganter B, Lipsick JS. Myb and oncogenesis. *Adv Cancer Res.* 1999;76:21-60.
117. Sakura H, Kanei-Ishii C, Nagase T, Nakagoshi H, Gonda TJ, Ishii S. Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene. *Proc Natl Acad Sci U S A.* 1989;86:5758-5762.
118. Weston K, Bishop JM. Transcriptional activation by the v-myb oncogene and its cellular progenitor, c-myb. *Cell.* 1989;58:85-93.

119. Tanikawa J, Yasukawa T, Enari M et al. Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain. *Proc Natl Acad Sci U S A*. 1993;90:9320-9324.
120. Ogata K, Morikawa S, Nakamura H et al. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*. 1994;79:639-648.
121. Dini PW, Eltman JT, Lipsick JS. Mutations in the DNA-binding and transcriptional activation domains of v-Myb cooperate in transformation. *J Virol*. 1995;69:2515-2524.
122. Andersson KB, Kowenz-Leutz E, Brendeford EM, Tygsett AH, Leutz A, Gabrielsen OS. Phosphorylation-dependent down-regulation of c-Myb DNA binding is abrogated by a point mutation in the v-myb oncogene. *J Biol Chem*. 2003;278:3816-3824.
123. Mo X, Kowenz-Leutz E, Laumonnier Y, Xu H, Leutz A. Histone H3 tail positioning and acetylation by the c-Myb but not the v-Myb DNA-binding SANT domain. *Genes Dev*. 2005;19:2447-2457.
124. Baker AM, Fu Q, Hayward W, Lindsay SM, Fletcher TM. The Myb/SANT domain of the telomere-binding protein TRF2 alters chromatin structure. *Nucleic Acids Res*. 2009;37:5019-5031.
125. Boyer LA, Langer MR, Crowley KA, Tan S, Denu JM, Peterson CL. Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol Cell*. 2002;10:935-942.

126. Mizuguchi G, Nakagoshi H, Nagase T et al. DNA binding activity and transcriptional activator function of the human B-myb protein compared with c-MYB. *J Biol Chem.* 1990;265:9280-9284.
127. Ma XP, Calabretta B. DNA binding and transactivation activity of A-myb, a c-myb-related gene. *Cancer Res.* 1994;54:6512-6516.
128. Bessa M, Joaquin M, Tavner F, Saville MK, Watson RJ. Regulation of the cell cycle by B-Myb. *Blood Cells Mol Dis.* 2001;27:416-421.
129. Golay J, Broccoli V, Borleri GM et al. Redundant functions of B-Myb and c-Myb in differentiating myeloid cells. *Cell Growth Differ.* 1997;8:1305-1316.
130. Golay J, Cusmano G, Introna M. Independent regulation of c-myc, B-myb, and c-myb gene expression by inducers and inhibitors of proliferation in human B lymphocytes. *J Immunol.* 1992;149:300-308.
131. Raschella G, Cesi V, Amendola R et al. Expression of B-myb in neuroblastoma tumors is a poor prognostic factor independent from MYCN amplification. *Cancer Res.* 1999;59:3365-3368.
132. Thorner AR, Hoadley KA, Parker JS, Winkel S, Millikan RC, Perou CM. In vitro and in vivo analysis of B-Myb in basal-like breast cancer. *Oncogene.* 2009;28:742-751.
133. Schmit F, Korenjak M, Mannefeld M et al. LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G2/M genes. *Cell Cycle.* 2007;6:1903-1913.
134. Golay J, Broccoli V, Lamorte G et al. The A-Myb transcription factor is a marker of centroblasts in vivo. *J Immunol.* 1998;160:2786-2793.

135. Golay J, Loffarelli L, Luppi M, Castellano M, Introna M. The human A-myb protein is a strong activator of transcription. *Oncogene*. 1994;9:2469-2479.
136. DeRocco SE, Iozzo R, Ma XP, Schwarting R, Peterson D, Calabretta B. Ectopic expression of A-myb in transgenic mice causes follicular hyperplasia and enhanced B lymphocyte proliferation. *Proc Natl Acad Sci U S A*. 1997;94:3240-3244.
137. Golay J, Luppi M, Songia S et al. Expression of A-myb, but not c-myb and B-myb, is restricted to Burkitt's lymphoma, slg+ B-acute lymphoblastic leukemia, and a subset of chronic lymphocytic leukemias. *Blood*. 1996;87:1900-1911.
138. Arsura M, Hofmann CS, Golay J, Introna M, Sonenshein GE. A-myb rescues murine B-cell lymphomas from IgM-receptor-mediated apoptosis through c-myc transcriptional regulation. *Blood*. 2000;96:1013-1020.
139. Chen RH, Fields S, Lipsick JS. Dissociation of transcriptional activation and oncogenic transformation by v-Myb. *Oncogene*. 1995;11:1771-1779.
140. Fu SL, Lipsick JS. FAETL motif required for leukemic transformation by v-Myb. *J Virol*. 1996;70:5600-5610.
141. Nomura T, Sakai N, Sarai A et al. Negative autoregulation of c-Myb activity by homodimer formation through the leucine zipper. *J Biol Chem*. 1993;268:21914-21923.
142. Kasper LH, Boussouar F, Ney PA et al. A transcription-factor-binding surface of coactivator p300 is required for haematopoiesis. *Nature*. 2002;419:738-743.

143. Zor T, De Guzman RN, Dyson HJ, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J Mol Biol.* 2004;337:521-534.
144. Dai P, Akimaru H, Tanaka Y et al. CBP as a transcriptional coactivator of c-Myb. *Genes Dev.* 1996;10:528-540.
145. Oelgeschlager M, Janknecht R, Krieg J, Schreek S, Luscher B. Interaction of the co-activator CBP with Myb proteins: effects on Myb-specific transactivation and on the cooperativity with NF-M. *EMBO J.* 1996;15:2771-2780.
146. Favier D, Gonda TJ. Detection of proteins that bind to the leucine zipper motif of c-Myb. *Oncogene.* 1994;9:305-311.
147. Sano Y, Ishii S. Increased affinity of c-Myb for CREB-binding protein (CBP) after CBP-induced acetylation. *J Biol Chem.* 2001;276:3674-3682.
148. Tomita A, Towatari M, Tsuzuki S et al. c-Myb acetylation at the carboxyl-terminal conserved domain by transcriptional co-activator p300. *Oncogene.* 2000;19:444-451.
149. Bies J, Feikova S, Bottaro DP, Wolff L. Hyperphosphorylation and increased proteolytic breakdown of c-Myb induced by the inhibition of Ser/Thr protein phosphatases. *Oncogene.* 2000;19:2846-2854.
150. Matre V, Nordgard O, Alm-Kristiansen AH, Ledsaak M, Gabrielsen OS. HIPK1 interacts with c-Myb and modulates its activity through phosphorylation. *Biochem Biophys Res Commun.* 2009;388:150-154.
151. Pani E, Ferrari S. p38MAPK delta controls c-Myb degradation in response to stress. *Blood Cells Mol Dis.* 2008;40:388-394.

152. Dahle O, Andersen TO, Nordgard O, Matre V, Del Sal G, Gabrielsen OS. Transactivation properties of c-Myb are critically dependent on two SUMO-1 acceptor sites that are conjugated in a PIASy enhanced manner. *Eur J Biochem.* 2003;270:1338-1348.
153. Morita Y, Kanei-Ishii C, Nomura T, Ishii S. TRAF7 sequesters c-Myb to the cytoplasm by stimulating its sumoylation. *Mol Biol Cell.* 2005;16:5433-5444.
154. Sramko M, Markus J, Kabat J, Wolff L, Bies J. Stress-induced inactivation of the c-Myb transcription factor through conjugation of SUMO-2/3 proteins. *J Biol Chem.* 2006;281:40065-40075.
155. Kitagawa K, Kotake Y, Kitagawa M. Ubiquitin-mediated control of oncogene and tumor suppressor gene products. *Cancer Sci.* 2009;100:1374-1381.
156. Beug H, Leutz A, Kahn P, Graf T. Ts mutants of E26 leukemia virus allow transformed myeloblasts, but not erythroblasts or fibroblasts, to differentiate at the nonpermissive temperature. *Cell.* 1984;39:579-588.
157. Graf T, v. Kirchbach A, Beug H. Characterization of the hematopoietic target cells of AEV, MC29 and AMV avian leukemia viruses. *Exp Cell Res.* 1981;131:331-343.
158. Ness SA, Beug H, Graf T. v-myb dominance over v-myc in doubly transformed chick myelomonocytic cells. *Cell.* 1987;51:41-50.
159. Queva C, Ness SA, Grasser FA, Graf T, Vandenbunder B, Stehelin D. Expression patterns of c-myb and of v-myb induced myeloid-1 (mim-1) gene during the development of the chick embryo. *Development.* 1992;114:125-133.

160. Burk O, Mink S, Ringwald M, Klempnauer KH. Synergistic activation of the chicken mim-1 gene by v-myb and C/EBP transcription factors. *EMBO J.* 1993;12:2027-2038.
161. Ness SA, Kowenz-Leutz E, Casini T, Graf T, Leutz A. Myb and NF-M: Combinatorial activators of myeloid genes in heterologous cell types. *Genes Dev.* 1993;7:749-759.
162. Leutz A, Damm K, Sterneck E et al. Molecular cloning of the chicken myelomonocytic growth factor (cMGF) reveals relationship to interleukin 6 and granulocyte colony stimulating factor. *EMBO J.* 1989;8:175-181.
163. Frampton J, McNagny K, Sieweke M, Philip A, Smith G, Graf T. v-Myb DNA binding is required to block thrombocytic differentiation of Myb-Ets-transformed multipotent haematopoietic progenitors. *EMBO J.* 1995;14:2866-2875.
164. Metz T, Graf T. v-myb and v-ets transform chicken erythroid cells and cooperate both in trans and in cis to induce distinct differentiation phenotypes. *Genes Dev.* 1991;5:369-380.
165. Metz T, Graf T. Fusion of the nuclear oncoproteins v-Myb and v-Ets is required for the leukemogenicity of E26 virus. *Cell.* 1991;66:95-105.
166. Kraut N, Frampton J, Graf T. Rem-1, a putative direct target gene of the Myb-Ets fusion oncoprotein in haematopoietic progenitors, is a member of the recoverin family. *Oncogene.* 1995;10:1027-1036.
167. Hogg A, Schirm S, Nakagoshi H et al. Inactivation of a c-Myb/estrogen receptor fusion protein in transformed primary cells leads to

granulocyte/macrophage differentiation and down regulation of c-kit but not c-myc or cdc2. *Oncogene*. 1997;15:2885-2898.

168. Melotti P, Calabretta B. Induction of hematopoietic commitment and erythromyeloid differentiation in embryonal stem cells constitutively expressing c-myb. *Blood*. 1996;87:2221-2234.

169. Melotti P, Ku DH, Calabretta B. Regulation of the expression of the hematopoietic stem cell antigen CD34: role of c-myb. *J Exp Med*. 1994;179:1023-1028.

170. Melotti P, Calabretta B. Ets-2 and c-Myb act independently in regulating expression of the hematopoietic stem cell antigen CD34. *J Biol Chem*. 1994;269:25303-25309.

171. Melotti P, Calabretta B. The transcription factors c-myb and GATA-2 act independently in the regulation of normal hematopoiesis. *Proc Natl Acad Sci U S A*. 1996;93:5313-5318.

172. Ness SA. Myb protein specificity: evidence of a context-specific transcription factor code. *Blood Cells Mol Dis*. 2003;31:192-200.

173. Calabretta B, Sims RB, Valtieri M et al. Normal and leukemic hematopoietic cells manifest differential sensitivity to inhibitory effects of c-myb antisense oligodeoxynucleotides: an in vitro study relevant to bone marrow purging. *Proc Natl Acad Sci U S A*. 1991;88:2351-2355.

174. Lieu YK, Reddy EP. Conditional c-myb knockout in adult hematopoietic stem cells leads to loss of self-renewal due to impaired proliferation and accelerated differentiation. *Proc Natl Acad Sci U S A*. 2009;106:21689-21694.

175. Orlic D, Anderson S, Biesecker LG, Sorrentino BP, Bodine DM. Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-myb and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7. *Proc Natl Acad Sci U S A*. 1995;92:4601-4605.
176. Chu TY, Besmer P. Characterization of the promoter of the proto-oncogene c-kit. *Proceedings of the National Science Council, Republic of China - Part B, Life Sciences*. 1995;19:8-18.
177. Krause DS, Mucenski ML, Lawler AM, May WS. CD34 expression by embryonic hematopoietic and endothelial cells does not require c-Myb. *Exp Hematol*. 1998;26:1086-1092.
178. Papathanasiou P, Tunningley R, Pattabiraman DR et al. A recessive screen for genes regulating hematopoietic stem cells. *Blood*. 2010
179. Sandberg ML, Sutton SE, Pletcher MT et al. c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Dev Cell*. 2005;8:153-166.
180. Burns CE, Galloway JL, Smith AC et al. A genetic screen in zebrafish defines a hierarchical network of pathways required for hematopoietic stem cell emergence. *Blood*. 2009;113:5776-5782.
181. Kissa K, Murayama E, Zapata A et al. Live imaging of emerging hematopoietic stem cells and early thymus colonization. *Blood*. 2008;111:1147-1156.

182. Hamilton B, Feng Q, Ye M, Welstead GG. Generation of induced pluripotent stem cells by reprogramming mouse embryonic fibroblasts with a four transcription factor, doxycycline inducible lentiviral transduction system. *J Vis Exp*. 2009
183. Meissner A, Wernig M, Jaenisch R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol*. 2007;25:1177-1181.
184. Sun Y, Li H, Liu Y, Mattson MP, Rao MS, Zhan M. Evolutionarily conserved transcriptional co-expression guiding embryonic stem cell differentiation. *PLoS One*. 2008;3:e3406.
185. Britos-Bray M, Friedman AD. Core binding factor cannot synergistically activate the myeloperoxidase proximal enhancer in immature myeloid cells without c-Myb. *Mol Cell Biol*. 1997;17:5127-5135.
186. Lang G, White JR, Argent-Katwala MJ, Allinson CG, Weston K. Myb proteins regulate the expression of diverse target genes. *Oncogene*. 2005;24:1375-1384.
187. Lei W, Liu F, Ness SA. Positive and Negative Regulation of c-Myb by Cyclin D1, Cyclin-Dependent Kinases and p27 Kip1. *Blood*. 2005;105:3855-3861.
188. Sandoval R, Pilkinton M, Colamonici OR. Deletion of the p107/p130-binding domain of Mip130/LIN-9 bypasses the requirement for CDK4 activity for the dissociation of Mip130/LIN-9 from p107/p130-E2F4 complex. *Exp Cell Res*. 2009;315:2914-2920.

189. Bartsch O, Horstmann S, Toprak K, Klempnauer KH, Ferrari S. Identification of cyclin A/Cdk2 phosphorylation sites in B-Myb. *Eur J Biochem.* 1999;260:384-391.
190. Muller-Tidow C, Wang W, Idos GE et al. Cyclin A1 directly interacts with B-myb and cyclin A1/cdk2 phosphorylate B-myb at functionally important serine and threonine residues: tissue-specific regulation of B-myb function. *Blood.* 2001;97:2091-2097.
191. Sinclair P, Harrison CJ, Jarosova M, Foroni L. Analysis of balanced rearrangements of chromosome 6 in acute leukemia: clustered breakpoints in q22-q23 and possible involvement of c-MYB in a new recurrent translocation, t(6;7)(q23;q32 through 36). *Haematologica.* 2005;90:602-611.
192. Sobieszczuk PW, Gonda TJ, Dunn AR. Structure and biological activity of the transcriptional initiation sequences of the murine c-myb oncogene. *Nucleic Acids Res.* 1989;17:9593-9611.
193. Bender TP, Thompson CB, Kuehl WM. Differential expression of c-myb mRNA in murine B lymphomas by a block to transcription elongation. *Science.* 1987;237:1473-1476.
194. Murati A, Gervais C, Carbuccioni N et al. Genome profiling of acute myelomonocytic leukemia: alteration of the MYB locus in MYST3-linked cases. *Leukemia.* 2009;23:85-94.
195. Wang ET, Sandberg R, Luo S et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008;456:470-476.

196. Shen-Ong GLC. Alternative internal splicing in c-myb RNAs occurs commonly in normal and tumor cells. *EMBO J.* 1987;6:4035-4039.
197. Shen-Ong GL. Alternate forms of myb: consequences of virus insertion in myeloid tumorigenesis and alternative splicing in normal development. *Current Topics in Microbiology & Immunology.* 1989;149:71-76.
198. Shen-Ong GLC, Lüscher B, Eisenman RN. A second c-myb protein is translated from an alternatively spliced mRNA expressed from normal and 5'-disrupted myb loci. *Mol Cell Biol.* 1989;9:5456-5463.
199. Shen-Ong GLC, Skurla Jr. RM, Owens JD, Mushinski JF. Alternative Splicing of RNAs Transcribed from the Human c-myb Gene. *Mol Cell Biol.* 1990;10:2715-2722.
200. Westin EH, Gorse KM, Clarke MF. Alternative splicing of the human c-myb gene. *Oncogene.* 1990;5:1117-1124.
201. Rosson D, Dugan D, Reddy EP. Aberrant splicing events that are induced by proviral integration: implications for myb oncogene activation. *Proc Natl Acad Sci U S A.* 1987;84:3171-3175.
202. Shen-Ong GL, Luscher B, Eisenman RN. A second c-myb protein is translated from an alternatively spliced mRNA expressed from normal and 5'-disrupted myb loci. *Mol Cell Biol.* 1989;9:5456-5463.
203. Shen-Ong GL, Skurla RMJ, Owens JD, Mushinski JF. Alternative splicing of RNAs transcribed from the human c-myb gene. *Mol Cell Biol.* 1990;10:2715-2722.

204. Woo CH, Sopchak L, Lipsick JS. Overexpression of an alternatively spliced form of c-Myb results in increases in transactivation and transforms avian myelomonoblasts. *J Virol*. 1998;72:6813-6821.
205. Kumar A, Baker SJ, Lee CM, Reddy EP. Molecular mechanisms associated with the regulation of apoptosis by the two alternatively spliced products of c-Myb. *Mol Cell Biol*. 2003;23:6631-6645.
206. Baker SJ, Kumar A, Reddy EP. p89c-Myb is not required for fetal or adult hematopoiesis. *Genesis*. 2010
207. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol*. 2007;14:185-193.
208. Ambros V. The functions of animal microRNAs. *Nature*. 2004;431:350-355.
209. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281-297.
210. Guglielmelli P, Tozzi L, Pancrazzi A et al. MicroRNA expression profile in granulocytes from primary myelofibrosis patients. *Exp Hematol*. 2007;35:1708-1718.
211. Garcia P, Frampton J. Hematopoietic lineage commitment: miRNAs add specificity to a widely expressed transcription factor. *Dev Cell*. 2008;14:815-816.
212. Navarro F, Gutman D, Meire E et al. miR-34a contributes to megakaryocytic differentiation of K562 cells independently of p53. *Blood*. 2009;114:2181-2192.

213. Fahl SP, Crittenden RB, Allman D, Bender TP. c-Myb is required for pro-B cell differentiation. *J Immunol.* 2009;183:5582-5592.
214. Ramsay RG, Gonda TJ. MYB function in normal and cancer cells. *Nat Rev Cancer.* 2008;8:523-534.
215. Shen-Ong GL. Alternative internal splicing in c-myb RNAs occurs commonly in normal and tumor cells. *EMBO J.* 1987;6:4035-4039.
216. Schuur ER, Rabinovich JM, Baluda MA. Distribution of alternatively spliced chicken c-myb exon 9A among hematopoietic tissues. *Oncogene.* 1994;9:3363-3365.
217. Ramsay RG, Ishii S, Nishina Y, Soe G, Gonda TJ. Characterization of alternate and truncated forms of murine c-myb proteins. *Oncogene Res.* 1989;4:259-269.
218. Zhu J, Shendure J, Mitra RD, Church GM. Single molecule profiling of alternative pre-mRNA splicing. *Science.* 2003;301:836-838.
219. Zhou Y, Ness SA. Myb proteins: angels and demons in normal and transformed cells. *Front Biosci.* 2011;16:1109-1131.
220. Quintana AM, Liu F, O'Rourke JP, Ness SA. Identification and Regulation of c-Myb Target Genes in MCF-7 Cells. *BMC Cancer.* 2011;11:30.
221. Rozovskaia T, Ravid-Amir O, Tillib S et al. Expression profiles of acute lymphoblastic and myeloblastic leukemias with ALL-1 rearrangements. *Proc Natl Acad Sci U S A.* 2003;100:7853-7858.
222. Hayashi Y. Gene expression profiling in childhood acute leukemia: progress and perspectives. *Int J Hematol.* 2003;78:414-420.

223. Corcione A, Arduino N, Ferretti E et al. Chemokine receptor expression and function in childhood acute lymphoblastic leukemia of B-lineage. *Leuk Res.* 2006;30:365-372.
224. Crazzolara R, Kreczy A, Mann G et al. High expression of the chemokine receptor CXCR4 predicts extramedullary organ infiltration in childhood acute lymphoblastic leukaemia. *Br J Haematol.* 2001;115:545-553.
225. Kang H, Chen IM, Wilson CS et al. Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. *Blood.* 2010;115:1394-1405.
226. Sun S, Zhang Z, Sinha R, Karni R, Krainer AR. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat Struct Mol Biol.* 2010;17:306-312.
227. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. *Science.* 2010;327:996-1000.
228. Jin S, Zhao H, Yi Y, Nakata Y, Kalota A, Gewirtz AM. c-Myb binds MLL through menin in human leukemia cells and is an important driver of MLL-associated leukemogenesis. *J Clin Invest.* 2010;120:593-606.
229. Somervaille TC, Cleary ML. Grist for the MLL: how do MLL oncogenic fusion proteins generate leukemia stem cells? *Int J Hematol.* 2010;91:735-741.

230. Kramer A, Hochhaus A, Saussele S, Reichert A, Willer A, Hehlmann R. Cyclin A1 is predominantly expressed in hematological malignancies with myeloid differentiation. *Leukemia*. 1998;12:893-898.
231. Muller C, Yang R, Park DJ, Serve H, Berdel WE, Koeffler HP. The aberrant fusion proteins PML-RAR alpha and PLZF-RAR alpha contribute to the overexpression of cyclin A1 in acute promyelocytic leukemia. *Blood*. 2000;96:3894-3899.
232. Albiñ A, Johnsen JI, Henriksson MA. MYC in oncogenesis and as a target for cancer therapies. *Adv Cancer Res*. 2010;107:163-224.
233. Furusato B, Mohamed A, Uhlen M, Rhim JS. CXCR4 and cancer. *Pathol Int*. 2010;60:497-505.
234. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25:402-408.
235. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A*. 2003;100:8817-8822.
236. Nakano M, Komatsu J, Matsuura S, Takashima K, Katsura S, Mizuno A. Single-molecule PCR using water-in-oil emulsion. *J Biotechnol*. 2003;102:117-124.
237. Macquarrie KL, Fong AP, Morse RH, Tapscott SJ. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet*. 2011

238. Lefrancois P, Zheng W, Snyder M. ChIP-Seq using high-throughput DNA sequencing for genome-wide identification of transcription factor binding sites. *Methods Enzymol.* 2010;470:77-104.
239. Cullum R, Alder O, Hoodless PA. The next generation: using new sequencing technologies to analyse gene regulation. *Respirology.* 2011;16:210-222.
240. Fullwood MJ, Wei CL, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* 2009;19:521-532.
241. Zhao J, Grant SF. Advances in whole genome sequencing technology. *Curr Pharm Biotechnol.* 2011;12:293-305.
242. Twine NA, Janitz K, Wilkins MR, Janitz M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One.* 2011;6:e16266.
243. Becker J, Semler O, Gilissen C et al. Exome Sequencing Identifies Truncating Mutations in Human SERPINF1 in Autosomal-Recessive Osteogenesis Imperfecta. *Am J Hum Genet.* 2011
244. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008;26:1135-1145.
245. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.

246. Xu MY, Aragon AD, Mascarenas MR, Torrez-Martinez N, Edwards JS. Dual primer emulsion PCR for next- generation DNA sequencing. *Biotechniques*. 2010;48:409-412.
247. Biedenkapp H, Borgmeyer U, Sippel AE, Klempnauer KH. Viral myb oncogene encodes a sequence-specific DNA-binding activity. *Nature*. 1988;335:835-837.
248. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*. 1999;17:1030-1032.
249. Huang L, Baldwin MA, Maltby DA et al. The identification of protein-protein interactions of the nuclear pore complex of *Saccharomyces cerevisiae* using high throughput matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry. *Mol Cell Proteomics*. 2002;1:434-450.
250. Knuesel M, Wan Y, Xiao Z et al. Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. *Mol Cell Proteomics*. 2003;2:1225-1233.
251. Oh IH, Reddy EP. The myb gene family in cell growth, differentiation and apoptosis. *Oncogene*. 1999;18:3017-3033.
252. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126:663-676.
253. Wernig M, Meissner A, Foreman R et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature*. 2007;448:318-324.

254. Cogswell JP, Cogswell PC, Kuehl WM et al. Mechanism of c-myc regulation by c-Myb in different cell lineages. *Mol Cell Biol.* 1993;13:2858-2869.
255. Ciznadija D, Tothill R, Waterman ML et al. Intestinal adenoma formation and MYC activation are regulated by cooperation between MYB and Wnt signaling. *Cell Death Differ.* 2009;16:1530-1538.
256. Takahashi K, Tanabe K, Ohnuki M et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell.* 2007;131:861-872.
257. Atlasi Y, Mowla SJ, Ziaee SA, Gokhale PJ, Andrews PW. OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells.* 2008;26:3068-3074.
258. Cauffman G, Liebaers I, Van Steirteghem A, Van de Velde H. POU5F1 isoforms show different expression patterns in human embryonic stem cells and preimplantation embryos. *Stem Cells.* 2006;24:2685-2691.
259. Lee J, Kim HK, Rho JY, Han YM, Kim J. The human OCT-4 isoforms differ in their ability to confer self-renewal. *J Biol Chem.* 2006;281:33554-33565.
260. Quintana AM, Zhou YE, Pena JJ, O'Rourke JP, Ness SA. Dramatic Repositioning of c-Myb to Different Promoters during the Cell Cycle Observed by Combining Cell Sorting with Chromatin Immunoprecipitation. *PLoS One.* 2011;6:e17362.
261. Thomas MD, Kremer CS, Ravichandran KS, Rajewsky K, Bender TP. c-Myb is critical for B cell development and maintenance of follicular B cells. *Immunity.* 2005;23:275-286.

262. Wang QF, Lauring J, Schlissel MS. c-Myb binds to a sequence in the proximal region of the RAG-2 promoter and is essential for promoter activity in T-lineage cells. *Mol Cell Biol.* 2000;20:9203-9211.
263. Kishi H, Jin ZX, Wei XC et al. Cooperative binding of c-Myb and Pax-5 activates the RAG-2 promoter in immature B cells. *Blood.* 2002;99:576-583.
264. Puig O, Caspary F, Rigaut G et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods.* 2001;24:218-229.
265. Boldt K, van Reeuwijk J, Gloeckner CJ, Ueffing M, Roepman R. Tandem affinity purification of ciliopathy-associated protein complexes. *Methods Cell Biol.* 2009;91:143-160.
266. Deng M, Li F, Ballif BA et al. Identification and functional analysis of a novel cyclin e/cdk2 substrate ankrd17. *J Biol Chem.* 2009;284:7875-7888.
267. Ness SA, Kowenz-Leutz E, Casini T, Graf T, Leutz A. Myb and NF-M: combinatorial activators of myeloid genes in heterologous cell types. *Genes Dev.* 1993;7:749-759.
268. Tagwerker C, Flick K, Cui M et al. A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. *Mol Cell Proteomics.* 2006;5:737-748.
269. Petrescu-Danila E, Voicu PM, Poitelea M, Stoica B, Stanescu R, Rusu M. Method for protein tagging in *Schizosaccharomyces pombe*. *Rev Med Chir Soc Med Nat Iasi.* 2006;110:403-408.

270. Burckstummer T, Bennett KL, Preradovic A et al. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat Methods*. 2006;3:1013-1019.