Secondary Somatic Mutations in ETV6-RUNX1 Acute Lymphoblastic Leukemia Pathogenesis

Secondary Somatic Mutations in ETV6-RUNX1 Acute Lymphoblastic Leukemia Pathogenesis

Matthew Leon

A Thesis Submitted to the Faculty of

The Harvard Medical School

in Partial Fulfillment of the Requirements

for the Degree of Master of Medical Sciences in Immunology

Harvard University

Boston, Massachusetts.

May, 2016

Thesis Advisor: Dr. Hanno Reinhard Hock Author: Matthew Leon

Secondary Somatic Mutations in ETV6-RUNX1 Acute Lymphoblastic Leukemia Pathogenesis

Abstract

ETV6-RUNX1, also known as TEL-AML1, is the most frequent chromosomal translocation in childhood Acute Lymphoblastic Leukemia (ALL) with an incidence of approximately 25%. This translocation occurs in hematopoietic stem cells and leads to the establishment of pre-leukemic progenitor B-cell clones that persist in the bone marrow for several years. It has been shown that the translocation itself is insufficient to generate ALL and that secondary somatic mutations are necessary to activate the leukemic phenotype. Recent genome analyses of patients exhibiting this type of leukemia have identified a host of mutated genes that may be working cooperatively to induce this hematologic malignancy. Previously, the cooperative action of these secondary mutations has been difficult to characterize due to a lack of ex vivo models that allow for targeted mutagenesis of multiple genes after the initial translocation. However, with the identification and development of CRISPR-Cas9 as a genome-editing tool, this is now a feasible undertaking and the crux of this project. Here we have generated a lentiviral CRISPR library composed of 59 unique guide RNAs (gRNA) to target 19 genes of interest that represent likely secondary hits responsible for activating the ETV6-RUNX1^{+} pre-leukemic clones. We will use this library to infect an ex vivo pro B-cell line and select for transformed mutants that have insertions/deletions at our loci of interest.

Table of Contents

Figures

Tables

Acknowledgements

This work would not have been possible without the guidance, mentorship and resources of my thesis advisor, Dr. Hanno Hock. I would like to thank Ryan Legraw, the Hock Lab Technician, for all of his help in learning specific techniques and assays. I would like to thank the Postdocs from the Hochedlinger Lab including Ryan Walsh, Justin Brumbaugh, and Ori Bar-Nur for their expertise in lentivirus production and tissue culture. I would like to thank the lab managers, administrative staff, and lab personnel in the MGH Cancer Center, 4th Floor Simches Building. I would like to thank the entire Master of Medical Sciences in Immunology program at Harvard Medical School. A special thanks to the program directors, Dr. Shiv Pillai and Dr. Michael Carroll, for their support of my work and decision to admit me to this program. I would like to thank the previous and current program coordinators, Carlien Frijlink and Selina Labriola, who have kept this program running smoothly and ensured success throughout. Lastly, I would like to thank Dr. Kevin Bonham, who's methods course was a big help in preparing for my lab work over the past year.

"This work was conducted with support from Students in the Master of Medical Sciences in Immunology program of Harvard Medical School. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University and its affiliated academic health care centers."

Chapter One: Background

1.1: Introduction

Leukemias, Lymphomas, and Myelomas represent the unique intersection between oncology and hematology. According to the American Cancer Society's 2015 Facts and Figures, these hematologic malignancies account for roughly 10% of the 1.6 million new cancer diagnoses in the United States[1]. These 160,000 individuals join nearly 1.2 million Americans who are currently in remission from or living with one of these blood cancers[1]. Hematologic malignancies represent a subsection of an overarching category of "Hematologic diseases", which includes a vast array of defects in blood cell formation, bleeding, clotting, and the development of blood vessels, bone marrow and lymph nodes. As a fluid tissue, blood has distinctive properties that make targeting its associated diseases different than other body systems. With respect to treatment for hematologic malignancies, there have been significant advances over the past few decades. Immunotherapies, chemotherapeutic agents, and even hematopoietic stem cells transplants have become a regular approach to treating and curing these cancers. However, blood neoplasms still contribute a significant portion (10%) to the total number of mortalities due to cancer in the United States.

1.2: Hematopoiesis, Lymphopoiesis and Leukemogenesis

Hematopoiesis is the complex, stepwise process responsible for the cellular compartment of blood. This process begins approximately one month into prenatal development and transitions through different microenvironments in the fetus. The first signs of hematopoiesis in humans occur at one month in the fetal yolk sac but, as development advances, blood cell formation can be seen in the fetal liver, spleen and developing lymph nodes[2]. These two different

time points and locations are separated into distinct classifications, primitive hematopoiesis and definitive hematopoiesis. The former is thought to give rise to red blood cell progenitors that can sustain the fetus until the latter is reached, at which time all components of blood are produced[2,3]. By 9 months, hematopoiesis is definitive and occurs predominantly in the bone marrow with a smaller amount seen in secondary lymphoid organs including the spleen and thymus[3]. Postnatal hematopoiesis is concentrated in the red marrow of long bones during childhood and the sternum, vertebrae and pelvis later in life. The bone marrow is a diverse niche composed of stromal cells such as adipocytes, osteoprogenitors reticular cells and mesenchymal stem cells that provide support to the linchpin of this system, the hematopoietic stem cell (HSC)[4]. Purified from mice as $Lin^{-}Kit^{+}Sca1^{+}FLT3^{-}CD34^{-}$, the HSC is the multipotent progenitor that gives rise to the three major blood cell lineages. HSCs exhibit two important properties that allow them to sustain the production of nearly 10^{12} new blood cells each day. Through a distinctive process called asymmetric division, hematopoietic stem cells have the ability to simultaneously self-renew and produce more differentiated progeny[5]. More specifically, the division of a hematopoietic stem cell can produce two daughter cells with different functional fates. While the exact mechanism of asymmetric division is uncertain, it is clear that HSCs give rise to a series of intermediate progenitors that ultimately become erythrocytes, lymphocytes and myelocytes.

Genetic control is a fundamental aspect of HSC self-renewal and the eventual lineage commitment of the aforementioned intermediate progenitors. While isolation of HSCs requires sophisticated methods, many genetic targeting experiments have been conducted over the years to help shed light on the major transcription factors (TFs) involved in HSC biology and blood cell lineage commitment. GATA-2, Notch-1, HoxB4 and Ikaros have all been identified as crucial transcription factors in HSC self-renewal[5]. The first step down the lineage tree is

formation of a Multipotent Progenitor (MPP)[6]. The MPP has less self-renewal capacity than an HSC but is still able to give rise all mature blood cells. Moving further into a state of differentiation, a critical split occurs where the MPP gives rise to committed lineage progenitors that lose their multipotency. The Megakaryocyte-Erythrocyte Progenitors (MEP) is the committed lineage progenitor that gives rise to platelets and red blood cells (RBCs) while mature white blood cells are derived from two separate progenitors. The Common Myeloid Progenitor (CMP) gives rise to precursors that produce neutrophils, basophils, eosinophils, monocytes, macrophages and dendritic cells[6]. The Common Lymphoid Progenitor (CLP) gives rise to precursors that produce B Cells, T cells and NK cells. The formation of these three lymphocytes from the CLP is known as lymphopoiesis and is a key concept to our work as dysfunction of this process leads to the formation of our target disease.

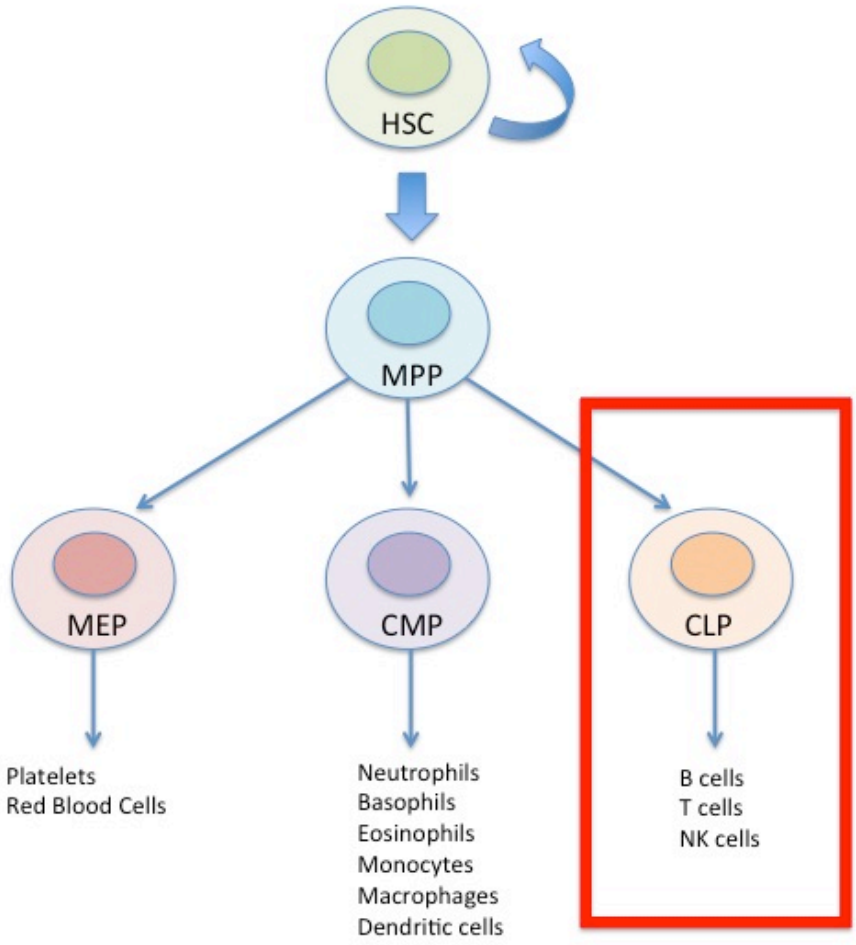


Figure 1. Hematopoiesis. *The Hematopoietic Stem Cell has the ability to self renew and differentiate within a given cell division. This asymmetric division helps maintain the supply of stem cells while still producing enough progeny to meet the needs of the organism. As differentiation occurs down the different lineages, potency is lost and cell fates become defined.*

Much the way early hematopoiesis exists under genetic control, lymphopoiesis is a highly regulated process that utilizes growth factors and transcription factors to differentiate the CLP into one of the three major lymphocytes. As this project focuses on B-precursor Acute Lymphoblastic Leukemia, we will examine the specific steps and regulators that control the transition from a CLP to a mature B cell. B lymphopoiesis occurs in the bone marrow and begins with the expression of Pu.1. A part of the Ets transcription factor family, Pu.1 has been shown to regulate transcription of B cell lineage-specific genes such λ5 and VpreB, which are components of the surrogate light chain needed in the generation of antibodies[7,8]. Knockout experiments of Pu.1 have shown a loss of T cells and myeloid-lineage cells in addition to the expected loss of B cells. This result was helpful in eliminating Pu.1 as the transcription factor specific for B cell lineage commitment. Following the initial upregulation of Pu.1, EBF and E2A are expressed and directly interact with a host of B cell genes including RAG1 and RAG2, which are crucial in the production of antibody diversity[3,7]. Knockouts of EBF and E2A in early B cell progenitors results in differentiation arrest at the pro-B cell stage of development. Additionally, these arrested cells have been shown to reverse their fate and differentiate into T cells[8]. Thus, similar to Pu.1, EBF and E2A are important in B cell development but are not the transcription factors that guarantee mature B cell formation. It is Pax5, a paired box domain protein, which is ultimately responsible for an irreversible commitment to the B cell lineage and is a gene of interest in our experimental undertaking.

Disruption of lymphopoiesis and myelopoiesis can lead to leukemogenesis. As its name implies, leukemogenesis is the formation of leukemia, one of the three major hematologic malignancies mentioned previously. At a basic level, leukemia is the expansion of abnormal, immature white blood cells in the bone marrow. These immature cells, often referred to as

"blasts", have undergone mutagenesis that prevents them from differentiating into mature cells. Blast expansion in the bone marrow disturbs the stromal microenvironment and disrupts the formation of platelets and red blood cells[8]. Leukemias typically fall under four major categories based on which lineage the cancer occurs in and the rate at which the malignancy progresses. Myeloid leukemias arise in immature myelocytes and myeloid precursors while lymphoblastic/lymphocytic leukemias develop in immature lymphocytes and lymphoid precursors[1]. Both myeloid and lymphoid leukemias are then divided into either acute or chronic. The acute classification is characterized by rapid proliferation of blasts in the bone marrow and more commonly seen in children. Chronic leukemias, on the other hand, usually take years to develop and are more common in adults as a consequence of an aging hematopoietic system. Within this general classification system, there are subtypes that correspond to specific cell types. For example, Acute Lymphoblastic Leukemia (ALL) can develop in B-cell progenitors or T-cell progenitors, which produce different disease phenotypes. Another important characteristic of leukemias, when compared to solid tumor malignancies, is the high incidence of chromosomal translocations and cytogenetic aberrations that are involved in their pathogenesis. This unique feature of leukemias and other hematologic malignancies make targeting them a highly specialized task.

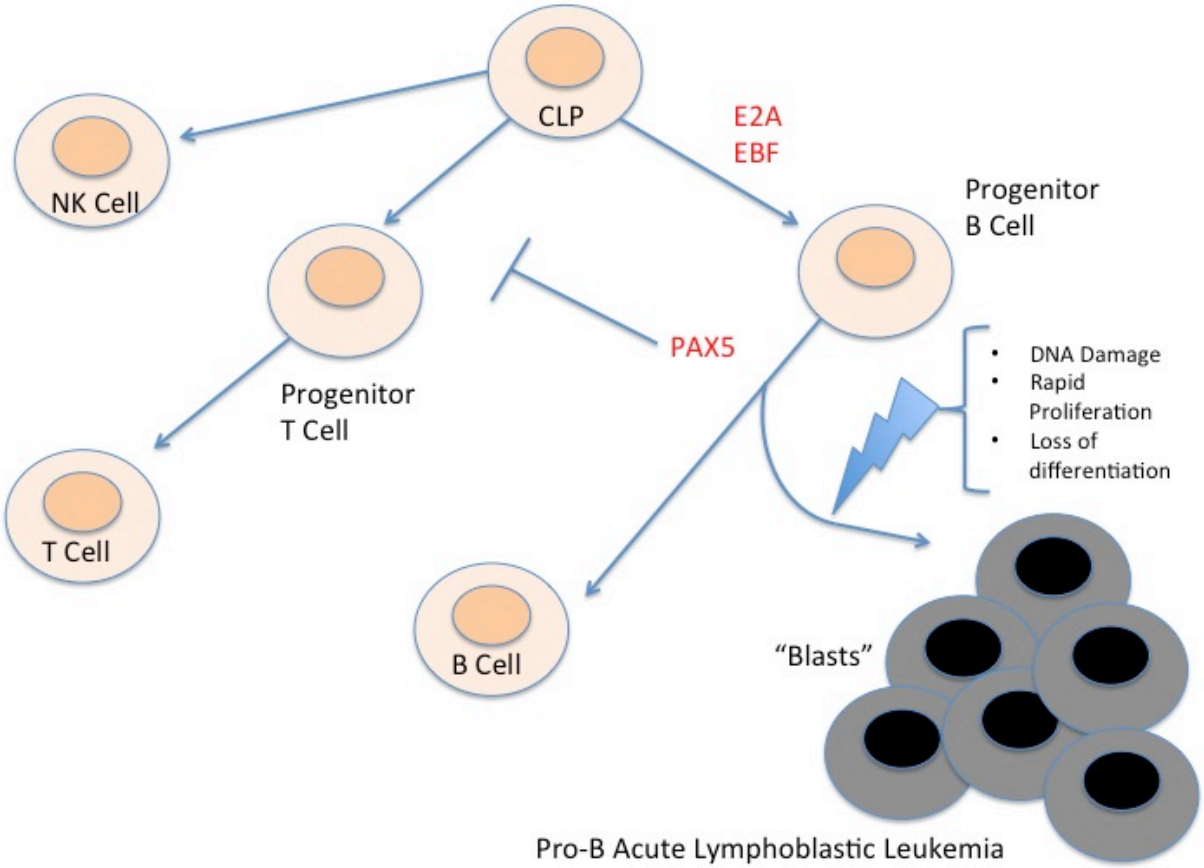


Figure 2. B-Lymphopoiesis and Leukemogenesis. *The Common Lymphoid Progenitor will differentiate into an NK cell, T cell or B cell depending on external signals and particular transcription factor profiles. While E2A and EBF are important drivers in the B cell lineage, PAX5 expression is the irreversible commitment step to continue down this path. During B cell development, oncogenic mutations in early progenitors result in abnormal, immature white blood cells that are unable to differentiate but rapidly proliferate. These "blasts" accumulate in the bone marrow and are the hallmark of pro-B Acute Lymphoblastic Leukemia*

1.3: ETV6-RUNX1 (TEL-AML1) B-Cell Precursor Acute Lymphoblastic Leukemia

Cytogenetics is the branch within genetics that focuses on chromosome structure and function. Technological advances in this field, such as the development of Fluorescence In Situ Hybridization (FISH), have been instrumental in identifying and studying the abundance of chromosomal abnormalities that give rise to leukemia. Chromosomal abnormalities include translocations, rearrangements, loss (monosomy), gain (trisomy), interstitial deletions, and inversions. Decades of cytogenetic data have given prognostic value to certain aberrations with respect to disease progression and survival rate. Additionally, the collection of this data has been instrumental in determining the prevalence of certain genetic events amongst all those diagnosed with leukemia.

Acute Lymphoblastic Leukemia (ALL) accounts for almost 80% of all diagnosed leukemias in children and represents the most common childhood cancer[9]. With such a high incidence in children, a vast array of genetic abnormalities have been identified and causally linked to disease pathogenesis. For example, ALL patients with trisomy 10, 17 or 18 have historically better outcomes than ALL patients with trisomy 5 and monosomy 7[10]. In addition to chromosome copy number, the presence of certain chromosome translocations has been a defining hallmark in both B-cell ALL and T-cell ALL. A translocation is a structural event between two non-homologous chromosomes where a portion from each chromosome breaks apart and switches places[11,12]. This molecular rearrangement often results in one of two situations. The first is the relocation of an enhancer or promoter from one gene to a region upstream of the coding sequence of another gene. This mechanism is often how certain proto-oncogenes become activated oncogenes. One such example in a patient diagnosed with pre-B cell ALL was the translocation t(5;14)(q31;q32), in which the promoter for IL-3 was rearranged

in front of the Immunoglobulin heavy chain joining region[12]. The result was the production of pre-B lymphoblasts and peripheral eosinophilia.

The other situation, which will be the focus for the remainder of this work, is a translocation that produces a novel fusion gene. In some cases, these new genes are capable of producing chimeric proteins that can impair normal cell operations and contribute to malignant transformation. In other situations, the resultant protein may be nonfunctional and have little effect on the cell's normal physiology. Much the way rearrangements of promoters/enhancer can activate proto-oncogenes, the fusing of two genetic elements can produce a fusion oncogene. In ALL, the four most common fusion oncogenes are ETV6-RUNX1 t(12; 21), E2A-PBX1 t(1; 19), BCR-ABL t(22; 9), and MLL-AF4 (t 4; 11)[9-11]. In one study conducted in 2012 that looked at 101 children with ALL, these four fusion genes along with a fifth (SIL-TAL1) were found in 88% of the patient population[13]. The genes represented in these fusion products are responsible for a variety of cellular functions including chromatin remodeling (MLL), cell division (ABL), and transcriptional regulation (E2A).

Of the four major fusion oncogenes seen in ALL, ETV6-RUNX1 presents a unique rearrangement that has garnered significant attention over the past two decades. Also referred to as TEL-AML1, the ETV6-RUNX1 fusion gene is the most common chromosomal translocation in pediatric ALL with a prevalence of approximately 25%[14-17]. This rearrangement is specific to early B-cell leukemia and is typically associated with a positive prognosis in children with B-cell precursor ALL[14]. The two genes that comprise this fusion product, ETV6 (TEL) and RUNX1 (AML1), have long been studied because of their importance in regulating hematopoiesis. TEL is part of the ETS family of transcription factors, which are ubiquitously expressed during embryonic development[18,19]. Several early studies provided important insight into the role this

transcription factor plays in development. Through knockout mouse models, it was first shown that TEL is not required for primitive hematopoiesis, but rather, is required to maintain the vascular system in the yolk sac where mesenchymal and neural stem cells are located early in development[19]. Later studies revealed that TEL is required for definitive hematopoiesis in the bone marrow and that is has the ability to regulate HSC self-renewal and activity[14,20,21]. In addition to its crucial role in hematopoiesis, later studies indicated that TEL acts as a tumor suppressor[18]. The evidence for this claim was that TEL, like many tumor suppressors, experiences a loss of heterozygosity. When TEL is involved in a rearrangement with another gene, the wild type TEL allele on the unaffected chromosome is often mutated or lost. Like TEL, RUNX1 (AML1) is a transcription factor and belongs to the Runt-related transcription factor family of genes[22]. Also similar to TEL, early knockout studies in mice revealed that AML1 is absolutely required during definitive hematopoiesis[23]. Functionally speaking, AML1 is a DNA-binding subunit that complexes with CBFβ to form the Core Binding Factor (CBF)[22]. The CBF complex has the unique ability to both activate and repress the transcription of genes related to cell differentiation. In the TEL/AML1 fusion oncogene, however, AML1 becomes impaired and is unable to interact with its normal target genes[24]. Additionally, it has been shown that the TEL/AML1 chimeric protein will bind the corepressor NCOR1 and downregulate the genes typically needed for hematopoietic progenitor differentiation[25]. Cytogenetic analysis of the TEL/AML1 fusion gene has identified the key break points for the two genes and the domains that remain intact when the two come together. Essentially, the 5' end of the TEL gene breaks off and combines with the 3' end of the AML1 gene. For TEL, the repressive PNT domain is retained and for AML1, the DNA-Binding Runt domain and the Transactivation domain persist but with impaired function[15,16,24].

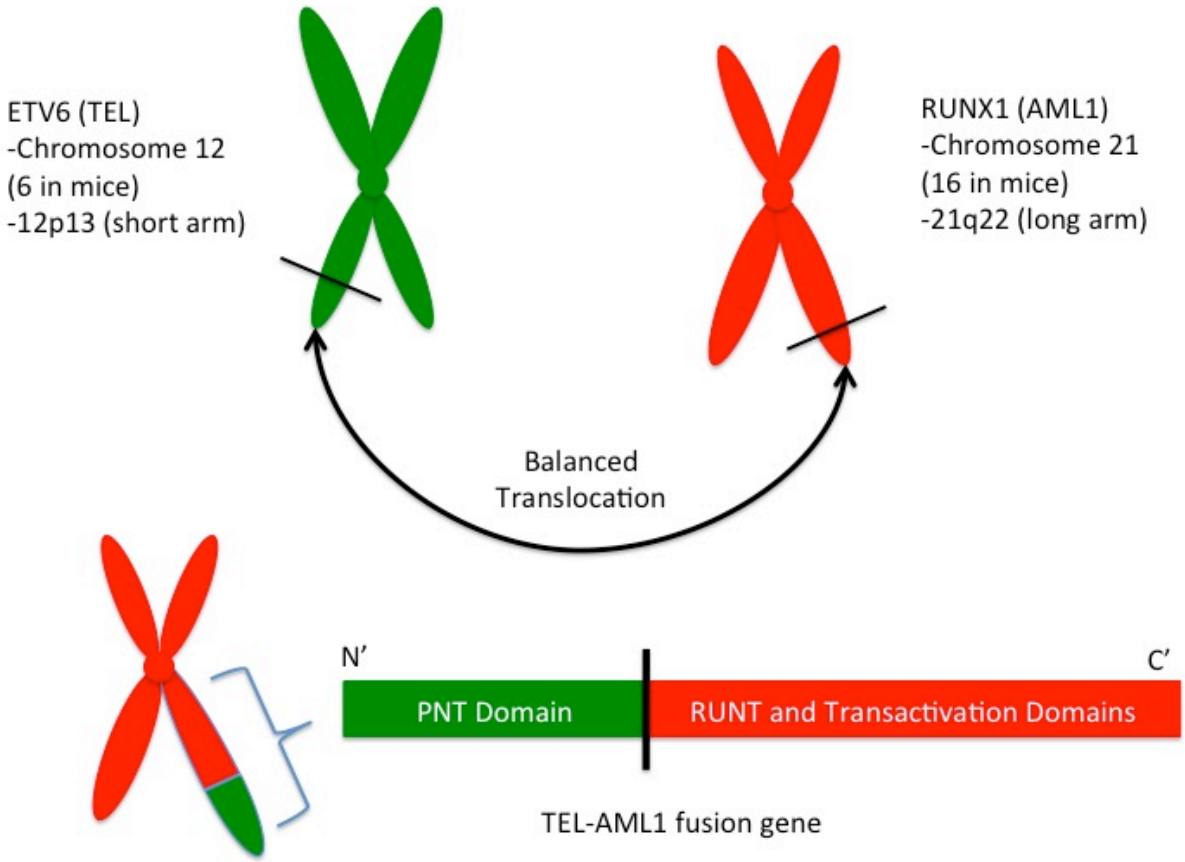


Figure 3. ETV6-RUNX1 chromosomal translocation. *While slight variations have been noted for the TEL and AML1 breakpoints, the general mechanism is constant throughout. The telomeric end of short arm of chromosome 12 breaks off and combines with the long arm of chromosome 21. The result is fusion gene that includes the Pointed (PNT) domain from TEL and the RUNT and Transactivation domains from AML1. The fusion protein product from this translocation disrupts hematopoiesis and creates a pre-leukemic state in the bone marrow that can eventually lead to pathogenic leukemia with the addition of secondary driver mutations.*

The major reason why this fusion gene has warranted and received so much attention lies in its role in leukemia pathogenesis. Unlike the BCR-ABL fusion, which has overwhelming evidence supporting its ability to directly induce transformation, TEL-AML1 is insufficient to produce the leukemic phenotype[12,14-17,26]. This discovery came as the result of several important studies that focused on the timing of the initial translocation and the latency period that exists before disease onset. Through a review of twin concordance studies, Greaves et al. were able to demonstrate that TEL-AML1 rearrangement can occur in early development. Their proof of principle was that a high percentage of identical twins that were TEL-AML1$^+$ had the same breakpoints in their rearrangements, which supports the hypothesis of a monoclonal origin during gestation[27]. This discovery prompted subsequent groups to examine the fusion gene in utero. Using a transgenic TEL-AML1 mouse model Schindler et al. demonstrated that the translocation affects HSCs and lymphoid progenitors during embryonic development, but does not directly establish leukemia[14]. The conclusion in the field was that secondary somatic mutations must arise during the post-translocation latency period capable of driving the affected TEL-AML1 progenitors to a lymphoblastic state.

With the introduction of exome and whole-genome sequencing, several groups began surveying the DNA of individuals diagnosed with TEL-AML1$^+$ ALL in search of secondary driver mutations. In 2013, an instrumental study was published in Nature Genetics that identified the most frequently mutated genes in patients exhibiting the t(12;21) translocation[28]. Additionally, they noted several mutated genes that had not been noted in previous TEL-AML1 screens. In conducting a genome-wide screen in patients that already have the disease, one cannot determine which mutations arose first and whether or not they were directly involved in activating the preleukemic clones. However, this study was a pivotal step in narrowing down the

number of possible genes that could be directly involved in this stepwise disease mechanism. Further, their analysis revealed the presence of multiple mutations within individual patient samples. This intriguing finding alludes to the possibility that a cooperative set of mutated genes, as opposed to a single gene, is required for malignant transformation. Ultimately, their findings produced a list of almost two-dozen potential targets for anyone in need of a jumping off point to study these somatic mutations in a disease model. The difficulty lies in the latter portion of the previous statement. A challenge to those who study the molecular events in TEL-AML1 ALL has been the ability to develop a disease model that captures the initial translocation, the latency period, the secondary somatic mutations and finally the disease initiation. This challenge is the motivation behind our project design. For the first time, our lab has combined the aforementioned pool of knowledge regarding TEL-AML1 driver mutations and the recent discovery of CRISPR-Cas9 genome-editing technology to develop an ex vivo model of precursor B-cell acute lymphoblastic leukemia pathogenesis.

Chapter Two: Materials and Methods

2.1: CRISPR-Cas9 Technology

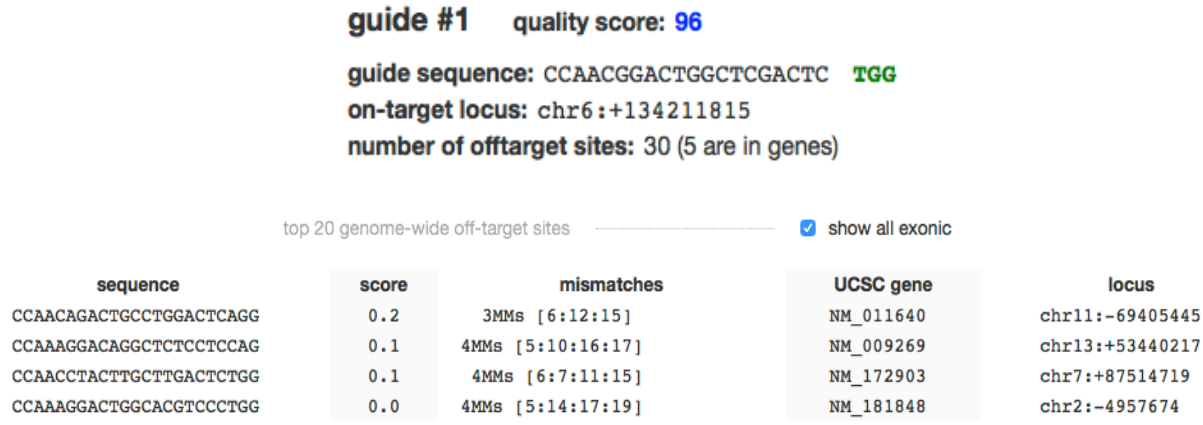
The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Cas9 endonuclease system is a novel technology that has fostered significant growth in the field of genome editing over the past three years. This system was first identified in bacteria over a decade ago as a defense mechanism against viral infection and the incorporation of foreign genetic material into the host genome. In 2013, Feng Zhang and his group at MIT were able to harness this natural system as a means to edit the genome of eukaryotic cells[29,30]. This system was a departure from the commonly used TALEN and zinger finger approaches to genetic engineering. There are two important elements that combine to make a functional CRISPR system. The first is the single guide RNA (sgRNA), a 20 base pair sequence that is complementary to a unique target sequence in the genome that will undergo editing. The second is the Cas9 endonuclease, which is an enzyme isolated from particular bacterial species responsible for introducing double-strand breaks in the DNA. When these two components are expressed in the target cell, the guide RNA combines with the Cas9 enzyme producing an activated functional unit. The key to successful operation of the guideRNA-Cas9 complex is the Protospacer Adjacent Motif (PAM). The PAM is short base pair sequence in the genome that the Cas9 enzyme recognizes and binds. This sequence varies depending on the bacteria from which the Cas9 enzyme originated. The most commonly utilized Cas9 comes from Streptococcus Pyogenes and recognizes the PAM sequence 5'-NGG-3', where N is any of the 4 DNA nucleotides[29]. Upon Cas9 binding of the PAM, the guide RNA will attempt to complex with the genomic target sequence immediately upstream from the PAM. The binding proceeds from 3' to 5' and if enough homology exists, Cas9 will undergo a conformational change. The new

configuration exposes the endonuclease domain, which cleaves the DNA 3-4 base pairs upstream of the PAM sequence. The result is a DNA double-strand break that will undergo one of two DNA repair mechanisms: Homology Directed Repair (HDR) or Non-Homologous End Joining (NHEJ). The latter is an error-prone mechanism that occurs when no repair template is available and is responsible for nucleotide insertions and deletions (indels) at the repair site. Indels that occur in coding exons alter the mRNA transcript and subsequently shift the reading frame during translation. The major consequence of frameshift mutations is the production of truncated and elongated proteins which may result in a gain-of-function or loss-of-function for a particular gene. Insertions and deletions via NHEJ are fundamental to our project and the means by which we introduce driver mutations to our genes of interest.

2.2: Guide RNA Design Tool

The CRISPR design tool, made available by the Zhang lab at MIT, was used in the selection of gRNAs against our 19 target genes. This software takes unique genomic inputs and produces a list of potential guide RNAs and the associated PAM needed for Cas9 recognition. Additionally, the software scans the genome for possible off-targets associated with a given guide. For each guide, a "quality score" from 0 to 100 is assigned based on complementarity and potential off-targets. According to their program algorithm, a guide with a quality score of 50 or higher is a strong prospect for use in the system. In total, 59 gRNAs were selected from the options generated by the design tool. Each guide selected was verified to contain a unique 20-nucleotide sequence followed by a 3 base pair PAM of the form NGG. All 59 guides had a quality score above 50 and were therefore considered strong candidates to be included in our CRISPR library. For each guide, potential off-targets were noted based on homology and the

number of mismatches. Figure 4 is an example of the readout obtained for an individual gRNA. This particular guide was selected to target Exon 6 of the ETV6 locus unaffected by the TEL-AML1 translocation. As is apparent from this figure, this guide had a robust quality score and the potential off-targets all contain mismatches from the original guide. The off-targets provided are not guaranteed to undergo double-strand breaks. Rather, they represent other sequences in the genome that have similar homology and should therefore be noted when analyzing results.



guide #1 **quality score: 96**

guide sequence: CCAACGGACTGGCTCGACTC **TGG**
on-target locus: chr6:+134211815
number of offtarget sites: 30 (5 are in genes)

top 20 genome-wide off-target sites — show all exonic

sequence	score	mismatches	UCSC gene	locus
CCAACAGACTGCCTGGACTCAGG	0.2	3MMs [6:12:15]	NM_011640	chr11:-69405445
CCAAAGGACAGGCTCTCCTCCAG	0.1	4MMs [5:10:16:17]	NM_009269	chr13:+53440217
CCAACCTACTTGCTTGACTCTGG	0.1	4MMs [6:7:11:15]	NM_172903	chr7:+87514719
CCAAAGGACTGGCACGTCCCTGG	0.0	4MMs [5:14:17:19]	NM_181848	chr2:-4957674

Figure 4. CRISPR Design Tool Guide RNA Readout. *In addition to the sequence and genomic location for each designed gRNA, a quality score was provided based on potential off-targets. As is evident from this example, off-targets typically deviate from the target sequence by 3 or 4 basepair mismatches.*

2.3: CRISPR Lentiviral Vector System

To express the 59 guide RNAs in our target cell line, we utilized the lentiviral vector system established by Kosuke Yusa at the Wellcome Trust Sanger Institute in Cambridge, UK. In their 2013 Nature Biotechnology publication, "Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library", the Yusa lab constructed an empty gRNA expression vector for the purpose of transfecting mammalian cells with CRISPR guide RNAs. Yusa and his colleagues targeted genes related to GPI-anchor biosynthesis in an embryonic stem cell line that constitutively expressed Cas9[31]. Of the 52 guides they designed and transfected, 50 were able to produce double strand breaks in their target genes. This high efficiency validated their lentiviral vector system and was the basis for our decision to use this approach. The pKLV-U6gRNA(BbsI)-PGKpuro2ABFP plasmid, pictured in figure 5, was a gift from Kosuke Yusa (Addgene plasmid # 50946). The key features of this plasmid are the gRNA scaffold, the BbsI cloning site, puromycin selection marker, BFP tag, and the ampicillin resistance for growth in bacteria.

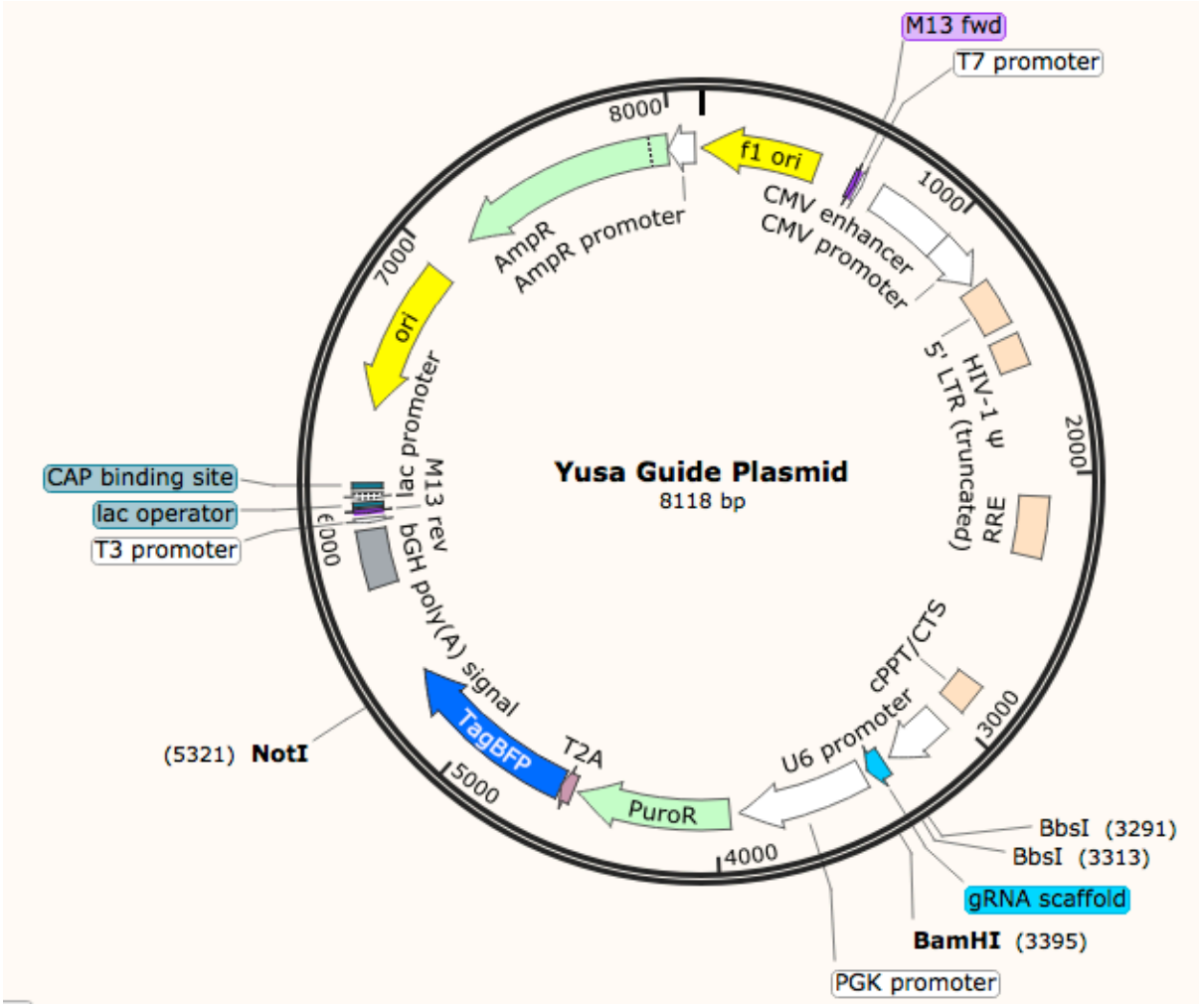


Figure 5. Yusa Guide RNA Plasmid Map. *Guide RNA duplex oligos were inserted into this plasmid via the BbsI cloning sites. This plasmid contains a BFP fluorescent tag and a Puromycin resistance marker. The HIV-1Ψ motif allows for lentiviral packaging and the SIN sequence inactivates the virus upon integration. The key link bringing Cas9 together with our designed guides is the gRNA scaffold just downstream of the cloning site. (Plasmid Map created with SnapGene® Viewer)*

2.4: Guide RNA cloning into the Yusa Plasmid

The pKLV-U6gRNA(BbsI)-PGKpuro2ABFP plasmid was obtained from Addgene (plasmid #50946) in the form of a bacterial agar stab. The culture was streaked out on an LB-Agar plate with Ampicillin and incubated overnight at 37°C. Individual colonies were isolated and used to inoculate 2 mL LB-Ampicillin liquid cultures. Following overnight incubation, plasmid DNA was extracted via Qiagen Miniprep kit and diagnostics were performed to confirm the identity of the plasmid. Plasmid identity was initially confirmed with two separate enzyme double-digests (SacII/NcoI and BamHI/NotI) and later by full plasmid sequencing. The isolated Yusa plasmid was then transformed in One Shot® Stbl3™ Chemically Competent *E. coli* because its ability to accurately clone DNA that will be used in lentiviral production. A large batch of the plasmid was prepared using a Qiagen Maxiprep kit. The key to cloning the gRNAs into the Yusa Plasmid is a three-step digestion, ligation and transformation.

1. Digestion: The Yusa plasmid in its unmodified form has an 18-base pair spacer consisting of two BbsI enzyme recognition sites. BbsI is a rare enzyme that recognizes the sequence 5'GAAGAC3' and cuts 2 bases pairs downstream, crosses over, and cuts after 4 more base pairs on the opposite strand. Figure 6 illustrates the specific cut pattern.

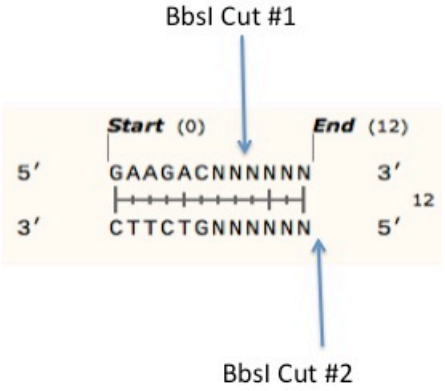


Figure 6. BbsI Cut Pattern. *Following recognition of the sequence GAAGAC, the BbsI enzyme cuts 2 base pairs downstream, crosses to the other strand and cuts a second time 4 more base pairs downstream.*

Enzymatic digestion of the Yusa Plasmid with BbsI results in a linearized vector with 2 noncomplementary sticky ends available for DNA subcloning. This linearized form of the vector is what we utilized for cloning our gRNAs. To isolate the linear fragment, we ran several 20 ug digests of the Yusa backbone vector. Following digestion, the fragments were run on a 0.8% agarose gel and the linearized band was extracted from gel. A Qiaex II Gel Extraction Kit was used to isolate the DNA from the gel. The linearized DNA concentration was measured by spec and ran on a gel to ensure the isolated DNA was, in fact, linear. Figure 7 demonstrates the linearization process for the large prep we created.

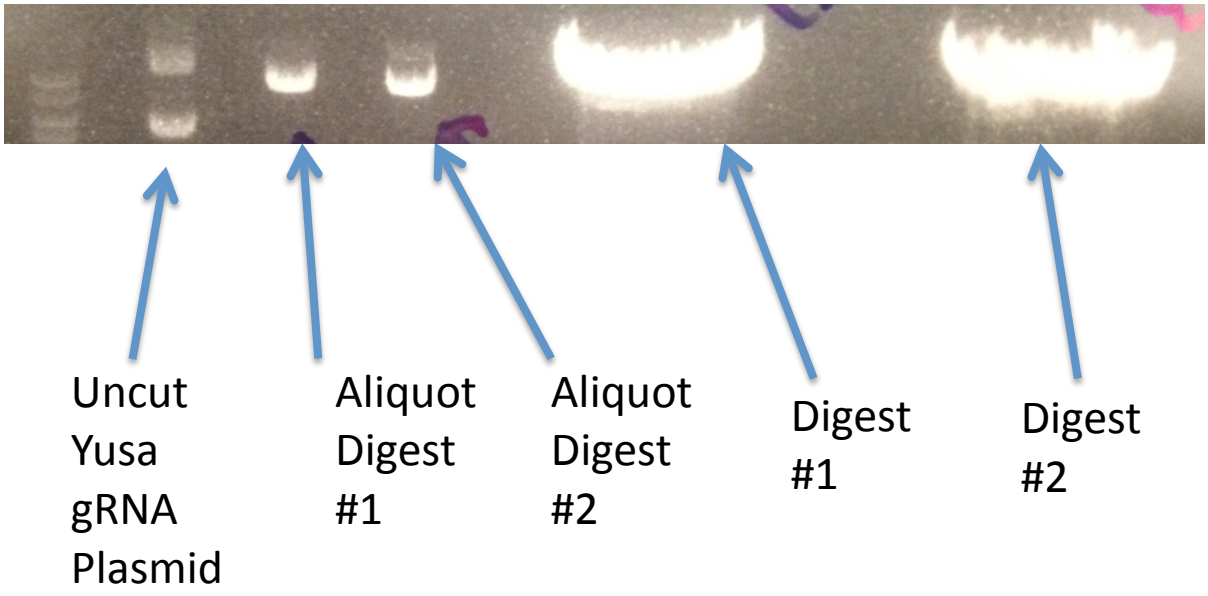


Figure 7. Yusa Plasmid Linerization. *BbsI digestion of the Yusa gRNA expression vector was required in order to clone our guide RNAs. Following digestion and gel electrophoresis, the linearized plasmid (digest #1 and digest #2) was extracted and used in ligation reactions.*

2. Ligation: Following the preparation of stock linear pKLV-U6gRNA(BbsI)-PGKpuro2ABFP, we designed and ordered DNA oligos with our target gRNAs that could be cloned into the vector. In order for successful ligation to occur, the gRNAs had to be in duplex form with complementary sticky ends to the overhangs left in the Yusa Plasmid following BbsI digestion. Thus, two oligos were designed per guide: a top strand in the form 5'-CACC(20 BP guide)GT-3' and a reverse-complement bottom strand in the form 5'-TAAAAC(reverse complement)-3'. When the two oligos anneal to form a duplex, the resultant fragment is the insert for the linearized Yusa Plasmid. 118 custom oligos were ordered through Invitrogen and resuspended in TE Buffer to obtain a stock concentration of 100 uM. Successful ligation of each guide oligo and its complementary bottom oligo into the linearized vector was achieved as follows. Intermediate 40 uM dilutions for all top and bottom oligos were made. In TempAssure PCR tubes 3.5 uL of top oligo (40 uM), 3.5 uL of bottom oligo (40 uM), 2 uL T4 Ligase, 2 uL T4 Ligase Buffer, 1 ug of linearized Yusa plasmid and DH_2O were combined to a total volume of 20 uL. PCR tubes with the ligation mixtures were placed in the thermocycler at 16°C for 8 – 12 hours. Figure 8 depicts the linearization and ligations steps.

3. Transformation: For each guide ligation, 2.5 uL of the reaction mixture was added to 25 uL of thawed Stbl3 E. Coli. The bacteria was incubated on ice for 30 minutes and then heat-shocked for 45 seconds at 42°C. Following heat shock, the transformation mixture was placed back on ice and 125 uL of SOC medium was added. Competent cells were then placed in 37°C shaker at 225 RPM for 1 hour. Transformation efficiencies for ligation mixtures are generally lower than control backbone plasmids. For this reason, the entire transformation volume was spread on an LB-Ampicillin plate and placed in the 37°C incubator for overnight

growth. Successful ligations would typically produce 15-20 colonies per plate, of which 3 would were picked and used to inoculate 2 mL LB-Ampicillin liquid cultures. Following growth in liquid culture, minipreps were performed to isolate the ligation products

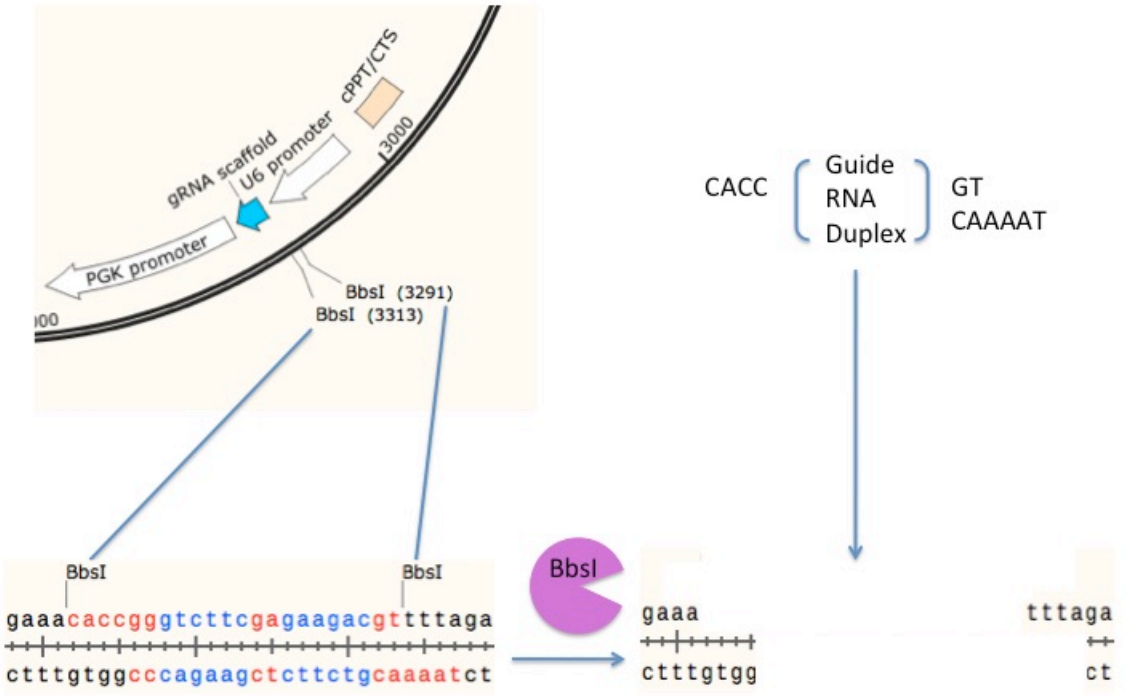


Figure 8. gRNA Duplex Ligation Schematic. *We used a two-step protocol to clone each of our 59 unique gRNAs into the Yusa guide RNA expression vector. Several large digests of the backbone plasmid with BbsI were performed to remove the spacer from the cloning site. The linearized plasmid was run on a gel to separate it from the cut out spacer. Following gel extraction and purification, the linearized form was incorporated in a ligation reaction along with two complementary oligos that would form the duplex insert into the cloning site.*

Upon isolation of the ligated plasmid described above, two steps were taken to confirm that the gRNA adapter was cloned into the linearized Yusa Plasmid. An issue that arose during the cloning process was that a small portion of the linearized Yusa Plasmid that was extracted from the gel contained the uncut, circular backbone. Early partial plasmid sequencing of DNA isolated following ligation and transformation revealed a small portion of colonies that contained the original plasmid. To combat this, a diagnostic BbsI/NcoI digest was performed on all minipreps following transformation of ligation products. If a gRNA was successfully cloned into the linear Yusa Plasmid, the BbsI site was lost. We took advantage of this fact and conducted the double digest to distinguish between backbone plasmid and successful cloning. With a single NcoI cutsite on the original plasmid, a single linear band at 8100 bp following digestion would indicate the loss of the BbsI cutsite and successful ligation of the target adapter. Subsequently, if two bands were present (2200 bp and 5800 bp) it would indicate that the BbsI site was still present and that the transformed plasmid was the original backbone plasmid. Figure 9 demonstrates this differential process:

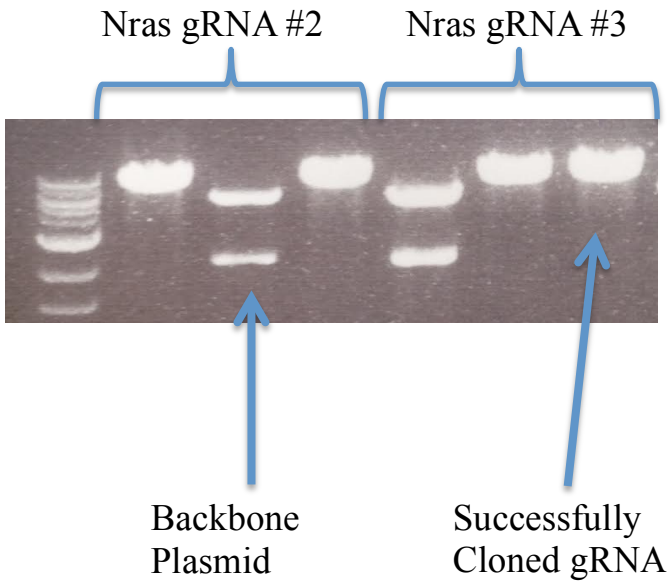


Figure 9. Cloning Conformational Digest. *For NRAS Guide #2, the first and the third lanes represent minipreps with a successfully cloned insert whereas the second lane is the expected band pattern for the circularized, original Yusa backbone plasmid. Likewise, the second and third lanes for NRAS Guide #3 correspond to successful clonings while the first lane is the original Yusa plasmid*

When a miniprep was found to produce a linear band following BbsI/NcoI digestion, it was suspected to be a successfully cloned gRNA. To confirm this, partial plasmid sequencing was performed at the adapter insertion site. The presence of a U6 promoter upstream of the insertion site and gRNA scaffold provided a target for sequencing primer use. The LKO.1 5' U6 primer (5'-GACTATCATATGCTTACCGT-3') designed by the Weinberg Lab was used to sequence from the U6 primer at the 5' end to the gRNA scaffold at the 3' end. 10 uL of sequencing primer at a concentration of 3M was combined with 10 uL of the cloned plasmid and sent to the Massachusetts General Hospital DNA Core Facility.

2.5: Cas9 Vectors

The other key component to the CRISPR system is the incorporation of a Cas9 enzyme into our target cell line. To optimize our experimental design and ensure broad coverage, we selected 3 unique Cas9 vectors that have been used successfully in gene targeting experiments. pL-CRISPR.EFS.GFP (Addgene plasmid # 57818) and pL- CRISPR.EFS.tRFP (Addgene plasmid # 57819) were gifts from Benjamin Ebert[32]. LentiCas9-EGFP (Addgene plasmid # 63592) was a gift from Phil Sharp & Feng Zhang[33]. These three plasmids were ordered from Addgene and delivered as bacteria in an agar stab. Each culture was streaked out on an LB-Agar plate with Ampicillin and incubated overnight at 37°C. Individual colonies were isolated and used to inoculate 2 mL LB-Ampicillin liquid cultures. Following overnight incubation, plasmid DNA was extracted via Qiagen Miniprep kit and diagnostics were performed to confirm the identities of the three plasmids. The Feng Zhang Lenticas9 underwent two diagnostic double-digests, NheI/SpeI and SacII/BamHI, to confirm the correct plasmid was isolated. The Ebert RFP (11,701 bp) and Ebert GFP (11,707 bp) plasmids are nearly identical with the major difference being their individual fluorescent tags. Both were digested with a NotI/BamHI double-digest to confirm they were Ebert plasmids. To distinguish between the two, a unique single-cutter sequence was identified in the Ebert GFP plasmid not seen in the Ebert RFP plasmid. This enzyme, BsrGI, was used to digest both plasmids and confirm that the two were unique from each other. Following these initial diagnostics, maxipreps of all three Cas9 vectors were completed in preparation for lentiviral production. To confirm that no mutational events or rearrangements had occurred, full plasmid sequencing was conducted on these Cas9 vectors and input sequences were confirmed.

2.6: Pooled gRNA library and MGA gene gRNAs

To produce the library, we combined 3 uL of each of the 59 gRNA minipreps into a single eppendorf [164 ng/uL]. Next, the combined miniprep was transformed in 50 uL of Stbl3 E. Coli and dispersed across 10 LB-Ampicillin plates for overnight growth. Following incubation, sufficient colony coverage was noted and all of the individual colonies were scraped together and used to inoculate a 250 mL LB-Ampicillin liquid culture. The culture was incubated at 37°C overnight and a maxiprep was performed to isolate the pooled DNA.

2.7: Lentiviral Production and Titration

To integrate the gRNAs and Cas9 cDNAs into our target cell's genome, we utilized the lentiviral approach seen in the Yusa paper described earlier. Lentiviral packaging was performed for the three Cas9 vectors, the pooled gRNA library and the four MGA gRNAs. The following protocol details the transfection method used to obtain the viral supernatants. 293T cells were used as our packaging cell line. Media for the cells was prepared using 440 ml DMEM, 50 ml Fetal Bovine Serum, 5 ml penstrep and 5 ml glutamine. The cells were plated and expanded across eighteen 10cm plates. Two plates were used for each of the eight vectors (3 Cas9 vectors + 1 Library + 4 MGA gRNAs) requiring a total of 16 plates. The remaining two plates were for transfection of a positive control RFP/GFP plasmid that has been used repeatedly in our lab with high efficiency. The morning of the transfection a media change was performed and 10 ml of fresh media was added to each of the 18 plates. For our packaging system vectors, we used the Δ8.9 plasmid (tat, rev, gag/pol) and vsv-g. For each plate, the following steps were performed. 6 ug of backbone plasmid, 5.4 ug Δ8.9 plasmid, and 0.6 ug of vsv-g were added to an eppendorf with 170 ul of optiMEM and incubated at room temperature for 5 minutes. Meanwhile, 40 ul of

of PEI (transfection agent) was added to a separate eppendorf containing 170 ul of optiMEM and incubated at room temperature for 5 minutes. The two eppendorfs were then combined and incubated at room temperature for 20 minutes. The reaction mixture was then pipetted on top of the target 293Ts in a dropwise fashion. At 24 hours post-transfection, a media change was completed. At 48 and 72 hours post-transfection, viral media was collected, filtered through a 0.45 uM filter, and added to ultracentrifuge tubes. Following the two collections, samples were spun down in a Beckman Coulter Ultracentrifuge for 2 hours at 20,000 RPM and 4°C. Following centrifugation, samples were decanted only leaving a small volume (~200 ul) of media to settle and were placed in the 4°C fridge. Following 2 hours of incubation, the samples were brought out and pipetted carefully up and down 10 times to loosen the viral pellet. The duplicate viral supernatants were pooled and 50 ul aliquots were prepared and placed in -80°C.

To check our viral titers, we performed a virus titration using NIH 3T3 Fibroblasts. 3T3 media was prepared as follows: 440 ml DMEM, 50 ml Inactivated Calf Serum, 5 ml penstrep, and 5 ml glutamine. Cells were seeded over four 10cm plates and expanded for 3 days. The day before the infection, the 3T3s were counted and plated across 3, 12-well plates. 30 of the 36 total wells were used and $1*10^5$ cells were seeded per well. 3 wells were dedicated to each of the 8 viral vectors, 3 for the positive control and the remaining three were left uninfected (negative control). For each lentiviral supernatant, 3 dilutions were prepared – 1:1, 1:5, and 1:25. 10 ul were taken from each dilution and added to an eppendorf with 250 ul of 3T3 media and 2.1 ul of Polybrene (infection agent at 250X). Virus/Media/Polybrene mixture was incubated at room temperature for 30 minutes. At this point, media was removed from the wells of the 12-well plates and the viral mixture was added. Plates were lightly swerled and placed in 4°C for 1 hour. After this incubation, the samples were moved to the 37°C incubator for another 6 hours. After

this time point, 1 ml of 3T3 media was added to each well and 12-plates were left overnight. The following day, media was changed and after 48 hours, infected cells were visible under the fluorescent microscope as red, green or blue depending on the tag. To quantify the infection efficiency, the 3T3 cells were run through FACS and percent infection was determined based on the corresponding fluorescent tags (MGA gRNAs – Blue, CRISPR library – Blue, Ebert RFP, Ebert GFP, Feng Zhang Lenticas9 GFP). FACS plots for 1:1 samples for the positive control, 8 vectors and the negative control can found in the Supplemental Figures and Tables. These plots confirm that virus was successfully produced and that target cell lines could now be infected with our Cas9 vectors and CRISPR library.

2.8: Cell Lines and CRISPR-Cas9 Lentiviral Infection Protocol

For our proof-of-principle experiment, two cells lines were utilized. ST2, a stromal cell line from mouse bone marrow, was seeded on a Corning T25 flask with 8 ml of ST2 media (RPMI, 5% FBS, 1% Penstrep, 1% Glutamine, 1.8 ul β-Mercaptoethanol). These adherent cells served as feeders for Clone 8, a mouse pro-B cell line. Clone 8 pro-B cells were thawed, spun down and resuspended in ST2/B-Cell media (RPMI, 10% FBS, 5% Recombinant IL-7, 1% Penstrep, 1% Glutamine, 1.8 ul β-Mercaptoethanol). Media from the ST2 feeder flask was removed and replaced with 1 ml Clone 8 pro-B cells + 7 ml of ST2/B-Cell media. The culture was expanded until the pro-B cells were approximately 90% confluent. On the day of infection, 100 ul of MGA gRNA #1 lentiviral supernatant, 100 ul of MGA gRNA #3 lentiviral supernatant, 100 ul of MGA gRNA #4 lentiviral supernatant, 200 ul of FZ LentiCas eGFP lentiviral supernatant and 28 ul of Polybrene were combined in an eppendorf. 5 ml of ST2/B-Cell media was removed from the culture and the lentiviral supernatant/polybrene mixture was added to the flask. Following overnight incubation at 37°C, 7 ml of fresh ST2/B-Cell media was added to the culture. 2 days post-infection, the pro-B cells were sorted using FACS into three populations: GFP^+BFP^+, BFP^+ only, and GFP^+ only. Uninfected Clone 8 Pro-B cells were used as a negative control during sorting.

Chapter Three: Results

3.1: Selection of Gene Target Sites

The 2013 Nature Genetics publication, "RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia", was a helpful guide in our selection of genes to target using CRISPR-Cas9. In this work, Papaemmanuil et al. conducted exome and whole-genome sequencing on samples from 57 ALL patients with the ETV6-RUNX1 translocation[28]. As part of their analysis, they identified 14 genes with the highest frequency of somatic mutations across their cohort. In many instances, an individual patient would exhibit mutations in more than one of these high frequency genes. The mutations themselves included deletions, frameshifts and missense mutations. Table 1 depicts the top 14 genes identified and the number of patients that exhibited a mutational event at each particular gene.

Table 1. Top 14 Genes with Secondary Somatic Mutations.

Gene	Frequency of Mutational Event (Out of 57)
ETV6 (Locus unaffected by the translocation)	14
TBL1XR1	13
PAX5	12
ATF7IP	11
BTG1	11
RAG2	12
BTLA	9
NR3C2	8
CDKN2A and CDKN2B	7 (Between the two genes)
KRAS	6
STAG2	5
ZMYM2	4
MGA	4

Table is a summarization of data from Papaemmanuil et al. 2013, figure six[28].

These genes are responsible for a diversity of cellular functions. Some specifically relate to B-development while others are involved in general cycle cell and cell differentiation processes. Along with these 14 targets, we decided to include 5 additional genes that have also been noted to undergo somatic mutagenesis following the TEL-AML1 rearrangement. These genes are NRAS, NSD2, SAE1, SMC5 and SMC1A. Similar to the first 14, these 5 genes are responsible for varying functions including tissue signaling, chromatin modifications, and Cohesin complex formation.

Protein domains and their corresponding coding exons were analyzed for the 19 genes of interest to determine the appropriate genomic locations for Cas9-induced mutations. As our target cell line is murine-derived, we focused on the GRCm38.p4 genome using Ensembl genome database and UniProt protein database. Particular attention was given to domains that define the functional role of these genes in normal cellular processes such as the DNA-binding domain in transcription factors and Ligand-binding domain in nuclear hormone receptors. Additionally, literature review was conducted for each gene to determine if previous studies had identified mutational hotspots linked to gene disruption and disease. Based on our findings, specific coding exons were selected for Cas9 targeting. For genes with little available information, early exons were chosen because indels introduced early in the coding sequence have a higher likelihood of knocking out the gene via frameshifts and early stop codons. Figures 10 through 28 provide an in-depth analysis of each gene, its function, and the differential reasoning for targeting specific exons.

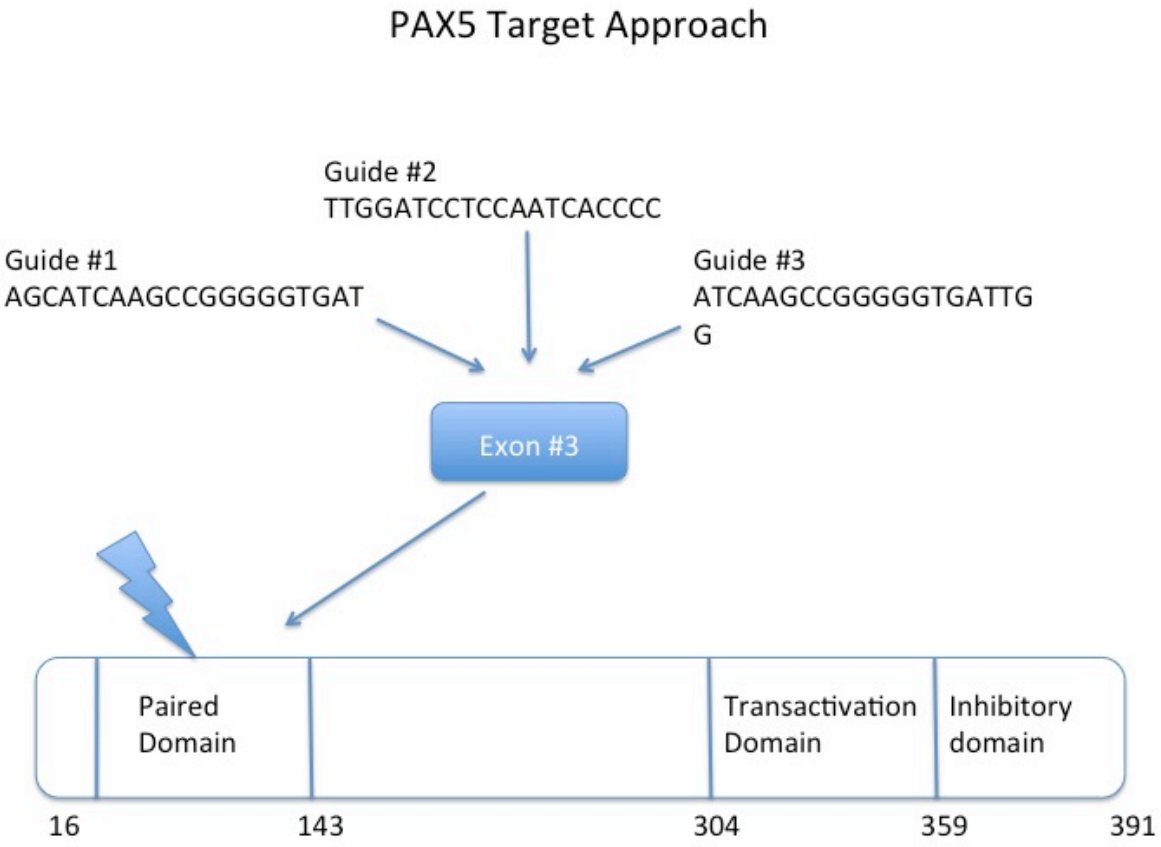


Figure 10. PAX5 Gene. *A member of the Paired Box transcription factor family, PAX5 is present in early lymphoid progenitors and its expression is the key step in B-cell lineage commitment. Multiple studies have also provided evidence that it serves as a tumor suppressor and that haploinsufficiency of PAX5 can lead to ALL. In Dang et al. 2015, it was demonstrated through exome and Sanger sequencing of leukemic B-cells that knockout mutations of PAX5 occur almost exclusively in the Paired Box domain*[34]*. This domain is responsible for binding the DNA-binding domains of other transcription factors including those of the ETS family such as ETV6 and those of the RUNX family such as RUNX1. This pool of findings provides strong evidence that PAX5 is a secondary hit in TEL-AML1*$^{+}$ *ALL as it interacts with the two critical genes involved in the translocation. Additionally, it has been shown that the Paired Box domain is split into two sub-domains that can function independently*[35]*. Taking into account all of this information, we decided to target Exon 3 as it covers both subdomains and includes amino acid*

85 which has been deemed the most common location for mutagenesis in the Paired Domain of PAX5.

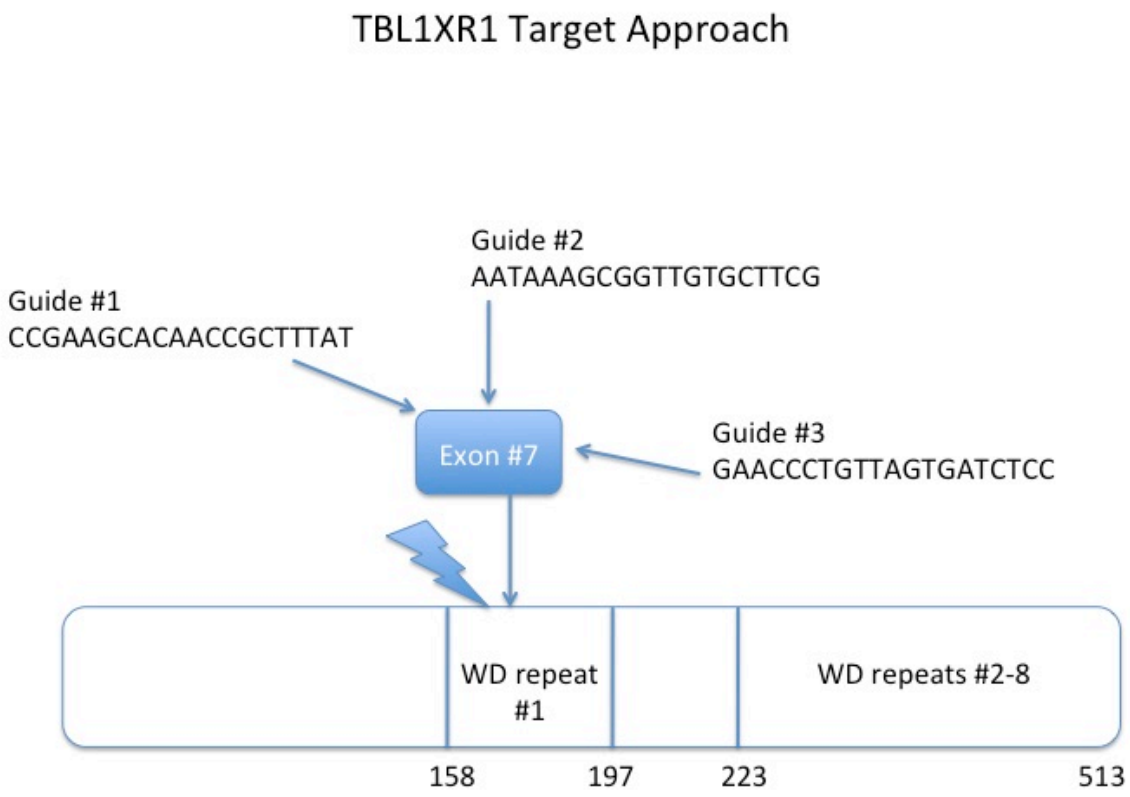


Figure 11. TBL1XR1 (TBLR1) Gene. *Transducin (beta)-like 1X-linked receptor 1 encodes for a nuclear protein involved in gene regulation. TBL1XR1 interacts with the NCoR (nuclear receptor corepressor)/SMRT (silencing mediator of retinoic acid and thyroid hormone receptors) repressor complex which is responsible for regulating glucocorticoid hormone receptor expression[36]. It has been shown that this process relies on protein-protein interactions between a specific WD40 domain on TBL1XR1 and the RD4 domain on NCoR. Once bound to the NCoR/SMRT complex, TBL1XR1 binds to the DNA encoding for the hormone receptor of interest and blocks transcription[37]. A 2015 paper examining the physiological role of TBL1XR1 identified the first WD40 domain as the one that directly binds NCoR[36]. With respect to ALL, the link between this gene and the TEL-AML1 translocation was made as early as 2008 when Parker*

et al. noted a novel somatic deletion of TBL1XR1 in 15% of their ETV6-RUNX1^{+} pre-B-ALL patients[38]. Additionally, TBL1XR1 has been shown to have a regulator function in hematopoiesis and been implicated in important signaling pathways such as Wnt/β-catenin and Nf-κB[39]. Taking into consideration the importance of the first WD40 repeat, we decided to target Exon #7 (the 4th coding exon) which encodes this particular domain. While the goal is to knock this gene, we hope that targeting this particular repeat will at the very least prevent TBL1XR1 interaction with the NCoR/SMRT repressor complex in order to recapitulate the findings of these previous studies.

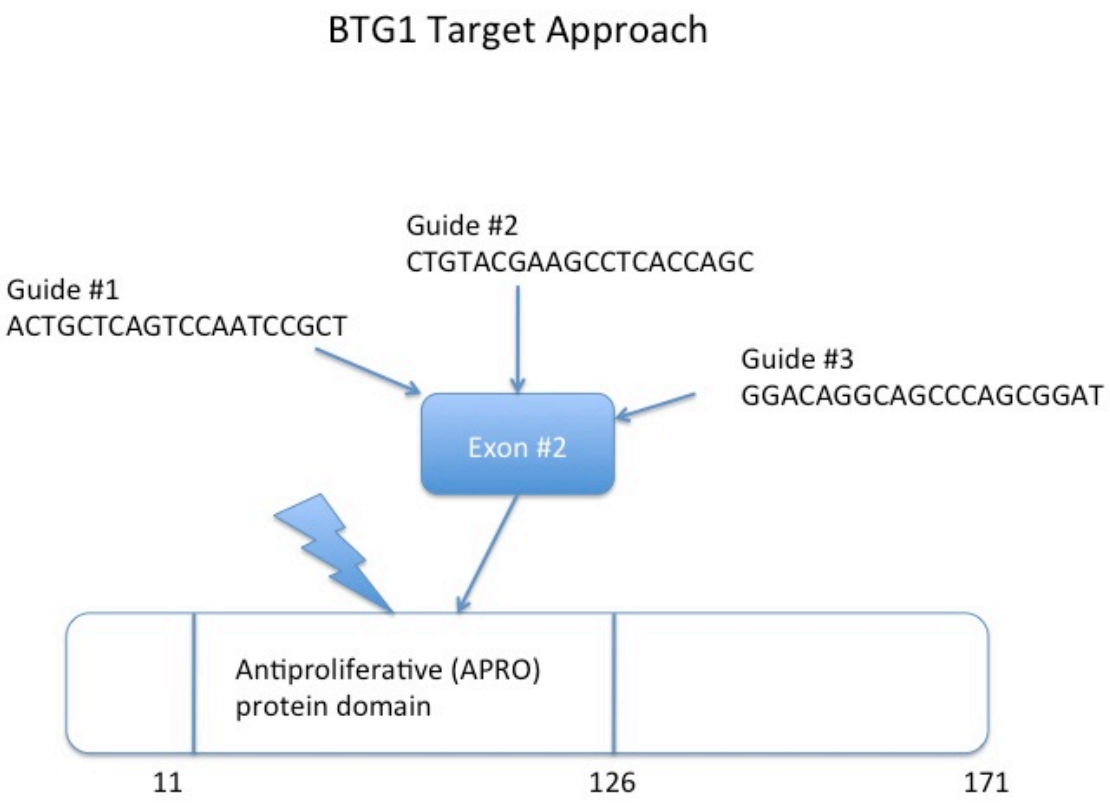


Figure 12. BTG1 Gene. *B-cell Translocation Gene 1 plays an important role in the negative regulation of the cell cycle. It is most highly expressed during the G0 and G1 phases and decreases dramatically once the G1 checkpoint has been passed. Through its antiproliferative protein domain, it is able to bind regulatory transcription complexes that have been shown to negatively control cell proliferation*[40]*. More specifically, BTG1 has been shown to interact with the CCR4-NOT (CNOT) transcription complex and the Homeobox Protein HOXB9*[41]*. Overexpression CNOT subunits and HOXB9 have been linked to blood cancers as well as other malignancies. In their 2007 publication that looked at genetic abnormalities in TEL-AML1*$^{+}$ *ALL, Tsuzuki et al. noted a recurring loss of BTG1 in a significant percentage (25%) of their patient samples. BTG1 was the fourth most commonly deleted gene behind PAX5 (25%),*

CDKN2A (29%), and the TEL gene not involved in the translocation (89%)[42]. The BTG1 gene has two major coding exons. Based on the work of Waanders et al., we decided to target exon 2. In their study of precursor B-cell ALL, they noted 8 different deletion sizes of BTG1 but all of them mapped to the second exon[43]. The second exon is also the larger of the two and presents a wider genomic range to induce double-strand breaks.

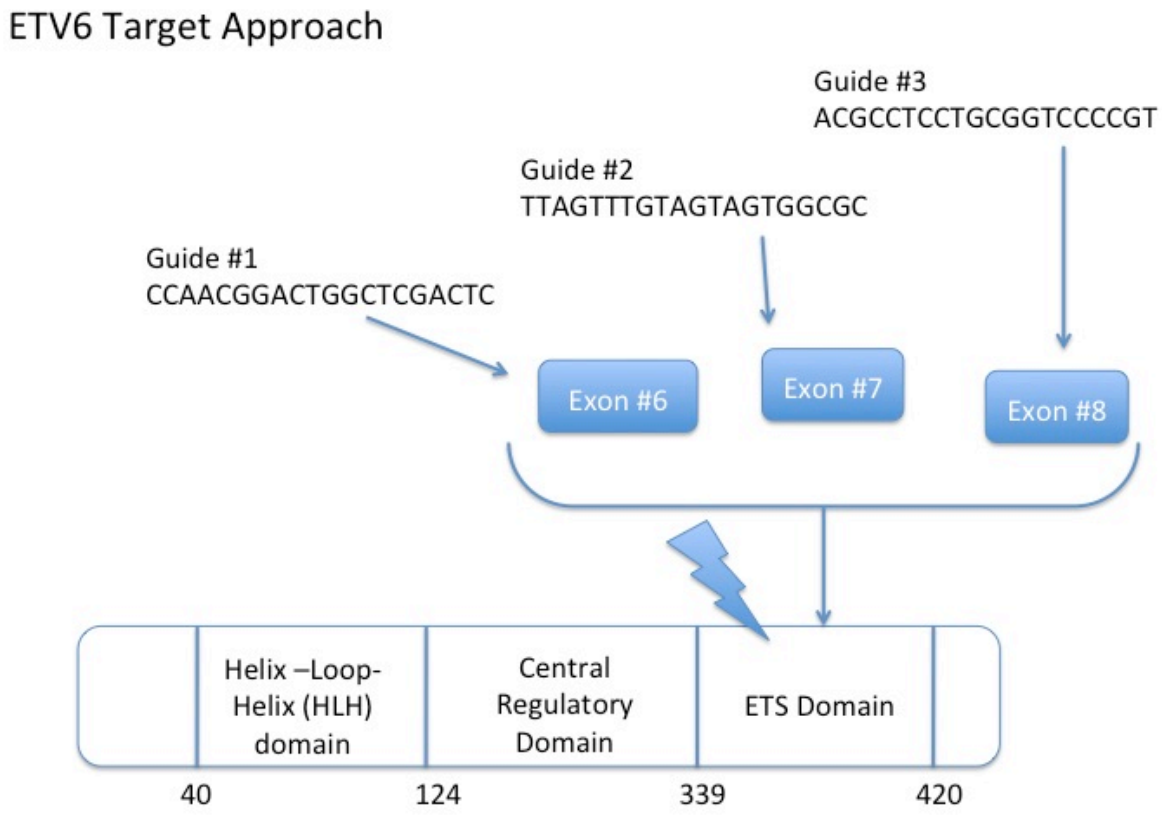


Figure 13. ETV6 (Tel) Gene. *The loss of the unaffected TEL allele is the most common secondary event following the initial TEL-AML1 gene fusion. In Papaemmanuil et al., 14 of the 57 patient samples exhibited TEL loss, which was the highest of any gene*[28]*. It is not fully understood why this molecular event occurs but speculation is that TEL may act as a tumor suppressor or be involved in the regulation of a closely related tumor suppressor*[44]*. With respect to structure, the TEL gene encodes for a transcription factor with three important domains; the N-terminus Pointed (PNT) domain, a Central Regulatory Domain, and a C-terminus DNA-binding domain of the ETS-type. It has been previously demonstrated that blocking of the ETS DNA-binding domain impairs the functionality of the TEL gene. Additionally, mutations have*

been identified in the ETS domain that induce red cell macrocytosis and thrombocytopenia and establish a preleukemic state[45]*. Eight coding exons give rise to a 57-kDa TEL protein. Exons 6,7 and 8 correspond to the DNA-binding domain and, in order to ensure coverage of the entire domain, each was incorporated into our gRNA design. Genomic sequences from each of the three exons were input into the CRISPR Design Tool and the top gRNA was selected for each to be included in our library.*

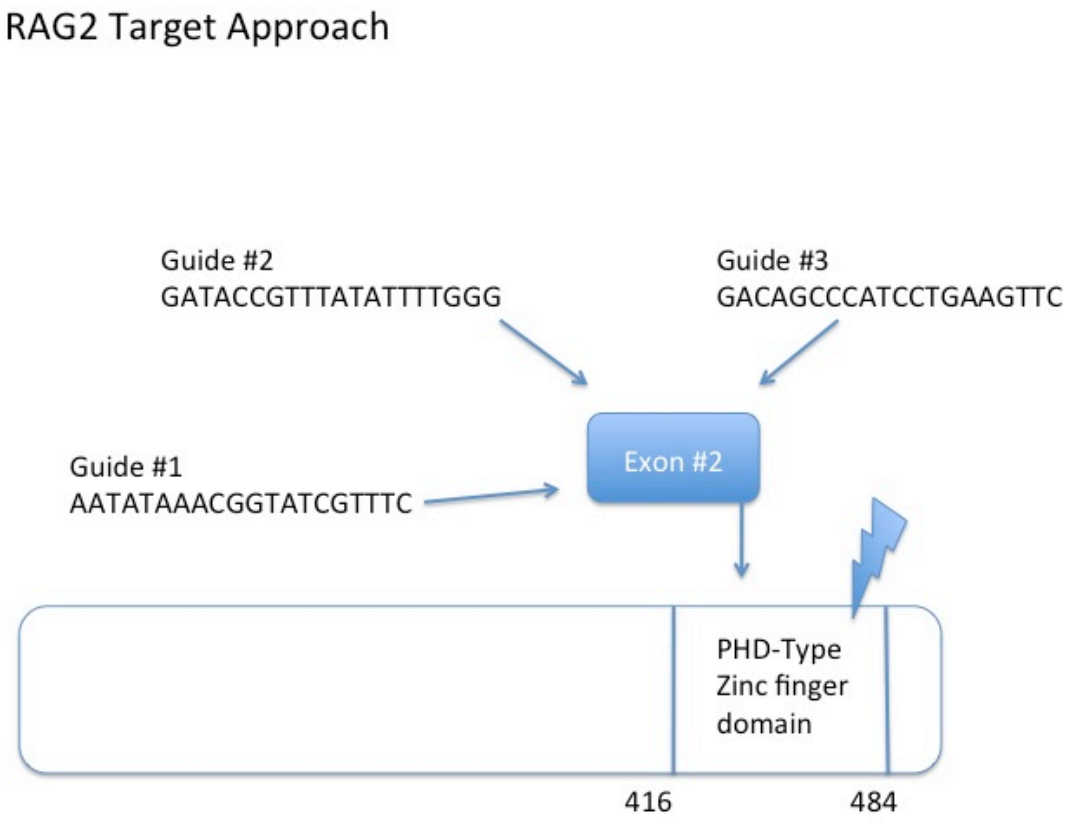


Figure 14. RAG2 Gene. *Recombination Activating Gene 2 encodes for a component of the functional enzyme responsible for VDJ rearrangement of B-cell receptor (BCR) and T-Cell receptor (TCR) genes. RAG2 complexes with RAG1 and induces double-strand breaks in the germline DNA corresponding to the antigen-binding site for these receptors. These breaks are repaired by Non-homologous end joining (NHEJ) which results in receptor diversity and the production of unique clones during B-cell and T-cell development*[46]*. RAG1 is responsible for the catalytic activity while RAG2 serves as a scaffold and binds to the DNA. The important functional domain that makes this possible is the PHD-type Zinc Finger. The N terminus of the domain complexes with RAG1 while the C terminus has been shown to interact with histone H3*

modification, trimethylated lysine 4 (H3K4me3) which assists in directing the RAG complex to the correct genomic DNA[47]. Because this domain is so crucial to proper RAG2 function, we decided to target it with 3 guide RNAs. The RAG2 gene has two splice variants with a single coding exon corresponding to the Zinc Finger domain. RAG2 deletions have been previously noted in childhood ALL and are thought to be responsible for aberrant RAG1 cutting due to dysregulation of the enzymatic complex.

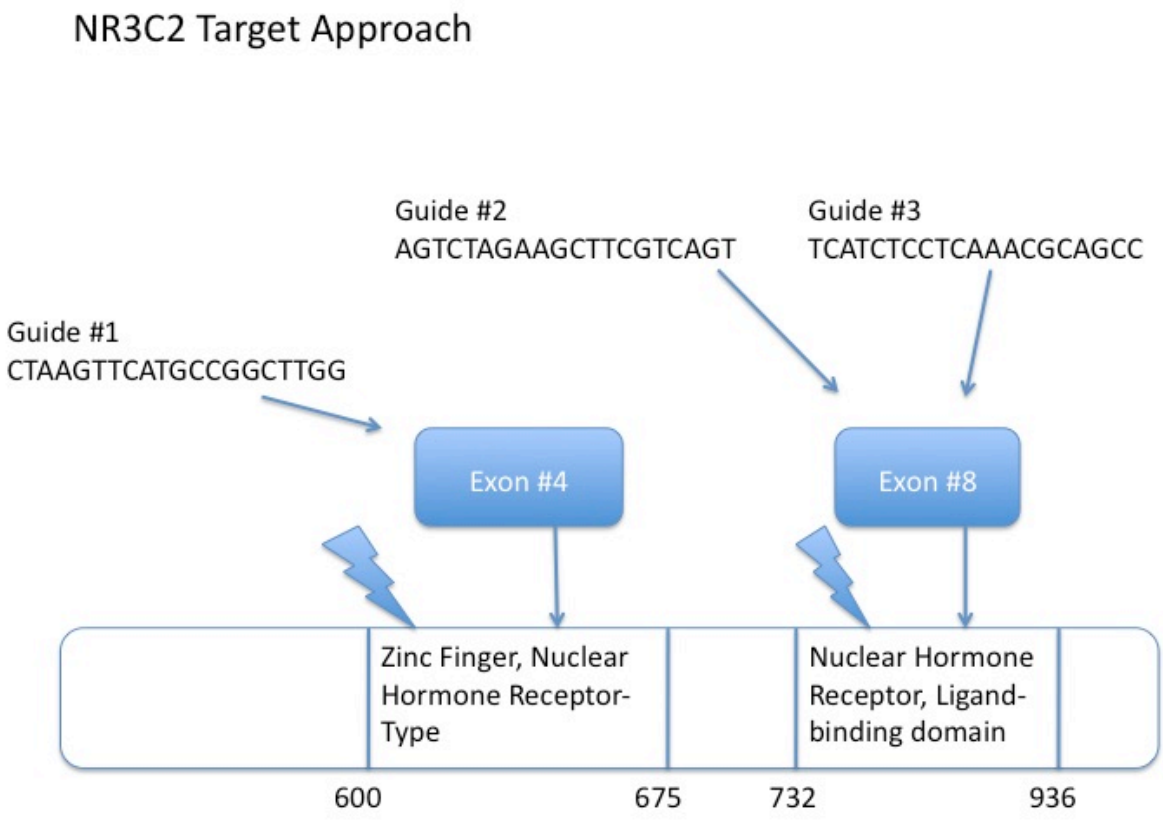


Figure 15. NR3C2 Gene. *Commonly referred to as Mineralocorticoid Receptor (MR), the nuclear receptor subfamily 3, group C, member 2 is expressed ubiquitously in the body and is a key regulator of electrolyte balance. This nuclear receptor has been shown to bind both mineralocorticoids such as Aldosterone and glucocorticoids such as Cortisol*[48]*. However, MR function is most well understood in the context of regulating salt concentration and blood pressure. When MR binds Aldosterone, the ligand-receptor complex moves from the cytoplasm to the nucleus where it regulates hormone response elements that control proteins involved in ion transport. Of importance is MR's effect on the production of the Sodium-Potassium ATPase. Dysfunction in the Mineralocorticoid Receptor has been implicated in numerous cardiovascular*

and metabolic diseases[48]. With respect to leukemia, Nr3c2 deletion has been noted across a few studies as being a common occurrence in TEL-AML1^{+} ALL patients[28,42]. There are two important protein domains on the Mineralocorticoid Receptor that are required for proper functioning. The Zinc Finger domain (amino acids 600-675) is responsible for complexing with the target hormone response element and the Ligand-binding domain is necessary for the initial interaction with Aldosterone. There are 8 splice variants associated with MR and only two of the coding exons are conserved across all 8, exon 4 (3rd coding exon) and exon 8 (7th coding exon). Exon 4 encodes part of the Zinc Finger Domain while exon 8 encodes for part of the Ligand-binding domain. Thus, we decided to target both of these exons in hopes of completely disrupting the gene.

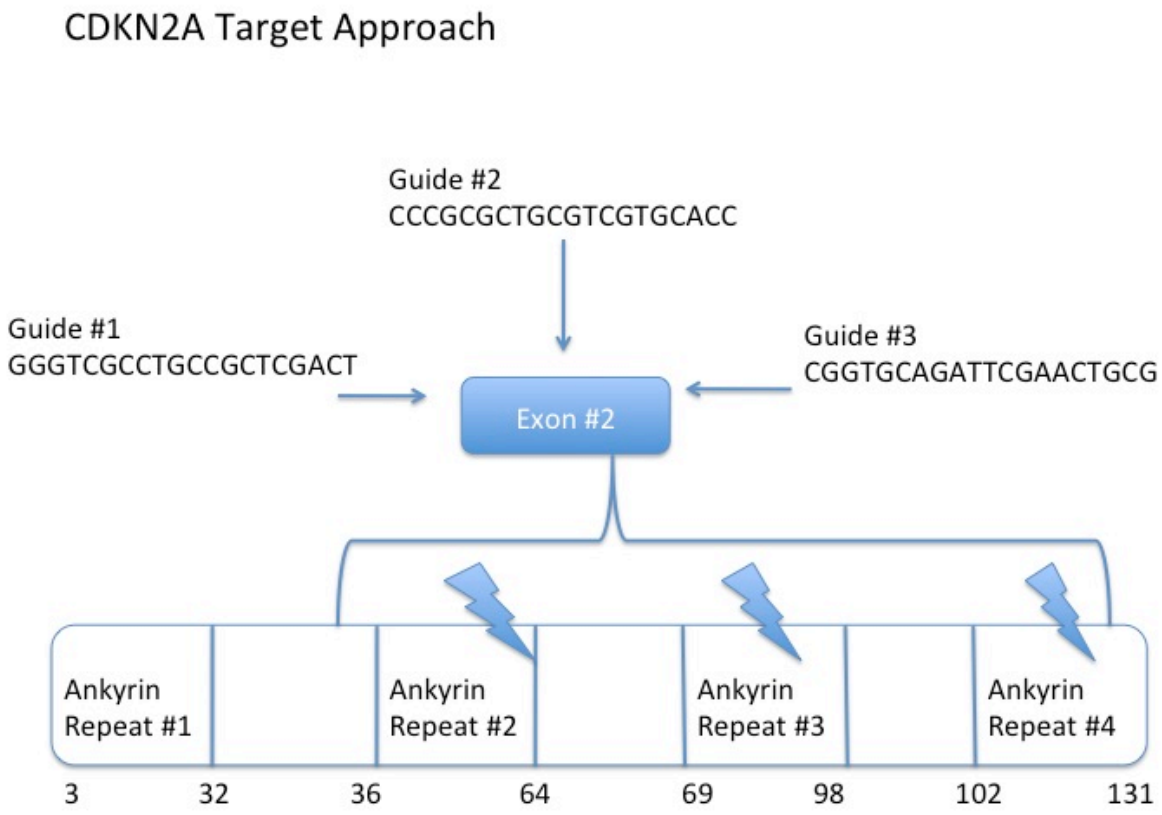


Figure 16. CDKN2A (p16INK4a/ P14ARF) Gene. *Cyclin-Dependent Kinase Inhibitor 2A is a critical cell cycle gene and encodes 2 tumor suppressors, P16 and P14ARF. Alternative splicing of the coding mRNA transcript allows for the production of two unique tumor suppressors from the same gene. P16 indirectly activates the RB tumor suppressor, which prevents the transition from G1 to S in the cell cycle*[49]*. It does this by inhibiting Cyclin Dependent Kinases 4 and 6 (CDK4 and CDK6), which normally phosphorylate and inactivate the RB protein. However, when CDK4/CDK6 are blocked by CDKN2A, RB remains unphosphorylated and is able to stop cell cycling. P14ARF acts in similar way by indirectly activating the p53 tumor suppressor*[49]*. MDM2 is a negative regulator of the p53 tumor suppressor and inactivates it during cell cycling.*

P14ARF will block MDM2, which frees up p53 and allows it to stop the cyclin-CDK signaling pathway needed to move through the different phases of the cell cycle. With regards to structure, a series of Ankyrin Repeats represent the major domains. Ankryin repeats are involved in protein-protein interactions and are seen throughout cell cycle signaling. We decided to target Exon 2 as it encodes for 3 of the 4 Ankyrin repeats and is the only exon conserved between the two alternative splice variants. Additionally, previous studies of CDKN2A reveal a high propensity for mutations in the second coding exon and that polymorphisms in this exon create susceptibility to ALL[50].

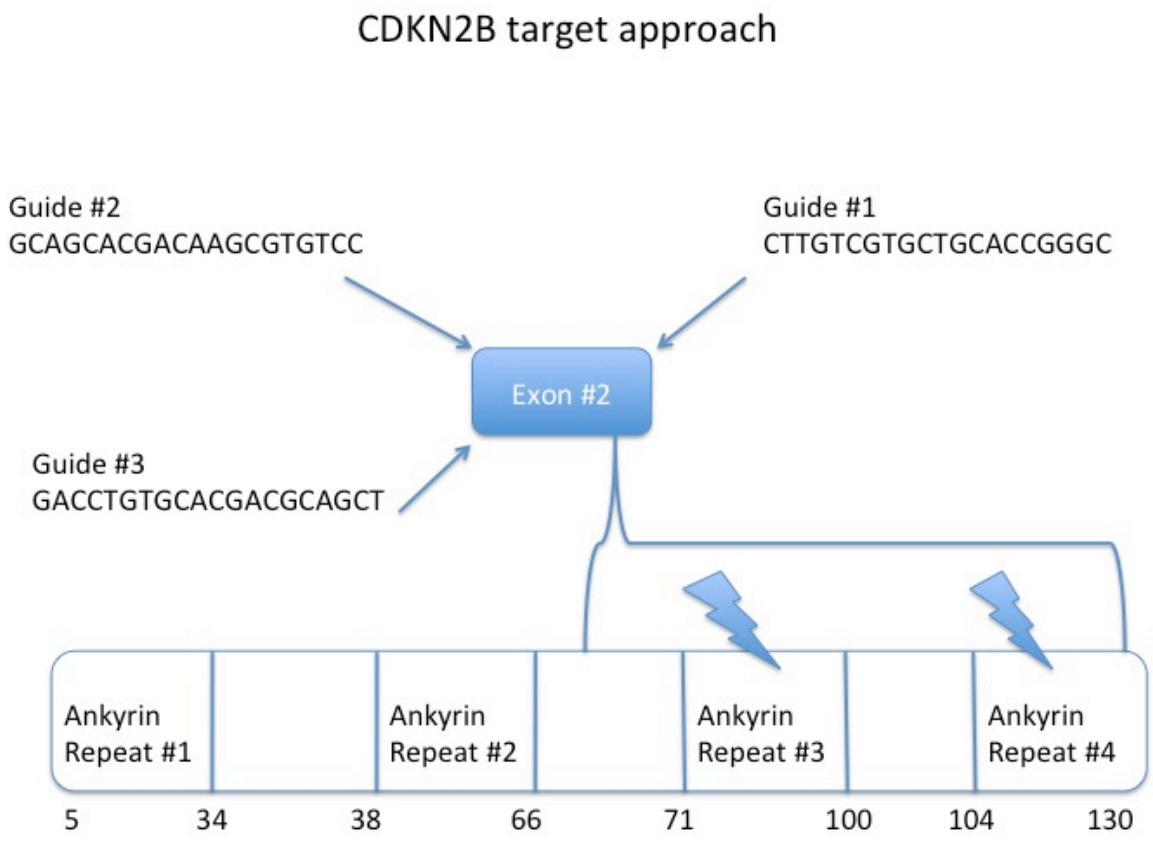


Figure 17. CDKN2B (p15INK4B) Gene. *The Cyclin-Dependent Kinase Inhibitor 2B gene lies adjacent to CDKN2A in the genome. It encodes the p15INK4B tumor suppressor, which works in a mechanistically similar fashion to p16. P15INK4B binds directly to CDK4 and CDK6 preventing the phosphorylation of Cyclin D and the transition from G1 to S in the cell cycle. A unique feature of this tumor suppressor, when compared to P16 and P14ARF, is its role as a link between TGFβ expression and negative regulation of cell proliferation*[51]*. It has been shown that cells exposed to TGFβ have an increase in P15 expression resulting in growth arrest due to CDK4/CDK6 inhibition. This effector pathway provides an interesting link between an immune cytokine and the cell cycle. Additionally, several independent patient studies have revealed that*

p15 is deleted in a high number of pediatric ALL patients[52-54]*. Like CDKN2A, this gene is composed of 4 Ankryin repeat domains, which are needed to bind CDK4/CDK6 and inhibit Cyclin D activation. We decided to target exon #2 as it as the larger of the two coding exons and is commonly deleted in heme malignancies.*

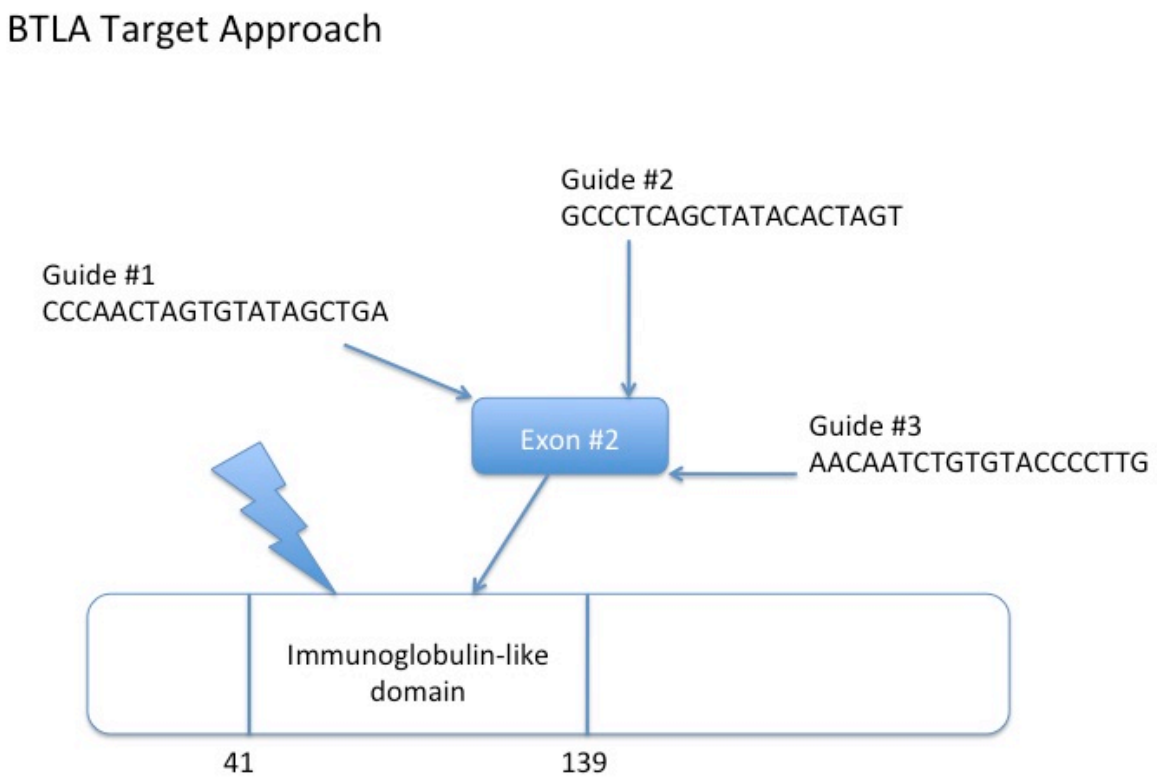


Figure 18. BTLA Gene. *B and T Lymphocyte Associated gene is a member of the Immunoglobulin gene superfamily and is expressed on the surface of T helper 1 (TH1) cells. This protein is a co-inhibitory receptor containing a single Ig-like domain, a transmembrane domain and two Immunoreceptor tyrosine-based inhibitory motifs (ITIMs)*[55]*. BTLA is absent on the surface of immature T cells but, upon activation, is expressed exclusively on the surface of TH1 cells. Two major ligands have been identified for BTLA, B7-H4 and the Tumor Necrosis Factor Receptor (TNFR). When bound to either of these, the BTLA receptor initiates inhibitory signals that block TH1 proliferation and responsiveness. In a cohort study of 1154 patients with precursor B-cell ALL, BTLA was deleted in 5% of the cases and was associated with poorer*

prognosis and increased propensity for relapse following treatment[56]. Because of its importance in ligand binding and the initiation of inhibitory signals, the Ig-like domain was chosen as our target for CRISPR-Cas9 mutagenesis. With six coding exons and two splice variants, exon 2 was the perfect target as it conserved across both variants, is an early exon, is the largest exon and directly translates into the Ig-like domain.

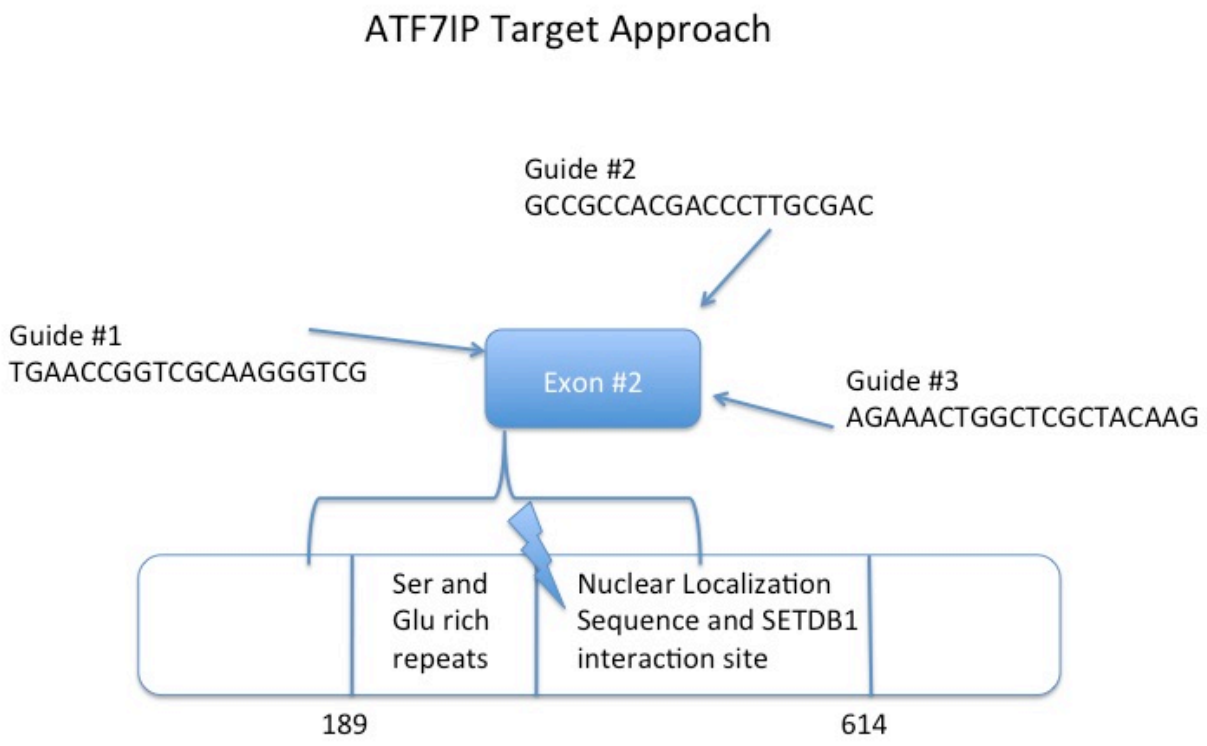


Figure 19. ATF7IP (AM/MCAF1) gene. *Activating transcription factor 7-interacting protein 1 is a nuclear protein involved in chromatin modification of histone H3K9. ATF7IP forms a repressor complex with Methyl-CpG-binding domain protein 1 (MBD1) and Histone-lysine N-methyltransferase (SETDB1)*[57]*. As indicated through their respective nomenclature, SETDB1 is methyltransferase and MBD1 is a nuclear protein that binds methylated DNA. Using two different binding domains, ATF7IP is responsible for bringing these two proteins together and shuttling them to the target H3K9. MBD1 binds H3K9 dimethyl and SETDB1 adds an addition methyl group to produce a trimethylated histone H3K9*[57]*. H3K9me3 is often a measurement of heterochromatin and transcriptional repression. Additionally, aberrant H3K9 methylation can*

lead to disruptions in NOTCH signaling and have been implicated in B-cell and T-cell leukemogenesis[58]. In Kobayashi et al. 2014, a patient with precursor B-cell ALL was noted to have a novel fusion gene composed of ATF7IP and Beta-type platelet-derived growth factor receptor (PDGFRB)[59]. As for a link between ATF7IP and TEL-AML1^{+} acute lymphoblastic leukemia, Papaemmanuil et al. was the first to establish that this gene undergoes recurrent somatic mutations across a patient population[60]. Selecting a target genomic sequence was difficult for this gene, as it has been identified to produce 8 different protein-coding splice variants. We decided to target an early, conserved exon that encodes a series of serine and glutamine rich repeats as well as the protein's nuclear localization sequence and the beginning of the SETDB1-binding domain. In targeting this exon, we hope to introduce a frameshift mutation that effectively eliminates the nuclear localization sequence, the SETDB1-binding domain and the MBD1-binding domain at the C terminus of the protein.

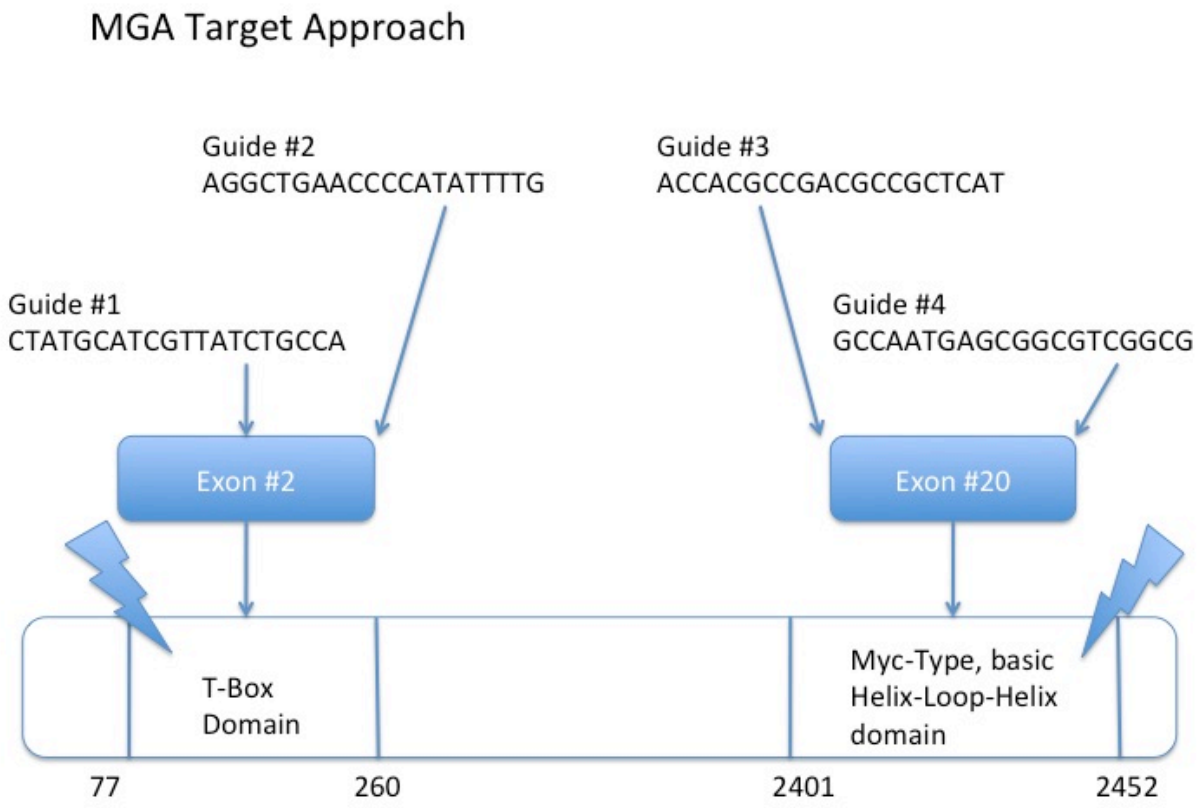


Figure 20. MGA Gene. *The MAX Gene Associated gene encodes for a large transcription factor that is involved in both transcriptional activation and repression*[60]*. The key to understanding the MGA gene's function is an examination of its two critical protein domains. At the C terminus of the protein is a basic helix-loop-helix (bHLH) leucine-zipper domain. The MGA gene belongs to a family of transcription factors called the MAX network, which share this bHLH leucine-zipper domain and a common binding sequence of CACGTG*[61]*. The other transcription factors in this family include MAX, MAD, MNT, and c-MYC and are involved in cell fate, proliferation and programmed cell death. At the heart of the network is the Myc-Associated factor X (MAX) gene, which heterodimerizers with MGA and the other members of this network to form either*

coactivator or corepressor complexes depending on the binding partner. When MAX binds MGA at the C terminus, the dimeric complex serves as a transcriptional repressor by localizing to MYC-specific Enhancer Box DNA response elements and preventing c-MYC transcriptional activation[62]. Functionally speaking, c-MYC is involved in transcriptional activation of proliferation genes and has been intimately linked to B-cell proliferation[63]. If MGA underwent a loss of function mutation at this domain, one could postulate that c-MYC-induced proliferation would persist uncontrolled and could result in malignant transformation. At the N terminus of the protein is a T-Box Domain, which binds the T-box DNA response elements that are involved in early embryonic development and germ layer differentiation[62]. T-box family transcription factors, referred to as Brachyury in mice, are seen throughout early development and recognize the sequence TCACACCT. When the MAX protein binds MGA, the dimeric complex can also localize to T-box response elements and serve as a transcriptional activator during the development of the mesoderm in utero[64]. To effectively eliminate MGA's dual role as both an activator and a repressor, the T-domain and the bHLH domain were targeted with two guide RNAs each. Papaemmanuil et al. 2013 was the first study to identify secondary mutations in the MGA gene in TEL-AML1^{+} acute lymphoblastic leukemia[28]. Prior to this work, however, MGA loss was noted in patients with high-risk chronic lymphoblastic leukemia (CLL) indicating its importance in lymphopoiesis and potential role in precursor B-cell ALL[65].

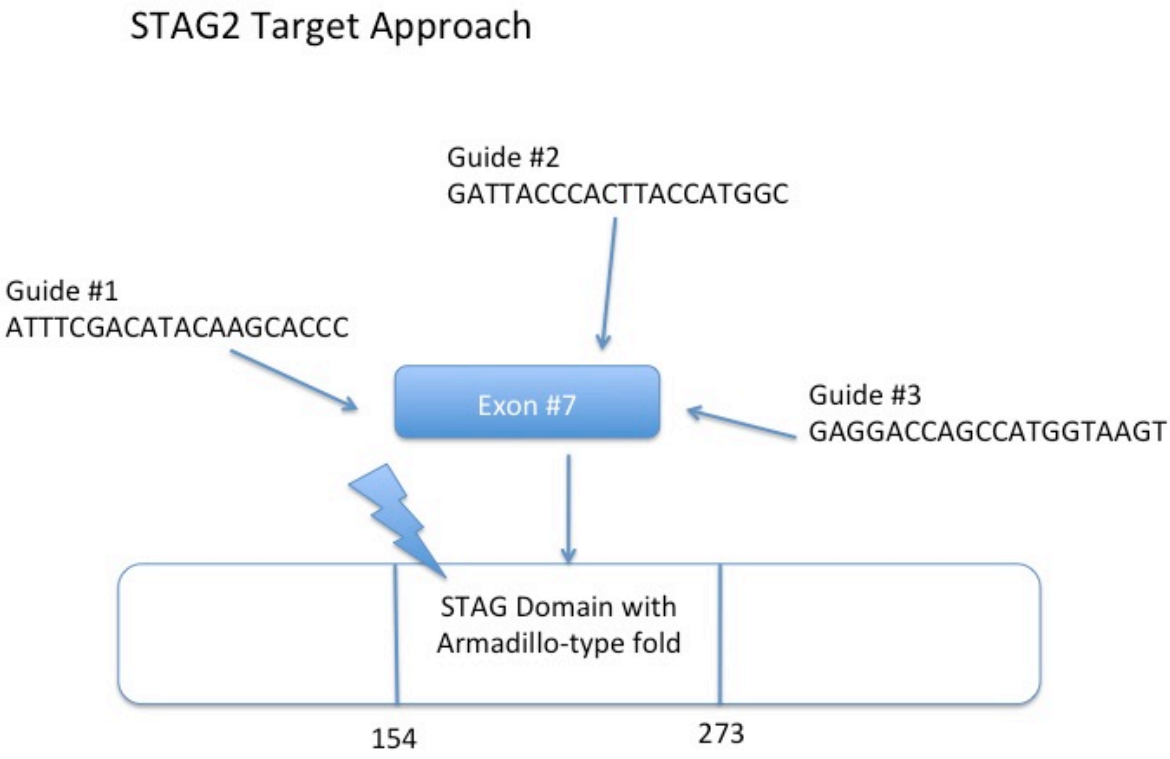


Figure 21. STAG2 (SA2) Gene. *The Stromal Antigen 2 gene encodes for a protein in the Cohesin complex. This ring-like complex is composed of 4 major subunits and plays a critical role in the cell cycle. During S-phase, when DNA is replicated, the Cohesin complex is formed and is responsible for cohesion of sister chromatids*[66]*. Later in both mitosis and meiosis, this complex ensures that sister chromatids separate evenly. STAG2 is an adapter that binds RAD21 on the Cohesin ring and brings it in proximity to the chromatin-binding site (see supplementary figure 1)*[67]*. Dysfunction of the Cohesin complex leads to aneuploidy, an abnormal chromosome number in a cell resulting from a breakdown in chromosome separation mechanisms during cell division*[68]*. Aneuplodies are a hallmark of many cancers and result in genomic instability capable of driving transformation. STAG2 deletions and*

mutations have been previously identified in hematologic malignancies but mostly in the myeloid lineage[69]. Papaemmanuil et al. 2013 was the first study to identify secondary mutations in the STAG2 gene in TEL-AML1^{+} acute lymphoblastic leukemia[28]. Each Stromal Antigen protein, of which there are three, contains a unique STAG domain. This domain is critical to binding within the Cohesin complex[70]. As such, we decided to target this domain via exon 7 (coding exon 6), which is conserved across the different splice variants. Without a functional STAG domain, the Stromal Antigen 2 protein will be unable to link the Cohesin ring with the chromatin.

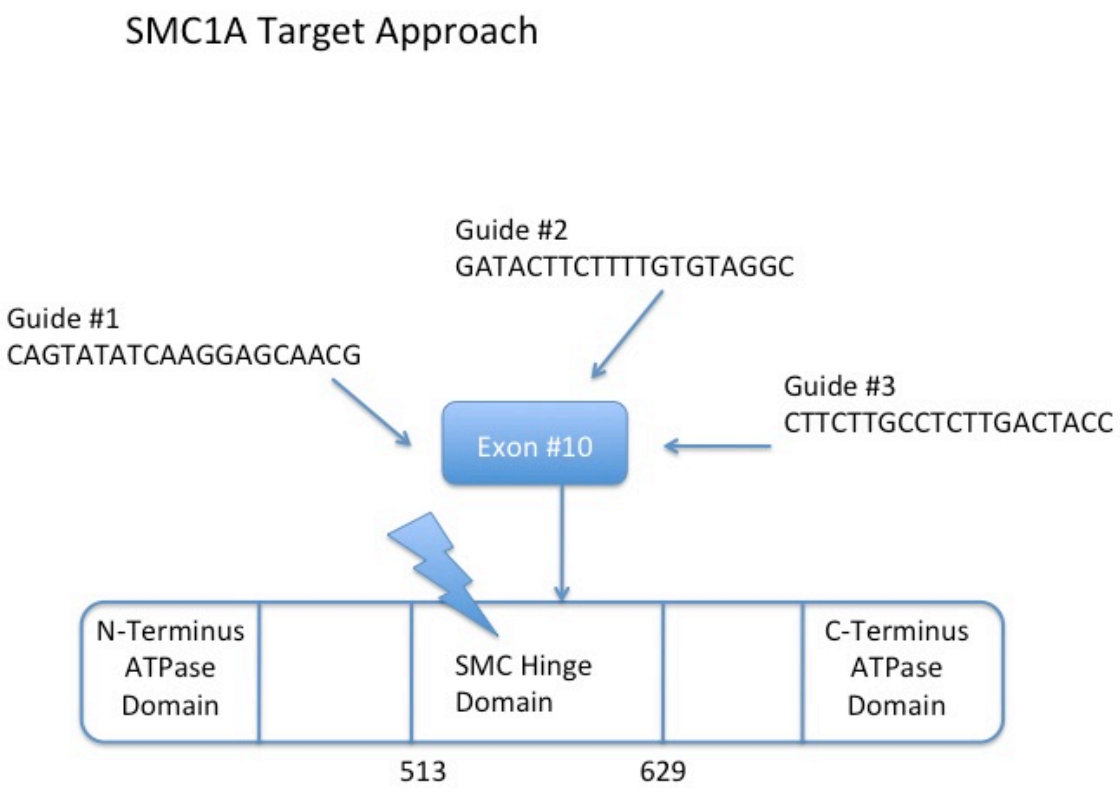


Figure 22. SMC1a Gene. *Structural Maintenance of Chromosomes protein 1A is another important subunit in the Cohesin complex. Whereas STAG2 serves as a linker between the Cohesin complex and the target chromatin, SMC1a is a structural unit within the Cohesin ring itself*[69,70]*. The SMC protein family has a conserved primary structure with ATPase domains at the N-terminus and C-terminus and a SMC hinge domain in the middle of the protein. SMC proteins are protomers that fold onto themselves to form a secondary structure with the N and C termini at one end of the protein and the hinge domain at the other. This secondary structure allows SMC proteins to heterodimerize with each other via hinge-hinge interactions. To form the Cohesin complex ring, SMC1 binds SMC3 through their respective hinge domains and together they bind RAD21 to form a triangular ring (see supplemental*

figure 1)[70]*. Cohesin encapsulates sister chromatids and keeps them bound until their separation during anaphase of mitosis and anaphase II of meiosis. Additionally, SMC1a has been implicated in DNA damage repair via the Ataxia Telangiectasia Mutated (ATM) kinase pathway. Kitagawa et al. demonstrated that SMC1 is a downstream phosphorylation target of ATM-NBS1-BRCA1, a key repair pathway that maintains genomic stability and promotes cell survival following DNA damage*[71]*. In 2014, an impactful study was published that examined recurrent somatic mutations in epigenetic regulators across 1,000 pediatric cancer genomes. Out of the 633 genes analyzed, SMC1a was one of the top 15 most frequently mutated genes*[72]*. This study centered on the most common pediatric cancers including leukemias, gliomas and medulloblastomas. As our target, we focused on the SMC hinge domain. If Cas9 cleavage disrupted this domain, the Cohesin ring would fail to form and could result in aneuploidy due to a failure in chromatid separation.*

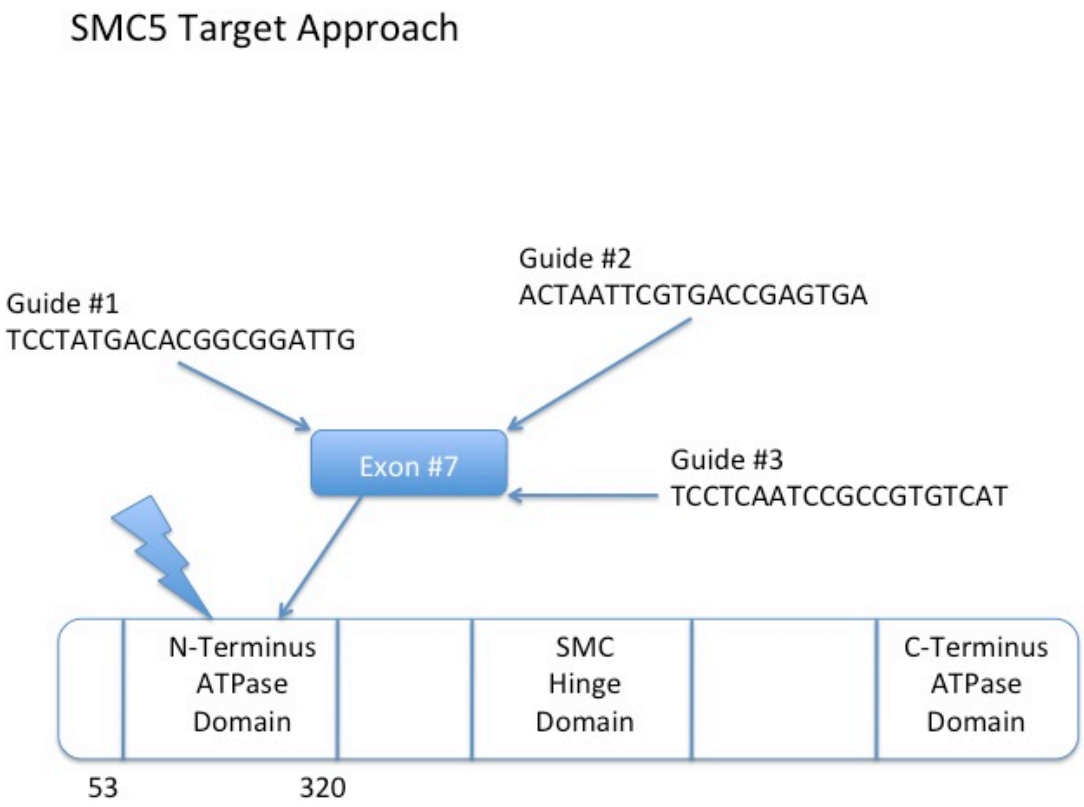


Figure 23. SMC5 Gene. *Structural Maintenance of Chromosomes protein 5 is another member of the SMC family of proteins and is involved in chromosome stabilization and DNA repair. SMC5 dimerizes with SMC6 to form a complex similar in structure to Cohesin (SMC1/SMC3) and Condensin (SMC2/SMC4)[73]. What makes the SMC5/SMC6 complex unique is the presence of 6 non-SMC subunit proteins (NSE1-NSE6) that associate with the complex and aid in its functionality[74]. Over the past few years, researchers have elucidated the function and importance of some of these NSE proteins. NSE1 has been classified as a ubiquitin ligase and NSE2 as a SUMO ligase involved in DNA repair[74]. More recently, NSE3 was identified as the subunit responsible for SMC5/SMC6 binding to double stranded DNA[75]. The SMC5/SMC6 complex has*

been shown to localize to double-strand breaks (DSBs) in the genome and mediate repair via the RAD51 pathway and recruitment of Cohesin. Due to its role in DSB repair, SMC5/SMC6 has been linked to homologous recombination in meiosis that contributes to genetic diversity. Additionally, there is mounting evidence that cancer cells can utilize this complex to lengthen telemomeres via the Alternative Lengthening of Telomeres (ALT) pathway and PML bodies in the nucleus[76]. Rather than target the hinge domain like in SMC1a, we decided to disrupt the N-terminus ATPase domain of SMC5 because this domain binds directly to NSE3, the subunit protein responsible for linking the SMC5/SMC6 complex with genomic DNA. SMC5 has only 1 splice variant so we selected exon 7, which corresponds directly with the N-terminal ATPase domain.

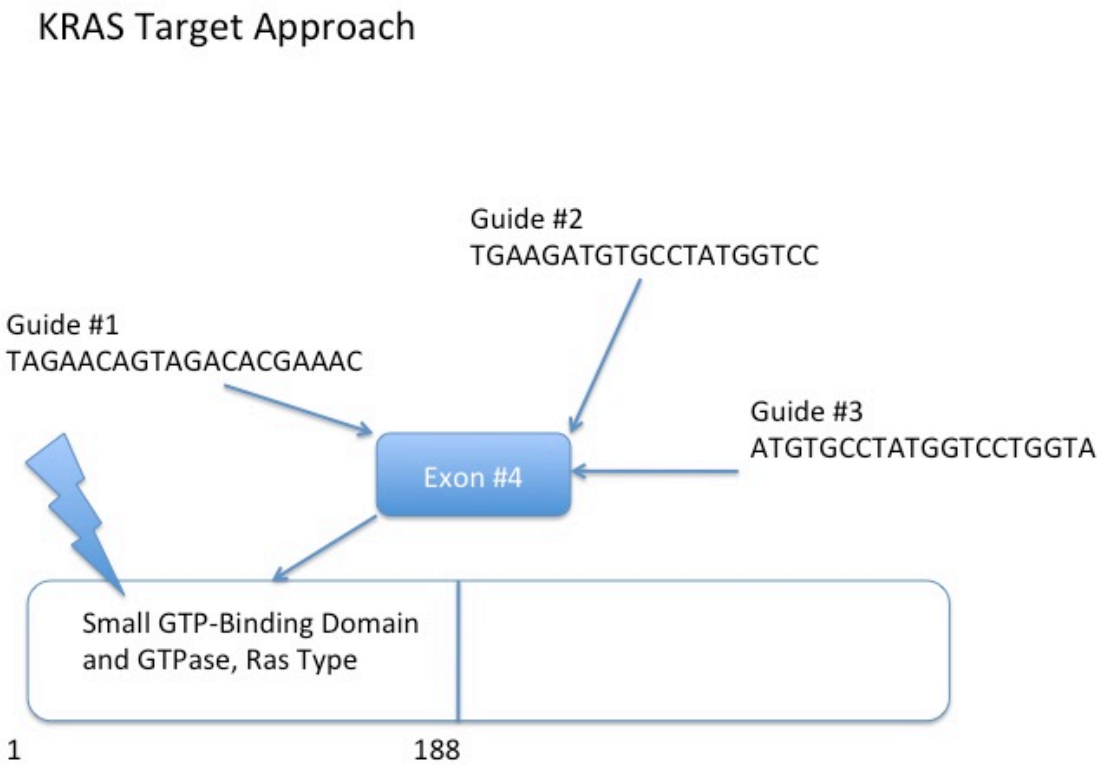


Figure 24. KRAS Gene. *The Kirsten rat sarcoma viral oncogene homolog gene encodes for a Small GTPase, which is a G-protein in the cytosol that binds guanosine triphosphate (GTP) and hydrolyzes it to guanosine diphosphate (GDP). Small GTPases are active when bound to GTP and inactive when bound to GDP, which is similar to the alpha subunit of heterotrimeric G-proteins. In the activate state, GTPases are involved in signaling pathways within the cell that control proliferation, differentiation, metabolism and motility. KRAS belongs to the RAS family of small GTPases, which are commonly mutated in a variety of cancers. In a normal homeostatic state, KRAS is involved in tissue signaling and cell proliferation pathways. More specifically, activated KRAS transmits signals in the RAF/MAPK and PI3K/AKT pathways as*

well as in several growth factor signaling pathways. Additionally, KRAS has been shown to act as tumor suppressor in the lymphoid lineages and mutations in this gene have been implicated in lymphoid leukemias[77,78]. In February of 2016, Chen et al. demonstrated that KRAS is absolutely required in B-cell lymphopoiesis and that a deficiency in this gene results in defective development at the pre-B stage[79]. To try and recapitulate this deficiency, we targeted exon 4 (coding exon 3), which gives rise to a portion of the GTPase and one of the GTP-binding domains (amino acids 116-119) on the KRAS protein.

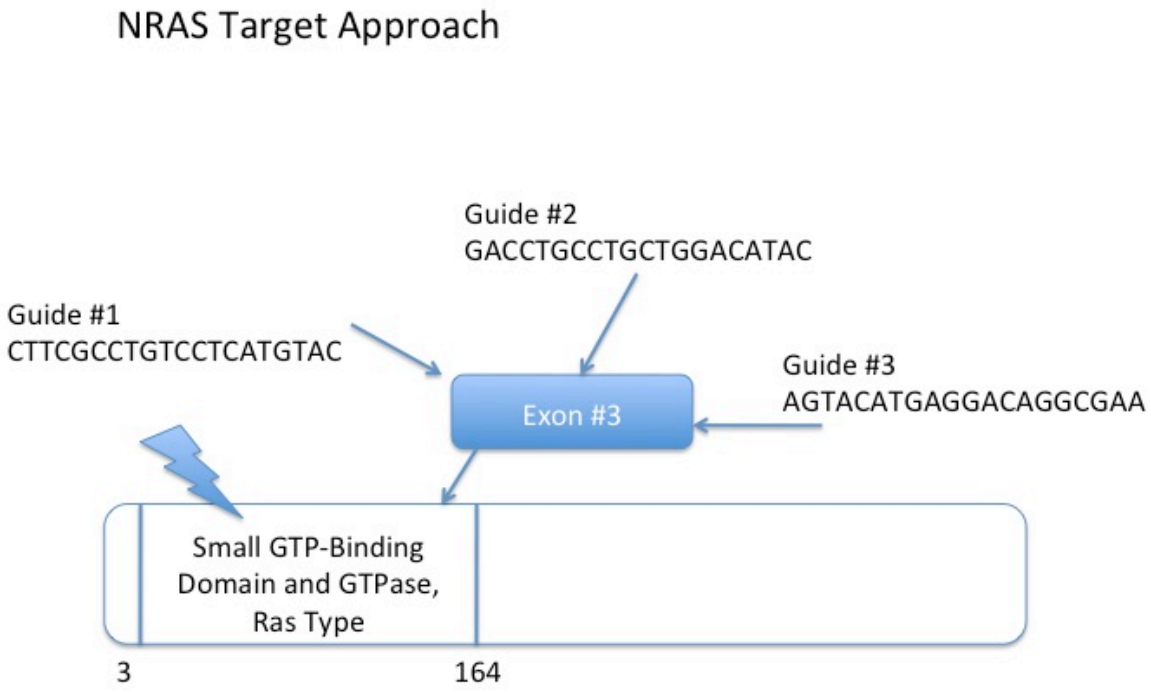


Figure 25. NRAS Gene. *Neuroblastoma RAS viral oncogene homolog is another member of the RAS family of Small GTPases. While NRAS and KRAS are structurally similar, they have been shown to signal through different intracellular pathways and effect distinct cellular functions. Through transformation studies, it has been shown that KRAS signals through Raf/RhoA and regulates cell adhesion while NRAS signals through Akt/Cdc42 and influences cell motility*[80]*. Additionally, the loss of one of these GTPases results in a skewed response by the other. In normal tissues, KRAS and NRAS balance each other because the processes they regulate both deal with cytoskeleton structure and function. The RAS GTPases are proto-oncogenes that are mutated in nearly a third of all cancers*[81]*. NRas mutations have been identified in both myeloid*

and lymphoid leukemias and result in disturbance of the MAPK pathway. While RAS mutations are generally missense and activating in nature, deletions have also been described in patient samples. From a clinical perspective, NRAS mutagenesis has been linked to poorer prognosis in patients with precursor B-Cell ALL[82]. Our decision to target exon 3 (coding exon 2) was based on a 2007 study that identified recurrent mutational hotspots in the NRAS genomic DNA. In this work, Paulsson et al. found that codon 61, which is found in exon 3, is a common mutation site in children with high hyperdiploid ALL[83]. This exon also codes for a portion of the GTPase domain and includes one of the three DNA binding domains in NRAS (Amino Acids 57-61).

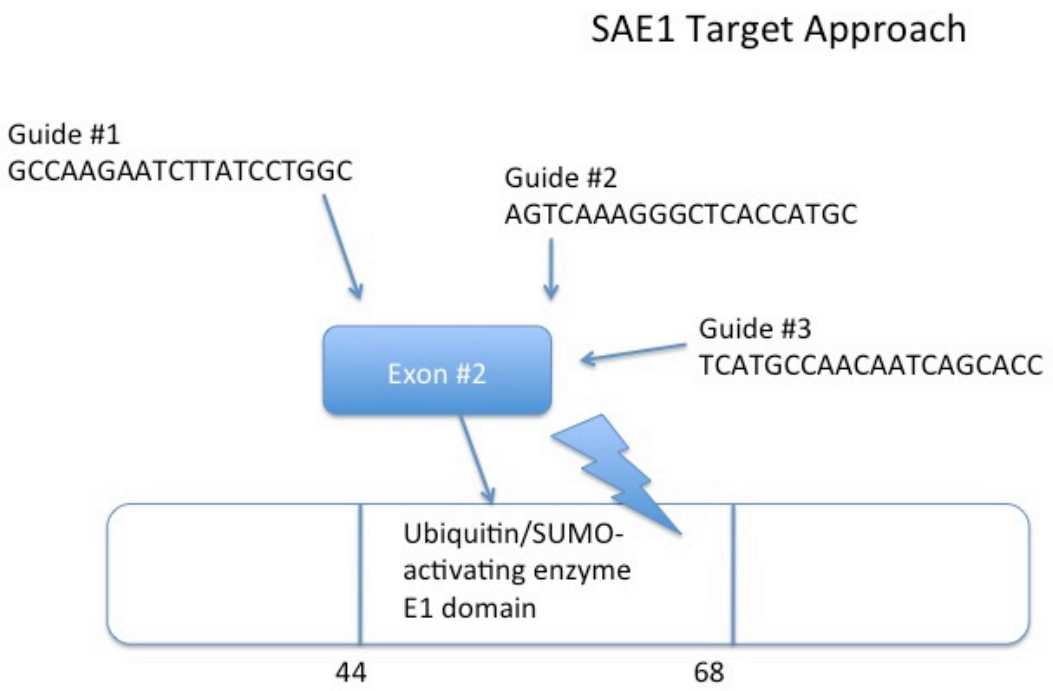


Figure 26. SAE1 Gene. *SUMOylation is a post-translational modification in which a SUMO (Small Ubiquitin-like Modifier) protein is covalently joined to other cellular proteins. This process is analogous to ubiquitination and a way to regulate protein transport and function. Similar to ubiquitination, SUMOylation is an enzymatic process that occurs in three steps and is facilitated by a cascade of enzymes*[84]*. SUMO-Activating Enzyme subunit 1 (SAE1) combines with SAE2 to form the SUMO-Activating Enzyme E1, the first enzyme in the multistep process. SAE1, through an ATP-dependent catalytic reaction, attaches the SUMO protein to SAE2 via a thioester bond. From there, the SUMO protein is transferred to SUMO E2 and E3 enzymes ultimately forming a covalent bound with the target protein. While the complexities of the SUMO*

pathway remain largely unknown, it is certain that SUMOlyation and cancer are intertwined. Studies have linked the SAE1-SAE2 complex to a variety of cancers[85]. Of particular interest to our work is the recent evidence that the SUMO E1 complex can enhance mutant KRAS-induced malignancy[86]. This cooperative relationship may be a potential pathway in TEL-AML1^{+} ALL as both KRAS and SAE1 were mutated genes of interest in the Papaemmanuil study. To disrupt SAE1 and the larger E1 complex, we have targeted Exon 2, which codes for the E1 domain that brings SAE1 and SAE2 together and initiates the thioesterification between SAE2-SUMO.

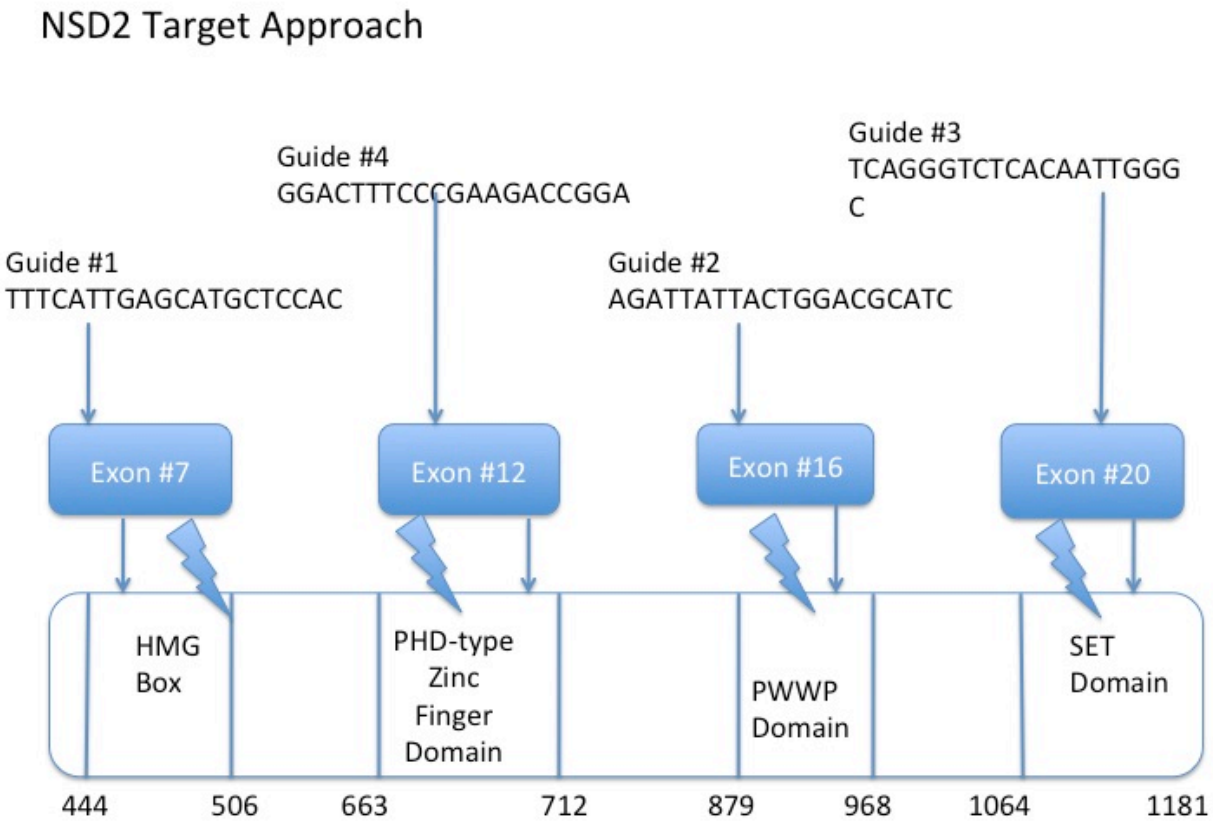


Figure 27. NSD2 (WHSC1) Gene. *The Wolf-Hirschhorn Syndrome Candidate 1 (WHSC1) gene codes for a histone-lysine N-methyltransferase called NSD2. Based on the fact that several chromatin modification genes have been identified as recurrent mutagenic targets, there is convincing evidence that disruption of epigenetic regulatory pathways plays a role in leukemia pathogenesis. Nuclear SET Domain-Containing Protein 2 (NSD2) functions by transferring two methyl groups to Histone 3 Lysine 36 (H3K36). It was believed based on early studies that the NSD2 methyltransferase served as a transcriptional repressor*[87]*. However, more recent work has demonstrated that NSD2 methylation of H3K36 results in transcriptional activation of important immunologic pathways such as the inflammatory NF-κB*[88]*. Overexpression of NSD2 has been*

shown to activate silent oncogenes including TGFA, PAK1, MET, and RRAS2. With respect to disease, NSD2 is overexpressed in prostate cancer and has been identified in a translocation in multiple myeloma, another hematologic malignancy[89]. In 2014, Jaffe et al. published an important paper examining the link between NSD2 and pediatric acute lymphoblastic leukemia. In this work the authors demonstrated, both in vivo and in vitro, that dysregulated NSD2 produced a "hyperactivated" H3K36 that resulted in leukemic transformation. More specifically, the dysregulated NSD2 had a mutation in the SET domain and was noted in 14% of ALL patients with the ETV6-RUNX1 translocation[90]. Further, they were able to show that knockdown of this mutant NSD2 resulted in a loss of the hyperactivated H3K36 state and death of the malignant leukemic cell line. Along with the SET domain, there are three other major domains that constitute the NSD2 enzyme. The HMG box is the DNA binding domain and PHD-type Zinc Fingers assist in histone binding. As for the PWWP domain, a recent 2016 publication demonstrated that this domain allowed NSD2 the unique ability to bind to H3K36me2 and directly regulate transcription[91]. Taking into account the importance of each of these domains, we decided to create CRISPR guides against each of the 4 domains. This involved targeted 4 separate exons that correspond to the 4 domains and are conserved across the major coding splice variants.

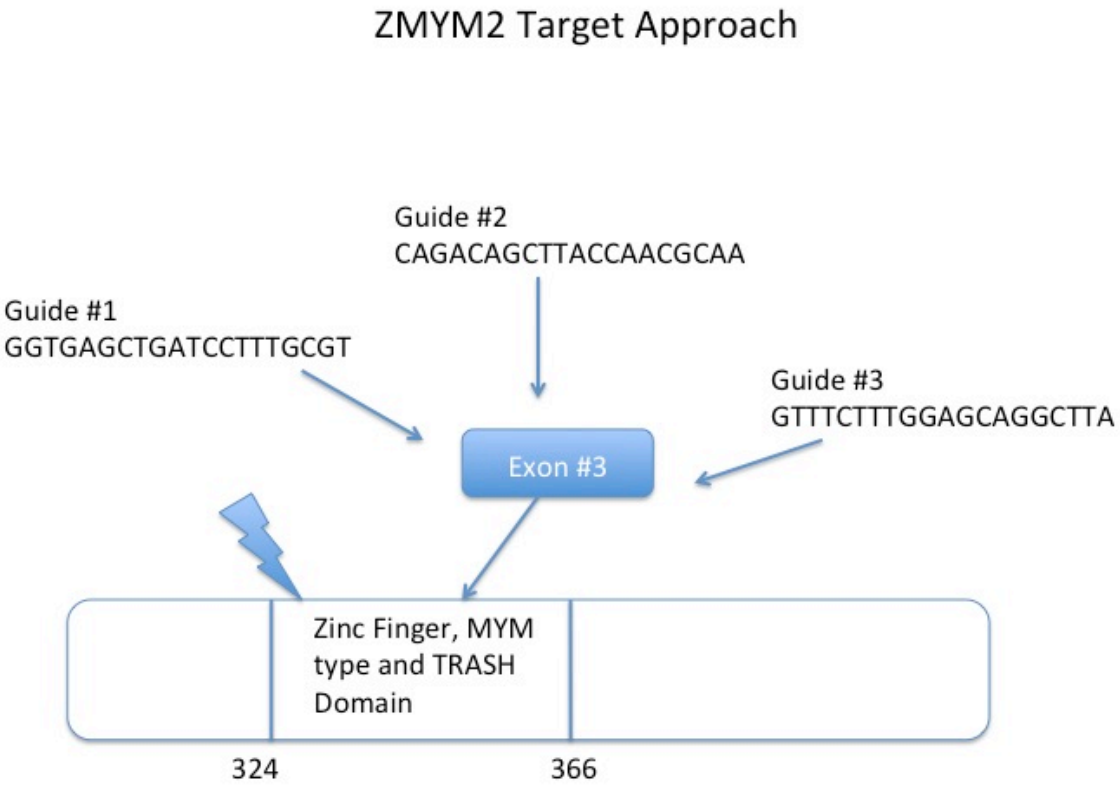


Figure 28. ZMYM2 (FIM/ ZNF198) Gene. *Zinc finger MYM-type protein 2 was first discovered in patients with Stem Cell Leukemia/Lymphoma (SCLL) as a gene fusion partner with Fibroblast Growth Factor Receptor-1 (FGFR1)*[92]*. SCLL is a myeloproliferative syndrome that, if left untreated, can progress to leukemia or lymphoma. Because of the manner in which it was discovered, ZMYM2 is generally referred to in the literature as FIM (Fused In Myeloproliferative disorders). FIM is ubiquitously expressed during development except for in hematopoietic tissue. Interestingly enough, it has been demonstrated that forced expression of FIM in the Aorta-Gonad-Mesonephros (AGM) region, a site of early embryonic hematopoiesis, prevents hemangioblast differentiation into HSCs*[92]*. Since this finding, subsequent studies have*

confirmed that ZMYM2 is involved in chromatin modification and transcriptional repression. More specifically, FIM is part of the LSD1/COREST/HDAC1 (LCH) transcriptional corepressor complex, which functions by removing activating acetyl groups from histones. FIM serves as a scaffold for the repressor complex and supports it during chromatin binding at Histone 3, Lysine 9 (H3K9) and Histone 3, Lysine 4 (H3K4). In 2008, Gocke et al. elucidated the structural biochemistry behind the protein-protein interactions between FIM and the LSD1/COREST/HDAC1 complex[93]. They determined that a string of 9 MYM-type Zinc fingers were critical in maintaining the complex integrity. For our target, we designed guides against the first MYM-type zinc finger because it is the initial functional domain in the protein and corresponds to an early exon. Additionally, this domain includes a conserved cysteine repeat known as TRASH (Trafficking, Resistance, And Sensing Heavy metal). This repeat is important in metal coordination within the enzyme complex. Lastly, ZMYM2 is downstream SUMOlyation target of the SUMO-1. As we have already seen, disruption of SAE1 and the SUMO1 complex is a pathway of interest in leukemia. There may be a potential cooperative disruption of SUMO1 and FIM that results in our phenotype of interest.

3.2: Guide RNA Library Construction and Lentiviral Production

Following a detailed examination of each gene, the next step in our model was the production of custom guide RNAs for CRISPR-based mutagenesis. The genomic sequences corresponding to our selected coding exons (see supplementary table S2) were analyzed using the CRISPR Design Tool (see materials and methods 2.2). For 17 of our 19 genes, the top 3 non-redundant guides were selected for each sequence input. The non-redundant stipulation was included because some guides produced double-strand breaks in the same location. To avoid this and ensure broad coverage of our input sequences, we selected 3 guides per gene that would introduce different breakpoints to our region of interest. For MGA and NSD2, 4 guides were selected for each because they are larger genes and contain a greater number of key functional domains. In total, we selected 59 guide RNAs covering 26 exons across 19 genes. All 59 guides were deemed "high quality" choices based on the design tool's computational algorithm (quality scores > 50%).

With a defined set of endogenous target sequences, our subsequent task was finding a way to deliver these guide RNAs into our target cell line. Lentiviral vector systems have been an effective method for introducing both Cas9 and the gRNA components of CRISPR into cell lines[29-33]. This technique requires the synthesis of duplex gRNA oligos and a gRNA expression vector that can be packaged and integrated into the target genome. Through a series of enzymatic digests, ligation reactions, and transformations (outlined in materials and methods 2.4), we were able to clone all 59 guide RNAs into the Yusa lentiviral gRNA expression vector (figure 5). For all plasmid DNA isolated from these ligation/transformation reactions, two steps were taken to confirm successful gRNA insertion into the Yusa plasmid. The first was a BbsI/NcoI diagnostic double-digest. If a gRNA was successfully cloned into the Yusa expression vector, the BbsI cut

site was lost and the digested DNA appeared as a single band at 8100 BP (see materials and methods 2.4). For DNA samples that exhibited this band pattern, partial plasmid sequencing was performed at the gRNA adapter insertion site. The presence of a U6 promoter upstream of the insertion site and gRNA scaffold provided an ideal region for sequencing primer use. The LKO.1 5' U6 primer (5'-GACTATCATATGCTTACCGT-3') designed by the Weinberg Lab at MIT was used to sequence from the U6 promoter at the 5' end to the gRNA scaffold at the 3' end of the Yusa expression vector. This two-step check was performed on all minipreps until it was confirmed that we had successfully cloned all 59 custom gRNAs.

A looming question in the study of ETV6-RUNX1 ALL is whether individual secondary mutations are capable of producing the disease phenotype or if several cooperative mutations are needed. As a first pass, we decided to use a library approach to target our cell line. This method requires pooling the 59 cloned guides into a single large prep for lentiviral transduction (see materials and methods 2.6). In doing this, we intend to introduce all 59 guides at once and allow selection to take over. In addition to the pooled library, maxipreps of the 4 guide RNAs targeting the MGA gene were performed in parallel. This gene is of particular interest to our lab as it relates to TEL-AML1 leukemia and transcriptional regulation in hematopoiesis. As was mentioned in the target approach to MGA (figure 20), this transcription factor localizes to MYC-specific Enhancer Box DNA response elements and prevents c-MYC transcriptional activation. Dysregulated c-MYC is a hallmark of many solid and hematologic malignancies and may have a role in TEL-AML1^{+} ALL as a result of somatic mutations to the MGA gene.

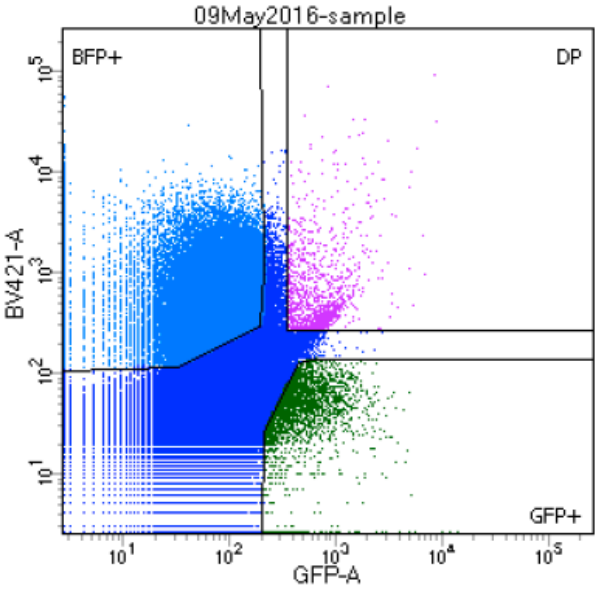
With large preps of our gRNA library, 3 Cas9 vectors (materials and methods 2.5), and 4 MGA gRNA transfer vectors, our next step was transfection of HEK 293T cells to produce 8 lentiviral supernatants (see materials and methods 2.7). To measure our titers, we infected NIH

3T3 fibroblasts at three different dilutions and determined the infection efficiency. FACS plots of our viral titers are included in the supplementary figures and tables.

3.3: Proof-of-Principle Experiment

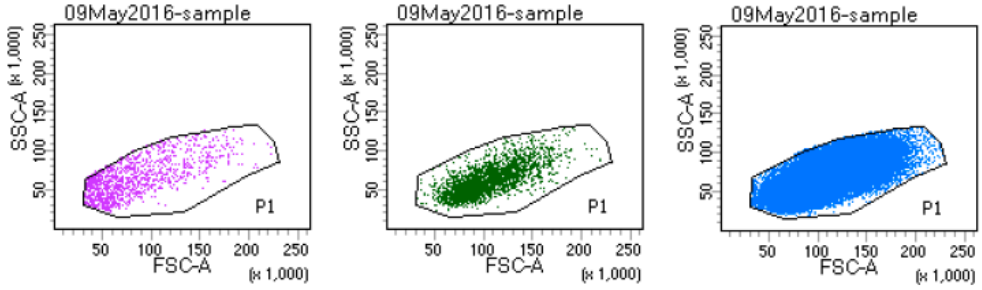
To optimize the infection protocol for our target TEL-AML1^{+} cell line and evaluate our system's efficiency, we conducted a proof-of-principle experiment using a wild-type Pro-B cell line ("Clone 8"), our high-titer MGA lentiviral vectors and the Feng Zhang GFP LentiCas9 vector (see materials and methods 2.8). First, we pooled the viral supernatants for MGA gRNA #1 (95.5% transduction efficiency), MGA gRNA #3 (94.5% transduction efficiency), and MGA gRNA #4 (85.4% transduction efficiency). Next, we coinfected our Pro-B cell suspension culture with the pooled MGA supernatants and the Feng Zhang Cas9 GFP supernatant (64.3% transduction efficiency). After two days, we sorted the viable Pro-B cells using FACS into three populations: GFP^{+}BFP^{+}, GFP^{+} only, and BFP^{+} only. Figure 29 is the FACS plot from this sorting experiment and contains the percent of total for each population as well as the absolute number of cells sorted. Of particular interest to us were the GFP^{+}BFP^{+} and GFP^{+} only populations. GFP^{+}BFP^{+} cells are ones that have been successfully coinfected with both Cas9 and one or multiple of the MGA gRNAs. The GFP^{+} cells only incorporated the Zhang Cas9 vector into their genome but are able to constitutively express the enzyme. These cells could be useful in future experiments that require a Cas9-expressing target cell line. Both of these populations were seeded onto fresh ST2 stromal cells and are currently being expanded in vitro.

Figure 29. Proof-of-Principle Infection Results.



Tube: sample

Population	#Events	%Parent	%Total
All Events	1,010,000	####	100.0
P1	918,291	90.9	90.9
P2	895,962	97.6	88.7
P3	895,414	99.9	88.7
DP	1,201	0.1	0.1
GFP+	2,915	0.3	0.3
BFP+	80,233	9.0	7.9



Double Positive cells expressing both Cas9 and MGA gRNAs

20,000 cells sorted

GFP$^+$ cells expressing Cas9 only

50,000 cells sorted

BFP$^+$ cells expressing only MGA gRNAs

1 million cells sorted

Chapter 4: Discussion

4.1: Next Steps

We will continue to expand the GFP^+BFP^+ and GFP^+ only populations from our proof-of-principle experiment. For the GFP^+BFP^+ cells, we are interested in seeing if our gRNAs + Cas9 produced indel mutations within the MGA gene. To do this, we will isolate a portion of the culture and first determine which MGA guide RNA(s) have been integrated into the cells via a PCR primer pair developed against the original Yusa Guide RNA Plasmid. Since all of the guide RNAs were cloned into the same expression vector, only one primer pair is needed and based on sequence results, we will be able to determine which guide(s) are present in our selected cells. We have already created this gRNA primer pair in preparation for this step:

- Forward Primer (5'-CAGGGACAGCAGAGATCCAG-3')
- Reverse Primer (5'-CCGGTGGATGTGGAATGTGT-3')

Figure 30 depicts the region of the integrated Yusa Plasmid that we will be analyzing by PCR

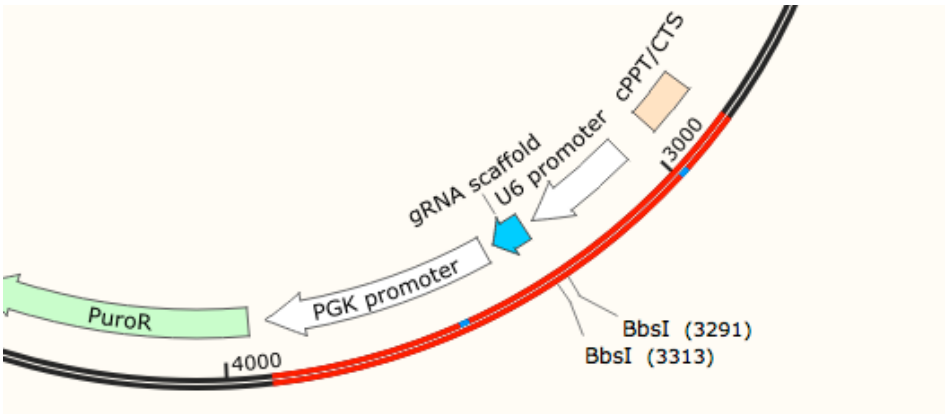


Figure 30. gRNA PCR Primer Pair. *Using the primer pair we designed against the Yusa Guide Plasmid, we will be able to sequence our selected cells for the region shown in red above. This will allow us to determine which of the MGA gRNAs were successfully incorporated into the genome of our Pro-B cell line.*

Once we have determined which MGA guides are being expressed, we will conduct deep sequencing of the genomic DNA corresponding to our target exons (exons #2 and #20) for the MGA gene. For this, we will use custom PCR primers that amplify the predicted cut site for each gRNA. Supplementary Table S3 includes the PCR primer pairs that we developed for all 59 guide RNA target sites included in our library. Primer pairs were designed based on guidelines set forth by the Mass Geneneral Hospital CCIB DNA core for CRISPR Sequencing. The two major restrictions were that amplicons must be between 200-280 BP and our site of interest (cut site) must be within 100 BP from either the 5' or 3' end of the amplicon to ensure high throughput results. In addition to Supplementary Table S3, visual depictions of our primer pairs and amplicons have been included in the supplementary figures and tables.

If we are able to demonstrate that Cas9 and our custom MGA gRNAs introduced indels within the MGA gene, we will then put our CRISPR library to the test. For this, we will complete two separate infections. One will be a coinfection of "Clone 8" Pro-B cells with our CRISPR library and the Ebert RFP Cas9 vector (see materials and methods 2.5). The other will be a Library infection of the GFP^+ "Clone 8" Pro-B cells from our proof-of-principle experiment, which constitutively express Cas9. There are two advantages to this experimental design. The first is we will be able to gauge if there is a significant difference between coinfection of separate Cas9 + gRNA vectors versus gRNA vector infection into cells already expressing Cas9. The second is that we will be able to compare the efficiency of two separate Cas9 vectors (Ebert RFP Cas9 vs. Feng Zhang Lenticas9 eGFP). Figure 31 details our experimental design.

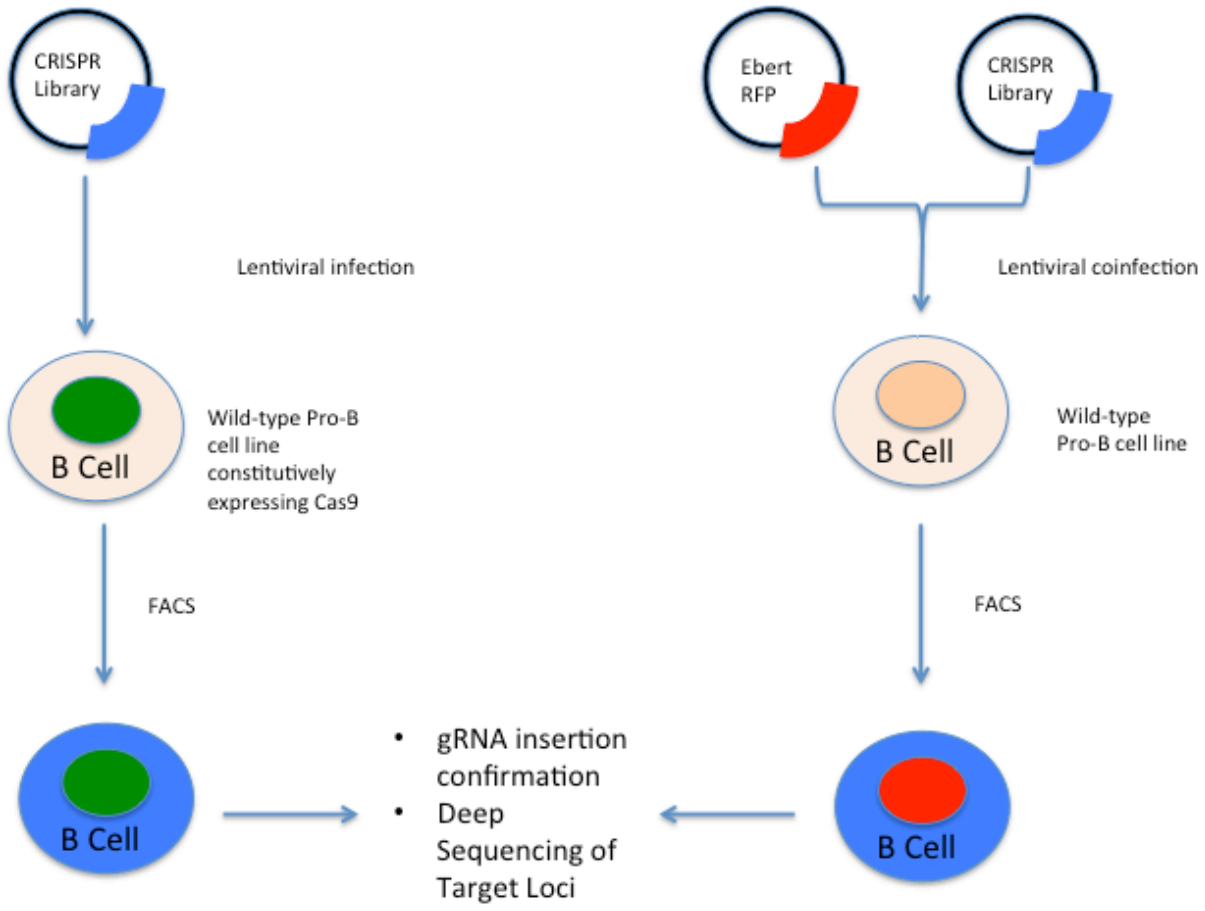


Figure 31. Pro-B cell line Infection and Genetic Screening.

A few days after infection, we will sort for BFP^+RFP^+ and BFP^+GFP^+ populations, as these will correspond to cells that include our library and a Cas9 vector. We will then follow the steps outlined above including determining which guides have been incorporated into our selected population and what, if any, mutations have occurred at our endogenous target sequences. To do this, we will use the primer pairs in supplementary table S3, which cover the entire library. With assistance from the MGH Cancer Center bioinformatics core, we will look to see if certain genes and mutations have a selection bias.

If this step is successful, we will then introduce our CRISPR Library into a TEL-$AML1^+$ pro-B cell line to recapitulate the actually secondary events of ETV6-RUNX1 B-precursor Acute Lymphoblastic Leukemia. In addition to deep sequencing analysis, we will monitor our cultures for mutants that exhibit malignant transformation. Using transformation assays such as focus forming assay and growth factor removal, we will select for malignant cells and conduct deep sequencing to determine if a specific mutation or combination of mutations have driven our pre-leukemic clones to neoplasia. The next step would be to move to an in vivo transplantation model where we test our malignant clones in mice and determine if they produce the disease phenotype.

4.2: Limitations

With CRISPR-Cas9 genome-editing technology there are certainly limitations. As was noted earlier, Cas9 endonuclease activity can have off-target effects. When we conduct our deep sequencing, we will have to be cognizant of potential cuts at loci other than the ones we intended to modify. If we find that off-target cuts are overabundant and are confounding our results, we could potentially improve our experimental design by incorporating the recently developed High-fidelity Cas9 nuclease[94]. Engineered by the Joung Lab at MGH, the SpCas9-HF1 is a modified version of the wild-type SpCas9. Through studies of Cas9 energetics, Kleinstiver et al. determined that four non-specific contact points on the enzyme were contributing to off-target cuts. 4 amino acid substitutions were made at these contact points and the result was a modified version of the enzyme that produced virtually no off-target cuts. Another limitation to our model is the fact that ex vivo culture does not capture the complexity of the bone marrow niche. During leukemogenesis, the dynamic stromal environment in the bone marrow plays a key role in disease pathogenesis. Overcoming this limitation will require moving from ex vivo cell culture to in vivo transplantation.

4.3 Conclusions

In this work, we utilized CRISPR-Cas9 genome-editing technology to explore the mechanism behind ETV6-RUNX1 B-precursor acute lymphoblastic leukemia. Using the current knowledge of this disease, we developed a guide RNA library against 19 genes that have been implicated as potential secondary hits capable of transforming TEL-AML1^{+} preleukemic cells.. The genes that compose this library are involved in a variety of cellular processes; some specific to B-cell development while others are expressed across cell types. Many of these genes are involved in overlapping pathways such as chromatin modifications, cell differentiation, proliferation and transcriptional regulation. As the literature suggests, it may be that a single gene mutation or multiple gene mutations are necessary and sufficient for disease pathogenesis. In contrast to a generic screen approach, significant time was spent honing in on the existing literature and targeting genes that have recurrently appeared mutated in pre-B ALL and other hematologic malignancies. At the heart of this project was the individual attention given to each gene. With the design and production of our lentiviral library, we can now infect progenitor B cells ex vivo and begin to answer some of the looming questions that remain regarding the molecular mechanism of this type of leukemia. If we can definitely identify genes that are susceptible following the formation of the TEL-AML1 fusion gene, novel therapies can be developed to prevent the progression of this disease in children who have the inciting lesion. Learning more about the mechanism of this disease will also be helpful in approaching other leukemias that result from genomic rearrangements such as BCR-ABL and E2A-PBX1. Especially if CRISPR technology can be harnessed for gene therapy, we could effectively repair rearranged DNA with the wild-type sequences. This will take improvements in both the

technology and our understanding of the molecular events that cause genomic instability and chromosome translocations.

Chapter 5: Bibliography

1. Cancer Facts & Figures 2015. Atlanta, GA: American Cancer Society; 2015.
2. Kajikhina K, Tsuneto M, Melchers F. Environments of hematopoiesis and B-lymphopoiesis in fetal liver. *Clin Exp Rheumatol*. 2015 Jul-Aug; **33**(4 Suppl 92): S91-3.
3. Jagannathan-Bogdan M, Zon LI. Hematopoiesis. *Development.* 2013 Jun; **140** (12): 2463-7.
4. Anthony BA, Link DC. Regulation of hematopoietic stem cells by bone marrow stromal cells. *Trends Immunol*. 2014 Jan; **35** (1): 32-7.
5. Marciniak-Czochra A, Stiehl T, Ho AD, Jäger W, Wagner W. Modeling of asymmetric cell division in hematopoietic stem cells--regulation of self-renewal is essential for efficient repopulation. *Stem Cells Dev*. 2009 Apr; **18** (3): 377-85.
6. Zhu J, Emerson SG. Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene*. 2002 May 13; **21** (21): 3295-313.
7. Hagman J, Lukin K. Transcription factors drive B cell development. *Curr Opin Immunol*. 2006 Apr; **18** (2): 127-34.
8. Pelayo R, Dorantes-Acosta E, Vadillo E, *et al*. From HSC to B-lymphoid cells in normal and malignant hematopoiesis. In: Pelayo R, ed. *Advances in Hematopoietic Stem Cell Research*. 2011.
9. Purizaca J, Meza I, Pelayo R. Early lymphoid development and microenvironmental cues in B-cell acute lymphoblastic leukemia. *Arch Med Res*. 2012 Feb; **43** (2): 89-101.
10. Heerema NA. *et al*. Prognostic impact of trisomies of chromosomes 10, 17, and 5 among children with acute lymphoblastic leukemia and high hyperdiploidy (> 50 chromosomes). *J Clin Oncol*. 2000 May; **18** (9): 1876-87.
11. Zuckerman T, Rowe JM. Pathogenesis and prognostication in acute lymphoblastic leukemia. *F1000Prime Rep*. 2014 Jul 8; **6**: 59.
12. Nambiar M, Kari V, Raghavan SC. Chromosomal translocations in cancer. *Biochim Biophys Acta*. 2008 Dec; **1786** (2): 139-52.
13. Awan T. *et al*. Five most common prognostically important fusion oncogenes are detected in the majority of Pakistani pediatric acute lymphoblastic leukemia patients and are strongly associated with disease biology and treatment outcome. *Asian Pac J Cancer Prev*. 2012; **13** (11): 5469-75.
14. Schindler JW, Van Buren D, Foudi A, Krejci O, Qin J, Orkin SH, Hock H. TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell*. 2009 Jul 2; **5** (1): 43-53.
15. Fischer M. *et al*. Defining the oncogenic function of the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model. *Oncogene*. 2005 Nov 17; **24** (51): 7579-91.
16. Zelent A, Greaves M, Enver T. Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia. *Oncogene*. 2004 May 24; **23** (24): 4275-83.
17. Zhang J. *et al*. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood*. 2011 Sep 15; **118** (11): 3080-7.
18. Bohlander SK. ETV6: a versatile player in leukemogenesis. *Semin Cancer Biol*. 2005 Jun; **15** (3): 162-74.

19. Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR, Orkin SH. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J*. 1997 Jul 16; **16** (14): 4374-83.
20. Wang LC. *et al*. The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev*. 1998 Aug 1; **12** (15): 2392-402.
21. Hock H, Meade E, Medeiros S, Schindler JW, Valk PJ, Fujiwara Y, Orkin SH. Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev*. 2004 Oct 1; **18** (19): 2336-41.
22. Blyth K, Cameron ER, Neil JC. The RUNX genes: gain or loss of function in cancer. *Nat Rev Cancer*. 2005 May; **5** (5): 376-87.
23. Okuda T, Nishimura M, Nakao M, Fujita Y. RUNX1/AML1: a central player in hematopoiesis. *Int J Hematol*. 2001 Oct; **74** (3): 252-7.
24. Hiebert, S. W. *et al*. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Mol. Cell. Biol*. **16**, 1349–1355 (1996).
25. Guidez, F. *et al*. Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL–AML1 oncoprotein. *Blood*. 2000. **96**, 2557–2561
26. Hemmeryckx B, Reichert A, Watanabe M, Kaartinen V, de Jong R, Pattengale PK, Groffen J, Heisterkamp N. BCR/ABL P190 transgenic mice develop leukemia in the absence of Crkl. *Oncogene*. 2002 May 9; **21** (20): 3225-31.
27. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. *Blood*. 2003 Oct 1; **102** (7): 2321-33.
28. Papaemmanuil E. *et al*. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet*. 2014 Feb; **46** (2): 116-25.
29. Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet*. 2015 May; **16** (5): 299-311.
30. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013 Nov; **8** (11): 2281-308.
31. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera M del C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol*. 2014 Mar; **32** (3): 267-73.
32. Heckl D. *et al*. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat Biotechnol*. 2014 Sep; **32** (9): 941-6.
33. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ,Weissleder R, Lee H, Zhang F, Sharp PA. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*. 2015 Mar 12; **160** (6): 1246-60.
34. Dang J. *et al*. PAX5 is a tumor suppressor in mouse mutagenesis models of acute lymphoblastic leukemia. *Blood*. 2015 Jun 4; **125** (23): 3609-17.
35. Holmes ML, Pridans C, Nutt SL. The regulation of the B-cell gene expression programme by Pax5. *Immunol Cell Biol*. 2008 Jan; **86** (1): 47-53.
36. Li, Jian Yi et al. TBL1XR1 in Physiological and Pathological States. *American Journal of Clinical and Experimental Urology* **3.1** (2015): 13–23.

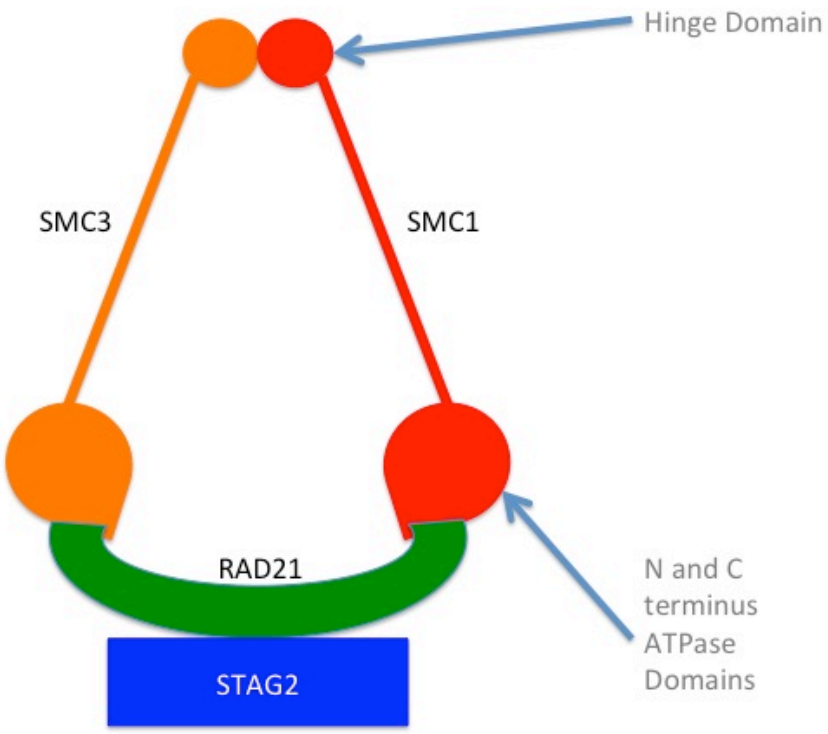
37. Jones CL. *et al*. Loss of TBL1XR1 disrupts glucocorticoid receptor recruitment to chromatin and results in glucocorticoid resistance in a B-lymphoblastic leukemia model. *J Biol Chem*. 2014 Jul 25; **289** (30): 20502-15.
38. Parker, H. *et al*. The complex genomic profile of ETV6-RUNX1 positive acute lymphoblastic leukemia highlights a recurrent deletion of TBL1XR1. *Genes Chromosomes Cancer*. 2008. **47**: 1118–1125.
39. Zhang XM, Chang Q, Zeng L, Gu J, Brown S, Basch RS. TBLR1 regulates the expression of nuclear hormone receptor co-repressors. *BMC Cell Biol*. 2006 Aug 7; **7**: 31.
40. Rouault JP. *et al*. BTG1, a member of a new family of antiproliferative genes. *EMBO J*. 1992 Apr; **11** (4): 1663-70.
41. Prévôt D, Voeltzel T, Birot AM, Morel AP, Rostan MC, Magaud JP, Corbo L. The leukemia-associated protein Btg1 and the p53-regulated protein Btg2 interact with the homeoprotein Hoxb9 and enhance its transcriptional activation. *J Biol Chem*. 2000 Jan 7; **275** (1): 147-53.
42. Tsuzuki S. *et al*. Genetic abnormalities involved in t(12;21) TEL-AML1 acute lymphoblastic leukemia: analysis by means of array-based comparative genomic hybridization. *Cancer Sci*. 2007 May; **98** (5): 698-706.
43. Waanders E. *et al*. The origin and nature of tightly clustered BTG1 deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. *PLoS Genet*. 2012; **8** (2): e1002533.
44. Kempski HM, Sturt NT. The TEL-AML1 fusion accompanied by loss of the untranslocated TEL allele in B-precursor acute lymphoblastic leukaemia of childhood. *Leuk Lymphoma*. 2000 Dec; **40** (1-2): 39-47.
45. Noetzli L. *et al*. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet*. 2015 May; **47** (5): 535-8.
46. Kuiper RP. *et al*. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia*. 2007 Jun; **21**(6): 1258-66.
47. Shimazaki N, Tsai AG, Lieber MR. H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. *Mol Cell*. 2009 Jun 12; **34** (5): 535-44.
48. Terker AS, Ellison DH. Renal mineralocorticoid receptor and electrolyte homeostasis. *Am J Physiol Regul Integr Comp Physiol*. 2015 Nov 1; **309** (9): R1068-70.
49. Hahn WC, Weinberg RA. Modeling the molecular circuitry of cancer. *Nat Rev Cancer*. 2002 May; **2** (5): 331-41.
50. Walsh KM. *et al*. A Heritable Missense Polymorphism in CDKN2A Confers Strong Risk of Childhood Acute Lymphoblastic Leukemia and Is Preferentially Selected during Clonal Evolution. *Cancer Res*. 2015 Nov 15; **75** (22): 4884-94.
51. Hannon GJ, Beach D. p15INK4B is a potential effector of TGF-beta-induced cell cycle arrest. *Nature*. 1994 Sep 15; **371** (6494): 257-61.
52. Takeuchi S. *et al*. Analysis of a family of cyclin-dependent kinase inhibitors: p15/MTS2/INK4B, p16/MTS1/INK4A, and p18 genes in acute lymphoblastic leukemia of childhood. *Blood*. 1995 Jul 15; **86** (2): 755-60.
53. Okuda T. *et al*. Frequent deletion of p16INK4a/MTS1 and p15INK4b/MTS2 in pediatric acute lymphoblastic leukemia. *Blood*. 1995 May 1; **85** (9): 2321-30.

54. Guney S. *et al*. Molecular characterization of 9p21 deletions shows a minimal common deleted region removing CDKN2A exon 1 and CDKN2B exon 2 in diffuse large B-cell lymphomas. *Genes Chromosomes Cancer*. 2011 Sep; **50** (9): 715-25.
55. Watanabe N. *et al*. BTLA is a lymphocyte inhibitory receptor with similarities to CTLA-4 and PD-1. *Nat Immunol*. 2003 Jul; **4** (7): 670-9.
56. Ghazavi F. *et al*. CD200/BTLA deletions in pediatric precursor B-cell acute lymphoblastic leukemia treated according to the EORTC-CLG 58951 protocol. *Haematologica*. 2015 Oct; **100** (10): 1311-9.
57. Wang H. *et al*. mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol Cell*. 2003 Aug; **12** (2): 475-87.
58. Kuang SQ. *et al*. Epigenetic inactivation of Notch-Hes pathway in human B-cell acute lymphoblastic leukemia. *PLoS One*. 2013 Apr 26; **8** (4): e61807.
59. Kobayashi K. *et al*. ATF7IP as a novel PDGFRB fusion partner in acute lymphoblastic leukaemia in children. *Br J Haematol*. 2014 Jun; **165** (6): 836-41.
60. Jones S. An overview of the basic helix-loop-helix proteins. *Genome Biol*. 2004; **5** (6): 226.
61. Grandori C, Cowley SM, James LP, Eisenman RN. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol*. 2000; **16**: 653-99.
62. Hurlin PJ, Steingrìmsson E, Copeland NG, Jenkins NA, Eisenman RN. Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *EMBO J*. 1999 Dec 15; **18** (24): 7019-28.
63. de Alboran IM. *et al*. Analysis of C-MYC function in normal cells via conditional gene-targeted mutation. *Immunity*. 2001 Jan; **14** (1): 45-55.
64. Washkowitz AJ, Schall C, Zhang K, Wurst W, Floss T, Mager J, Papaioannou VE. Mga is essential for the survival of pluripotent cells during peri-implantation development. *Development*. 2015 Jan 1; **142** (1): 31-40.
65. De Paoli L. *et al*. MGA, a suppressor of MYC, is recurrently inactivated in high-risk chronic lymphocytic leukemia. *Leuk Lymphoma*. 2013 May; **54** (5): 1087-90.
66. Peters JM, Tedeschi A, Schmitz J. The Cohesin complex and its roles in chromosome biology. *Genes Dev*. 2008 Nov 15; **22** (22): 3089-114.
67. Horsfield JA. *et al*. Cohesin-dependent regulation of Runx genes. *Development*. 2007 Jul; **134** (14): 2639-49.
68. Solomon DA. *et al*. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*. 2011 Aug 19; **333** (6045): 1039-43.
69. Thota S. *et al*. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*. 2014 Sep 11; **124** (11): 1790-8.
70. Losada A. Cohesin in cancer: chromosome segregation and beyond. *Nat Rev Cancer*. 2014 Jun; **14** (6): 389-93.
71. Kitagawa R, Bakkenist CJ, McKinnon PJ, Kastan MB. Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. *Genes Dev*. 2004 Jun 15; **18** (12): 1423-38.
72. Huether R. *et al*. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun*. 2014 Apr 8; **5**:3630.

73. Potts PR, Yu H. The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nat Struct Mol Biol*. 2007 Jul; **14** (7): 581-90.
74. De Piccoli G, Torres-Rosell J, Aragón L. The unnamed complex: what do we know about Smc5-Smc6? *Chromosome Res*. 2009; **17** (2): 251-63.
75. Zabrady K. *et al*. Chromatin association of the SMC5/6 complex is dependent on binding of its NSE3 subunit to DNA. *Nucleic Acids Res*. 2016 Feb 18; **44** (3): 1064-79.
76. Stephan AK, Kliszczak M, Morrison CG. The Nse2/Mms21 SUMO ligase of the Smc5/6 complex in the maintenance of genome stability. *FEBS Lett*. 2011 Sep 16; **585** (18): 2907-13.
77. Staffas A, Karlsson C, Persson M, Palmqvist L, Bergo MO. Wild-type KRAS inhibits oncogenic KRAS-induced T-ALL in mice. *Leukemia*. 2015 May; **29** (5): 1032-40.
78. Xiao H. *et al*. Mutations in epigenetic regulators are involved in acute lymphoblastic leukemia relapse following allogeneic hematopoietic cell transplantation. *Oncotarget*. 2015.
79. Chen Y. *et al*. Kras Is Critical for B Cell Lymphopoiesis. *J Immunol*. 2016 Feb 15; **196** (4): 1678-85.
80. Fotiadou PP, Takahashi C, Rajabi HN, Ewen ME. Wild-type NRas and KRas perform distinct functions during transformation. *Mol Cell Biol*. 2007 Oct; **27** (19): 6742-55.
81. Nonami A. *et al*. Identification of novel therapeutic targets in acute leukemias with NRAS mutations using a pharmacologic approach. *Blood*. 2015 May 14; **125** (20): 3133-43.
82. Dunna NR. *et al*. NRAS mutations in de novo acute leukemia: prevalence and clinical significance. *Indian J Biochem Biophys*. 2014 Jun; **51** (3): 207-10.
83. Paulsson K. *et al*. Mutations of FLT3, NRAS, KRAS, and PTPN11 are frequent and possibly mutually exclusive in high hyperdiploid childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer*. 2008 Jan; **47** (1): 26-33.
84. Eifler K, Vertegaal AC. SUMOylation-Mediated Regulation of Cell Cycle Progression and Cancer. *Trends Biochem Sci*. 2015 Dec; **40** (12): 779-93.
85. Mattoscio D, Chiocca S. SUMO pathway components as possible cancer biomarkers. *Future Oncol*. 2015; **11** (11): 1599-610.
86. Luo J. *et al*. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell*. 2009 May 29; **137** (5): 835-48.
87. Kim JY. *et al*. Multiple-myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity. *Mol Cell Biol*. 2008. Mar; **28** (6): 2023-34.
88. Yang P. *et al*. Histone methyltransferase NSD2/MMSET mediates constitutive NF-κB signaling for cancer cell proliferation, survival, and tumor growth via a feed-forward loop. *Mol Cell Biol*. 2012 Aug; **32** (15): 3121-31.
89. Kuo AJ. *et al*. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol Cell*. 2011 Nov 18; **44** (4): 609-20.
90. Jaffe JD. *et al*. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet*. 2013 Nov; **45** (11): 1386-91.

91. Sankaran SM, Wilkinson AW, Gozani O. A PWWP domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated Lys36 of histone H3 and regulates NSD2 function at chromatin. *J Biol Chem*. 2016 Feb 24.
92. Inoue H, Nobuhisa I, Okita K, Takizawa M, Pébusque MJ, Taga T. Negative regulation of hematopoiesis by the fused in myeloproliferative disorders gene product. *Biochem Biophys Res Commun*. 2004 Jan 2; **313** (1): 125-8.
93. Gocke CB, Yu H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One*. 2008 Sep 22; **3** (9): e3255.
94. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016 Jan 28; **529** (7587): 490-5.

Chapter 6: Supplemental Figures and Tables



Supplemental Figure S1. Cohesin Complex.

Supplementary Table S1. Summary table including list of genes, functions, gRNA target sequences and oligonucleotide sequences

Gene	Site	Function	Endogenous target sequence (20 nt + PAM)	Position (GRCm38)	Top strand oligo (5' > 3')	Bottom strand oligo (5' > 3')
Pax5	Site 1	Transcription Factor - Early B Cell Development	AGCATCAAGCCGGGGGTGAT TGG	4: 44691994 to 44692016	CACCAGCATCAAGCCGGGGGTGATGT	TAAAACATCACCCCCGGCTTGATGCT
Pax5	Site 2	Transcription Factor - Early B Cell Development	TTGGATCCTCCAATCACCCC CGG	4: 44691985 to 44692007	CACCTTGGATCCTCCAATCACCCCGT	TAAAACGGGGTGATTGGAGGATCCAA
Pax5	Site 3	Transcription Factor - Early B Cell Development	ATCAAGCCGGGGGTGATTGG AGG	4: 44691991 to 44692013	CACCATCAAGCCGGGGGTGATTGGGT	TAAAACCCAATCACCCCCGGCTTGAT
Tbl1xr1	Site 1	Subunit of the NCoR and SMRT repressor complexes.	CCGAAGCACAACCGCTTTAT TGG	3: 22190963 to 22190985	CACCCCGAAGCACAACCGCTTTATGT	TAAAACATAAAGCGGTTGTGCTTCGG
Tbl1xr1	Site 2	Subunit of the NCoR and SMRT repressor complexes.	AATAAAGCGGTTGTGCTTCG GGG	3: 22190965 to 22190987	CACCAATAAAGCGGTTGTGCTTCGGT	TAAAACCGAAGCACAACCGCTTTATT
Tbl1xr1	Site 3	Subunit of the NCoR and SMRT repressor complexes.	GAACCCTGTTAGTGATCTCC TGG	3: 22191018 to 22191040	CACCGAACCCTGTTAGTGATCTCCGT	TAAAACGGAGATCACTAACAGGGTTC
Btg1	Site 1	Antiproliferative Gene during G0/G1 of Cell Cycle	ACTGCTCAGTCCAATCCGCT GGG	10: 96618378 to 96618400	CACCACTGCTCAGTCCAATCCGCTGT	TAAAACAGCGGATTGGACTGAGCAGT
Btg1	Site 2	Antiproliferative Gene during G0/G1 of Cell Cycle	CTGTACGAAGCCTCACCAGC AGG	10: 96618497 to 96618519	CACCCTGTACGAAGCCTCACCAGCGT	TAAAACGCTGGTGAGGCTTCGTACAG
Btg1	Site 3	Antiproliferative Gene during G0/G1 of Cell Cycle	GGACAGGCAGCCCAGCGGAT TGG	10: 96618368 to 96618390	CACCGGACAGGCAGCCCAGCGGATGT	TAAAACATCCGCTGGGCTGCCTGTCC
Etv6	Site 1	ETS TF required for Hematopoiesis, oncogene	CCAACGGACTGGCTCGACTC TGG	6:134261798 to 134261820	CACCCCAACGGACTGGCTCGACTCGT	TAAAACGAGTCGAGCCAGTCCGTTGG
Etv6	Site 2	ETS TF required for Hematopoiesis, oncogene	TTAGTTTGTAGTAGTGGCGC AGG	6: 134263057 to 134263079	CACCTTAGTTTGTAGTAGTGGCGCGT	TAAAACGCGCCACTACTACAAACTAA
Etv6	Site 3	ETS TF required for Hematopoiesis, oncogene	ACGCCTCCTGCGGTCCCCGT GGG	6: 134266400 to 134266422	CACCACGCCTCCTGCGGTCCCCGTGT	TAAAACACGGGGACCGCAGGAGGCGT
Rag2	Site 1	Responsible for VDJ Rearrangment in T and B cell development	AATATAAACGGTATCGTTTC TGG	2: 101629978 to 101630000	CACCAATATAAACGGTATCGTTTCGT	TAAAACGAAACGATACCGTTTATATT
Rag2	Site 2	Responsible for VDJ Rearrangment in T and B cell development	GATACCGTTTATATTTTGGG AGG	2: 101629986 to 101630008	CACCGATACCGTTTATATTTTGGGGT	TAAAACCCCAAAATATAAACGGTATC
Rag2	Site 3	Responsible for VDJ Rearrangment in T and B cell development	GACAGCCCATCCTGAAGTTC TGG	2: 101629938 to 101629960	CACCGACAGCCCATCCTGAAGTTCGT	TAAAACGAACTTCAGGATGGGCTGTC
Nr3c2	Site 1	Mineralocorticoid Receptor regulating hormone response elements	CTAAGTTCATGCCGGCTTGG AGG	8: 77155293 to 77155315	CACCCTAAGTTCATGCCGGCTTGGGT	TAAAACCCAAGCCGGCATGAACTTAG
Nr3c2	Site 2	Mineralocorticoid Receptor regulating hormone response elements	AGTCTAGAAGCTTCGTCAGT TGG	8: 77217552 to 77217574	CACCAGTCTAGAAGCTTCGTCAGTGT	TAAAACACTGACGAAGCTTCTAGACT
Nr3c2	Site 3	Mineralocorticoid Receptor regulating hormone response elements	TCATCTCCTCAAACGCAGCC TGG	8: 77217450 to 77217472	CACCTCATCTCCTCAAACGCAGCCGT	TAAAACGGCTGCGTTTGAGGAGATGA

Cdkn2a	Site 1	Encodes the Tumor Suppressors ARF and INK4a	GGGTCGCCTGCCGCTCGACT TGG	4: 89276784 to 89276806	CACCGGGTCGCCTGCCGCTCGACTGT	TAAAACAGTCGAGCGGCAGGCGACCC
Cdkn2a	Site 2	Encodes the Tumor Suppressors ARF and INK4a	CCCGCGCTGCGTCGTGCACC GGG	4: 89276874 to 89276896	CACCCCCGCGCTGCGTCGTGCACCGT	TAAAACGGTGCACGACGCAGCGCGGG
Cdkn2a	Site 3	Encodes the Tumor Suppressors ARF and INK4a	CGGTGCAGATTCGAACTGCG AGG	4: 89276916 to 89276938	CACCCGGTGCAGATTCGAACTGCGGT	TAAAACCGCAGTTCGAATCTGCACCG
Cdkn2b	Site 1	Encodes the Tumor Suppressor p15INK4b	CTTGTCGTGCTGCACCGGGC AGG	4: 89307163 to 89307185	CACCCTTGTCGTGCTGCACCGGGCGT	TAAAACGCCCGGTGCAGCACGACAAG
Cdkn2b	Site 2	Encodes the Tumor Suppressor p15INK4b	GCAGCACGACAAGCGTGTCC AGG	4: 89307173 to 89307195	CACCGCCCGGTGCAGCACGACAAGGT	TAAAACCTTGTCGTGCTGCACCGGGC
Cdkn2b	Site 3	Encodes the Tumor Suppressor p15INK4b	GACCTGTGCACGACGCAGCT CGG	4: 89307204 to 89307226	CACCGACCTGTGCACGACGCAGCTGT	TAAAACAGCTGCGTCGTGCACAGGTC
Btla	Site 1	Gene product Negatively Regulates T-Cell Responses	CCCAACTAGTGTATAGCTGA GGG	16: 45629595 to 45629616	CACCCCCAACTAGTGTATAGCTGAGT	TAAAACTCAGCTATACACTAGTTGGG
Btla	Site 2	Gene product Negatively Regulates T-Cell Responses	GCCCTCAGCTATACACTAGT TGG	16: 45629595 to 45629615	CACCGCCCTCAGCTATACACTAGTGT	TAAAACACTAGTGTATAGCTGAGGGC
Btla	Site 3	Gene product Negatively Regulates T-Cell Responses	AACAATCTGTGTACCCCTTG AGG	16: 45239180 to 45239202	CACCAACAATCTGTGTACCCCTTGGT	TAAAACCAAGGGGTACACAGATTGTT
Atf7ip	Site 1	Nuclear Protein that associates with Heterochromatin	TGAACCGGTCGCAAGGGTCG TGG	6: 136560550 to 136560572	CACCTGAACCGGTCGCAAGGGTCGGT	TAAAACCGACCCTTGCGACCGGTTCA
Atf7ip	Site 2	Nuclear Protein that associates with Heterochromatin	GCCGCCACGACCCTTGCGAC CGG	6: 136560554 to 136560576	CACCGCCGCCACGACCCTTGCGACGT	TAAAACGTCGCAAGGGTCGTGGCGGC
Atf7ip	Site 3	Nuclear Protein that associates with Heterochromatin	AGAAACTGGCTCGCTACAAG TGG	6: 136560465 to 136560487	CACCAGAAACTGGCTCGCTACAAGGT	TAAAACCTTGTAGCGAGCCAGTTTCT
Mga	Site 1	Dual-specificity TF regulating MAX network and T-box genes	CTATGCATCGTTATCTGCCA AGG	2: 119903259 to 119903281	CACCCTATGCATCGTTATCTGCCAGT	TAAAACTGGCAGATAACGATGCATAG
Mga	Site 2	Dual-specificity TF regulating MAX network and T-box genes	AGGCTGAACCCCATATTTTG GGG	2: 119903112 to 119903134	CACCAGGCTGAACCCCATATTTTGGT	TAAAACCAAAATATGGGGTTCAGCCT
Mga	Site 3	Dual-specificity TF regulating MAX network and T-box genes	ACCACGCCGACGCCGCTCAT TGG	2: 119958254 to 119958276	CACCACCACGCCGACGCCGCTCATGT	TAAAACATGAGCGGCGTCGGCGTGGT
Mga	Site 4	Dual-specificity TF regulating MAX network and T-box genes	GCCAATGAGCGGCGTCGGCG TGG	2: 119958253 to 119958275	CACCGCCAATGAGCGGCGTCGGCGGT	TAAAACCGCCGACGCCGCTCATTGGC
Stag2	Site 1	Subunit of the Cohesin Complex	ATTTCGACATACAAGCACCC TGG	X: 42223425 to 42223447	CACCATTTCGACATACAAGCACCCGT	TAAAACGGGTGCTTGTATGTCGAAAT
Stag2	Site 2	Subunit of the Cohesin Complex	GATTACCCACTTACCATGGC TGG	X: 42223255 to 42223277	CACCGATTACCCACTTACCATGGCGT	TAAAACGCCATGGTAAGTGGGTAATC
Stag2	Site 3	Subunit of the Cohesin Complex	GAGGACCAGCCATGGTAAGT GGG	X: 42223260 to 42223282	CACCGAGGACCAGCCATGGTAAGTGT	TAAAACACTTACCATGGCTGGTCCTC

Supplementary Table S1 (continued)

Smc1a	Site 1	Subunit of the Cohesin Complex	CAGTATATCAAGGAGCAACG TGG	X: 152033825 to 152033847	CACCCAGTATATCAAGGAGCAACGGT	TAAAACCGTTGCTCCTTGATATACTG
Smc1a	Site 2	Subunit of the Cohesin Complex	GATACTTCTTTTGTGTAGGC TGG	X: 152033722 to 152033744	CACCGATACTTCTTTTGTGTAGGCGT	TAAAACGCCTACACAAAAGAAGTATC
Smc1a	Site 3	Subunit of the Cohesin Complex	CTTCTTGCCTCTTGACTACC TGG	X: 152033860 to 152033882	CACCCTTCTTGCCTCTTGACTACCGT	TAAAACGGTAGTCAAGAGGCAAGAAG
Smc5	Site 1	Involved in DNA repair and recruitment of Cohesin Complex	TCCTATGACACGGCGGATTG AGG	19: 23257535 to 23257557	CACCTCCTATGACACGGCGGATTGGT	TAAAACCAATCCGCCGTGTCATAGGA
Smc5	Site 2	Involved in DNA repair and recruitment of Cohesin Complex	ACTAATTCGTGACCGAGTGA AGG	19: 23257589 to 23257611	CACCACTAATTCGTGACCGAGTGAGT	TAAAACTCACTCGGTCACGAATTAGT
Smc5	Site 3	Involved in DNA repair and recruitment of Cohesin Complex	TCCTCAATCCGCCGTGTCAT AGG	19: 23257534 to 23257556	CACCTCCTCAATCCGCCGTGTCATGT	TAAAACATGACACGGCGGATTGAGGA
Kras	Site 1	small GTPase, tissue signaling, oncogene	TAGAACAGTAGACACGAAAC AGG	6: 145232157 to 145232179	CACCTAGAACAGTAGACACGAAACGT	TAAAACGTTTCGTGTCTACTGTTCTA
Kras	Site 2	small GTPase, tissue signaling, oncogene	TGAAGATGTGCCTATGGTCC TGG	6: 145232205 to 145232227	CACCTGAAGATGTGCCTATGGTCCGT	TAAAACGGACCATAGGCACATCTTCA
Kras	Site 3	small GTPase, tissue signaling, oncogene	ATGTGCCTATGGTCCTGGTA GGG	6: 145232200 to 145232222	CACCATGTGCCTATGGTCCTGGTAGT	TAAAACTACCAGGACCATAGGCACAT
Nras	Site 1	small GTPase, tissue signaling, oncogene	CTTCGCCTGTCCTCATGTAC TGG	3: 103060296 to 103060318	CACCCTTCGCCTGTCCTCATGTACGT	TAAAACGTACATGAGGACAGGCGAAG
Nras	Site 2	small GTPase, tissue signaling, oncogene	GACCTGCCTGCTGGACATAC TGG	3: 103060236 to 103060258	CACCGACCTGCCTGCTGGACATACGT	TAAAACGTATGTCCAGCAGGCAGGTC
Nras	Site 3	small GTPase, tissue signaling, oncogene	AGTACATGAGGACAGGCGAA GGG	3: 103060298 to 103060320	CACCAGTACATGAGGACAGGCGAAGT	TAAAACTTCGCCTGTCCTCATGTACT
Sae1	Site 1	Subunit of SUMO-activating enzyme	GCCAAGAATCTTATCCTGGC AGG	7: 16378399 to 16378421	CACCGCCAAGAATCTTATCCTGGCGT	TAAAACGCCAGGATAAGATTCTTGGC
Sae1	Site 2	Subunit of SUMO-activating enzyme	AGTCAAAGGGCTCACCATGC TGG	7: 16378376 to 16378398	CACCAGTCAAAGGGCTCACCATGCGT	TAAAACGCATGGTGAGCCCTTTGACT
Sae1	Site 3	Subunit of SUMO-activating enzyme	TCATGCCAACAATCAGCACC CGG	7: 16378442 to 16378464	CACCTCATGCCAACAATCAGCACCGT	TAAAACGGTGCTGATTGTTGGCATGA
Nsd2	Site 1	Histone methyltransferase - HMG Box domain	TTTCATTGAGCATGCTCCAC TGG	5: 33861090 to 33861112	CACCTTTCATTGAGCATGCTCCACGT	TAAAACGTGGAGCATGCTCAATGAAA
Nsd2	Site 2	Histone methyltransferase - PWWP domain	AGATTATTACTGGACGCATC AGG	5: 33882901 to 33882923	CACCAGATTATTACTGGACGCATCGT	TAAAACGATGCGTCCAGTAATAATCT
Nsd2	Site 3	Histone methyltransferase - SET domain	TCAGGGTCTCACAATTGGGC TGG	5: 33891565 to 33891587	CACCTCAGGGTCTCACAATTGGGCGT	TAAAACGCCCAATTGTGAGACCCTGA
Nsd2	Site 4	Histone methyltransferase - PHD-type Zinc Finger domain	GGACTTTCCCGAAGACCGGA AGG	5: 33880124 to 33880146	CACCGGACTTTCCCGAAGACCGGAGT	TAAAACTCCGGTCTTCGGGAAAGTCC
Zmym2	Site 1	Zinc-Finger Protein, stabilizes LSD1/COREST/HDAC1 complex on chromatin	GGTGAGCTGATCCTTTGCGT TGG	14: 56911262 to 56911284	CACCGGTGAGCTGATCCTTTGCGTGT	TAAAACACGCAAAGGATCAGCTCACC
Zmym2	Site 2	Zinc-Finger Protein, stabilizes LSD1/COREST/HDAC1 complex on chromatin	CAGACAGCTTACCAACGCAA AGG	14: 56911251 to 56911273	CACCCAGACAGCTTACCAACGCAAGT	TAAAACTTGCGTTGGTAAGCTGTCTG
Zmym2	Site 3	Zinc-Finger Protein, stabilizes LSD1/COREST/HDAC1 complex on chromatin	GTTTCTTTGGAGCAGGCTTA TGG	14: 56911319 to 56911341	CACCGTTTCTTTGGAGCAGGCTTAGT	TAAAACTAAGCCTGCTCCAAAGAAAC

Supplementary Table S1 (continued)

Supplementary Table S2. List of genes, target exons and genomic sequences used in the CRISPR design tool analysis

Gene	Site	Exon Number	CRISPR Design Tool Input Sequence
Pax5 (Chromosome 4: 44,524,757-44,710,487)	Site 1	Exon 3	GTATTATGAGACAGGAAGCATCAAGCCGGGGGGTGATTGGAGGATCCAAACCAAAGGTTGCCACTCCCAAAGT GGTGGAAAAAATCGCTGAGTACAAACGCCAAAACCCTACCATGTTTGCCTGGGAGATCAGGGACCGGCTGTT GGCAGAGCGAGTCTGTGACAATGACACTGTGCCCAGCGTCAGCTCCATCAACAG
Pax5 (Chromosome 4: 44,524,757-44,710,487)	Site 2	Exon 3	GTATTATGAGACAGGAAGCATCAAGCCGGGGGTGATTGGAGGATCCAAACCAAAGGTTGCCACTCCCAAAGT GGTGGAAAAAATCGCTGAGTACAAACGCCAAAACCCTACCATGTTTGCCTGGGAGATCAGGGACCGGCTGTT GGCAGAGCGAGTCTGTGACAATGACACTGTGCCCAGCGTCAGCTCCATCAACAG
Pax5 (Chromosome 4: 44,524,757-44,710,487)	Site 3	Exon 3	GTATTATGAGACAGGAAGCATCAAGCCGGGGGTGATTGGAGGATCCAAACCAAAGGTTGCCACTCCCAAAGT GGTGGAAAAAATCGCTGAGTACAAACGCCAAAACCCTACCATGTTTGCCTGGGAGATCAGGGACCGGCTGTT GGCAGAGCGAGTCTGTGACAATGACACTGTGCCCAGCGTCAGCTCCATCAACAG
Tbl1xr1 (Chromosome 3: 22,076,652-22,216,594)	Site 1	Exon 7	ATAATCACACTGACATGATGGAAGTAGATGGGGATGTGGAAATCCCTTCCAATAAAGCGGTTGTGCTTCGGG GCCATGAATCTGAAGTTTTCATCTGTGCCTGGAACCCTGTTAGTGATCTCCTGGCATCAGG
Tbl1xr1 (Chromosome 3: 22,076,652-22,216,594)	Site 2	Exon 7	ATAATCACACTGACATGATGGAAGTAGATGGGGATGTGGAAATCCCTTCCAATAAAGCGGTTGTGCTTCGGG GCCATGAATCTGAAGTTTTCATCTGTGCCTGGAACCCTGTTAGTGATCTCCTGGCATCAGG
Tbl1xr1 (Chromosome 3: 22,076,652-22,216,594)	Site 3	Exon 7	ATAATCACACTGACATGATGGAAGTAGATGGGGATGTGGAAATCCCTTCCAATAAAGCGGTTGTGCTTCGGG GCCATGAATCTGAAGTTTTCATCTGTGCCTGGAACCCTGTTAGTGATCTCCTGGCATCAGG
Btg1 (Chromosome 10: 96,617,001-96,622,813)	Site 1	Exon 2	TTCGCATCAACCATAAGATGGATCCTCTGATTGGACAGGCAGCCCAGCGGATTGGACTGAGCAGTCAGGAGTT GTTCAGGCTTCTCCCAAGTGAACTCACACTCTGGGTTGACCCCTACGAAGTGTCCTACCGGATTGGAGAGGAT GGCTCCATCTGCGTGCTGTACGAAGCCTCACCAGCAGGAGGTAGCACTCAAAACAGCACCAACGTGCAAATG GTAGACAGCAGAATCAGCTGTA
Btg1 (Chromosome 10: 96,617,001-96,622,813)	Site 2	Exon 2	TTCGCATCAACCATAAGATGGATCCTCTGATTGGACAGGCAGCCCAGCGGATTGGACTGAGCAGTCAGGAGTT GTTCAGGCTTCTCCCAAGTGAACTCACACTCTGGGTTGACCCCTACGAAGTGTCCTACCGGATTGGAGAGGAT GGCTCCATCTGCGTGCTGTACGAAGCCTCACCAGCAGGAGGTAGCACTCAAAACAGCACCAACGTGCAAATG GTAGACAGCAGAATCAGCTGTA
Btg1 (Chromosome 10: 96,617,001-96,622,813)	Site 3	Exon 2	TTCGCATCAACCATAAGATGGATCCTCTGATTGGACAGGCAGCCCAGCGGATTGGACTGAGCAGTCAGGAGTT GTTCAGGCTTCTCCCAAGTGAACTCACACTCTGGGTTGACCCCTACGAAGTGTCCTACCGGATTGGAGAGGAT GGCTCCATCTGCGTGCTGTACGAAGCCTCACCAGCAGGAGGTAGCACTCAAAACAGCACCAACGTGCAAATG GTAGACAGCAGAATCAGCTGTA
Etv6 (Chromosome 6: 134,035,700-134,270,155)	Site 1	Exon 6	ACTGTAGACTGCTTTGGGATTATGTCTATCAGTTGCTGTCTGACAGCCGGTACGAAAACTTCATCCGATGGGA GGACAAAGAATCCAAAATATTCCGGATAGTGGATCCCAACGGACTGGCTCGACTCTGGGGAAACCATAAG
Etv6 (Chromosome 6: 134,035,700-134,270,155)	Site 2	Exon 7	AACAGAACAAACATGACCTATGAGAAAATGTCCAGAGCCCTGCGCCACTACTACAAACTAAACATTATCAGGA AGGAGCCCGGACAAAGGCTTTTGTTCAG
Etv6 (Chromosome 6: 134,035,700-134,270,155)	Site 3	Exon 8	GTTCATGAAAACCCCAGATGAGATCATGAGTGGCCGGACAGACCGTCTAGAACACCTCGAGTCTCAAGTGCTG GATGAACAAACGTACCAAGAGGATGAACCTACCATAGCCTCACCGGTGGGCTGGCCAAGAGGAAACCTGCCC ACGGGGACCGCAGGAGGCGTGATGGAAGCAGGCGAGCTAGGGGTGGCTGTAAAGGAAGAGACCCGGGAAT AG
Rag2 (Chromosome 2: 101,624,718-101,632,529)	Site 1	Exon 2	GGGTGTTCTCTTTGGAGGACGTTCATACATGCCTTCTACCCAGAGAACCACAGAAAAATGGAATAGTGTAGCT GACTGCCTACCCCATGTTTTCTTGATAGATTTTGAATTTGGGTGTGCTACATCATATATTCTCCCAGAACTTCAG GATGGGCTGTCTTTTCATGTTTCTATTGCCAGAAACGATACCGTTTATATTTTGGGAGGACACTCACTTGCCAG TAATATACGCCCTGCTAA

Rag2 (Chromosome 2: 101,624,718-101,632,529)	Site 2	Exon 2	GGGTGTTCTCTTTGGAGGACGTTCATACATGCCTTCTACCCAGAGAACCACAGAAAAATGGAATAGTGTAGCT GACTGCCTACCCCATGTTTTCTTGATAGATTTTGAATTTGGGTGTGCTACATCATATATTCTCCCAGAACTTCAG GATGGGCTGTCTTTTCATGTTTCTATTGCCAGAAACGATACCGTTTATATTTTGGGAGGACACTCACTTGCCAG TAATATACGCCCTGCTAA
Rag2 (Chromosome 2: 101,624,718-101,632,529)	Site 3	Exon 2	GGGTGTTCTCTTTGGAGGACGTTCATACATGCCTTCTACCCAGAGAACCACAGAAAAATGGAATAGTGTAGCT GACTGCCTACCCCATGTTTTCTTGATAGATTTTGAATTTGGGTGTGCTACATCATATATTCTCCCAGAACTTCAG GATGGGCTGTCTTTTCATGTTTCTATTGCCAGAAACGATACCGTTTATATTTTGGGAGGACACTCACTTGCCAG TAATATACGCCCTGCTAA
Nr3c2 (Chromosome 8: 76,899,442-77,245,012)	Site 1	Exon 4	GACAACACAACTATCTGTGTGCTGGAAGAAATGACTGCATTATTGATAAGATTCGGAGAAAGAACTGTCCTGC CTGCAGGCTCCAGAAATGCCTCCAAGCCGGCATGAACTTAGGAG
Nr3c2 (Chromosome 8: 76,899,442-77,245,012)	Site 2	Exon 8	TTCCAAAAGATGGCCTCAAGAGCCAGGCTGCGTTTGAGGAGATGAGGACAAATTACATCAAAGAACTGAGGA AAATGGTCACCAAGTGTCCCAACAGTTCTGGACAGAGTTGGCAGAGGTTCTACCAACTGACGAAGCTTCTAGA CTCCATGCATGAT
Nr3c2 (Chromosome 8: 76,899,442-77,245,012)	Site 3	Exon 8	TTCCAAAAGATGGCCTCAAGAGCCAGGCTGCGTTTGAGGAGATGAGGACAAATTACATCAAAGAACTGAGGA AAATGGTCACCAAGTGTCCCAACAGTTCTGGACAGAGTTGGCAGAGGTTCTACCAACTGACGAAGCTTCTAGA CTCCATGCATGAT
Cdkn2a (Chromosome 4: 89,274,471-89,294,653)	Site 1	Exon 2	GTGATGATGATGGGCAACGTTCACGTAGCAGCTCTTCTGCTCAACTACGGTGCAGATTCGAACTGCGAGGACC CCACTACCTTCTCCCGCCCGGTGCACGACGCAGCGCGGGAAGGCTTCCTGGACACGCTGGTGGTGCTGCACG GGTCAGGGGCTCGGCTGGATGTGCGCGATGCCTGGGGTCGCCTGCCGCTCGACTTGGCCCAAGAGCGGGGA CATCAAGACATCGTGCGATATTTG
Cdkn2a (Chromosome 4: 89,274,471-89,294,653)	Site 2	Exon 2	GTGATGATGATGGGCAACGTTCACGTAGCAGCTCTTCTGCTCAACTACGGTGCAGATTCGAACTGCGAGGACC CCACTACCTTCTCCCGCCCGGTGCACGACGCAGCGCGGGAAGGCTTCCTGGACACGCTGGTGGTGCTGCACG GGTCAGGGGCTCGGCTGGATGTGCGCGATGCCTGGGGTCGCCTGCCGCTCGACTTGGCCCAAGAGCGGGGA CATCAAGACATCGTGCGATATTTG
Cdkn2a (Chromosome 4: 89,274,471-89,294,653)	Site 3	Exon 2	GTGATGATGATGGGCAACGTTCACGTAGCAGCTCTTCTGCTCAACTACGGTGCAGATTCGAACTGCGAGGACC CCACTACCTTCTCCCGCCCGGTGCACGACGCAGCGCGGGAAGGCTTCCTGGACACGCTGGTGGTGCTGCACG GGTCAGGGGCTCGGCTGGATGTGCGCGATGCCTGGGGTCGCCTGCCGCTCGACTTGGCCCAAGAGCGGGGA CATCAAGACATCGTGCGATATTTG
Cdkn2b (Chromosome 4: 89,306,289-89,311,032)	Site 1	Exon 2	AACTGCGCCGACCCTGCCACCCTTACCAGACCTGTGCACGACGCAGCTCGGGAAGGCTTCCTGGACACGCTTG TCGTGCTGCACCGGGCAGGGGCGCGGCTGGATGTGTGTGACGCCTGG
Cdkn2b (Chromosome 4: 89,306,289-89,311,032)	Site 2	Exon 2	AACTGCGCCGACCCTGCCACCCTTACCAGACCTGTGCACGACGCAGCTCGGGAAGGCTTCCTGGACACGCTTG TCGTGCTGCACCGGGCAGGGGCGCGGCTGGATGTGTGTGACGCCTGG
Cdkn2b (Chromosome 4: 89,306,289-89,311,032)	Site 3	Exon 2	AACTGCGCCGACCCTGCCACCCTTACCAGACCTGTGCACGACGCAGCTCGGGAAGGCTTCCTGGACACGCTTG TCGTGCTGCACCGGGCAGGGGCGCGGCTGGATGTGTGTGACGCCTGG
Btla (Chromosome 16: 45,224,315-45,252,895)	Site 1	Exon 2	GGTGTAAGCACAATGGAACAATCTGTGTACCCCTTGAGGTTAGCCCTCAGCTATACACTAGTTGGGAAGAAAA TCAATCAGTTCCGGTTTTTGTTCTCCACTTTAAACCAATACATCTCA
Btla (Chromosome 16: 45,224,315-45,252,895)	Site 2	Exon 2	GGTGTAAGCACAATGGAACAATCTGTGTACCCCTTGAGGTTAGCCCTCAGCTATACACTAGTTGGGAAGAAAA TCAATCAGTTCCGGTTTTTGTTCTCCACTTTAAACCAATACATCTCA
Btla (Chromosome 16: 45,224,315-45,252,895)	Site 3	Exon 2	GGTGTAAGCACAATGGAACAATCTGTGTACCCCTTGAGGTTAGCCCTCAGCTATACACTAGTTGGGAAGAAAA TCAATCAGTTCCGGTTTTTGTTCTCCACTTTAAACCAATACATCTCA
Atf7ip (Chromosome 6: 136,506,167-136,610,862)	Site 1	Exon 2	TTCCAGTGAGCCCTCTTCCAGTGAGCCCACCTGTAGTGAGCCCATCTCTGGAGATCCTGTTTCTGAAGAAGCCG CCTCTCATGATCTTGTCTCTGGTGATTCCACTTGTAGCGAGCCAGTTTCTGGTGAGCCGGTGTCTCATGAAGCA GCTTCGAGTGAGCCAGCTACAAGTGAGCCAGCTTCAGATGAACCGGTCGCAAGGGTCGTGGCGGCCTGCGAA CTGGCTCCTGGGGAGTCAGC

Supplementary Table S2 (continued)

Atf7ip (Chromosome 6: 136,506,167-136,610,862)	Site 2	Exon 2	TTCCAGTGAGCCCTCTTCCAGTGAGCCCACCTGTAGTGAGCCCATCTCTGGAGATCCTGTTTCTGAAGAAGCCG CCTCTCATGATCTTGTCTCTGGTGATTCCACTTGTAGCGAGCCAGTTTCTGGTGAGCCGGTGTCTCATGAAGCA GCTTCGAGTGAGCCAGCTACAAGTGAGCCAGCTTCAGATGAACCGGTCGCAAGGGTCGTGGCGGCCTGCGAA CTGGCTCCTGGGGAGTCAGC
Atf7ip (Chromosome 6: 136,506,167-136,610,862)	Site 3	Exon 2	TTCCAGTGAGCCCTCTTCCAGTGAGCCCACCTGTAGTGAGCCCATCTCTGGAGATCCTGTTTCTGAAGAAGCCG CCTCTCATGATCTTGTCTCTGGTGATTCCACTTGTAGCGAGCCAGTTTCTGGTGAGCCGGTGTCTCATGAAGCA GCTTCGAGTGAGCCAGCTACAAGTGAGCCAGCTTCAGATGAACCGGTCGCAAGGGTCGTGGCGGCCTGCGAA CTGGCTCCTGGGGAGTCAGC
Mga (Chromosome 2: 119,897,228-119,969,581)	Site 1	Exon 2	TGGTCGTTGGTGGGAACCCAGTGGGAAGGCTGAACCCCATATTTTGGGGAGAGTTTTCATTCATCCAGAATCT CCTTCTACAGGTCATTATTGGATGCATCAACCAGTGTCTTTCTATAAACTCAAACTTACTAACAATACACTGGAC CAGGAAGGACATATTATCTTGCACTCTATGCATCGTTATCTGCCAAGGCTTCATTTGGTGCCAGCAGAAAAGGC TACAGAAGTGATACAGTT
Mga (Chromosome 2: 119,897,228-119,969,581)	Site 2	Exon 2	TGGTCGTTGGTGGGAACCCAGTGGGAAGGCTGAACCCCATATTTTGGGGAGAGTTTTCATTCATCCAGAATCT CCTTCTACAGGTCATTATTGGATGCATCAACCAGTGTCTTTCTATAAACTCAAACTTACTAACAATACACTGGAC CAGGAAGGACATATTATCTTGCACTCTATGCATCGTTATCTGCCAAGGCTTCATTTGGTGCCAGCAGAAAAGGC TACAGAAGTGATACAGTT
Mga (Chromosome 2: 119,897,228-119,969,581)	Site 3	Exon 20	ACAGAAGCTTTTGCATATTACCGACGGACACACACTGCCAATGAGCGGCGTCGGCGTGGTGAAATGAGAGAT CTTTTTGAAAAATTGAAG
Mga (Chromosome 2: 119,897,228-119,969,581)	Site 4	Exon 20	ACAGAAGCTTTTGCATATTACCGACGGACACACACTGCCAATGAGCGGCGTCGGCGTGGTGAAATGAGAGAT CTTTTTGAAAAATTGAAG
Stag2 (Chromosome X: 42,149,317-42,277,185)	Site 1	Exon 7	GATAGTGGGGATTACCCACTTACCATGGCTGGTCCTCAGTGGAAGAAGTTCAAGTCCAGCTTCTGTGAGTTCA TTGGTGTGTTAGTACGGCAGTGTCAATACAGTATCATATATGATGAGTACATGATGGATACTGTCATTTCACTT CTTACTGGATTGTCTGACTCCCAAGTCAGAGCATTTCGACATACAAGCACCCTGGCAG
Stag2 (Chromosome X: 42,149,317-42,277,185)	Site 2	Exon 7	GATAGTGGGGATTACCCACTTACCATGGCTGGTCCTCAGTGGAAGAAGTTCAAGTCCAGCTTCTGTGAGTTCA TTGGTGTGTTAGTACGGCAGTGTCAATACAGTATCATATATGATGAGTACATGATGGATACTGTCATTTCACTT CTTACTGGATTGTCTGACTCCCAAGTCAGAGCATTTCGACATACAAGCACCCTGGCAG
Stag2 (Chromosome X: 42,149,317-42,277,185)	Site 3	Exon 7	GATAGTGGGGATTACCCACTTACCATGGCTGGTCCTCAGTGGAAGAAGTTCAAGTCCAGCTTCTGTGAGTTCA TTGGTGTGTTAGTACGGCAGTGTCAATACAGTATCATATATGATGAGTACATGATGGATACTGTCATTTCACTT CTTACTGGATTGTCTGACTCCCAAGTCAGAGCATTTCGACATACAAGCACCCTGGCAG
Smc1a (Chromosome X: 152,016,428-152,062,694)	Site 1	Exon 10	TATGGCCGCCTCATTGACCTCTGCCAGCCTACACAAAAGAAGTATCAGATTGCTGTGACCAAGGTTTTAGGGA AGAACATGGATGCTATTATTGTAGACTCAGAGAAGACAGGTCGGGACTGTATTCAGTATATCAAGGAGCAAC GTGGAGAGCCTGAGACCTTCTTGCCTCTTGACTACCTGGAG
Smc1a (Chromosome X: 152,016,428-152,062,694)	Site 2	Exon 10	TATGGCCGCCTCATTGACCTCTGCCAGCCTACACAAAAGAAGTATCAGATTGCTGTGACCAAGGTTTTAGGGA AGAACATGGATGCTATTATTGTAGACTCAGAGAAGACAGGTCGGGACTGTATTCAGTATATCAAGGAGCAAC GTGGAGAGCCTGAGACCTTCTTGCCTCTTGACTACCTGGAG
Smc1a (Chromosome X: 152,016,428-152,062,694)	Site 3	Exon 10	TATGGCCGCCTCATTGACCTCTGCCAGCCTACACAAAAGAAGTATCAGATTGCTGTGACCAAGGTTTTAGGGA AGAACATGGATGCTATTATTGTAGACTCAGAGAAGACAGGTCGGGACTGTATTCAGTATATCAAGGAGCAAC GTGGAGAGCCTGAGACCTTCTTGCCTCTTGACTACCTGGAG
Smc5 (Chromosome 19: 23,206,441-23,273,897)	Site 1	Exon 7	GAATATGAGAATGTACGTCAGGAATATGAAGGTGTGAAACTAATTCGTGACCGAGTGAAGGAAGAGGTCAG GAAACTTAAAGAAGGGCAGATTCCTATGACACGGCGGATTGAGGAGATTGACCGGCAGCGTCACACTTTGGA AGTTCGAATCAAAGAAAAG
Smc5 (Chromosome 19: 23,206,441-23,273,897)	Site 2	Exon 7	GAATATGAGAATGTACGTCAGGAATATGAAGGTGTGAAACTAATTCGTGACCGAGTGAAGGAAGAGGTCAG GAAACTTAAAGAAGGGCAGATTCCTATGACACGGCGGATTGAGGAGATTGACCGGCAGCGTCACACTTTGGA AGTTCGAATCAAAGAAAAG
Smc5 (Chromosome 19: 23,206,441-23,273,897)	Site 3	Exon 7	GAATATGAGAATGTACGTCAGGAATATGAAGGTGTGAAACTAATTCGTGACCGAGTGAAGGAAGAGGTCAG GAAACTTAAAGAAGGGCAGATTCCTATGACACGGCGGATTGAGGAGATTGACCGGCAGCGTCACACTTTGGA AGTTCGAATCAAAGAAAAG

Supplementary Table S2 (continued)

Kras (Chromosome 6: 145,216,699-145,250,239)	Site 1	Exon 4	AGAACAAATTAAAAGAGTAAAGGACTCTGAAGATGTGCCTATGGTCCTGGTAGGGAATAAGTGTGATTTGCC TTCTAGAACAGTAGACACGAAACAGGCTCAGGAGTTAGCAAGGAGTTACGGGATTCCGTTCATTGAGACCTCA GCAAAGACAAGACAG
Kras (Chromosome 6: 145,216,699-145,250,239)	Site 2	Exon 4	AGAACAAATTAAAAGAGTAAAGGACTCTGAAGATGTGCCTATGGTCCTGGTAGGGAATAAGTGTGATTTGCC TTCTAGAACAGTAGACACGAAACAGGCTCAGGAGTTAGCAAGGAGTTACGGGATTCCGTTCATTGAGACCTCA GCAAAGACAAGACAG
Kras (Chromosome 6: 145,216,699-145,250,239)	Site 3	Exon 4	AGAACAAATTAAAAGAGTAAAGGACTCTGAAGATGTGCCTATGGTCCTGGTAGGGAATAAGTGTGATTTGCC TTCTAGAACAGTAGACACGAAACAGGCTCAGGAGTTAGCAAGGAGTTACGGGATTCCGTTCATTGAGACCTCA GCAAAGACAAGACAG
Nras (Chromosome 3: 103,058,285-103,067,914)	Site 1	Exon 3	GATTCTTACCGAAAGCAAGTGGTGATTGATGGTGAGACCTGCCTGCTGGACATACTGGACACAGCTGGACAA GAGGAGTACAGTGCCATGAGAGACCAGTACATGAGGACAGGCGAAGGGTTCCTCTGTGTATTTGCCATCAAT AATAGCAAATCATTTGCAGATATTAACCTCTACAG
Nras (Chromosome 3: 103,058,285-103,067,914)	Site 2	Exon 3	GATTCTTACCGAAAGCAAGTGGTGATTGATGGTGAGACCTGCCTGCTGGACATACTGGACACAGCTGGACAA GAGGAGTACAGTGCCATGAGAGACCAGTACATGAGGACAGGCGAAGGGTTCCTCTGTGTATTTGCCATCAAT AATAGCAAATCATTTGCAGATATTAACCTCTACAG
Nras (Chromosome 3: 103,058,285-103,067,914)	Site 3	Exon 3	GATTCTTACCGAAAGCAAGTGGTGATTGATGGTGAGACCTGCCTGCTGGACATACTGGACACAGCTGGACAA GAGGAGTACAGTGCCATGAGAGACCAGTACATGAGGACAGGCGAAGGGTTCCTCTGTGTATTTGCCATCAAT AATAGCAAATCATTTGCAGATATTAACCTCTACAG
Sae1 (Chromosome 7: 16,327,053-16,387,797)	Site 1	Exon 2	GCTTCGAGCTTCCCGGGTGCTGATTGTTGGCATGAAAGGACTTGGAGCTGAAATTGCCAAGAATCTTATCCTG GCAGGAGTCAAAGGGCTCACCATGCTGGACCACGAACAG
Sae1 (Chromosome 7: 16,327,053-16,387,797)	Site 2	Exon 2	GCTTCGAGCTTCCCGGGTGCTGATTGTTGGCATGAAAGGACTTGGAGCTGAAATTGCCAAGAATCTTATCCTG GCAGGAGTCAAAGGGCTCACCATGCTGGACCACGAACAG
Sae1 (Chromosome 7: 16,327,053-16,387,797)	Site 3	Exon 2	GCTTCGAGCTTCCCGGGTGCTGATTGTTGGCATGAAAGGACTTGGAGCTGAAATTGCCAAGAATCTTATCCTG GCAGGAGTCAAAGGGCTCACCATGCTGGACCACGAACAG
Nsd2 (Chromosome 5: 33,820,725-33,897,975)	Site 1	Exon 7	GTTGTAGCTGAACATCCAGATGCCTCAGGGGAAGAGATTGAAGAATTGCTTGGGTCCCAGTGGAGCATGCTC AATGAAAAGCAGAAAGCACGATATAATACAAAGTTTTCCCTAATGATCTCTGCCCAGTCTGAAGAAGACTCTG
Nsd2 (Chromosome 5: 33,820,725-33,897,975)	Site 2	Exon 16	GCACGAGATTGGAGAATTCCCTGTATTTTTCTTTGGGTCTAAAGATTATTACTGGACGCATCAGGCACGAGTGT TCCCATACATG
Nsd2 (Chromosome 5: 33,820,725-33,897,975)	Site 3	Exon 20	GACCGCATAATTGATGCTGGCCCCAAAGGGAATTATTCACGATTCATGAACCATAGCTGCCAGCCCAATTGTG AGACCCTGAAGTGGACAGTGAATGGGGACACACGGGTTGGCCTGTTTGCTGTGTGTGACATTCCTGCAG
Nsd2 (Chromosome 5: 33,820,725-33,897,975)	Site 4	Exon 12	CTGTGTGAGAAGACAGGCAGTCTCTTACTGTGTGAGGGCCCCTGTTGTGGAGCATTCCACCTAGCCTGCCTTG GACTTTCCCGAAGACCGGAAGGAAGATTCACCTGCACCGAATGTGCCTCAG
Zmym2 (Chromosome 14: 56,887,795-56,962,579)	Site 1	Exon 3	CAAGGGGTGGATTCTTTATCACCAGTGGCCTCACTTCCTAAACAGATATTCCAGCCCTCGAACCAACAACCCAC TAAACCAGTTAAAGTCACTTGTGCAAACTGCAAAAAACCTCTACAGAAGGGGCAGACAGCTTACCAACGCAAA GGATCAGCTCACCTCTTCTGTTCAACCACCTGCCTTTCTTCTTTCTCCCATAAGCCTGCTCCAAAGAAACTCTGTG TTATGTGTAAAAA
Zmym2 (Chromosome 14: 56,887,795-56,962,579)	Site 2	Exon 3	CAAGGGGTGGATTCTTTATCACCAGTGGCCTCACTTCCTAAACAGATATTCCAGCCCTCGAACCAACAACCCAC TAAACCAGTTAAAGTCACTTGTGCAAACTGCAAAAAACCTCTACAGAAGGGGCAGACAGCTTACCAACGCAAA GGATCAGCTCACCTCTTCTGTTCAACCACCTGCCTTTCTTCTTTCTCCCATAAGCCTGCTCCAAAGAAACTCTGTG TTATGTGTAAAAA
Zmym2 (Chromosome 14: 56,887,795-56,962,579)	Site 3	Exon 3	CAAGGGGTGGATTCTTTATCACCAGTGGCCTCACTTCCTAAACAGATATTCCAGCCCTCGAACCAACAACCCAC TAAACCAGTTAAAGTCACTTGTGCAAACTGCAAAAAACCTCTACAGAAGGGGCAGACAGCTTACCAACGCAAA GGATCAGCTCACCTCTTCTGTTCAACCACCTGCCTTTCTTCTTTCTCCCATAAGCCTGCTCCAAAGAAACTCTGTG TTATGTGTAAAAA

Supplementary Table S2 (continued)

Supplementary Table S3. List of genes, endogenous target sequences, and PCR primer pairs to amplify sites of interest (cut sites and indels)

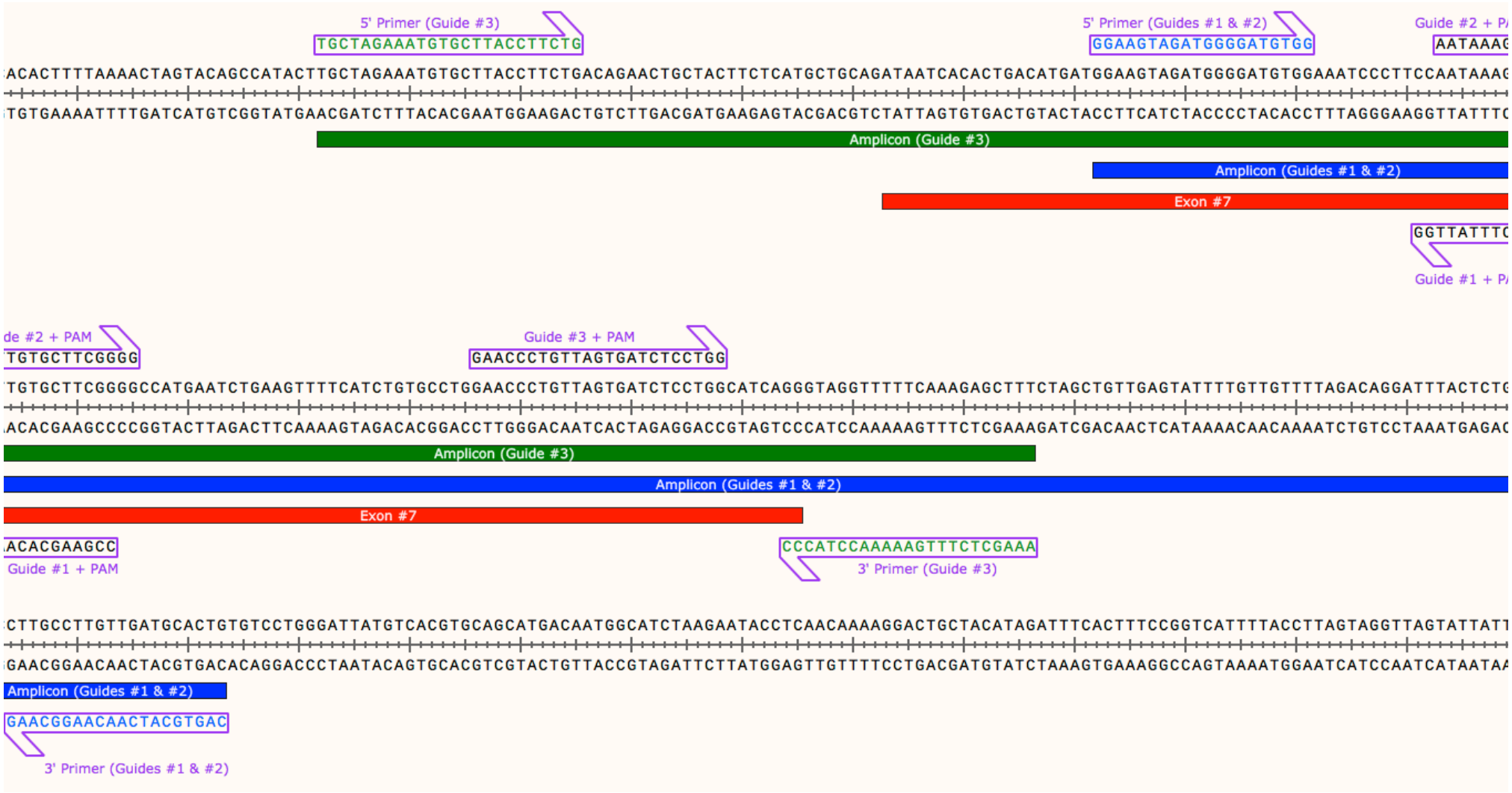
Gene	Site	Endogenous target sequence (20 nt + PAM)	Position (GRCm38)	Genome Locus 5' PCR Primer (5' -> 3')	Genome Locus 3' PCR Primer (5'-> 3')
Pax5	Site 1	AGCATCAAGCCGGGGGTGAT TGG	4: 44691994 to 44692016	GGTCCTCTCCTCTTGGGTCT	TCACCTGTTGATGGAGCTGA
Pax5	Site 2	TTGGATCCTCCAATCACCCC CGG	4: 44691985 to 44692007	GGTCCTCTCCTCTTGGGTCT	TCACCTGTTGATGGAGCTGA
Pax5	Site 3	ATCAAGCCGGGGGTGATTGG AGG	4: 44691991 to 44692013	GGTCCTCTCCTCTTGGGTCT	TCACCTGTTGATGGAGCTGA
Tbl1xr1	Site 1	CCGAAGCACAACCGCTTTAT TGG	3: 22190963 to 22190985	GGAAGTAGATGGGGATGTGG	CAGTGCATCAACAAGGCAAG
Tbl1xr1	Site 2	AATAAAGCGGTTGTGCTTCG GGG	3: 22190965 to 22190987	GGAAGTAGATGGGGATGTGG	CAGTGCATCAACAAGGCAAG
Tbl1xr1	Site 3	GAACCCTGTTAGTGATCTCC TGG	3: 22191018 to 22191040	TGCTAGAAATGTGCTTACCTTCTG	AAAGCTCTTTGAAAAACCTACCC
Btg1	Site 1	ACTGCTCAGTCCAATCCGCT GGG	10: 96618378 to 96618400	CGTTGTATTCGCATCAACCAT	CGTTGGTGCTGTTTTGAGTG
Btg1	Site 2	CTGTACGAAGCCTCACCAGC AGG	10: 96618497 to 96618519	CGTTGTATTCGCATCAACCAT	CGTTGGTGCTGTTTTGAGTG
Btg1	Site 3	GGACAGGCAGCCCAGCGGAT TGG	10: 96618368 to 96618390	CGTTGTATTCGCATCAACCAT	CGTTGGTGCTGTTTTGAGTG
Etv6	Site 1	CCAACGGACTGGCTCGACTC TGG	6:134261798 to 134261820	TGGGAGGACAAAGAATCCAA	CCTAATTGTGTGCCCAGTGA
Etv6	Site 2	TTAGTTTGTAGTAGTGGCGC AGG	6: 134263057 to 134263079	AAACAGAACAGAACAAACATGACC	CAATTGCAGTCATGGCTCTG
Etv6	Site 3	ACGCCTCCTGCGGTCCCCGT GGG	6: 134266400 to 134266422	TTCATGAAAACCCCAGATGAG	TATTCCCGGGTCTCTTCCTT
Rag2	Site 1	AATATAAACGGTATCGTTTC TGG	2: 101629978 to 101630000	GCCTTCTACCCAGAGAACCA	TTAGCAGGGCGTATATTACTGG
Rag2	Site 2	GATACCGTTTATATTTTGGG AGG	2: 101629986 to 101630008	GCCTTCTACCCAGAGAACCA	TTAGCAGGGCGTATATTACTGG
Rag2	Site 3	GACAGCCCATCCTGAAGTTC TGG	2: 101629938 to 101629960	GATTTTGAATTTGGGTGTGCT	GATTCCTCCTGGCAAGACTG
Nr3c2	Site 1	CTAAGTTCATGCCGGCTTGG AGG	8: 77155293 to 77155315	TTCGGAGAAAGAACTGTCCTG	CCCTGAGAAAACAAGTCACCA
Nr3c2	Site 2	AGTCTAGAAGCTTCGTCAGT TGG	8: 77217552 to 77217574	TGTCTTCCTTAACACAGTTCCAAA	TCACTGCACGTTGACTTCGT
Nr3c2	Site 3	TCATCTCCTCAAACGCAGCC TGG	8: 77217450 to 77217472	TGTCTTCCTTAACACAGTTCCAAA	TCACTGCACGTTGACTTCGT
Cdkn2a	Site 1	GGGTCGCCTGCCGCTCGACT TGG	4: 89276784 to 89276806	CGTAGCAGCTCTTCTGCTCA	TCGCACGATGTCTTGATGTC
Cdkn2a	Site 2	CCCGCGCTGCGTCGTGCACC GGG	4: 89276874 to 89276896	CGTAGCAGCTCTTCTGCTCA	TCGCACGATGTCTTGATGTC
Cdkn2a	Site 3	CGGTGCAGATTCGAACTGCG AGG	4: 89276916 to 89276938	CGTAGCAGCTCTTCTGCTCA	TCGCACGATGTCTTGATGTC
Cdkn2b	Site 1	CTTGTCGTGCTGCACCGGGC AGG	4: 89307163 to 89307185	GACCCTGCCACCCTTACC	GAACCCGCAGTCAATCTCC
Cdkn2b	Site 2	GCAGCACGACAAGCGTGTCC AGG	4: 89307173 to 89307195	GACCCTGCCACCCTTACC	GAACCCGCAGTCAATCTCC
Cdkn2b	Site 3	GACCTGTGCACGACGCAGCT CGG	4: 89307204 to 89307226	GACCCTGCCACCCTTACC	GAACCCGCAGTCAATCTCC
Btla	Site 1	CCCAACTAGTGTATAGCTGA GGG	16: 45629595 to 45629616	CCTAATGTGACTTGGTGTAAGCA	CACATGGATGGTTACTGAATGG
Btla	Site 2	GCCCTCAGCTATACACTAGT TGG	16: 45629595 to 45629615	CCTAATGTGACTTGGTGTAAGCA	CACATGGATGGTTACTGAATGG
Btla	Site 3	AACAATCTGTGTACCCCTTG AGG	16: 45239180 to 45239202	CCTAATGTGACTTGGTGTAAGCA	CACATGGATGGTTACTGAATGG
Atf7ip	Site 1	TGAACCGGTCGCAAGGGTCG TGG	6: 136560550 to 136560572	TCTCTGGAGATCCTGTTTCTGA	GGGCACAATCATCCAAGG
Atf7ip	Site 2	GCCGCCACGACCCTTGCGAC CGG	6: 136560554 to 136560576	TCTCTGGAGATCCTGTTTCTGA	GGGCACAATCATCCAAGG
Atf7ip	Site 3	AGAAACTGGCTCGCTACAAG TGG	6: 136560465 to 136560487	TCTCTGGAGATCCTGTTTCTGA	GGGCACAATCATCCAAGG

Mga	Site 1	CTATGCATCGTTATCTGCCA AGG	2: 119903259 to 119903281	AAGTGGAATGGTCGTTGGTG	CGGGGCCATTTAACTGTATC
Mga	Site 2	AGGCTGAACCCCATATTTTG GGG	2: 119903112 to 119903134	AAGTGGAATGGTCGTTGGTG	CGGGGCCATTTAACTGTATC
Mga	Site 3	ACCACGCCGACGCCGCTCAT TGG	2: 119958254 to 119958276	TTGCATATTACCGACGGACA	AACTCTGTTTCCACTTCCCTCA
Mga	Site 4	GCCAATGAGCGGCGTCGGCG TGG	2: 119958253 to 119958275	TTGCATATTACCGACGGACA	AACTCTGTTTCCACTTCCCTCA
Stag2	Site 1	ATTTCGACATACAAGCACCC TGG	X: 42223425 to 42223447	CTGGATTGTCTGACTCCCAAG	TGAGTTCAAGGCCAGCATAG
Stag2	Site 2	GATTACCCACTTACCATGGC TGG	X: 42223255 to 42223277	GGCACACATCTATCTTTGCAT	CACTGCCGTACTAACACACCA
Stag2	Site 3	GAGGACCAGCCATGGTAAGT GGG	X: 42223260 to 42223282	GGCACACATCTATCTTTGCAT	CACTGCCGTACTAACACACCA
Smc1a	Site 1	CAGTATATCAAGGAGCAACG TGG	X: 152033825 to 152033847	CAGTATGGCCGCCTCATT	ACCCCTTCACAAGCCTTACC
Smc1a	Site 2	GATACTTCTTTTGTGTAGGC TGG	X: 152033722 to 152033744	CAGTATGGCCGCCTCATT	ACCCCTTCACAAGCCTTACC
Smc1a	Site 3	CTTCTTGCCTCTTGACTACC TGG	X: 152033860 to 152033882	CAGTATGGCCGCCTCATT	ACCCCTTCACAAGCCTTACC
Smc5	Site 1	TCCTATGACACGGCGGATTG AGG	19: 23257535 to 23257557	CCTCTGCTTACATTTTGACTTTAGG	TCGAACTTCCAAAGTGTGACG
Smc5	Site 2	ACTAATTCGTGACCGAGTGA AGG	19: 23257589 to 23257611	TGCAAGGAATATGAGAATGTACGTC	TTCAGACAATCCCAAACTGAAC
Smc5	Site 3	TCCTCAATCCGCCGTGTCAT AGG	19: 23257534 to 23257556	CCTCTGCTTACATTTTGACTTTAGG	TCGAACTTCCAAAGTGTGACG
Kras	Site 1	TAGAACAGTAGACACGAAAC AGG	6: 145232157 to 145232179	AACCACTCAGTCAAGCTTTGTG	CCCGTAACTCCTTGCTAACTC
Kras	Site 2	TGAAGATGTGCCTATGGTCC TGG	6: 145232205 to 145232227	AACCACTCAGTCAAGCTTTGTG	CCCGTAACTCCTTGCTAACTC
Kras	Site 3	ATGTGCCTATGGTCCTGGTA GGG	6: 145232200 to 145232222	AACCACTCAGTCAAGCTTTGTG	CCCGTAACTCCTTGCTAACTC
Nras	Site 1	CTTCGCCTGTCCTCATGTAC TGG	3: 103060296 to 103060318	TAAAGATTGAGCCTGTCCTTGTC	ATTGATGGCAAATACACAGAGG
Nras	Site 2	GACCTGCCTGCTGGACATAC TGG	3: 103060236 to 103060258	CTGTTAGCGGGTTGAGGGTA	TCATGGCACTGTACTCCTCTTG
Nras	Site 3	AGTACATGAGGACAGGCGAA GGG	3: 103060298 to 103060320	TAAAGATTGAGCCTGTCCTTGTC	ATTGATGGCAAATACACAGAGG
Sae1	Site 1	GCCAAGAATCTTATCCTGGC AGG	7: 16378399 to 16378421	GTGTTTCTGTGGTGCATATAGTGA	AGATGGCACATACCTGTTCG
Sae1	Site 2	AGTCAAAGGGCTCACCATGC TGG	7: 16378376 to 16378398	GTGTTTCTGTGGTGCATATAGTGA	AGATGGCACATACCTGTTCG
Sae1	Site 3	TCATGCCAACAATCAGCACC CGG	7: 16378442 to 16378464	ATGTGCTTCACAGGCTTCG	GTGGCACTGAACCTGGATTT
Nsd2	Site 1	TTTCATTGAGCATGCTCCAC TGG	5: 33861090 to 33861112	ATGCCTCAGGGGAAGAGATT	CACACCTGCATGGGTGATAA
Nsd2	Site 2	AGATTATTACTGGACGCATC AGG	5: 33882901 to 33882923	GCAGTACAAGCCTTCTTGTTCAAA	CCCTCCATGTATGGGAACAC
Nsd2	Site 3	TCAGGGTCTCACAATTGGGC TGG	5: 33891565 to 33891587	TGGGTTCTCTTTGTGTTCTGG	ACACAGCAAACAGGCCAAC
Nsd2	Site 4	GGACTTTCCCGAAGACCGGA AGG	5: 33880124 to 33880146	TGTGGAGCATTCCACCTAGC	TGAAAGAGCAGCAACATACACC
Zmym2	Site 1	GGTGAGCTGATCCTTTGCGT TGG	14: 56911262 to 56911284	TCTTGGATCTCCCAGTCAGC	AAAGGCAGGTGGTTGAACAG
Zmym2	Site 2	CAGACAGCTTACCAACGCAA AGG	14: 56911251 to 56911273	TCTTGGATCTCCCAGTCAGC	AAAGGCAGGTGGTTGAACAG
Zmym2	Site 3	GTTTCTTTGGAGCAGGCTTA TGG	14: 56911319 to 56911341	CCACCTGCCTTTCTTCTTTC	CGAAGCCCCATCTATAACCA

Supplementary Table S3 (continued)

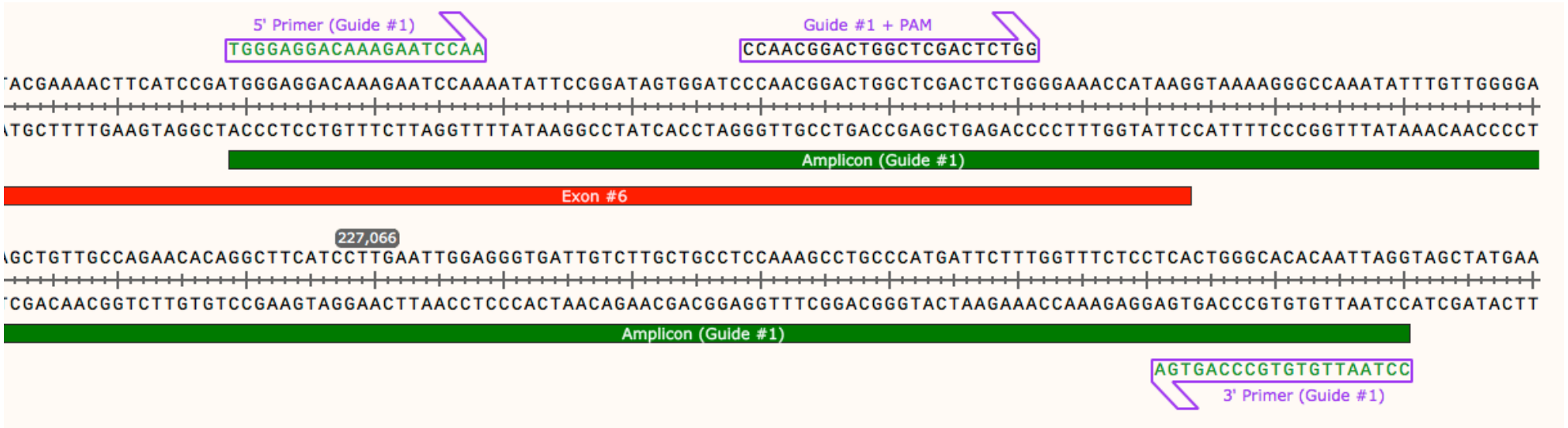
PAX5 PCR Primers and Amplicon

TBL1XR1 PCR Primers and Amplicon



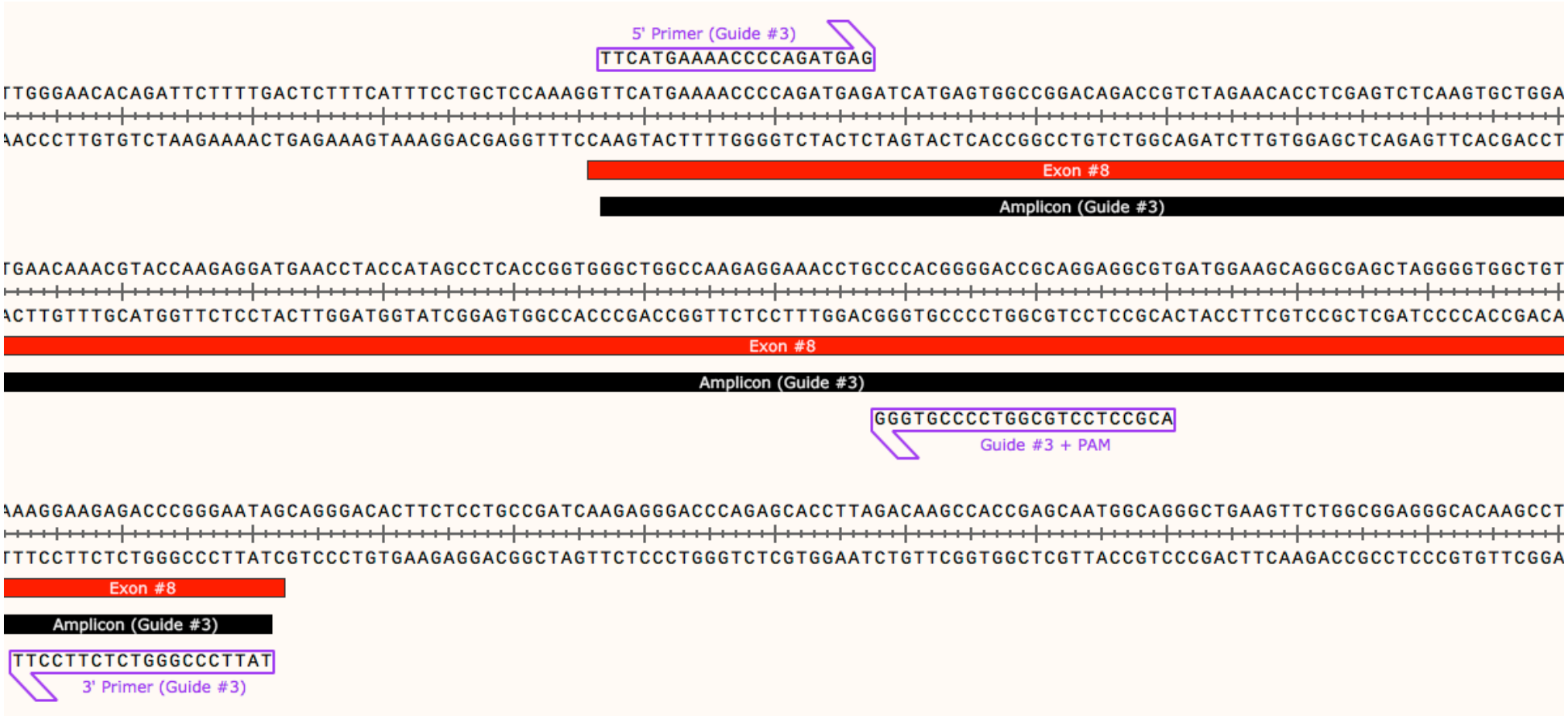
BTG1 PCR Primers and Amplicon

ETV6 Guide #1 PCR Primers and Amplicon

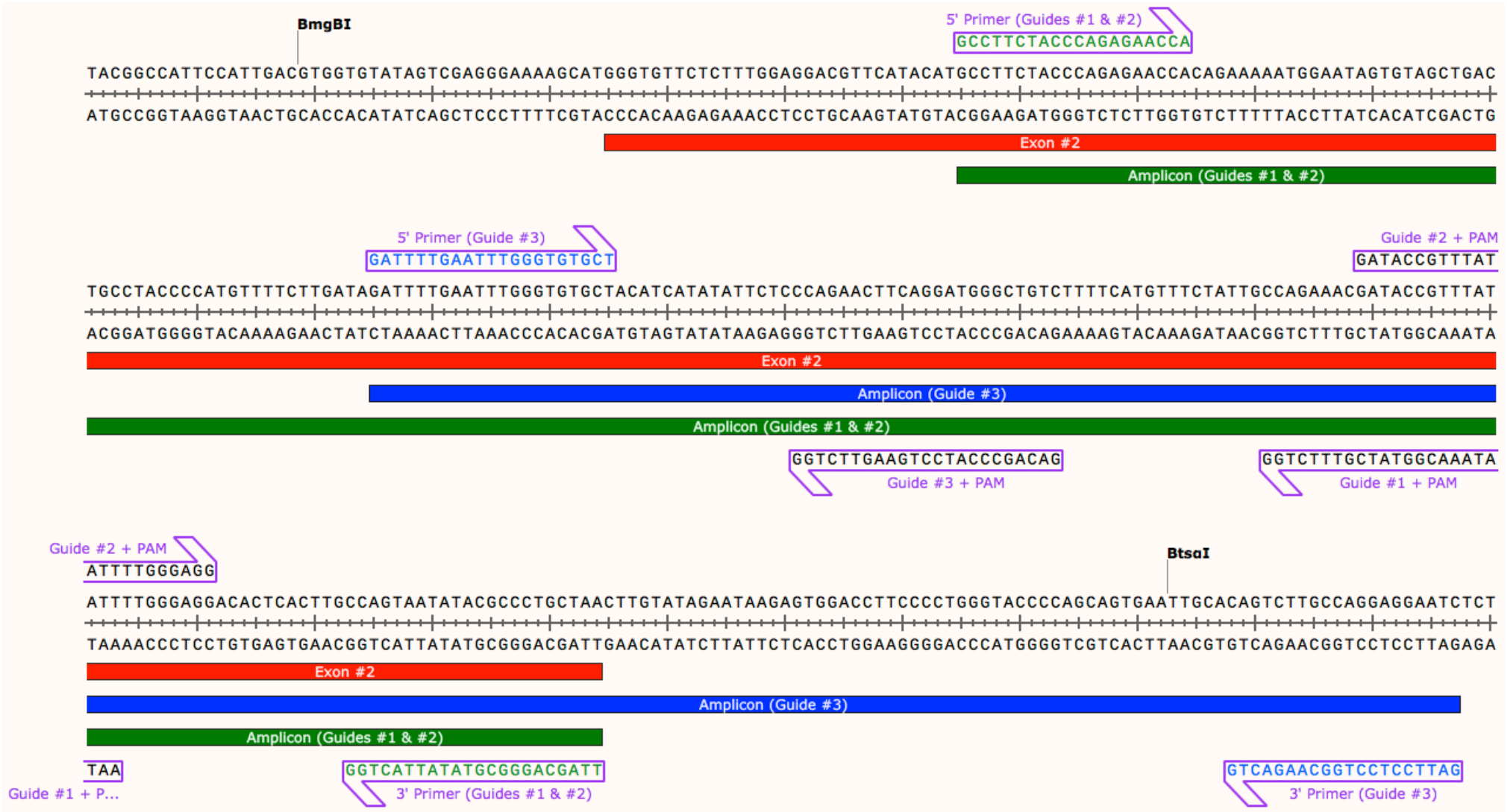


ETV6 Guide #2 PCR Primers and Amplicon

ETV6 Guide #3 PCR Primers and Amplicon

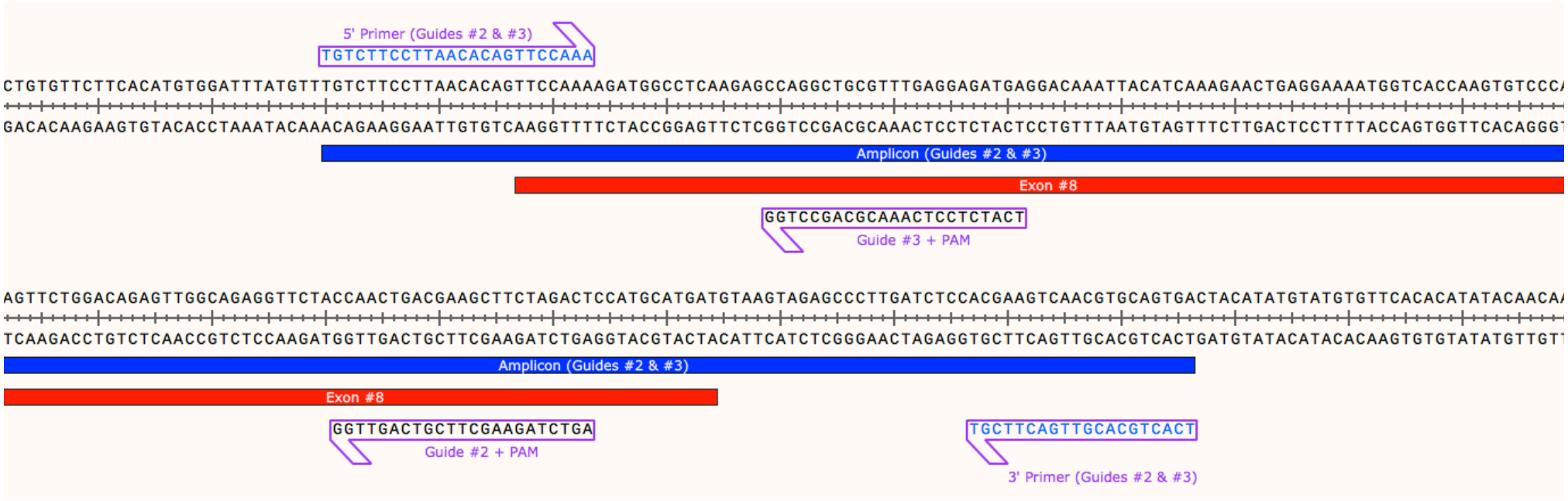


RAG2 PCR Primers and Amplicon



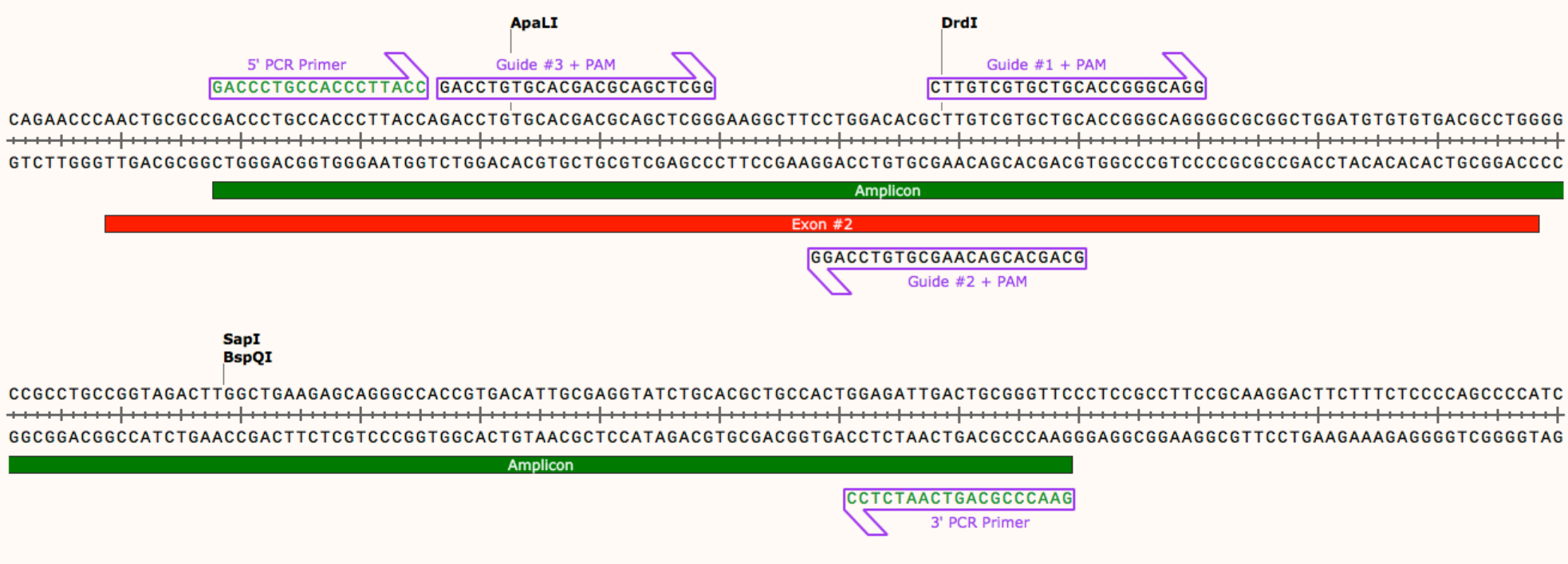
NR3C2 Guide #1 PCR Primers and Amplicon

NR3C2 Guides #2 & #3 PCR Primers and Amplicons

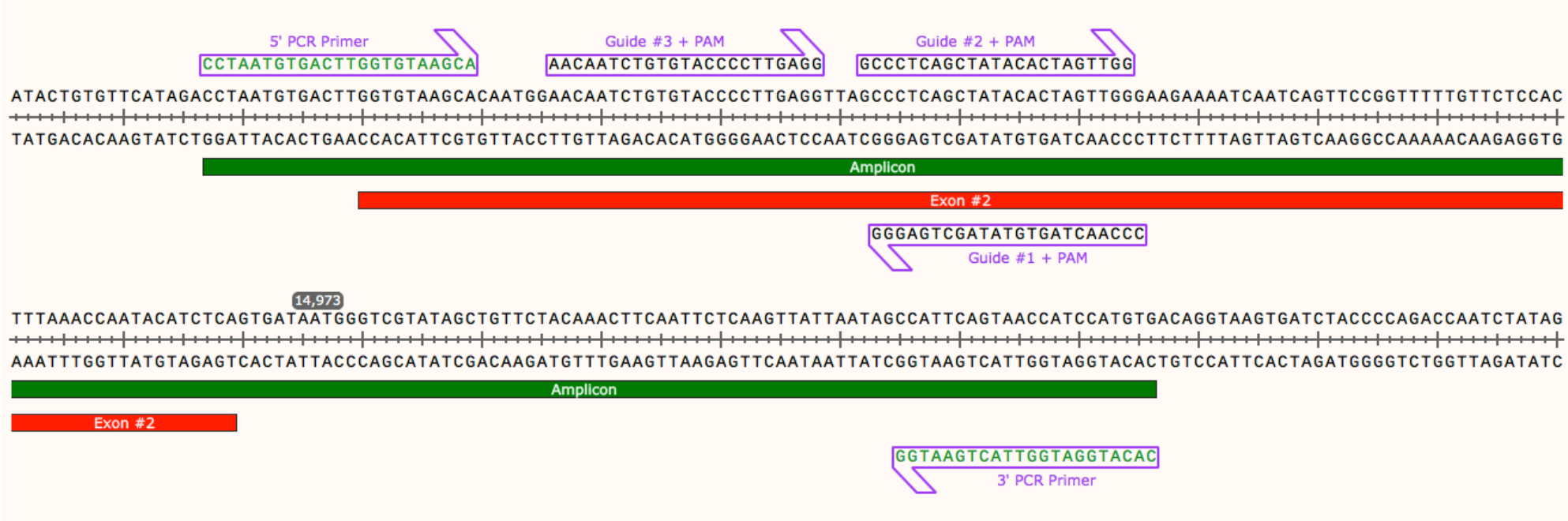


CDKN2A PCR Primers and Amplicon

CDKN2B PCR Primers and Amplicon



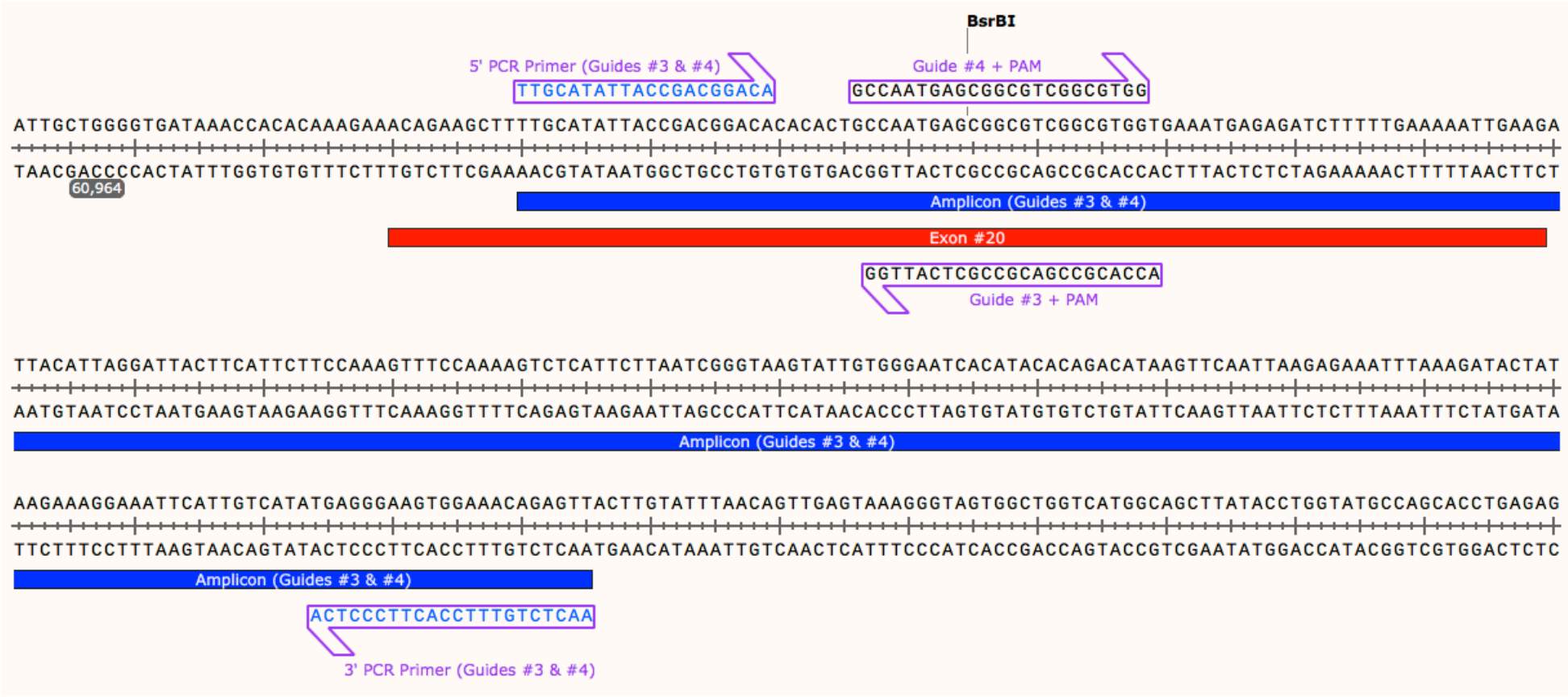
BTLA PCR Primers and Amplicon



ATF7IP PCR Primers and Amplicon

MGA Guides #1 & #2 PCR Primers and Amplicon

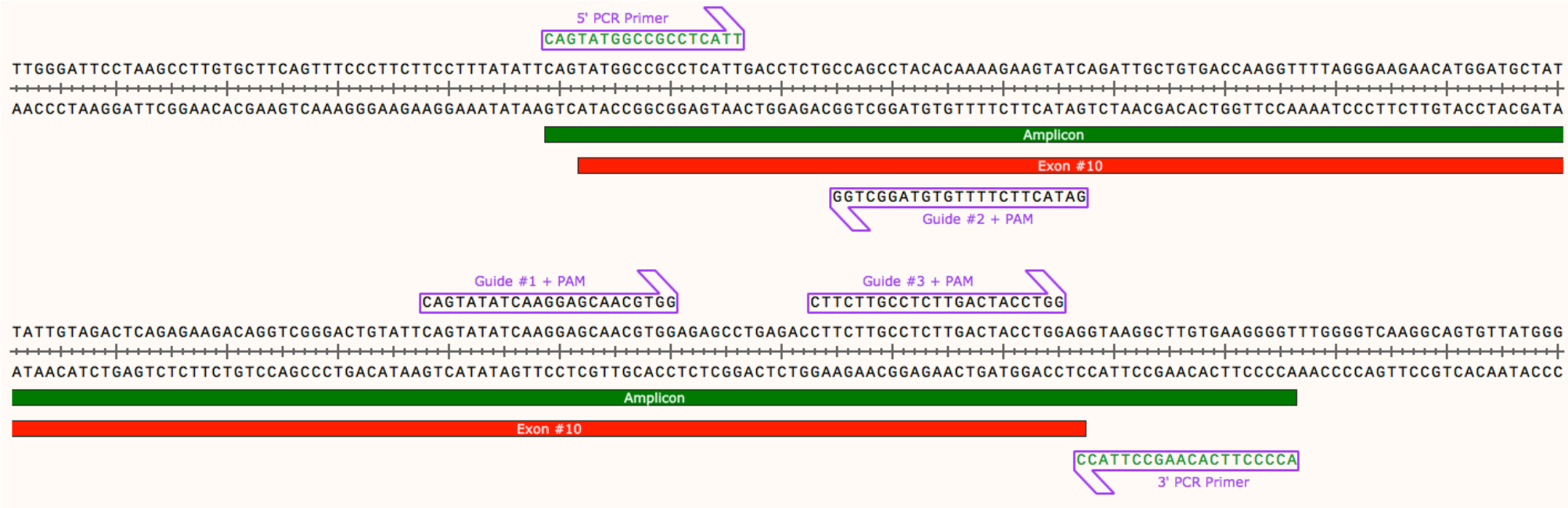
MGA Guides #3 & #4 PCR Primers and Amplicon



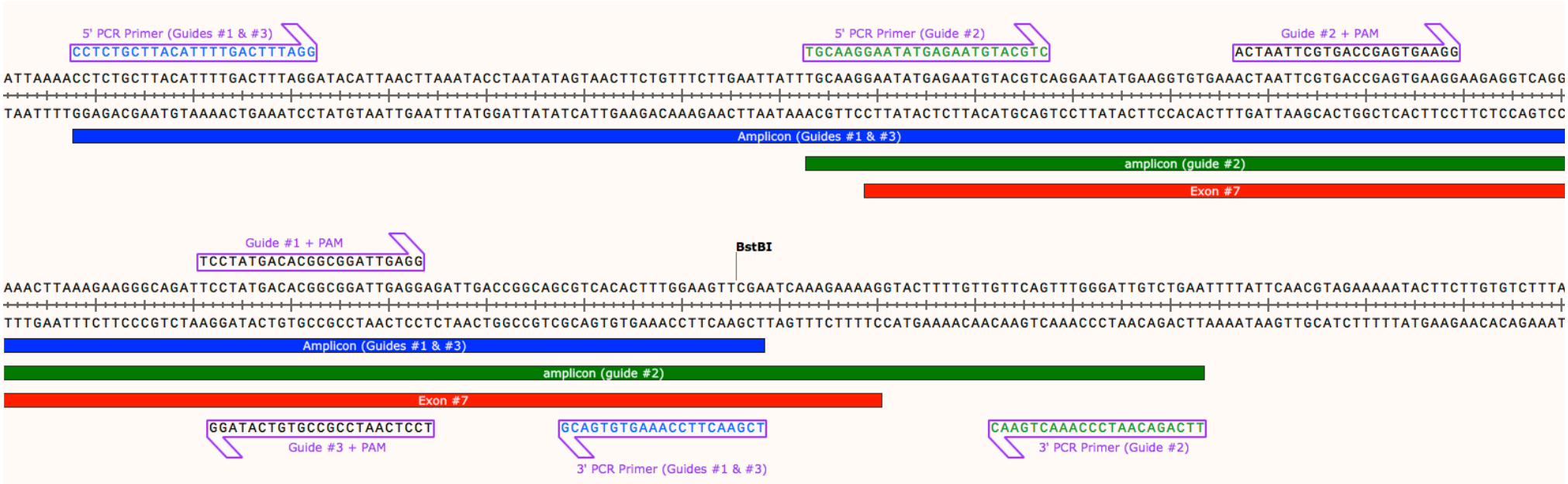
STAG2 Guide #1 PCR Primers and Amplicon

STAG2 Guides #2 & #3 PCR Primers and Amplicon

SMC1A PCR Primers and Amplicon

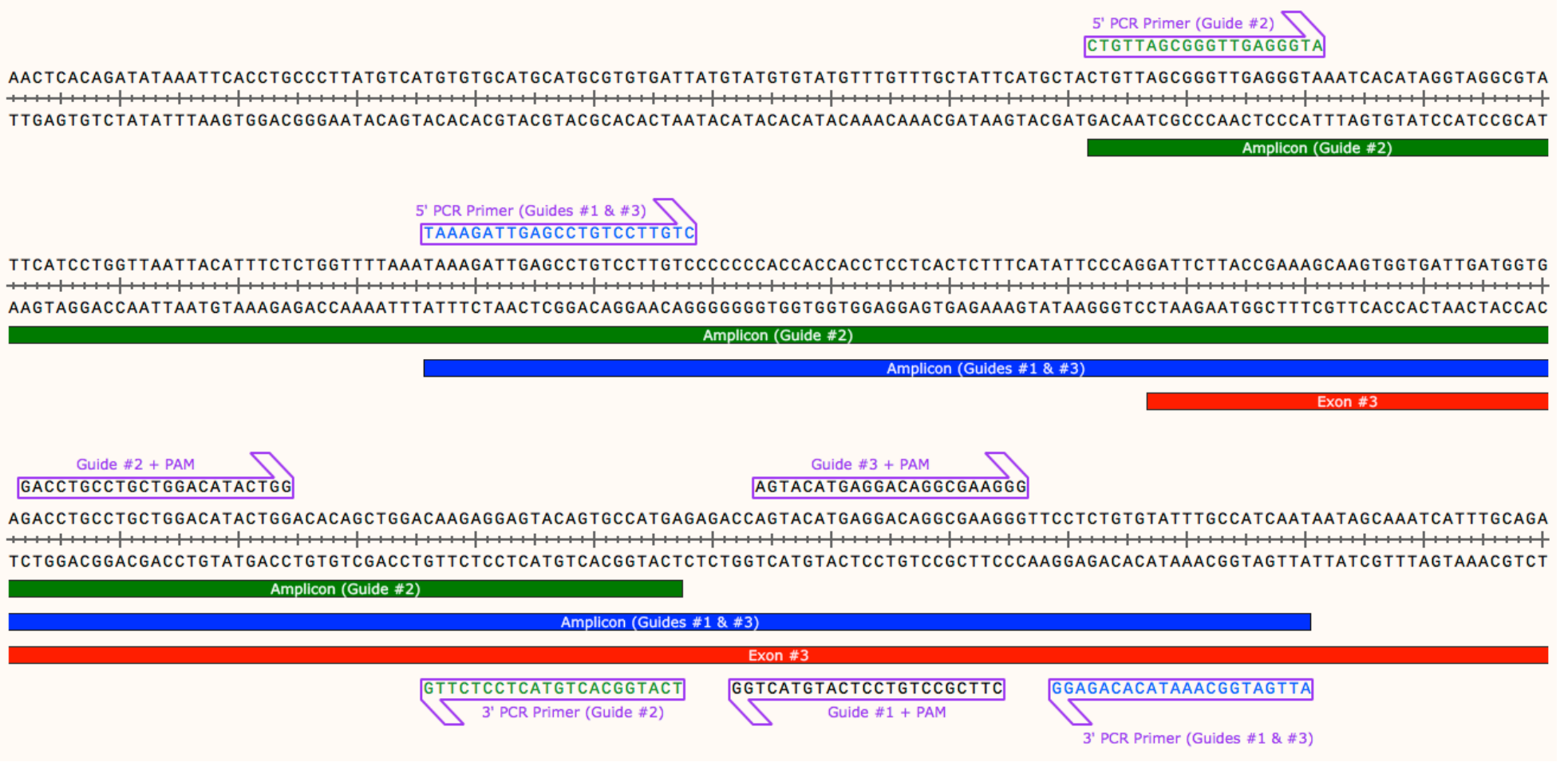


SMC5 PCR Primers and Amplicon

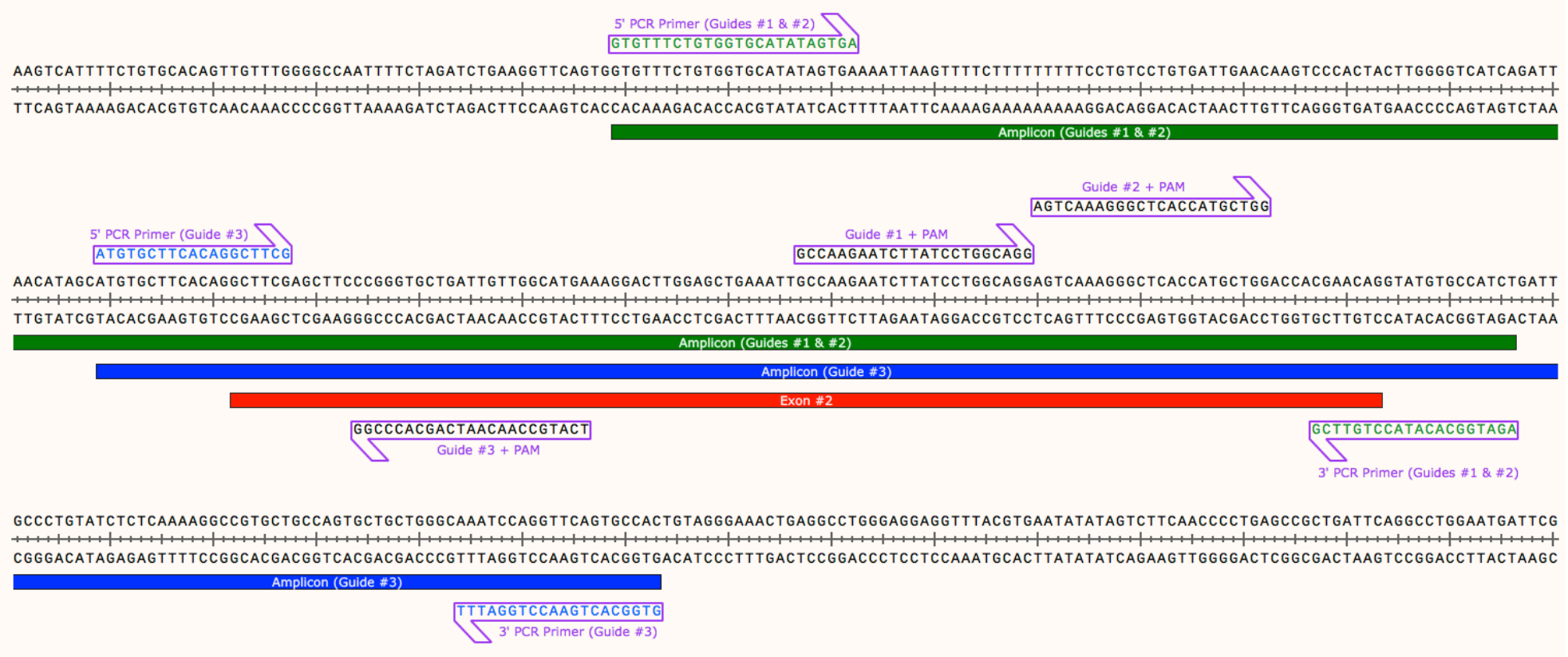


KRAS PCR Primers and Amplicon

NRAS PCR Primers and Amplicons

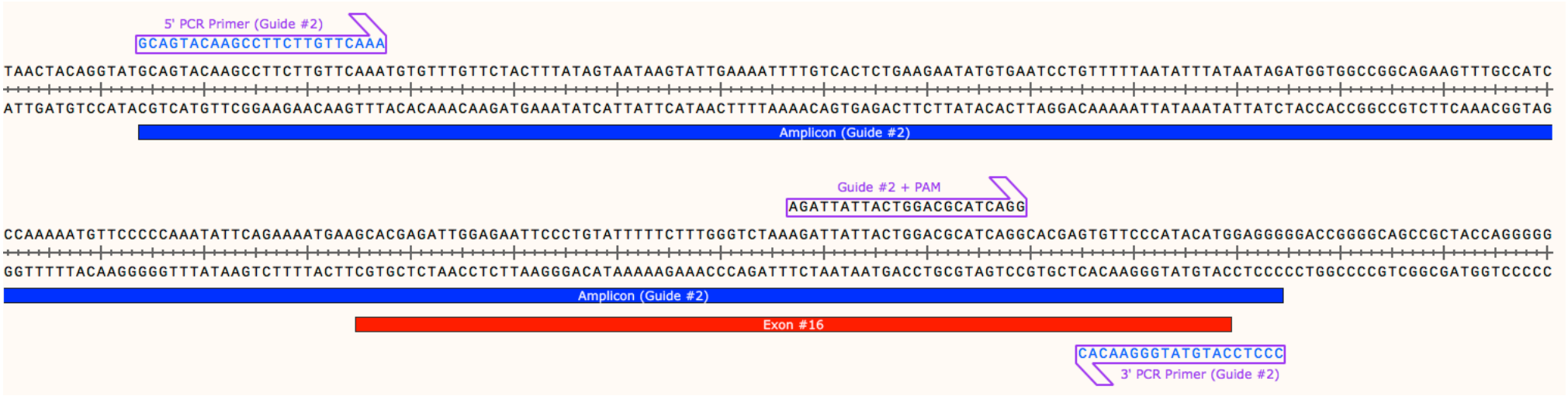


SAE1 PCR Primers and Amplicons



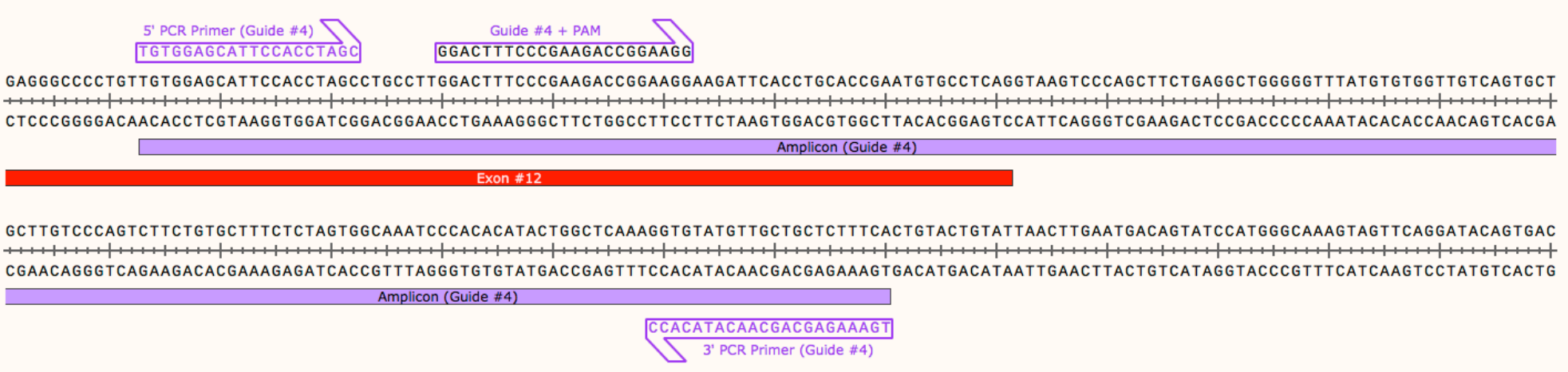
NSD2 Guide #1 PCR Primers and Amplicon

NSD2 Guide #2 PCR Primers and Amplicon

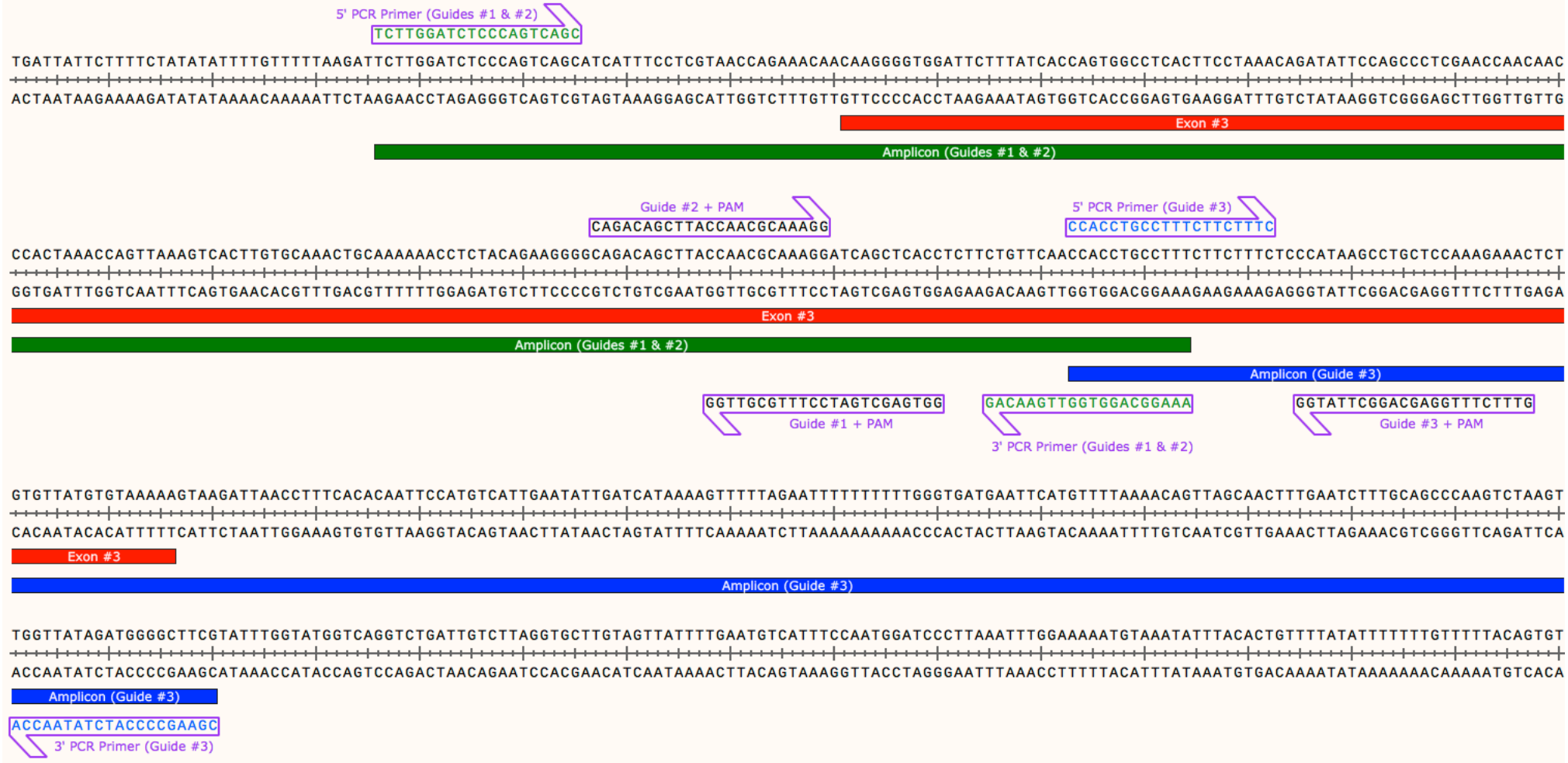


NSD2 Guide #3 PCR Primers and Amplicon

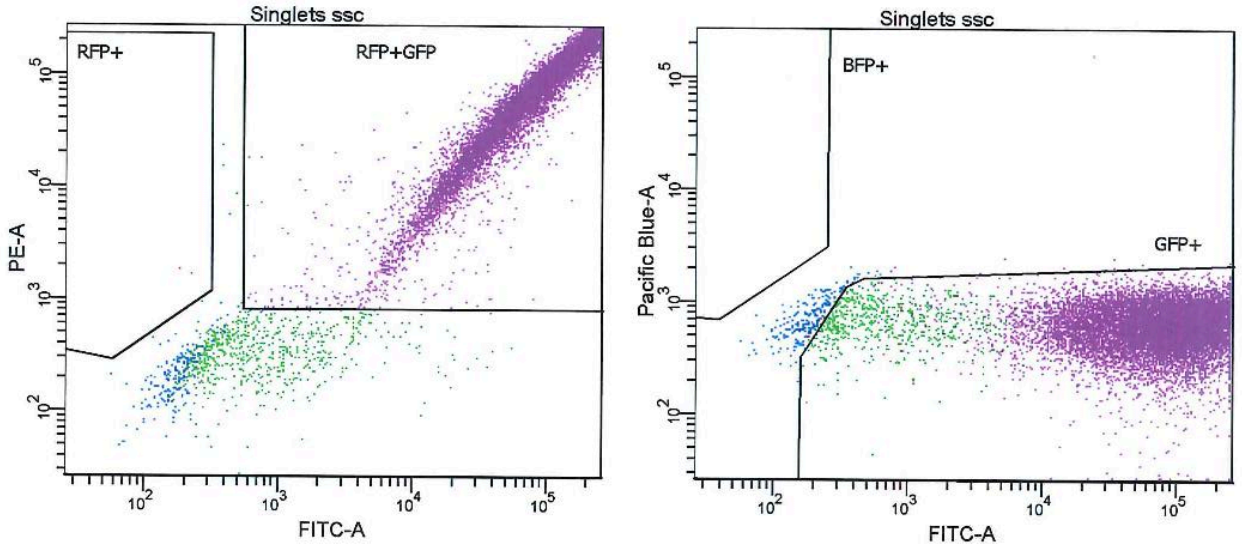
NSD2 Guide #4 PCR Primers and Amplicon



ZMYM2 PCR Primers and Amplicons



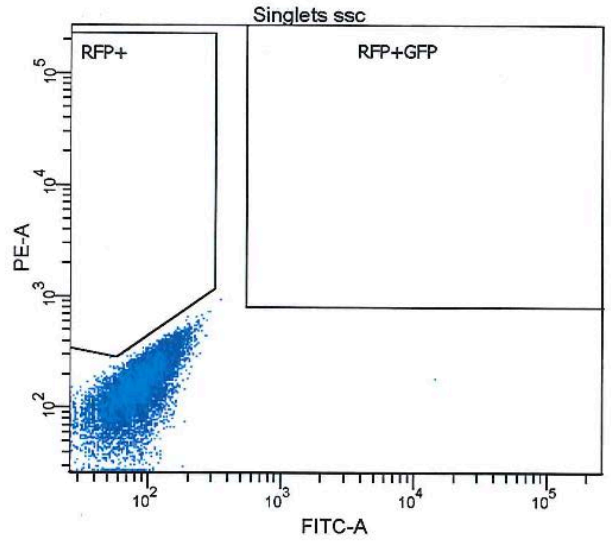
Supplemental FACS plots. NIH 3T3 titration to confirm successful lentiviral packaging and titers

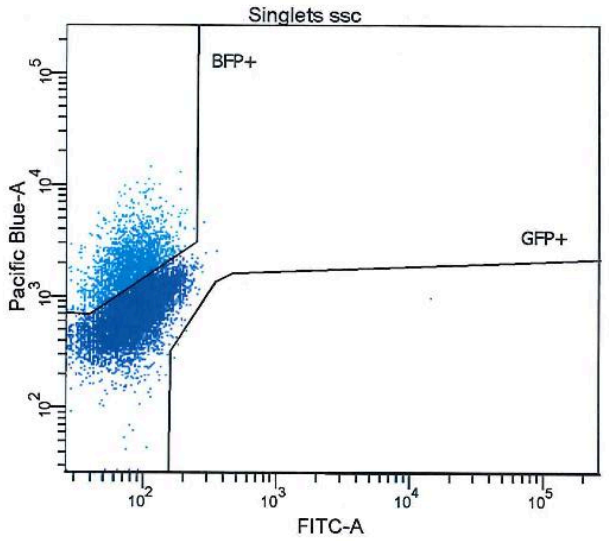


Tube: EFNa + RFP&GFP 1:1

Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,291	81.5	81.5
Singlets fsc	15,035	92.3	75.2
Singlets ssc	14,847	98.7	74.2
RFP+	2	0.0	0.0
GFP+	14,638	98.6	73.2
BFP+	0	0.0	0.0
RFP+GFP	13,981	94.2	69.9

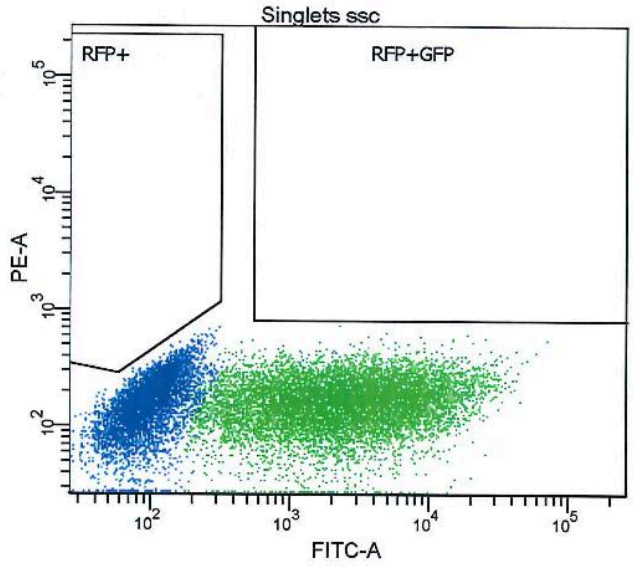
Positive Control Vector

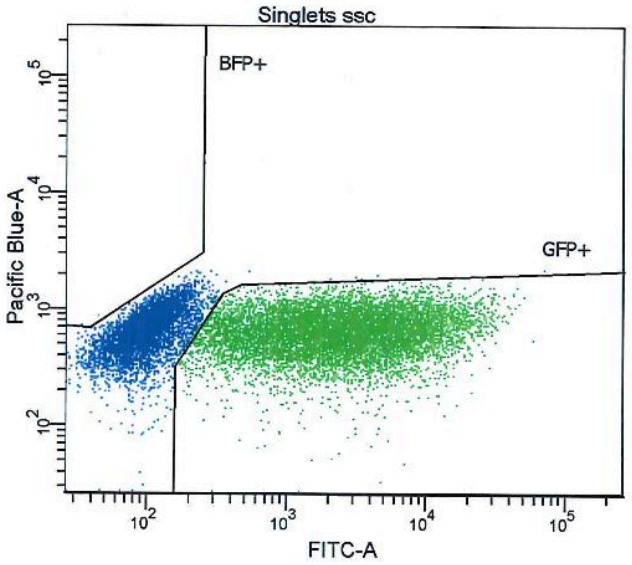




Tube: CRISPR Library BFP

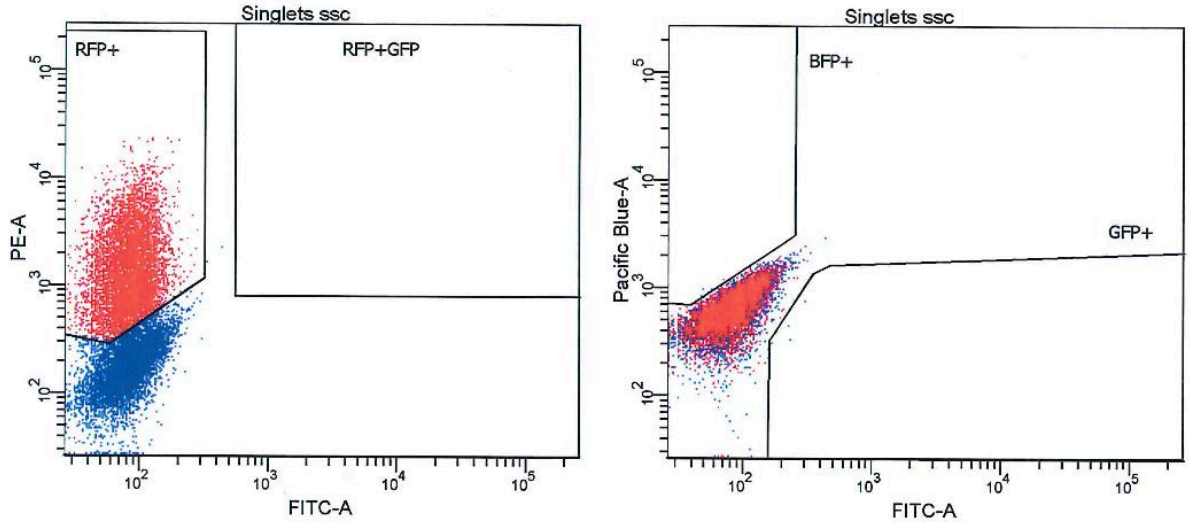
Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,437	82.2	82.2
Singlets fsc	15,784	96.0	78.9
Singlets ssc	15,704	99.5	78.5
RFP+	0	0.0	0.0
GFP+	4	0.0	0.0
BFP+	2,345	14.9	11.7
RFP+GFP	0	0.0	0.0





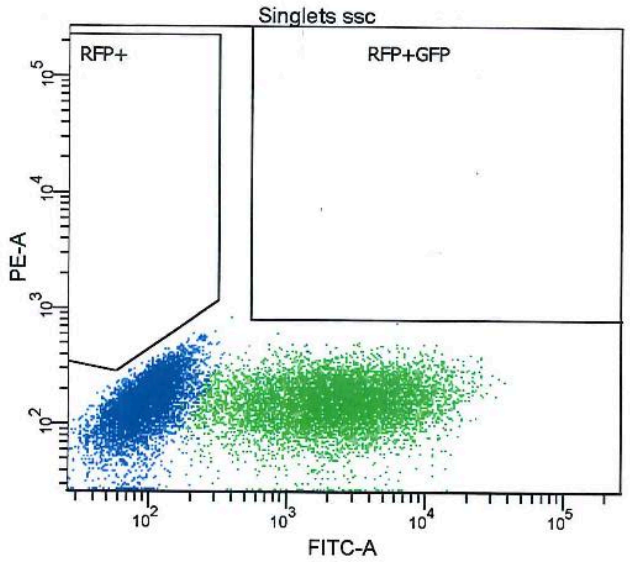
Tube: FZ Lenticas9 GFP

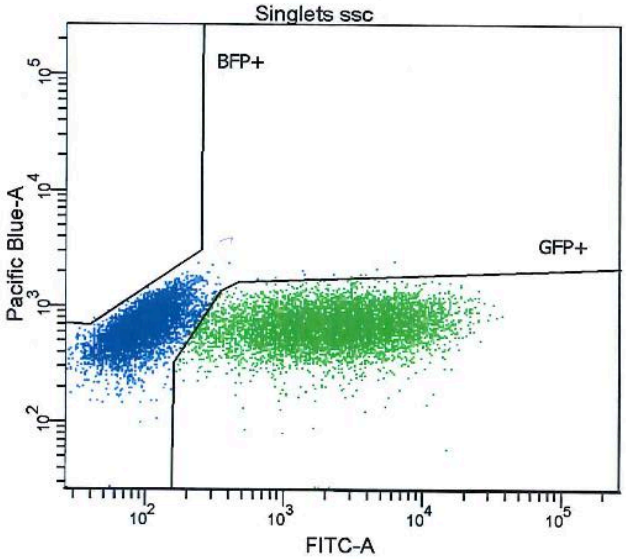
Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,595	83.0	83.0
Singlets fsc	15,636	94.2	78.2
Singlets ssc	15,494	99.1	77.5
RFP+	0	0.0	0.0
GFP+	9,964	64.3	49.8
BFP+	1	0.0	0.0
RFP+GFP	0	0.0	0.0



Tube: Ebert RFP

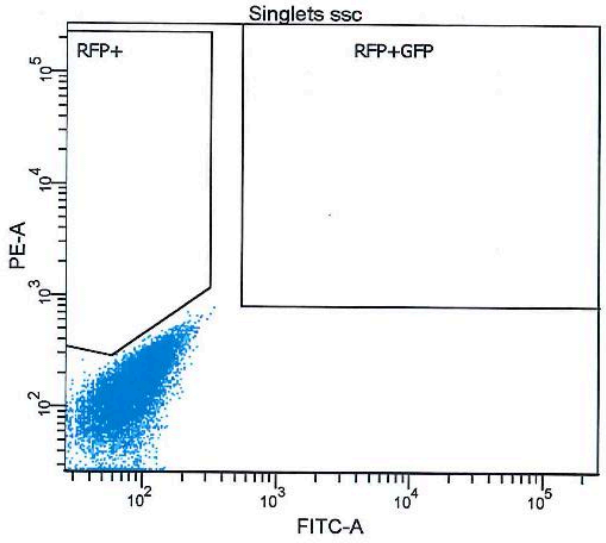
Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,563	82.8	82.8
Singlets fsc	15,753	95.1	78.8
Singlets ssc	15,675	99.5	78.4
RFP+	7,183	45.8	35.9
GFP+	9	0.1	0.0
BFP+	4	0.0	0.0
RFP+GFP	0	0.0	0.0

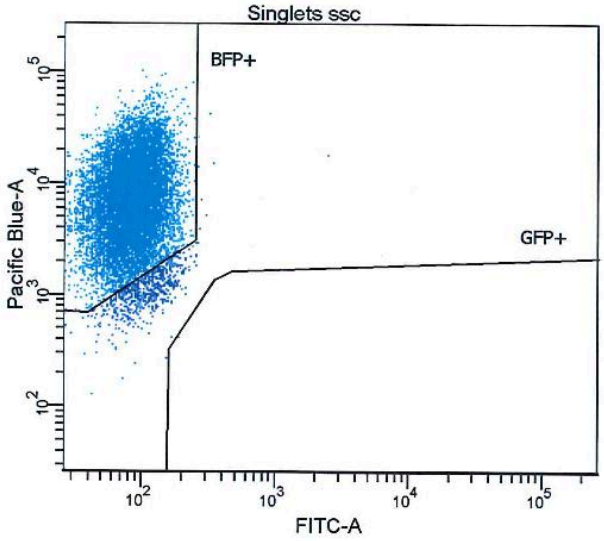




Tube: Ebert GFP

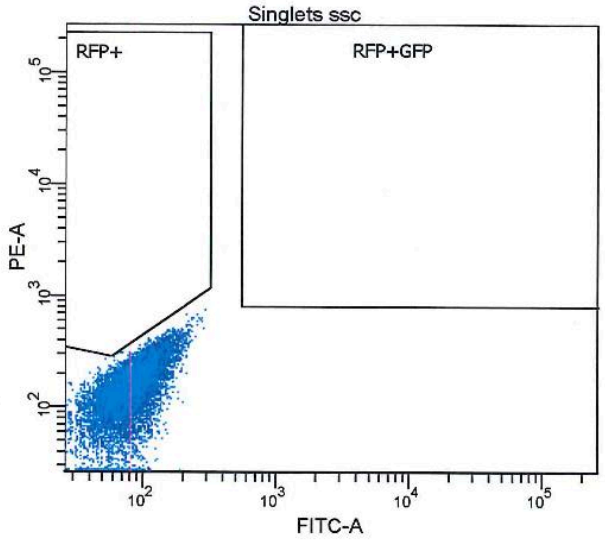
Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,960	84.8	84.8
Singlets fsc	16,158	95.3	80.8
Singlets ssc	16,089	99.6	80.4
RFP+	0	0.0	0.0
GFP+	8,176	50.8	40.9
BFP+	0	0.0	0.0
RFP+GFP	1	0.0	0.0

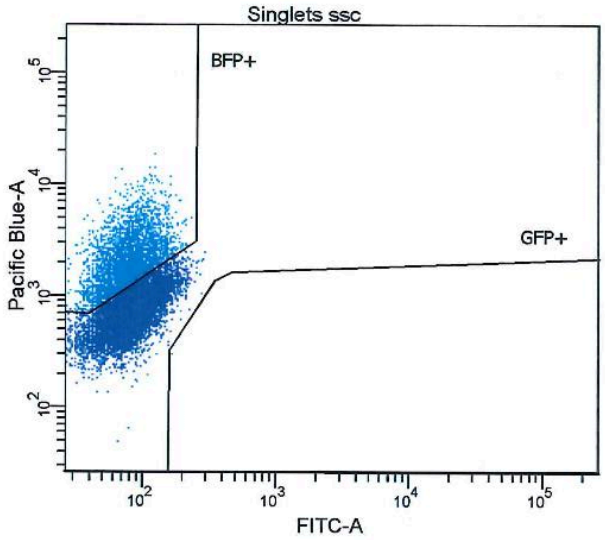




Tube: BFP Mga 1

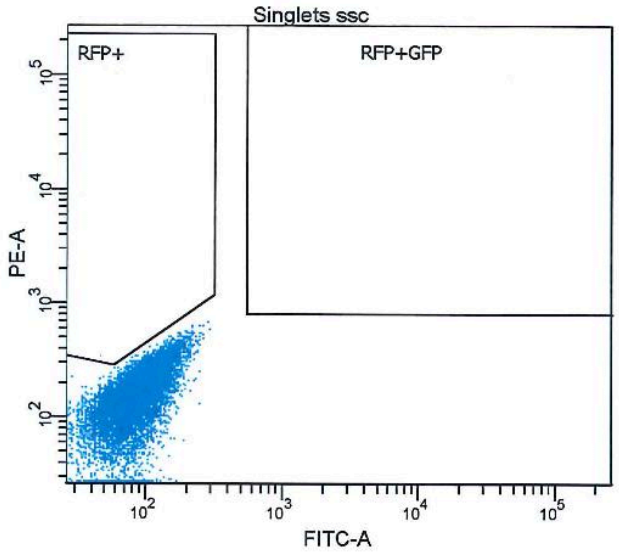
Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,103	80.5	80.5
Singlets fsc	14,575	90.5	72.9
Singlets ssc	14,379	98.7	71.9
RFP+	0	0.0	0.0
GFP+	1	0.0	0.0
BFP+	13,737	95.5	68.7
RFP+GFP	1	0.0	0.0

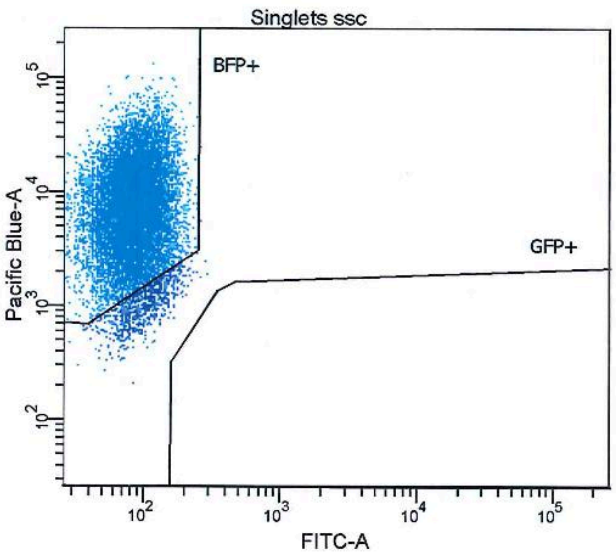




Tube: BFP Mga 2

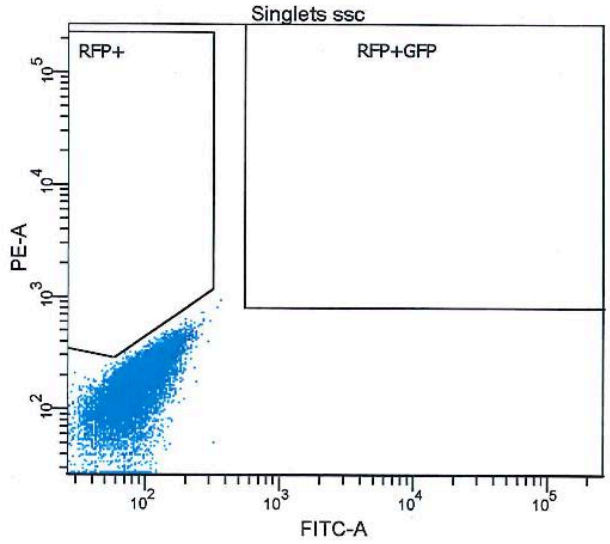
Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,341	81.7	81.7
Singlets fsc	14,892	91.1	74.5
Singlets ssc	14,684	98.6	73.4
RFP+	0	0.0	0.0
GFP+	2	0.0	0.0
BFP+	3,720	25.3	18.6
RFP+GFP	0	0.0	0.0

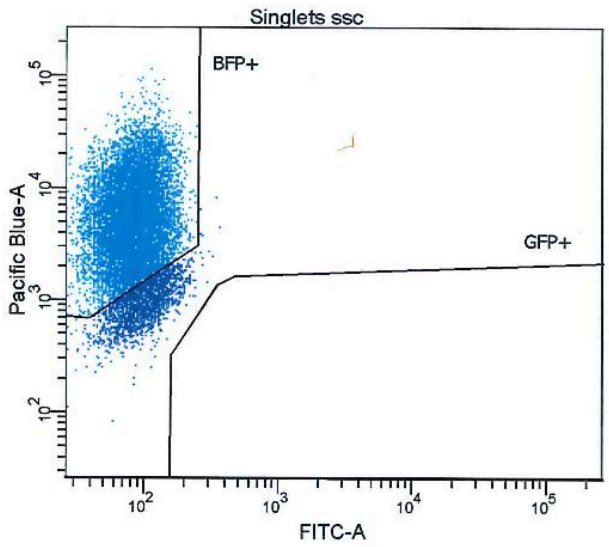




Tube: BFP Mga 3

Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	15,636	78.2	78.2
Singlets fsc	14,386	92.0	71.9
Singlets ssc	14,216	98.8	71.1
RFP+	0	0.0	0.0
GFP+	0	0.0	0.0
BFP+	13,435	94.5	67.2
RFP+GFP	0	0.0	0.0





Tube: BFP Mga 4

Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
Scatter	16,349	81.7	81.7
Singlets fsc	14,667	89.7	73.3
Singlets ssc	14,458	98.6	72.3
RFP+	0	0.0	0.0
GFP+	0	0.0	0.0
BFP+	12,344	85.4	61.7
RFP+GFP	0	0.0	0.0