Characterizing the role of ETV6 in lymphoid development and B-cell acute lymphoblastic leukemia predisposition

CHARACTERIZING THE ROLE OF ETV6 IN LYMPHOID DEVELOPMENT AND B-CELL ACUTE LYMPHOBLASTIC LEUKEMIA PREDISPOSITION

Jensen Joonho Pak

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, 2022

Thesis Advisor: Dr. Vijay Sankaran Jensen Joonho Pak

Characterizing the role of ETV6 in lymphoid development and B-cell acute lymphoblastic leukemia predisposition

Abstract

B cell acute lymphoblastic leukemia (ALL) is the most common form of pediatric cancer and one of the leading causes of death in children, characterized by clonal expansion of malignant lymphoid progenitors. While recent studies have identified a number of genetic risk factors for the development of ALL, our understanding of the mechanism of susceptibility to the disease remains incomplete. Sequencing studies have implicated germline *ETV6* mutations to predispose patients to development of blood cancers and particularly to B-cell ALL. However, the specific role for *ETV6* has not been characterized. Here, I investigated the role of germline *ETV6* mutations in human B cell lymphopoiesis and predisposition to ALL. Human cord blood-derived and adult peripherally mobilized $CD34^+$ cells were edited using a CRISPR-Cas9 editing strategy that has been established in our laboratory. I established an optimized *in vitro* co-culture system for B cell differentiation from hematopoietic stem cells. I examined the effects of this knockdown on cellular differentiation, including flow cytometry for phenotypic analysis of different stages of B cell lymphopoiesis. Perturbation of *ETV6,* particularly in the ETS domain, in $CD34^+$ HSPCs produced an altered lymphoid progenitor pool, notably a reduction in $CD45RA^+CD33^-$- and a subtle increase in $CD19^+$-expressing cells at one week of differentiation. Understanding the role of germline *ETV6* variation offers unique insight into the molecular etiology of and predisposition to ALL, providing future opportunities for treatment and potential prevention.

Table of contents

List of Figures

List of Tables

Acknowledgments

I would first like to thank my mentor and friend Dr. Lara Wahlster, without whom none of this thesis would have been possible. Despite your busy schedule with clinical practice and research projects, you somehow made time for a kid with limited genetics and hematology experience. Every step of the way, you have made this project productive, meaningful, and exciting, putting up with my questions and mistakes. I am truly grateful to have worked with you, and have learned so much from you about science, medicine, and life in general.

I would like to thank my PI Dr. Vijay Sankaran, who from our very first conversation shared his passion and support for me. Whether it was discussing data or offering career advice, you gave me confidence and inspired me to look forward to the future.

Thanks to everyone in the Sankaran Lab, who have helped me out and welcomed me this past year. It was truly a pleasure to be a part of this group. Special shout out to Susie, Alexis, Rick, Alex, Tre, and Mateusz!

Thanks to Dr. Shiv Pillai, Dr. Michael Carroll, and Dr. Gavin Porter for the opportunity and independence to study immunology at HMS. Thanks to all my classmates, especially Haim, Nate, Neal, and Steve!

A big thanks to my parents and my brother, who have always given me unconditional support. Thanks to all of my friends, especially those with whom I play too many video games late at night, and my roommate Bobby, who has to put up with me playing too many video games late at night.

Chapter 1
1.1 – Background

Human hematopoiesis and disease

Hematopoiesis, the process of blood formation, is a homeostatic process sustained by hematopoietic stem cells (HSCs). This system is responsible for the production of platelets necessary for blood clotting, red blood cells which transport oxygen, and all lineages of the innate and adaptive immune systems of the human body (Figure 1.1).

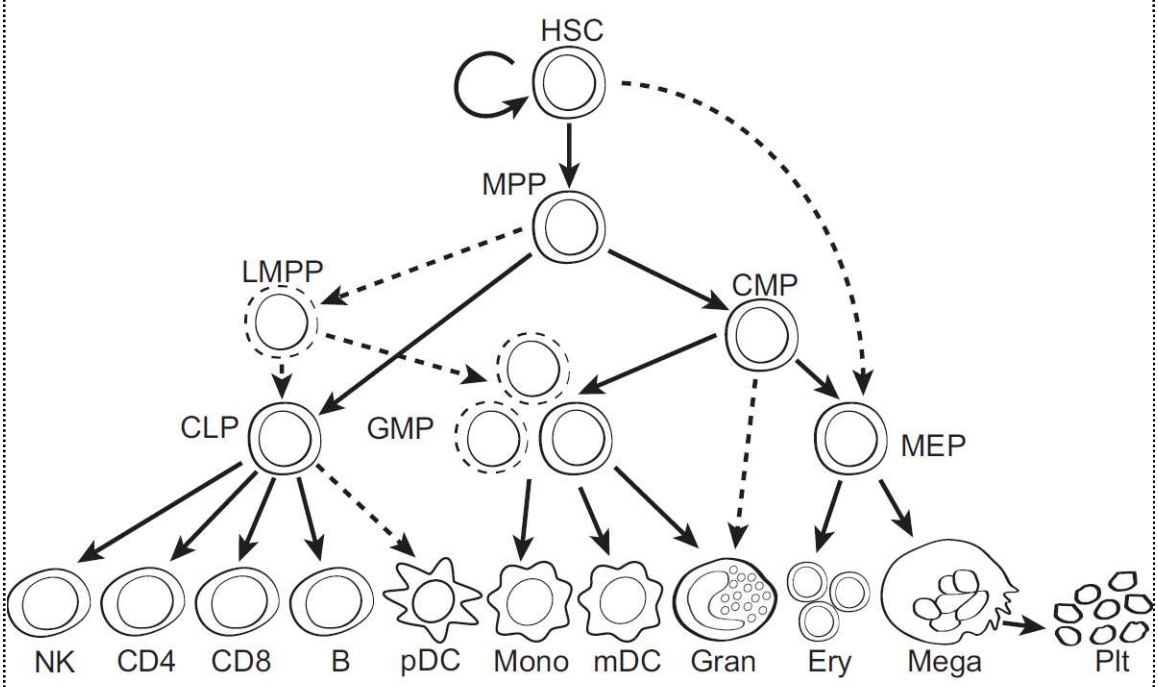

Figure 1.1. Overview of hematopoietic lineages deriving from human hematopoietic stem cells.

These pathways are controlled by transcription factors, which regulate self-renewal and differentiation through coordinated gene expression. While current and future studies uncover further resolution and heterogeneity within the

hematopoietic hierarchy, the classical understanding of hematopoiesis relies on uniform lineage commitment and discrete cell fate choices, a model that has shaped our understanding and investigation of pathological changes and disease progression[1,2].

Of note, B cells and other adaptive immune cells are derived from lymphoid progenitor cells. Aberrations in transcription factors can disrupt normal hematopoiesis and subsequently cause a range of blood cancers. In particular, there is increasing evidence for implicating genetic alterations in transcription factors affecting B-cell development in increasing risk for lymphoid-lineage diseases[3].

ETV6 overview

One such hematopoietic regulator, ETS variant transcription factor 6 (ETV6), is critically required for survival of bone marrow (BM) HSCs and development of the megakaryocyte lineage[4]. The *ETV6* gene resides on chromosome 12p13.2 and encodes a 57 kDa protein. ETV6 spans eight exons with two highly conserved functional domains: an N-terminal PNT domain that regulates dimerization and a C-terminal ETS DNA-binding domain[5,6]. Studies have shown that a motif responsible for nuclear localization resides within the ETS domain[7].

The *ETV6* gene, previously known as *TEL*, was first identified in a t(5;12) chromosomal translocation, expressing a fusion with the tyrosine kinase domain

of platelet-derived growth factor receptor β (PDGFRβ), implicated in the pathogenesis of chronic myelomonocytic leukemia (CMML)[8]. Early studies illustrated that ETV6 is required for hematopoiesis in the bone marrow[9]. TEL$^{-/-}$ mice display embryonic lethality due to a yolk sac angiogenic defect, also affecting neural and mesenchymal populations within the embryo[10]. In conditional knockout mouse models, ETV6 function was shown to be essential for HSC survival and the subsequent establishment of all hematopoietic lineages in the bone marrow, especially during first week of postnatal life. As a member of the ETS family, ETV6 acts as a transcriptional repressor, regulating a wide spectrum of target genes during hematopoiesis[11–13]. Notably, ETV6 has been reported to interact with co-repressors like HDAC3 and NCOR2 through its linker domain[14,15]. Through a variety of reporter assays, truncated ETV6 proteins retaining either essential domain were shown to exhibit a dominant-negative effect on wild-type ETV6 repressor activity, similarly as seen in familial germline mutations[5,16]. As mutants affecting different regions of ETV6 exhibit dominant negative activity, potential mechanisms include increased cytoplasmic retention of the wild-type protein to the cytoplasm, competition for co-repressors, and reduced affinity of repressor complexes due to altered DNA-binding activity of the ETS domain[5].

Chromosomal translocations in hematologic malignancies and leukemia

Many leukemias are characterized by recurrent structural or copy number aberrations. These alterations can be used to classify roughly 75-80% of pediatric ALL patients into prognostically and therapeutically relevant subgroups[17].

Chromosomal rearrangements such as genetic translocations result in an oncogenic factor being rearranged into a highly transcribed region or a fusion protein that greatly alters normal function. Implicated in numerous hematologic malignancies, ETV6 has been identified as a fusion partner in over 30 translocation oncogenes[5,18,19]. The most frequent chromosomal rearrangement, occurring in up to 25% of pediatric B cell acute lymphoblastic leukemia (B-ALL) patients, is t(12;21)(p13;q22), resulting in the ETV6-RUNX1 fusion gene[17,20]. It is likely that the mechanisms of pathogenesis of ETV6 translocations are different from that of other somatic and germline mutations. For example, the ETV6/RUNX1 fusion protein does not contain the ETV6 ETS binding domain and disrupts RUNX1-mediated transcriptional regulation, which may be distinct from the effects of ETV6 germline variants[21,22]. However, in several of these translocations, including MN1, BTL, and PAX5, the ETS DNA-binding domain of ETV6 is involved in the fusion protein. In the context of ETV6-RUNX1, additional somatic mutations involving the second *ETV6* allele have been reported[23].

Somatic mutations in ETV6

Genome-wide profiling studies have revealed components of cellular and signaling pathways that are frequently mutated in ALL, but are not necessarily considered founding or causal events[17]. In most ALL subtypes, multiple cooperating mutations are acquired or enriched for during leukemia progression. Among them, recurrent somatic ETV6 point mutations or deletions have been identified both in lymphoid

leukemias and myeloid malignancies such as myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML) (sources).

The somatic heterozygous *ETV6* mutations seen in *de novo* AML patients affected either the PNT or the ETS domain. Mutations resulting in truncated ETV6 proteins impaired transcriptional repression and affected repression of wild-type ETV6[16]. These studies suggested that ETV6 is post-transcriptionally regulated, and that loss of ETV6 expression separate from chromosomal translocations is a critical event for leukemogenesis.

Germline mutations in ETV6

In more recent years, rare germline *ETV6* mutations have been described, primarily identified among families exhibiting inherited thrombocytopenia and increased risk for ALL.

ETV6-related thrombocytopenia is an autosomal dominant inherited disease, primarily characterized by bleeding symptoms[24]. Among cases of inherited thrombocytopenia, the relative frequency of ETV6-related thrombocytopenia is estimated at ~5%[24]. Several studies have identified germline variation in the *ETV6* gene as responsible for a specific form of inherited thrombocytopenia[18,22,25,26]. These germline *ETV6* variants were absent from public genomic databases and were shown to be deleterious for protein function. For example, p.R369W affects

the ETS DNA-binding domain at a residue which is structurally involved in electrostatic and protein-protein interactions.

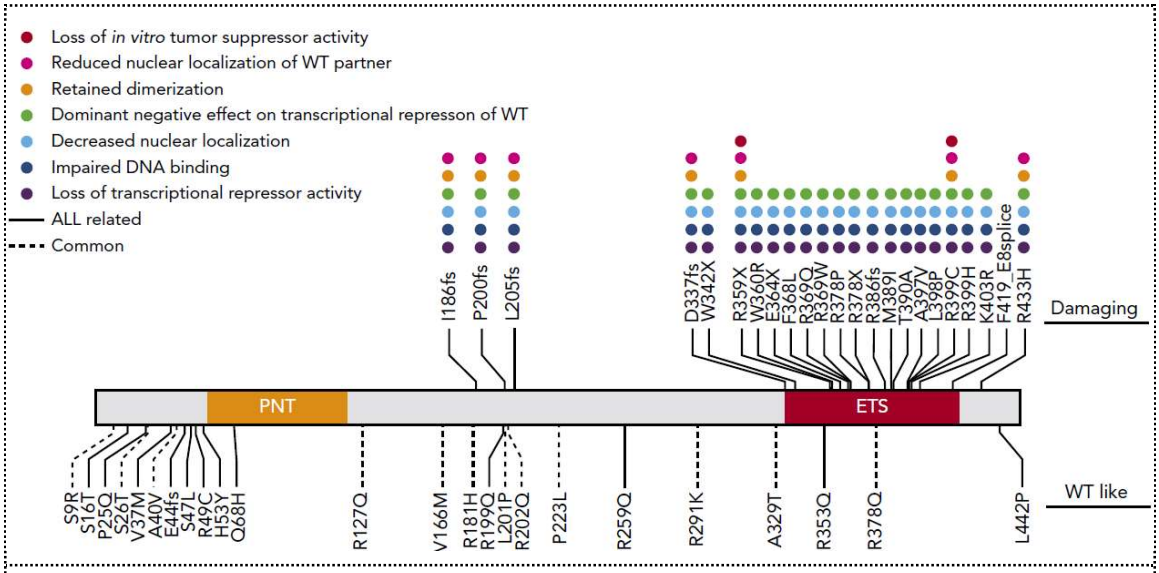

Figure 1.2. Functional characterization of ETV6 variants identified in patients with inherited thrombocytopenia. From Nishii et al., Blood (2021).

Thrombocytopenia results from defective megakaryocytes and impaired proplatelet formation, yielding a low platelet count. In murine studies, disruption of ETV6 after HSC commitment to the megakaryocyte lineage resulted in consistent platelet reduction[5]. The mice exhibited increased megakaryocyte colony formation from the bone marrow, where proplatelet formation after in vitro differentiation was defective[27].

The role of germline ETV6 mutations in B-ALL predisposition

In addition to presenting with thrombocytopenia, B-ALL was the most common malignancy among patients with germline *ETV6* mutations, accounting for approximately 66% of cases[25,28]. These germline variants offer an incompletely

characterized mechanism for predisposition to leukemia, and more specifically B-ALL (Figure 1.2).

B-ALL is the most common cancer in children and amongst the most frequent causes of death from cancer before 20 years of age[29,30]. While the national incidence rate of all ALL is estimated at approximately 1.6 per 100,000 individuals, with 80% of cases occur in children and 75% of patients younger than 6 years old[31,32]. The clinical outcome for pediatric B-ALL patients has dramatically improved, with chemotherapy yielding a greater than 80% 5-year overall survival[31].

B-ALL is usually diagnosed through a bone marrow aspirate and biopsy with morphology, cytogenetics, and flow cytometry. Patients with B-ALL commonly present with fever, fatigue, bone and joint pain, and lymphadenopathy. More than 90% of B-ALL patients present with hematologic abnormalities noted on peripheral blood sampling, including anemia, neutropenia, and thrombocytopenia. Today, chemotherapy has greatly improved survival rates to greater than 90% in many developed countries, with the exact rate varying by age and population[17]. Despite significant improvements in clinical outcomes, a considerable number of patients experience poor response to therapy[33]. Additionally, ~80% of survivors of childhood ALL will experience a post-treatment life-threatening medical event by age 45, an effect hypothesized to result from the intensity and duration of ALL treatment protocols[34]. The impact of long-term health complications of therapy emphasizes the need to define the underlying biology and root causes of this

disease. However, the molecular mechanisms underlying leukemia pathogenesis remain incompletely understood.

Genetic and familial studies have identified germline mutations in *ETV6, PAX5, CDKN2A,* and *IKZF1*, which can significantly increase risk for acquiring ALL[35]. However, the precise mechanisms of action for these predisposing risk variants remain enigmatic.

The development of ALL is thought to follow a two-hit model of leukemogenesis, with initial formation of a pre-leukemic clone and subsequent acquisition of secondary somatic mutations that drive progression to overt leukemia[36]. In pediatric B-ALL, a predisposing mutation is thought to lead to the generation of a preleukemic clone in which normal B cell development is disrupted, but does not necessarily lead to overt leukemic progression[37].

While much of the focus in ALL has been on somatic driver mutations that play a role in this disease, a heritable component for this disease is increasingly being recognized. Inherited genetic predisposition to ALL refers to an increased probability of developing cancer that can be attributed to germline variations. Population studies employing GWAS have robustly identified at least 12 genetic risk loci to date that increase the risk for acquiring ALL[38–42].

While a number of germline variants to ALL risk have been fairly well established through genome-wide association studies, the susceptibility conferred by commonly identified variants is around a 2-fold increase in ALL risk[43]. More challenging to identify, certain rare risk variants such as *ETV6* result in a more dramatic increase in disease risk, but often have incomplete penetrance and the extent to which they impact predisposition is still somewhat unknown: close to 1% of unselected sporadic ALL cases carry potentially damaging and highly penetrant germline variants in *ETV6*[22]. Overall, roughly 25-30% of germline *ETV6* variant carriers develop leukemia, illustrating the requirement for additional genomic abnormalities that must be acquired to induce leukemia.[6]

In a cohort of children enrolled in clinical trials for newly diagnosed ALL, 31 germline *ETV6* variants were observed only in children with ALL or were rare in non-ALL populations[22]. 15 of these variants – 4 nonsense mutations, 10 missense mutations, and 1 splice variant – were clustered within the ETS DNA-binding domain. The hyperdiploid leukemia karyotype was overrepresented in ALL cases harboring germline risk variants in *ETV6*, while the frequency of somatic ETV6/RUNX1 fusion was lower compared to wild-type *ETV6* genomes.

In another study systematically characterizing germline *ETV6* variants, they also identified an increased predisposition towards hematological malignancies, most prevalently B-ALL (Figure 1.2). 65% (22/34) of variants, deemed damaging, exhibited significant impairment of transcription repression compared to wild-type.

83% (5/6) of frameshift variants and all (4/4) nonsense variants predicted to truncate ETV6 resulted in nearly complete loss of repressor activity. As with the previous study, 92% (12/13) of damaging missense variants were located in the ETS domain and exhibited reduced DNA binding ability. 95% (21/22) of damaging variants exhibited a significant loss of nuclear localization. Additionally, a dose-dependent inhibition of wild-type ETV6 activity suggests a dominant negative mechanism of action on repressor activity.

In cell lines and overexpression models, missense mutations in the ETV6 conserved domains result in sustained cytoplasmic localization, decreased transcriptional repression, and impaired MK maturation[15]. The cytoplasmic mis-localization leads to upregulation of proinflammatory interferon response genes, which are regulated by HDAC3. Nuclear exclusion subsequently disrupts the cellular balance of histone acetylation and alter the epigenetic landscape.

As the specific mechanism of how these germline variants increase risk for B-ALL is not yet known, a potential role in regulating lymphoid development lays unexplored.

Role of ETV6 in lymphoid development

Lymphopoiesis, or lymphoid hematopoiesis, refers to the process in which lymphocytes, including B cells, T cells, and NK cells, are generated from HSCs along the lymphoid lineages. As B-ALL arises from improperly developed B-cell

progenitors, it is essential to understand the normal process of differentiation and interrogate how aberrations correlate with disease acquisition and progression. B cell developmental genes are some of the most frequently mutated in B-ALL, highlighting a role for disrupted lymphopoiesis in increasing lymphoid leukemia risk.

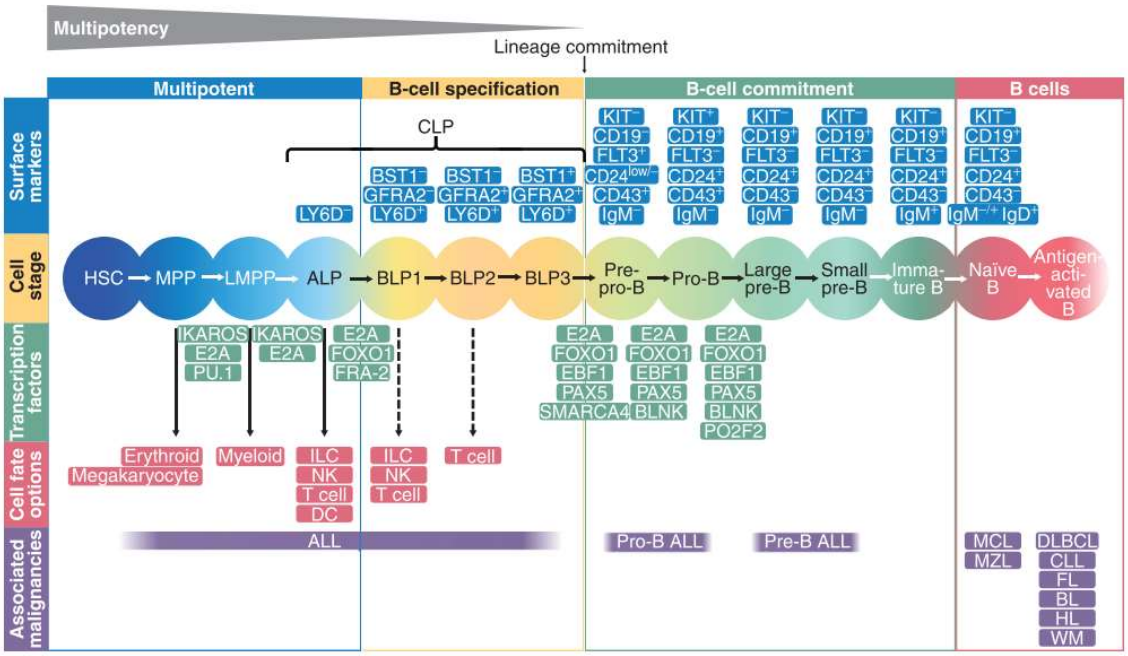

Figure 1.3. The trajectory of an HSC to a B cell is regulated by a complex network of transcription factors and signals that determine cell fate. From Fischer et al., Blood Cancer Discovery (2020).

B cells are derived from hematopoietic stem cells through a multi-step process tightly controlled by a series of transcription factors and signal transduction (Figure 1.3). As pluripotent stem cells can give rise to a variety of cell types, these signals regulate commitment, differentiation, and gene expression.

HSCs begin to lose their self-renewal capacity and transition to multipotent progenitors (MPPs). Subsequently, lymphoid-primed MPPs (LMPPs) still possess myeloid and lymphoid potential, losing erythroid and megakaryocyte potential. LMPPs that shed myeloid potential become common lymphoid progenitors (CLP).

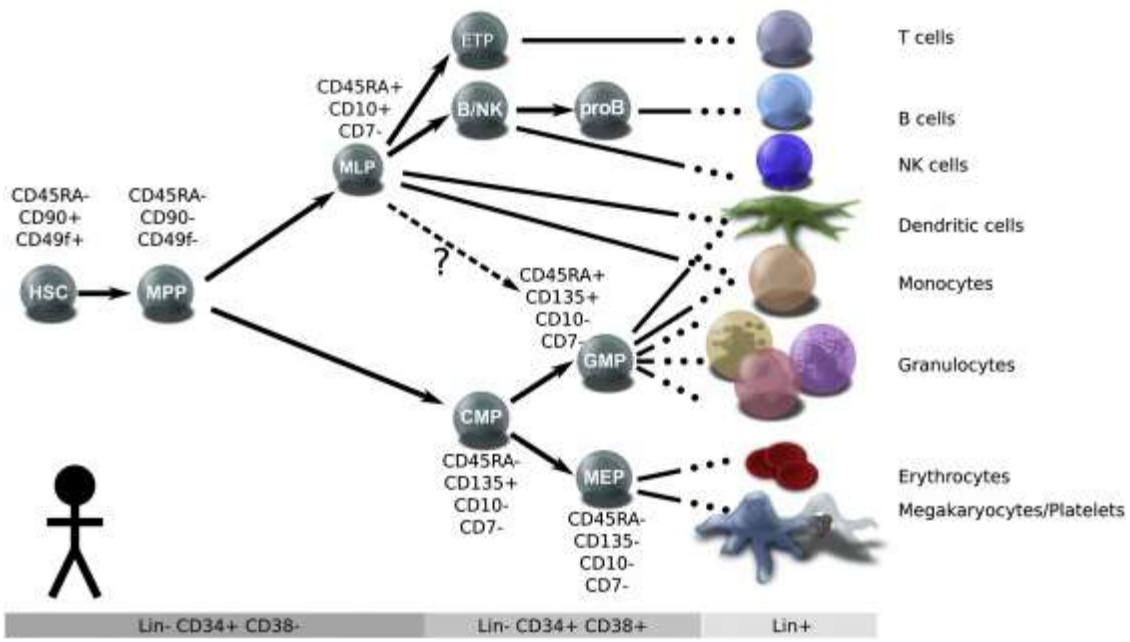

Figure 1.4. **The use of surface markers to classify and delineate stages of differentiation may not be as linear as classical models might suggest**. From Doulatov et al., Cell Stem Cell (2012).

Experimentally, stages of development can be characterized based on the expression of cell-surface markers and other genes, a process which has been both bolstered and hindered by studying mouse hematopoiesis[3]. For example, $CD34^+$ was the first marker found to enrich human HSCs and progenitors, and is used today as a hallmark of HSCs[44]. Subsequently, a number of multipotent intermediates between LT-HSCs and lineage-committed progenitors can be described by expression of select markers. Identifying the appropriate markers at

the corresponding time of differentiation has been difficult to capture both *in vitro* and *in vivo*.

Transcription factors involved in B cell differentiation contribute to the regulation of both lineage-specific gene expression programs and the three-dimensional genomic architecture[45]. Curiously, lineage-specific disruption of ETV6 in murine B cells was not associated with transformation or abnormalities in B cell development[4]. However, familial studies have identified infrequent germline variations in these transcription factors that are strongly associated with leukemia predisposition. Ultimately, the exact role of ETV6 in regulating lymphopoiesis is unknown, as to date no effects on terminal lymphoid maturation have been described.

1.2 – Open questions in the field

Given its identification through chromosomal translocations, somatic, and germline mutations, ETV6 is an intriguing candidate to study germline leukemia predisposition and other hematologic malignancy predisposition. While the function of ETV6 in HSC survival and maintenance has been described, its direct impact on more differentiated cells has not. More specifically, other transcription factors involved in hematopoiesis and more specifically B cell lineages like PAX5 have demonstrated predisposition to B-cell ALL by disrupting key developmental stages[46]. A similar mechanism of action for ETV6 has not yet been confirmed.

In addition, while some groups have described culturing techniques to generate B cells from stem cells *in vitro*, there is little consensus on a reliable method to differentiate and characterize B cells and their progenitors[47–49]. No animal or cell line models of hematologic malignancy resulting from mutant ETV6 have been reported.

As a single genetic alteration is usually insufficient to develop B-ALL outright, the role or mechanism of how it leads to a secondary oncogenic mutation, which can include many types of genetic abnormalities, is not fully understood. Some downstream targets and interaction partners with ETV6 have been described, but more work is needed to comprehensively describe its function as a transcriptional repressor.

Chapter 2

2.1 – Methodology and rationale

The majority of patients with germline ETV6 variants harbor heterozygous mutations in the ETS DNA-binding domain, resulting in loss of function. To recapitulate this effect, peripherally mobilized $CD34^+$ cells were edited with CRISPR/Cas9 single guide RNAs, introducing indels to biologically mimic the heterozygous loss of function seen in patients. I designed single guide RNAs targeting different exons of the *ETV6* gene, including the ETS domain and an early exon to achieve the most efficient knockdown. *ETV6* knockdown was confirmed via Sanger sequencing, qPCR, and Western blot.

I subsequently tested different culture systems to model *in vitro* differentiation from HSPCs to B cells. I characterized edited cells during this differentiation process using flow cytometry to measure expression of classical surface markers.

In more recent work, not fully included in this thesis, I have utilized single cell RNA sequencing to gain further clarity on the effects of *ETV6* editing on lymphoid development.

2.2 – Materials and methods

Culturing stem cells

Primary $CD34^+$ cells were obtained from the Fred Hutchinson Cancer Research Center. Cryopreserved cells were thawed in PBS with 10% FBS and cultured 24 hours prior to nucleofection. Cells were cultured in StemSpan II (STEMCELL #09655) supplemented with L-glutamine (1%), Pen-Strep (1%), StemSpan CC100 (1:100, STEMCELL #02690), UM171 (1:1000), and TPO (1:2000).

Culturing OP9 and MSC stromal cells

OP9 cells were obtained from the Dr. Stuart Orkin's lab at Boston Children's Hospital. Cells were cultured and passaged twice a week in α-MEM (Gibco #) supplemented with FBS (20%) and L-glutamine (1%).

Human Bone Marrow-Derived Mesenchymal and Wharton's Jelly-Derived Stem Cells were purchased from Lifeline Cell Technology (#FC-0057). Cells were thawed and cultured in StemLife Basal Medium supplemented with FBS (10-20%), IGF-1 (15 ng/mL), L-Alanyl-L-Glutamine (2.4 mM), FGF (125 pg/mL), and Pen-Strep (1%). Passaged every 3-5 days, cells were trypsinized, centrifuged at 1500 RPM, and split into 75 cm^2 flasks.

CRISPR-Cas9 editing of AAVS

Cord blood- or mobilized peripheral blood-derived HSPCs were nucleofected 48 hours after thawing. Briefly, ribonucleoprotein (RNP) complex was prepared by

mixing 100 pmol Cas9 (IDT) and 100 pmol modified sgRNAs (Synthego) and incubating at room temperature for 15 minutes. HSPCs were resuspended in 20 uL P3 solution and mixed with RNP in 20 uL Nucleovette strips. RNPs were delivered through electroporation using a Lonza 4D-Nucleofector (Program DZ100). Cells were subsequently plated in 24-well plates for further culturing or analysis.

Table 1. Sequences of CRISPR sgRNAs

Location	Name	Sequence
AAVS	AAVS	GGGGCCACUAGGGACAGGA
ETV6 Exon 2	Sg2-495	GGCGUCGAGGAAGCGUAACU
ETV6 Exon 2	Sg2-534	UCCUCCAUCCUGAGCGCUCG
ETV6 Exon 2	Sg2-537	UUCCAGUGCCUCGAGCGCUC
ETV6 ETS	Sg6-510	GGUACGAAAACUUCAUCCGA
ETV6 ETS	Sg6-514	CGAAAACUUCAUCCGAUGGG

Differentiation co-culture

Stromal cells were plated on 6-well plates 1-2 days prior to co-culture in IMDM supplemented with L-glutamine (1%), Pen-Strep (1%), Human serum albumin (2.5%), IL-7 (30 ng/mL), Flt3L (30 ng/mL), and SCF (100 ng/mL). 48 hours after nucleofection, $CD34^+$ cells were layered on top of the stromal cells. Media was supplemented every 2-4 days. At Week 1 and 2 timepoints, cells were briefly trypsinized and filtered through 40 um filter cap tubes. Cells were transferred on top of a new feeder cell layer.

Sanger sequencing for editing efficiency

Genomic DNA was extracted from harvested cells using an AllPrep DNA/RNA Micro Kit (Qiagen #80284) and quantified using a NanoDrop 2000. Genomic PCR

was performed using Platinum II Hotstart Mastermix (Thermo) in a Bio-Rad T100 thermal cycler. DNA was purified using a Monarch PCR & DNA Cleanup Kit and submitted for Sanger sequencing through Genewiz (Azenta Life Sciences). Indel frequency was calculated using Synthego ICE analysis.

Table 2. Primers for PCR amplification

Name	Sequence
AAVS_FWD	TTCGGGTCACCTCTCACTCC
AAVS_REV	TCCAGGAAATGGGGGTGTGTCAC
Exon2_FWD	TAAGCCGGATTGCTTGGGAAGC
Exon2_REV	GCGTGGCGAAGTCCTGTGAAAA
Exon6_FWD	ACCCAAGCTAGGCAGAAGCAGT
Exon6_REV	ATGATGCAGCCCAGCAAACCAG

Quantitative polymerase chain reaction

RNA was extracted from harvested cells using an AllPrep DNA/RNA Micro Kit and quantified using a NanoDrop 2000. cDNA was generated using an iScript cDNA Synthesis kit (BioRad) and quantified via NanoDrop. qPCR was subsequently conducted using an iQ SYBR Green Supermix kit (BioRad). GADPH was used as a reference gene.

Flow cytometry

Cultured cells were harvested from feeder layers either through gentle pipetting or brief trypsinization. Cells were washed with PBS with 2% FBS and incubated with FcR Blocking Reagent (Miltenyi #130-059-901) for 30 minutes. Cells were subsequently stained with the antibodies listed in Tables 3-5, incubated for an hour dark on ice, and analyzed on a BD LSR Fortessa in the Boston Children's Hospital Flow Cytometry Core.

Table 3. Flow cytometry markers used for OP9 co-culture	
Marker	Fluorophore
Live/Dead	DAPI
CD33	APC (Biolegend, Clone WM53)
CD19	PE-Cy7 (Biolegend, Clone HIB19)
CD34	BV421, PerCP-Cy5.5 (Biolegend, Clone 561)
CD38	BV421 (Biolegend, Clone HB-7)
CD45RA	APC-H7 (BD, Clone HI100)
CD90	PE (BD, Clone 5E10)
CD10	FITC (BD, Clone HI10a)
IgM	PE (Biolegend, Clone MHM-88)
CD20	FITC (BD, Clone L27)

Table 4. Flow cytometry markers used for early lymphoid differentiation	
Marker	Fluorophore
Live/Dead	DAPI
CD34	BV421, AF488 (Biolegend, Clone 561)
CD45RA	APC-H7
CD10	FITC, PE-Cy7 (Miltenyi, Clone 97C5), PE (Beckman, Clone ALB1)
CD33	APC
CD19	PE-Cy7, FITC (Miltenyi, Clone LT19)
CD43	VioBlue (Miltenyi, Clone REA833)

Table 5. Flow cytometry markers used for terminal B cell differentiation	
Marker	Fluorophore
Live/Dead	DAPI
CD19	PE-Cy7, FITC
CD20	FITC
IgM	PE

Western blot

Protein pellets were obtained on the day of cell harvest and lysed in RIPA Lysis and Extraction Buffer (Thermo Fisher). Protein concentrations were quantified using a BioRad DC Protein Assay. Samples were boiled at 90 C for 10 minutes and loaded onto 4-12% Bis-Tris gels with a Precision Plus Protein Kaleidoscope Prestained Ladder (BioRad). Gels were run at 100-200 V for 30-45 minutes at RT. Proteins were transferred to membranes using the Trans-Blot Turbo Transfer

System (BioRad). Membranes were treated with LI-COR Blocking Buffer and incubated with primary anti-ETV6 antibody (ThermoFisher #PA5-81865) overnight at 4 C. After incubation with secondary antibody for 1 hour at RT membranes were imaged using a LI-COR Odyssey.

Single cell RNA sequencing

Follow differentiation co-culture, $CD34^+$ (BV421) and/or $CD19^+$ (PE-Cy7) cells were sorted on a BD FACSAria II at the Boston Children's Hospital Flow Cytometry Core. Gene expression libraries were constructed using Chromium Next GEM Single Cell 3' Reagent Kits (10x Genomics, v.31 Dual Index) and read on a 2100 Bioanalyzer (Agilent). Samples were sequenced on a NovaSeq SP (Illumina). Preliminary analysis was conducted using Cell Ranger (10x Genomics) and Seurat algorithms[50].

2.3 – Results

CRISPR editing of ETV6 replicated heterozygous loss of function

In order to study the effects of disrupting ETV6, we utilized synthetic single guide RNAs (sgRNAs) to genetically modify *ETV6*. My lab has previously described a system of genetically editing cultured HSPCs using sgRNAs[51,52].

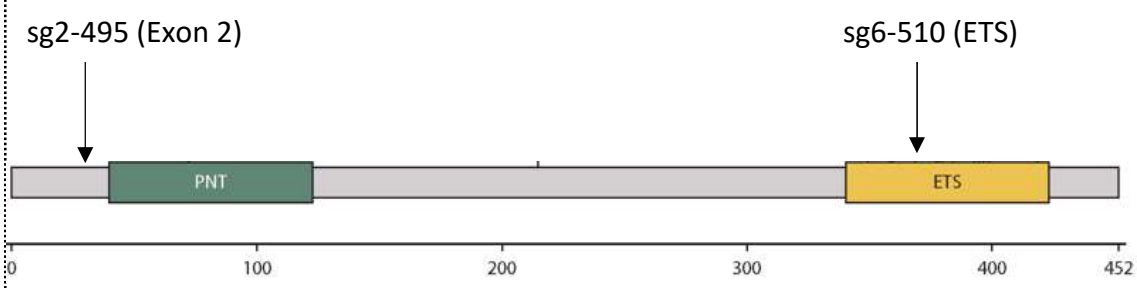

Figure 2.1. Targeted *ETV6* sites of sgRNAs for CRISPR editing. Synthetic single guide RNAs (sgRNAs) were generated from Synthego to target an early exon (sg2-495) and the ETS domain (sg6-510) of *ETV6*. Adapted from Di Paola et al., Blood (2019).

When Cas9 is introduced to DNA, it creates double-strand breaks that are subsequently repaired by non-homologous end joining. This process thus yields insertions and deletions (indels) in the targeted region, resulting in frameshift mutations that knock out the function of the gene. Ultimately, edits introduced by this knockdown served to biologically replicate the *ETV6* variants seen in patients. For my project, I utilized sgRNAs targeting two different locations in *ETV6*: exon 2 (sg2-495), and the ETS domain (sg6-510) (Figure 2.1). By targeting exon 2, I aimed at producing an efficient knockdown of the ETV6 gene. The ETS domain was specifically targeted, given this is a hotspot for pathogenic germline mutations reported in patients. As a control, an intronic safe harbor site (AAVS1) was edited.

HSPCs derived from human cord blood or mobilized peripheral blood were edited via nucleofection and cultured for 48 hours. Sanger sequencing revealed the percentage of indels, quantifying the efficiency of the editing.

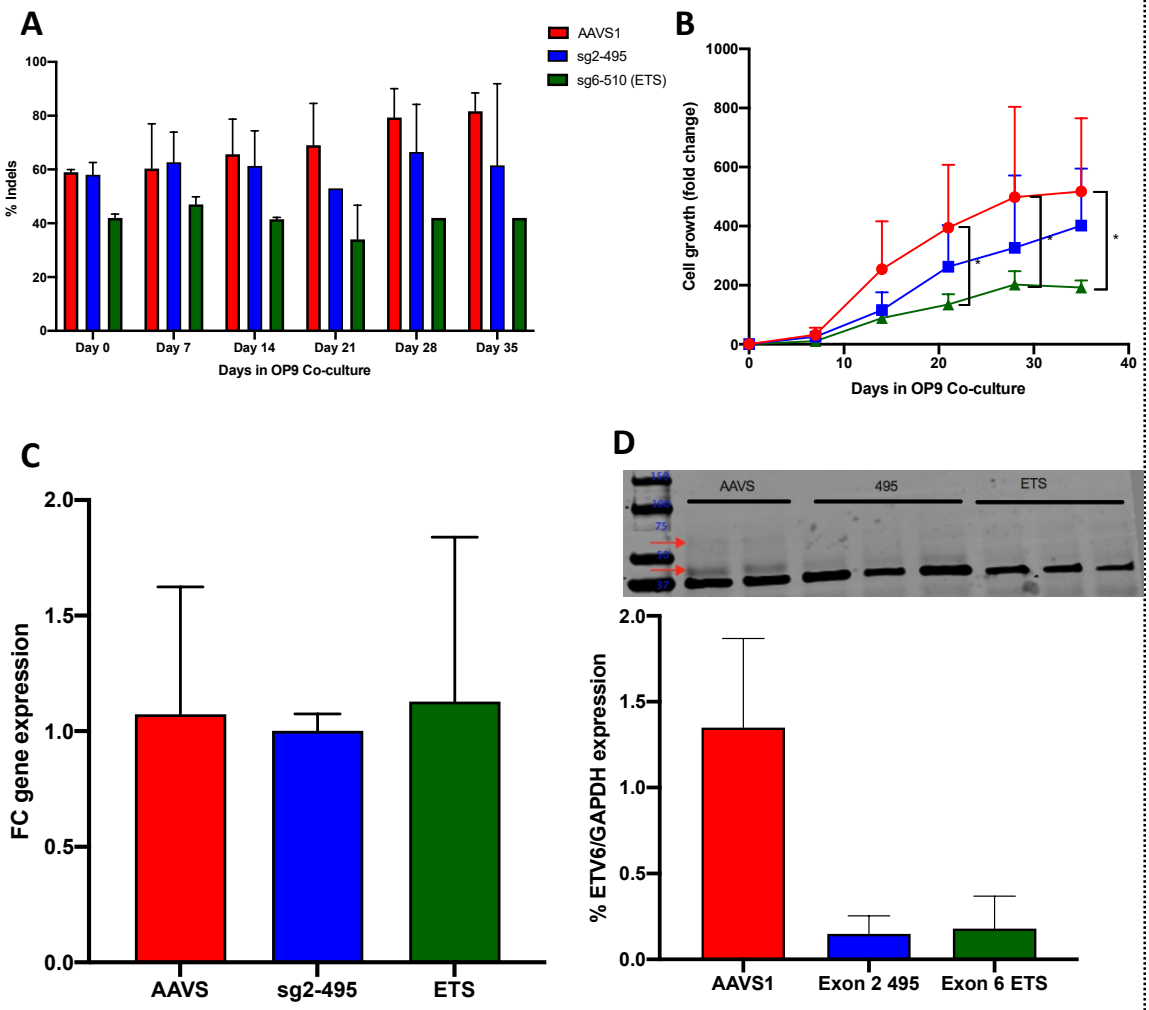

Figure 2.2. **CRISPR editing successfully disrupted ETV6 expression. A**. Sanger sequencing of HSPCs 48 hours after nuclofection confirmed the introduction of indels to the *ETV6 gene* both in an early exon and in the ETS domain. **B**. Initial cell growth analysis suggested a reduction in proliferation potential in edited cells, but subsequent results were variable. **C-D**. While qPCR analysis did not reveal any differences in ETV6 transcript levels as a result of editing, protein expression was significantly reduced, confirming CRISPR-mediated knockdown of ETV6.

Sanger sequencing confirmed the presence of indels in both early exons and the ETS domain of *ETV6* following Cas9 editing (Figure 2.2A). While early experiments suggested an effect of *ETV6* knockdown on overall cell proliferation (Figure 2.2B), the effect was not consistently seen. Western blot analysis confirmed knockdown of ETV6 protein expression, although mRNA levels seemed unaffected by editing (Figure 2.2C-D).

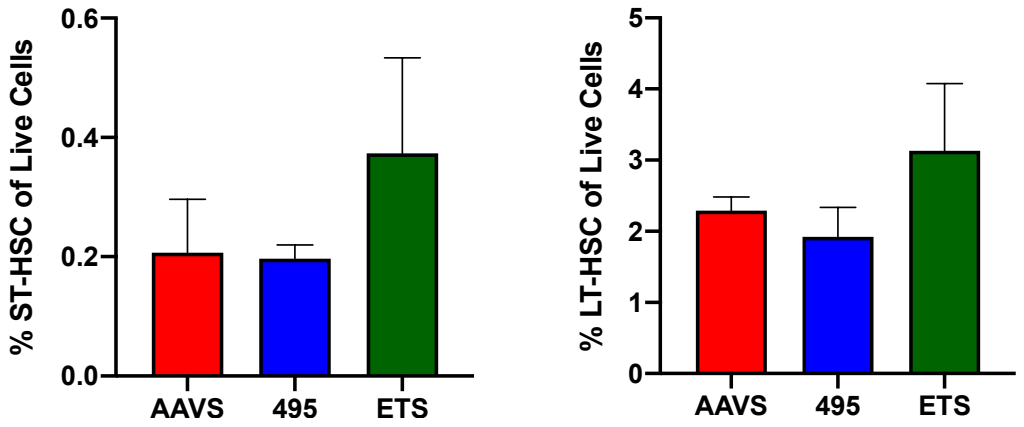

Figure 2.3. Editing of ETV6 did not significantly affect self-renewal capabilities of HSCs. At 72 hours after editing, no significant changes in the phenotypic long-term of short-term HSC compartments were noted by flow cytometry.

First, we assessed how ETV6 editing impacted the composition of the HSC pool. For this, a flow surface marker panel was utilized that examines the composition of the hematopoietic stem and progenitor cell pool[53]. CRISPR editing of the *ETV6* locus did not alter the short-term ($CD34^+CD45RA^-CD90^+ITGA3^-EPCR^+$) or long-term HSC compartments ($CD34^+/CD45RA^-/CD90^+/CD133^+/ITGA3^+/EPCR^+$) at 72 hours after editing (Figure 2.3).

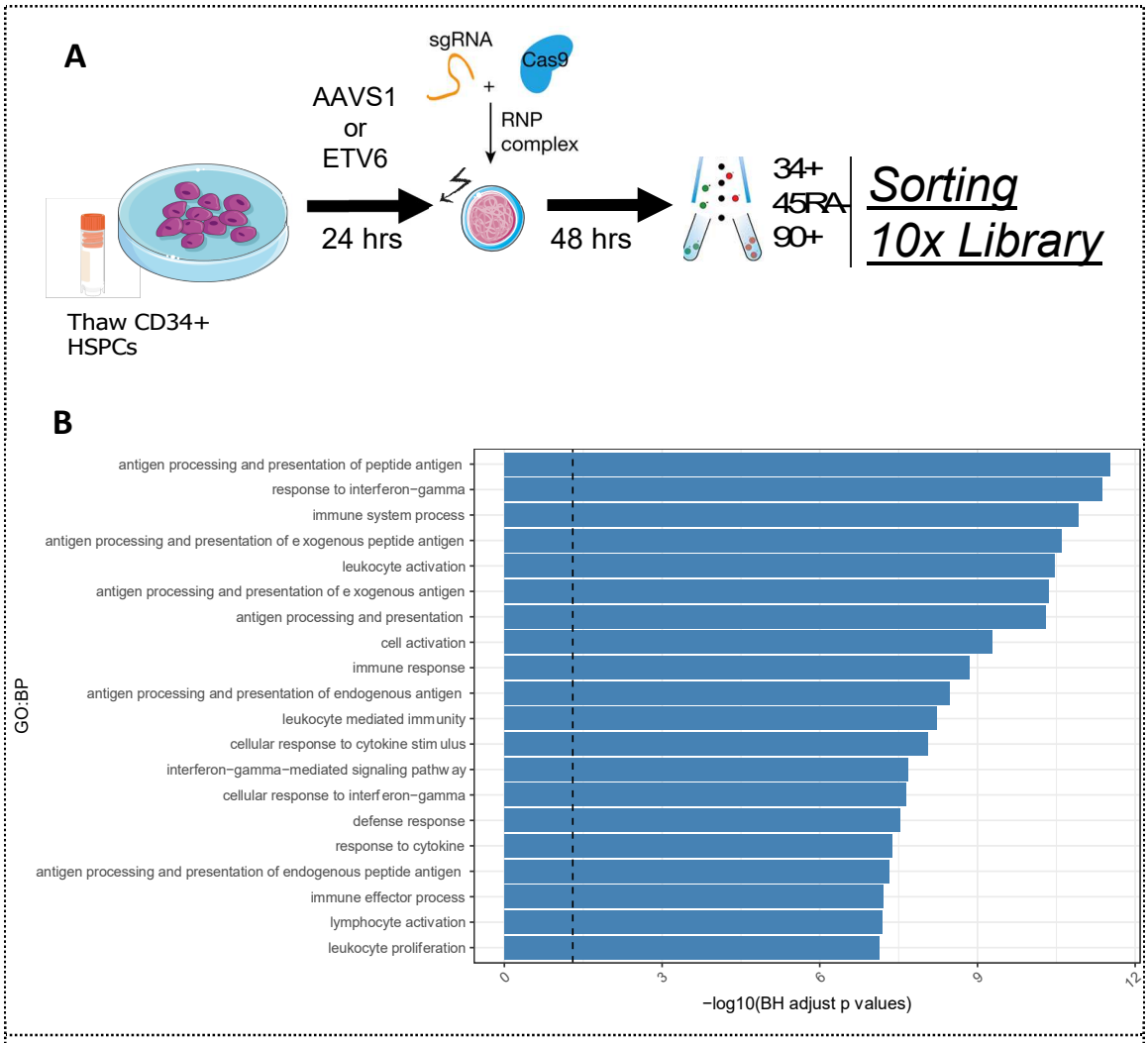

Figure 2.4. Single cell RNA sequencing of ETV6-edited CD34⁺ cells revealed gene expression differences. **A**. Overview of workflow setting up single cell RNA sequencing of *ETV6*-edited CD34⁺ HSPCs. **B**. GO cellular component analysis uncovered significant enrichment for genes involved in interferon response, leukocyte activation, and inflammation pathways.

Next, to evaluate the transcriptional consequences of ETV6 editing in human HSPCs, single cell RNA sequencing was conducted (Figure 2.4A). 48 hours after editing, cells were sorted to enrich for the HSPC compartment using the $CD34^{+}CD45RA^{-}CD90^{-}$ surface markers. Thereafter, single cell RNA sequencing libraries were prepped using the 10x Chromium Controller. Gene ontology analysis

identified a number of pathways of interest that were significantly enriched in edited cells, including interferon response, leukocyte/lymphocyte activation, and cellular immune response (Figure 2.4B).

Development of OP9 co-culture

Once the CRISPR system of genetically editing *ETV6* was established and validated to result in knockdown at the protein level, I designed a culture system that would take edited cells and induce them to differentiate them towards B cells. By comparing the differences between edited and non-edited cells, I could determine the direct effect of ETV6 loss on B cell differentiation.

In vivo, the microenvironment of the bone marrow provides signals for the development of lymphocyte progenitors and subsequent differentiation into B cells. Stromal cells form specific adhesive contacts and provide cytokine and chemokine signaling that control this differentiation and proliferation. Thus, most protocols designed at producing B cells in vitro are based on a co culture of HSPCs with stromal cells, which provide a similar stromal environment to replicate this development in vitro.

The use of OP9 stromal cells to model and drive B-cell differentiation has been previously used to investigate the role of a related ETS family transcription factor[49]. In vivo, MPPs express the receptor tyrosine kinase FLT3, which binds to its ligand on stromal cells[54]. FLT3 signaling is required for differentiation to the CLP stage.

IL-7 produced by stromal cells enables further development of B-lineage cells. Finally, stem-cell factor (SCF) induces proliferation of committed B-cell progenitors. These cytokines were exogenously added in addition to the stromal cell layer and replenished twice a week during co-culture.

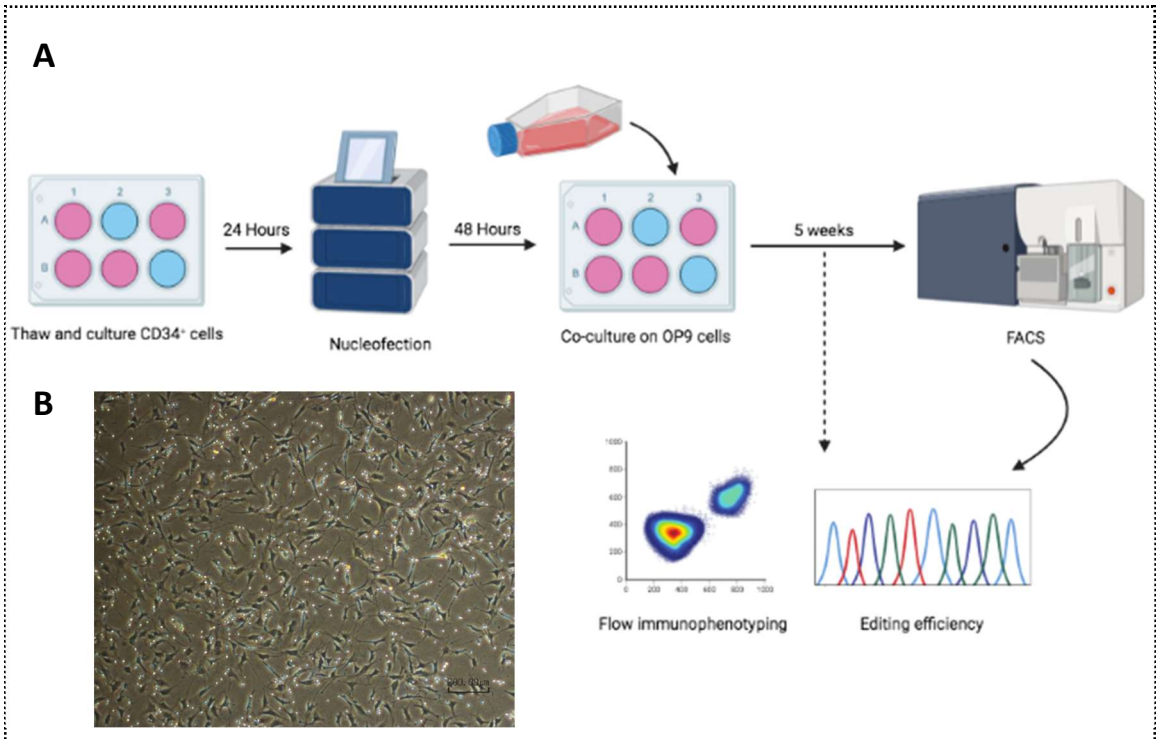

Figure 2.5. Overview of OP9 co-culture experiments. A. CRISPR editing of *ETV6* in CD34⁺ HSPCs was followed by five-week co-culture with OP9 stromal cells. Each week, cells were re-plated on a new feeder layer and assayed by flow cytometry for immunophenotyping and Sanger sequencing for editing efficiency. Created in BioRender. **B.** Microscopic image of differentiating HSPCs (bright, round) with OP9 stromal cells (darker, elongated).

Edited CD34+ cells were co-cultured with OP9 cells and monitored over the course of five weeks (Figure 2.5A). In order to prevent confluence of the feeder layer, differentiating cells were transferred to a fresh OP9 layer every week from the initial date of co-culture. While transferring on a weekly basis, cells were harvested for Sanger sequencing, qPCR, Western blot, and flow cytometry.

Characterizing *in vitro* differentiation of B cell progenitors through flow cytometry

The initial aim of the experiment was to identify and characterize the progenitor populations, based on cell surface markers, over the course of differentiation and compare the effects of ETV6 perturbation. However, the cell populations that arose from co-cultures did not adhere to the canonical and discrete stages of B cell development.

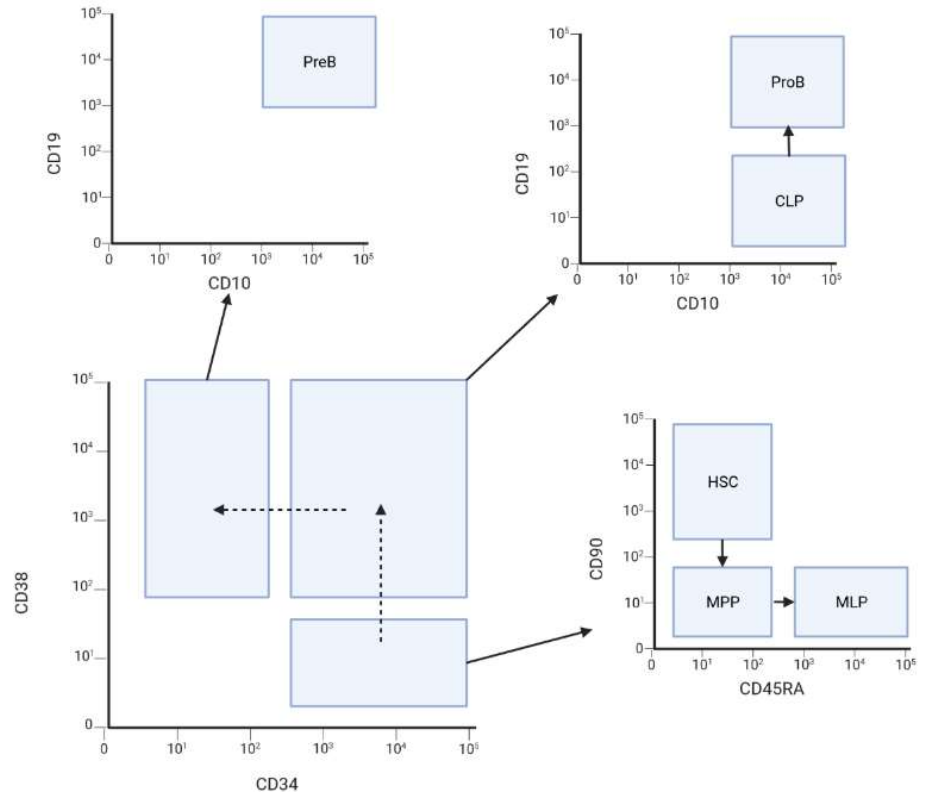

Figure 2.6. Canonical schematic for phenotyping B cell progenitor stages. Created in BioRender and adapted from STEMCELL Technologies.

Based on published, experimental, and computational models of lymphopoiesis, a panel of flow cytometry markers was used to characterize the stages of lymphoid development these cells were progressing through (Figure 2.6). Thus, I expected

to be able to differentiate discrete stages: HSC (Lineage$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^+$), MPP (Lineage$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^-$), MLP (Lineage$^-$CD34$^+$CD38$^-$CD45RA$^+$CD90$^-$), CLP (Lineage$^-$CD34$^+$CD38$^{-/lo}$CD45RA$^+$CD90$^-$), Pro-B (CD34$^+$CD38$^+$CD10$^+$CD19$^-$), and Pre-B (CD34$^-$CD38$^+$CD10$^+$CD19$^+$).

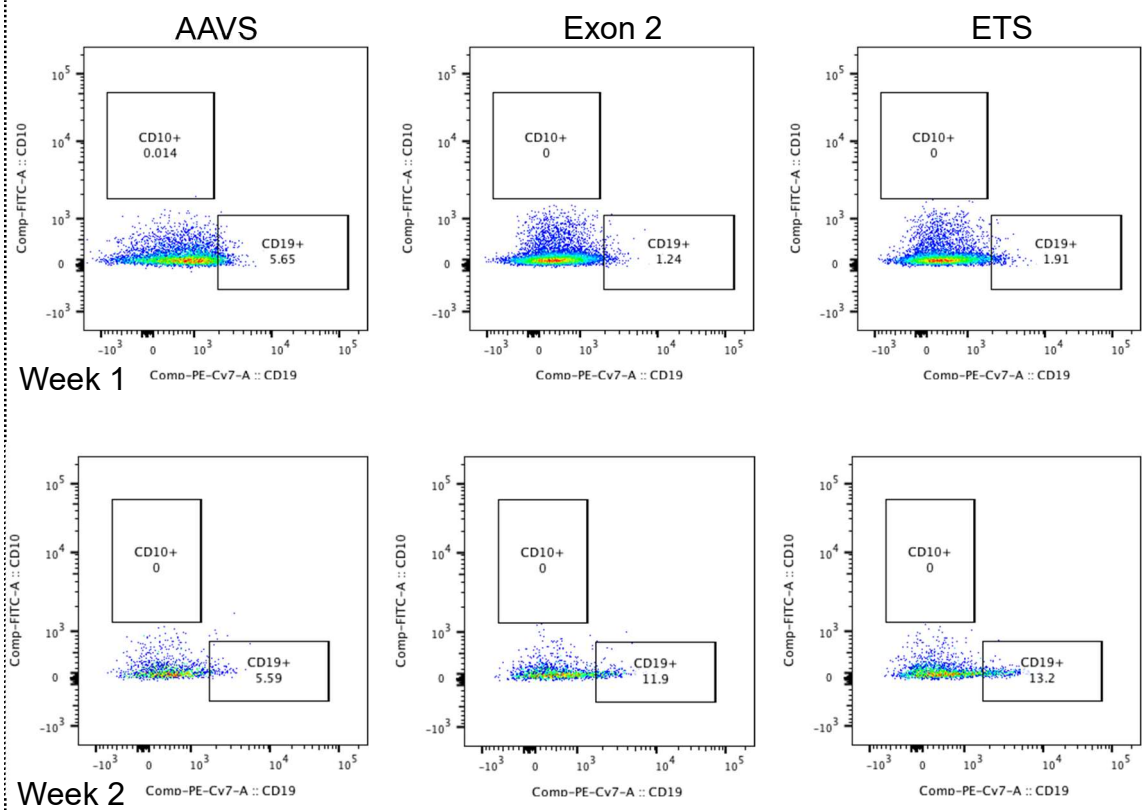

Figure 2.7. Early co-culture experiments did not yield substantial CD10 or CD19 positivity. HSPCs were edited and co-cultured with OP9 stromal cells. After 2 and 3 weeks in co-culture, cells were harvested and immunophenotyped via flow cytometry. After 3 weeks of co-culture with OP9 stromal cells, no CD10 and minimal CD19 is observed.

Based on early experiments, however, the expression of these markers did not neatly correspond with my schematic. At Weeks 2 and 3 in co-culture, minimal CD10 expression was observed. Fewer than 6% of cells (5.65% AAVS1, 1.24% Exon 2, 1.91% ETS) expressed CD19 after 2 weeks and less than 14% of cells

(5.59% AAVS1, 11.9% Exon 2, 13.2% ETS) expressed CD19 after 3 weeks (Figure 2.7). While I noted a potential effect on CD19 expression between edited and non-edited conditions, this effect was not reliably reproduced in OP9 co-culture.

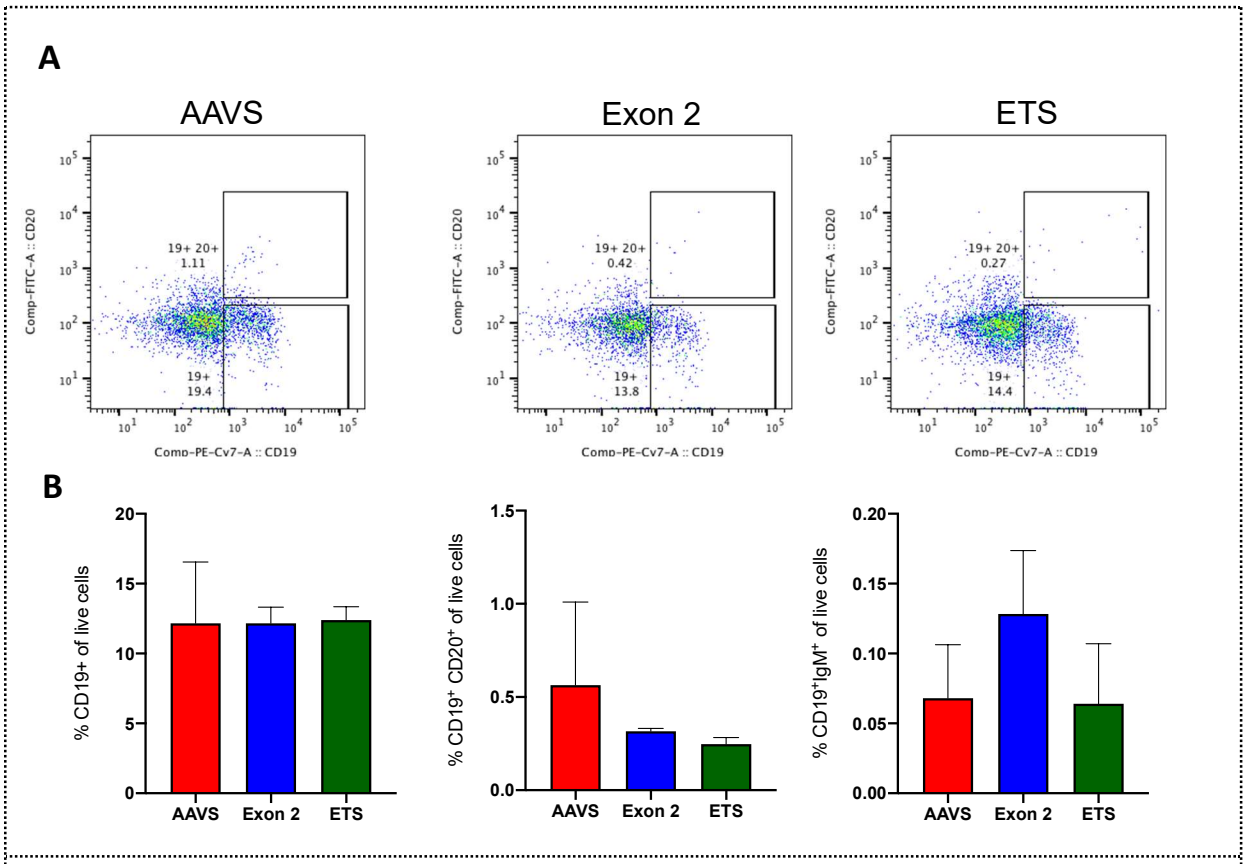

Figure 2.8. OP9 co-culture generated insufficient naïve B cells. After 4 weeks in OP9 co-culture cells were harvested and assayed for markers of more differentiated B cells. **A.** Representative flow cytometry plots of harvested cells illustrating CD19 expression but no CD20 expression. **B.** Flow cytometry revealed minimal CD20 and IgM expression.

Even after 4 weeks of co-culture with OP9 cells, the differentiating cells did not seem to display emergence of early B cell markers, such as CD10, and no terminal maturation was observed. While CD19 reaches almost 20% positivity by Week 4, no co-expression of CD20 or IgM was observed (Figure 2.8B). These results were somewhat discouraging, as combined with the lack of clarity with earlier

differentiation markers it was difficult to systematically track the development of these cells.

ETV6 perturbation in OP9 co-culture alters lymphoid progenitor population

While the overall flow results did not align with the initial panel and schematic design, there was a clear shift in populations at an early time point.

After 7 days of co-culture, 80.7% of ETS-edited cells were $CD34^-$, compared to 62.6% of non-edited cells (Figure 2.9C). The phenotypic effect seemed to reflect the specific targeting of the edit: $CD34^-CD45RA^+$ cells constituted 54.4% of ETS-edited cells, 38.5% of Exon 2-edited cells, and 29.2% of non-edited cells. There was a notable reduction in $CD38^+$, 34.3% in ETS-edited cells, 46.9% in Exon 2-edited cells, and 62.5% in non-edited cells (Figure 2.9D).

Since the results of early co-culture experiments suggested that the trajectory of differentiation *in vitro* was differing from a canonical model (Figure 2.6), I decided to revisit and reanalyze the culture using a different panel and
gating strategy (Figure 2.10). This strategy did identify a population that was significantly affected by ETV6 editing.

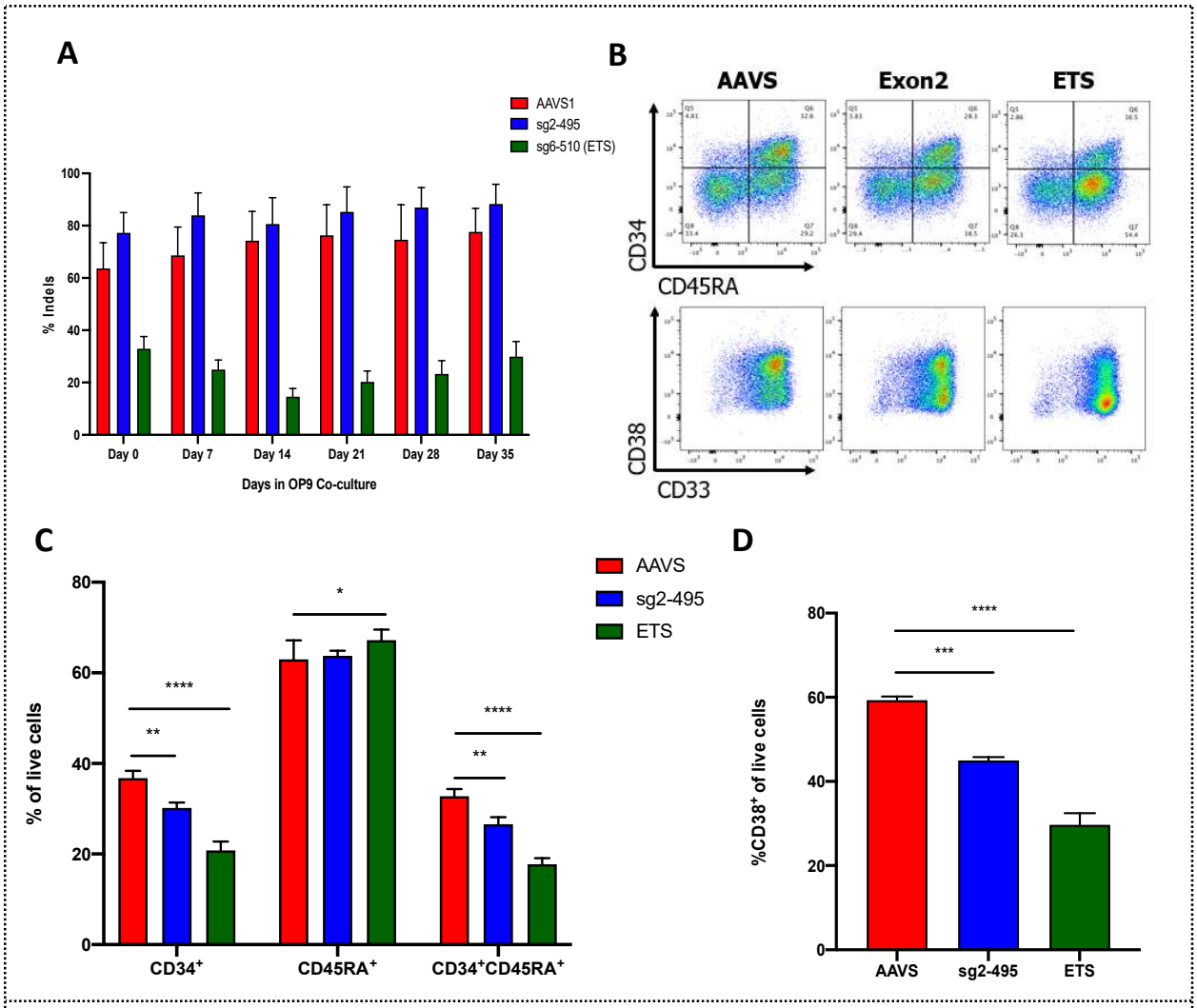

Figure 2.9. ETV6-edited cells exhibited shifts in early differentiation markers. A. Corresponding editing efficiencies of ETV6-edited HSPCs co-cultured with OP9 stromal cells. **B.** Representative flow cytometry plots of harvested cells expressing markers of early differentiation. **C.** After one week, ETV6 perturbation, especially in the ETS domain, resulted in a decrease in CD34-expressing cells, also expressing CD45RA. **D.** ETV6 perturbation also reduced CD38 positivity after one week in co-culture.

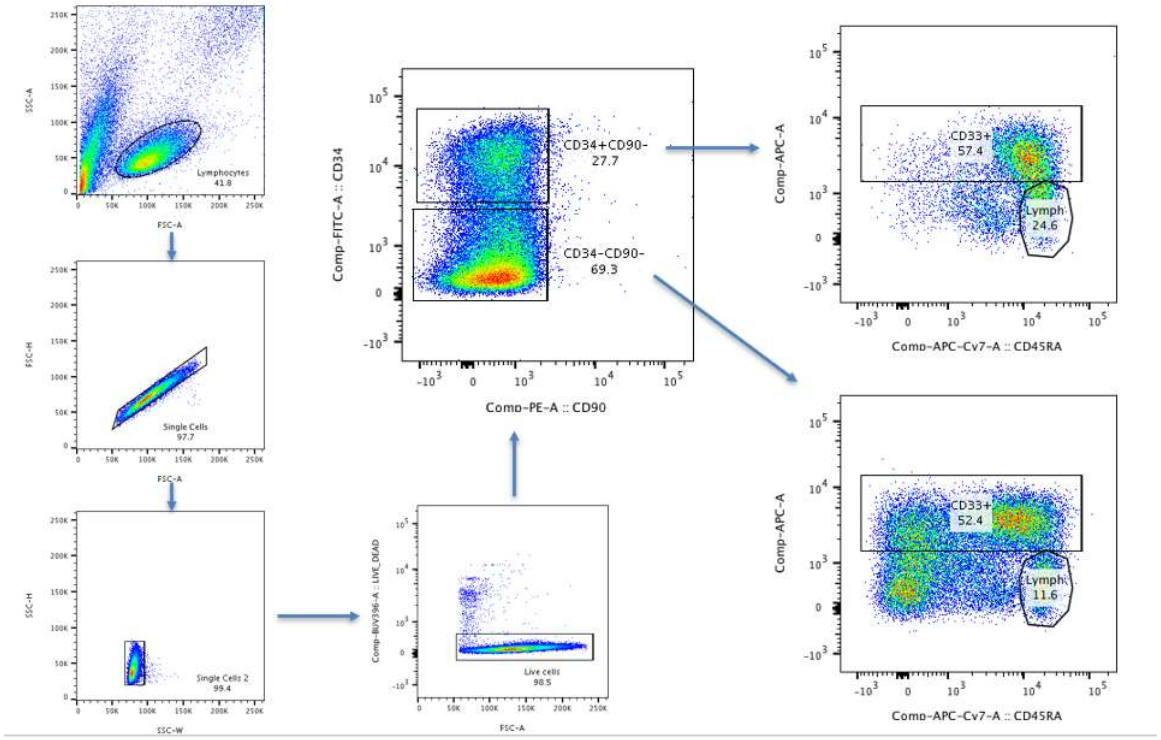

Figure 2.10. Gating strategy for identifying progenitor populations. Doublets and dead cells were initially excluded, followed by gating for CD90$^-$ cells. CD34$^+$ and CD34$^-$ were separately analyzed for differences as a result of ETV6 editing.

After 1 week in OP9 co-culture, a lymphoid progenitor population (CD34$^+$CD90$^-$ CD45RA$^+$CD33$^-$) was shown to be significantly reduced in ETV6-edited conditions (Figure 2.11B). These cells can loosely be described as progenitors that have lost multipotent “stemness” as well as myeloid potential. While this effect suggested a potential dose-dependent reduction, the most significant reduction occurred in ETS-edited conditions, independent of editing efficiency. When cells were edited in the ETS domain, a 2-fold to 10-fold decrease in this population was observed.

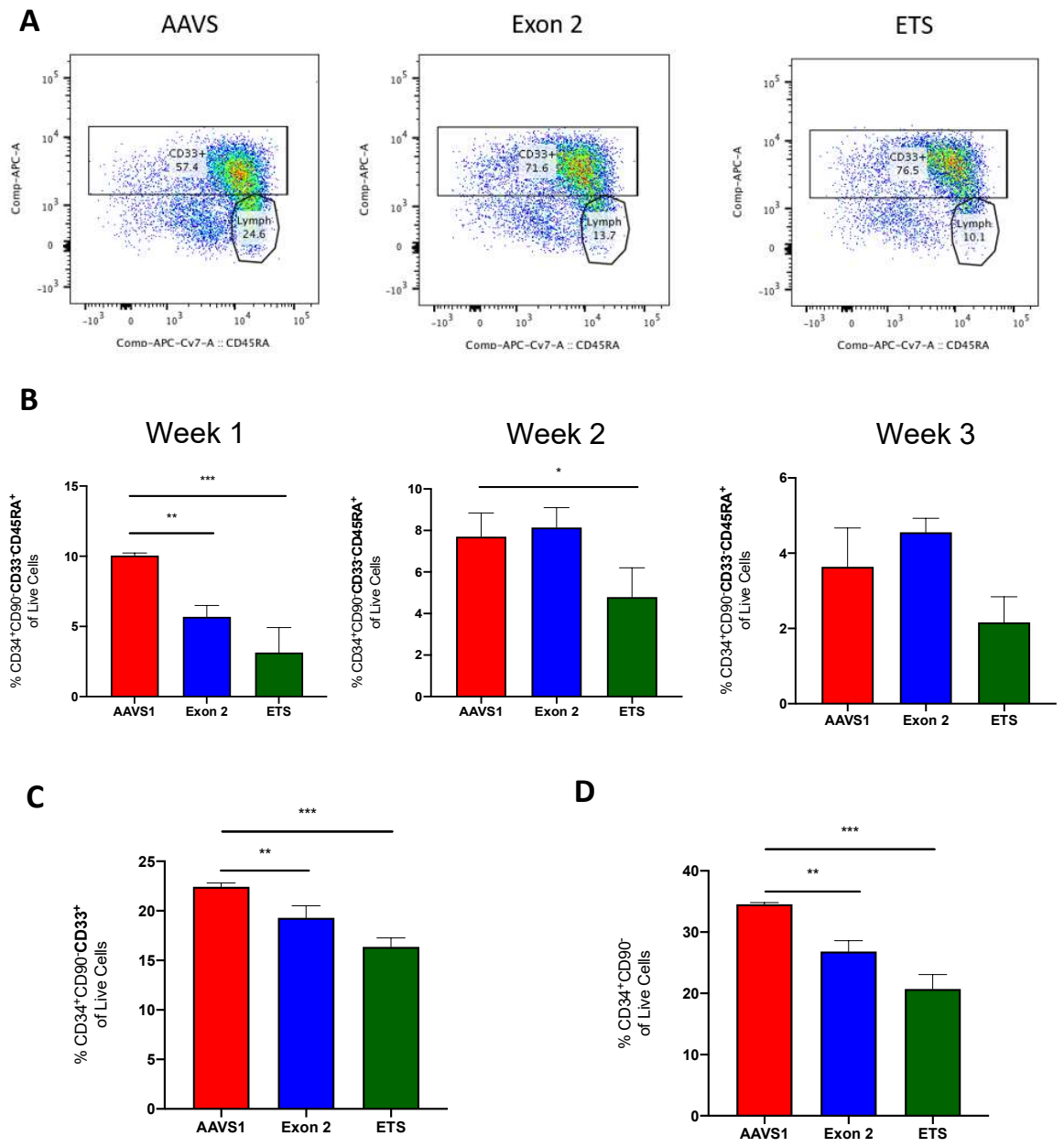

Figure 2.11. ETV6 editing resulted in a reduction in a specific lymphoid progenitor population. A. Representative flow plot illustrating the result of ETV6 knockdown on a lymphoid progenitor compartment, defined by $CD34^{+}CD90^{-}CD45RA^{+}CD33^{-}$. **B.** This phenotype is most significant at 1 week in co-culture, and subsequently no effect is seen. **C-D.** Myeloid progenitors and total progenitor populations are disproportionately reduced in ETV-edited conditions.

While this reduction might suggest an increase in myeloid progenitors, suggesting an imbalance in lineage potential, the observed proportion of myeloid progenitors also decreased in ETV6-edited conditions (Figure 2.11C). Given that the total number of progenitors exhibited a smaller decrease, it is likely that this phenotype is indeed a reduction specifically in the lymphoid lineage.

In subsequent weeks of differentiation, the difference between edited and non-edited conditions disappeared. This suggested that the effect ETV6 has on lymphoid differentiation occurs early on in the process and results in a transient phenotype. This reduction, ultimately, was seen in both cord blood-derived and mobilized peripheral blood-derived $CD34^+$ HSPCs.

Phenotype is recapitulated in MSC co-culture

Given the limited success of the OP9 co-culture in generating mature B cells at the end of 5 weeks, I turned to a different protocol. The use of mesenchymal stem cells (MSCs) as a feeder layer has been more recently tested, but lacks substantial supporting literature. Utilizing a different feeder layer, it was important to establish whether the cells would exhibit a similar phenotype while improving differentiation efficiency.

MSCs were derived from adult bone marrow (BM) or Wharton's jelly (WJ), a component of the umbilical cord. Given that the feeder cells were a human-derived

cell line rather than mouse-derived, I was optimistic about the potential of this stromal layer.

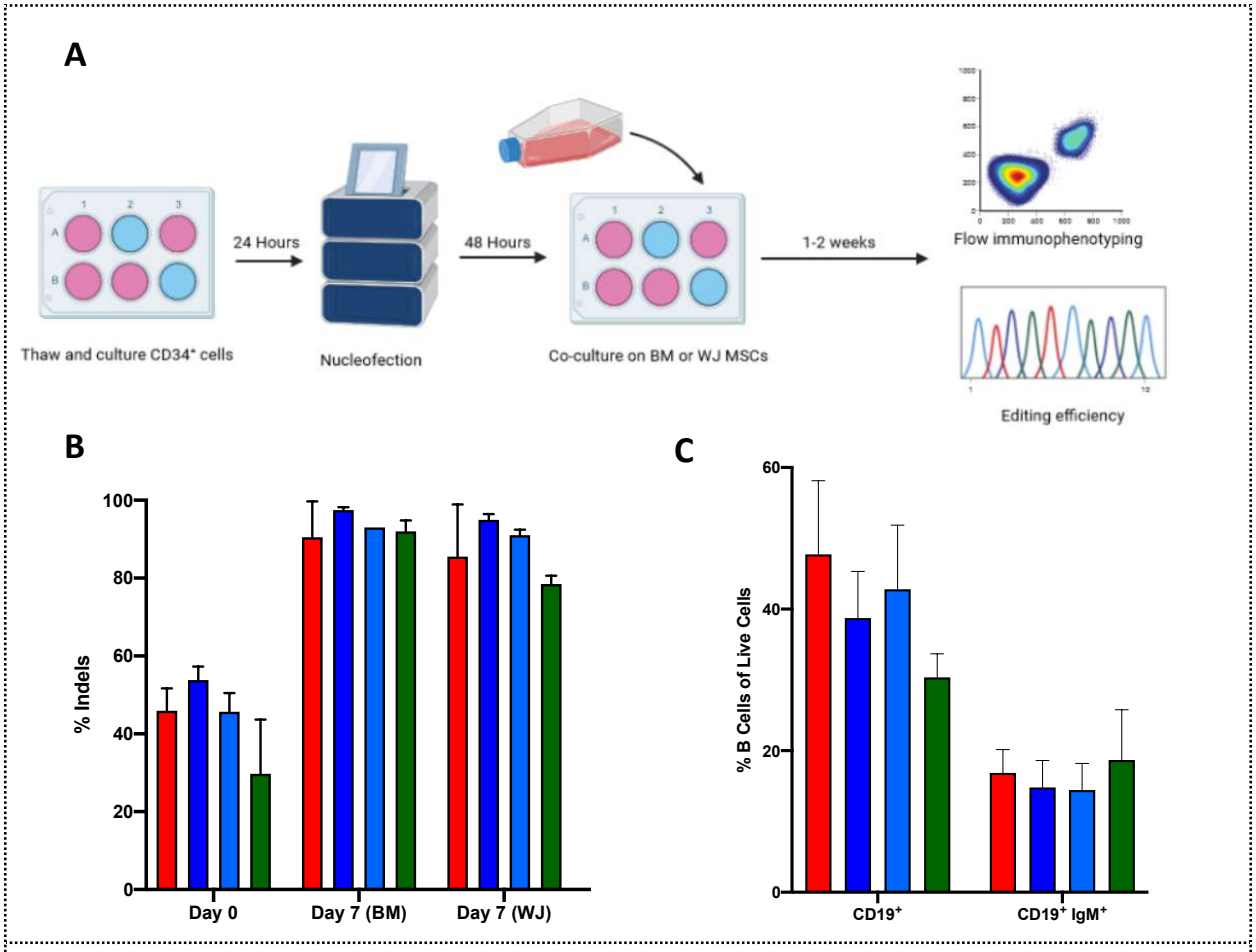

Figure 2.12. Edited cells co-cultured with BM-MSCs demonstrated increased potential for B cell differentiation. A. CD34⁺ HSPCs were similarly edited at an early exon and the ETS domain of *ETV6*. 48 hours after nucleofection, these cells were co-cultured with either bone marrow- or Wharton's jelly-derived mesenchymal stem cells. Differentiating cells were re-plated in co-culture every week. Created in BioRender. **B.** Representative editing efficiency of a BM-MSC experiment. **C.** Edited cells in MSC co-culture yielded higher B cell output.

The general workflow of using MSCs was very similar to that of OP9 cells, slightly altering the reagents used to culture them (Figure 2.12A). While the overall process only required 2-3 weeks, differentiating cells still needed to be transferred

to new feeder layers on a weekly basis, allowing for similar phenotyping via flow cytometry.

The overall output of later markers of B cell differentiation was significantly increased. Both edited and non-edited conditions yielded >30% CD19-expressing cells after 1 week in co-culture (Figure 2.12C). As an improvement to the OP9 co-culture, anywhere from 10-20% CD10 positivity was also observed after one week in differentiation co-culture. After 2-3 weeks, up to 20% of cells expressed IgM. Taken together, these results suggest that MSCs were a suitable replacement as a stromal layer to induce differentiation of B cells *in vitro* following CRISPR editing of *ETV6*.

Subsequently, I set out to determine whether the trajectory of differentiation using this new layer differed, and whether I could capture the same reduction in phenotype I had observed after 1 week of differentiation. While establishing this co-culture I was simultaneously troubleshooting issues regarding editing efficiency, so I included an additional sgRNA targeting an early exon of *ETV6*.

Cells co-cultured with BM-MSCs similarly presented a reduction in $CD34^{+}CD90^{-}CD45RA^{+}CD33^{-}$ cells after 1 week of differentiation (Figure 2.13A). Cells edited with sgRNA targeting early exons exhibited a roughly 2-fold decrease in this progenitor population, somewhat dependent on editing efficiency (Figure 2.13B). Using MSC co-cultures, the lymphoid progenitor population was nearly abolished

in ETS-edited conditions. While cells in WJ-MSC co-cultures yielded similar levels of immature B cells at the end of the co-culture, the phenotype was not clearly seen, prompting me to primarily use BM-MSCs for subsequent experiments (Figure 2.13C).

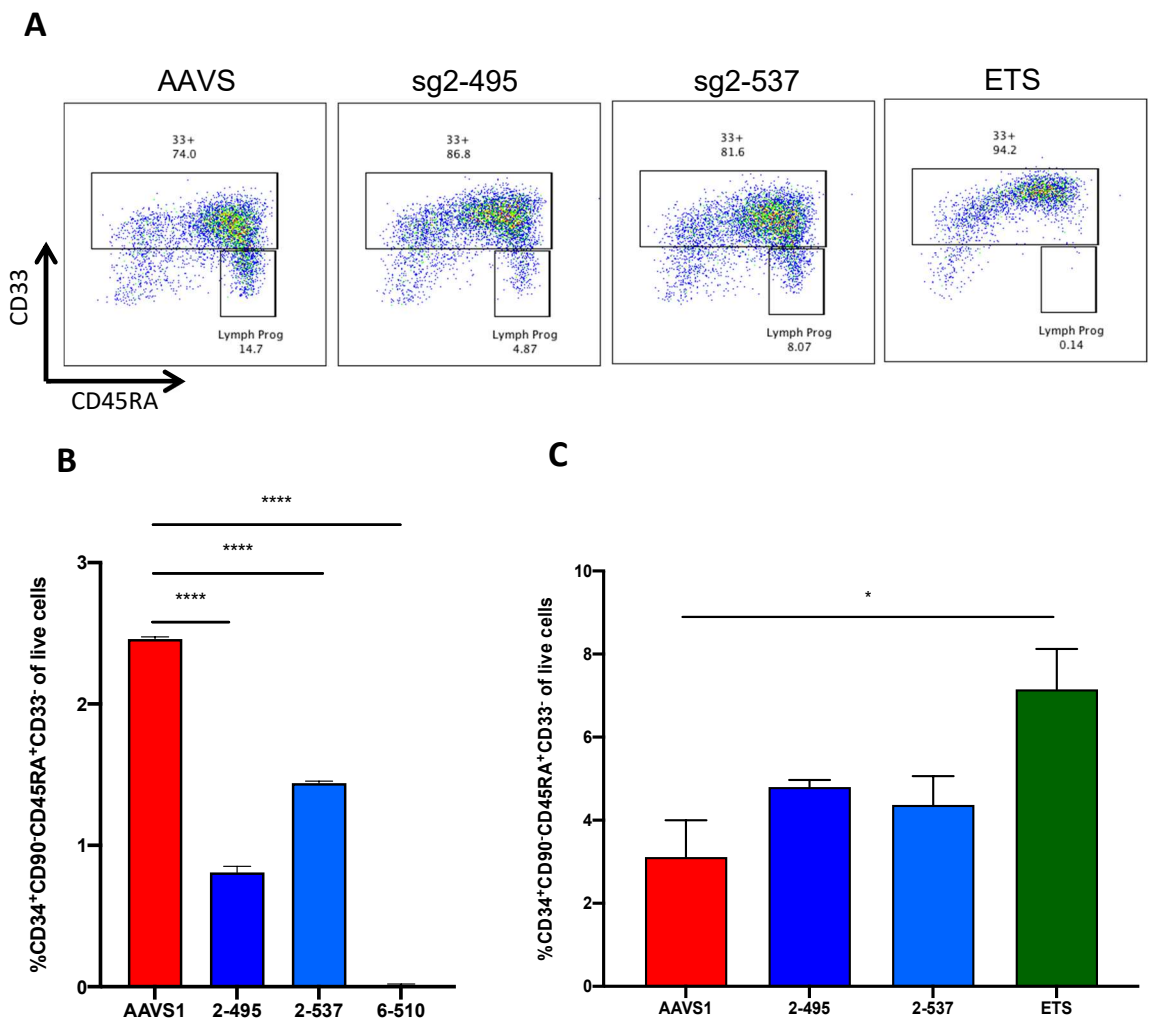

Figure 2.13. Edited cells co-cultured with BM-MSCs similarly exhibit a reduction in lymphoid progenitors. A. Representative flow cytometry plots of edited HSPCs after 1 week in co-culture with BM-MSCs. sg2-537 represents an additional sgRNA targeting an early exon of *ETV6* to confirm the effects of high editing efficiency. **B.** The BM-MSC co-culture exhibited a similar reduction in $CD34^{+}CD90^{-}CD45RA^{+}CD33^{-}$ cells after 1 week, most significantly reduced in the ETS condition. **C.** Cells co-cultured with WJ-MSCs did not recapitulate the previously observed phenotype.

Utilizing single cell RNA sequencing

Flow cytometry has provided evidence that coupling CRISPR knockdown with differentiation co-culture offers insight into the role of ETV6. However, the large variability and lack of discrete, stage-specific expression dictates the need for a higher resolution technique to identify intermediary cellular stages of differentiation.

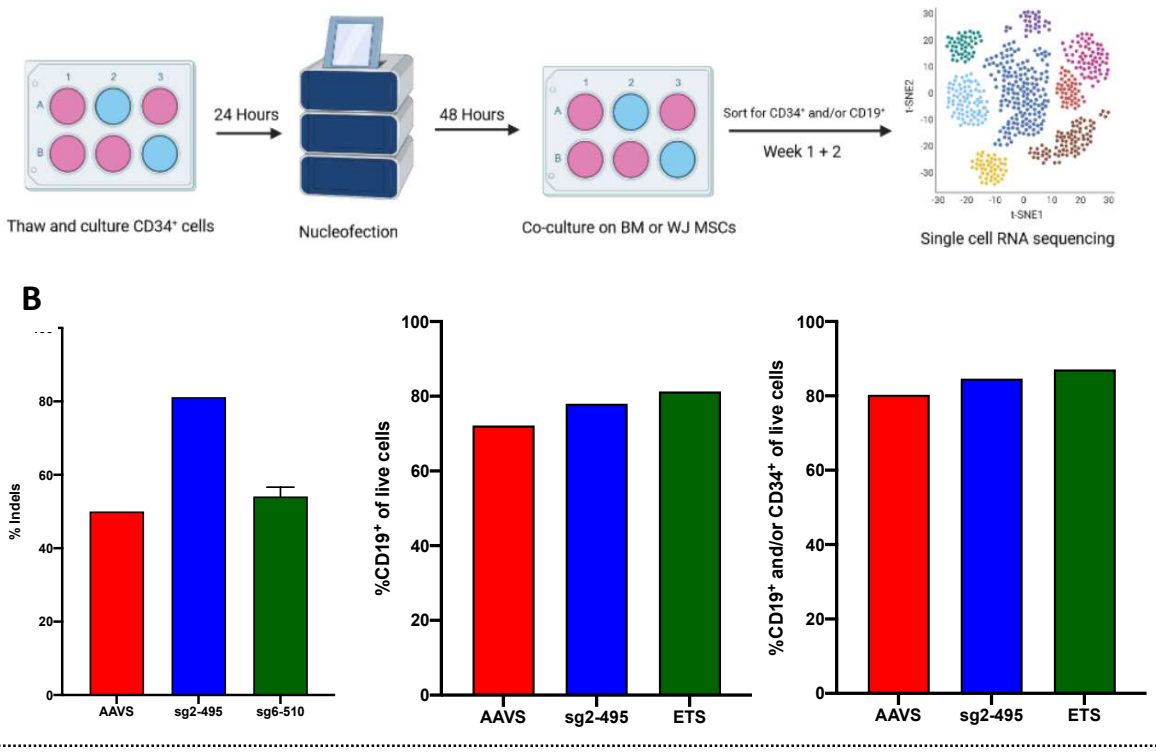

Figure 2.14. Preliminary set-up for single cell RNA sequencing to further elucidate effects of ETV editing. **A.** Overview of workflow coupling CRISPR editing of *ETV6* with single cell RNA sequencing. Created in BioRender. **B.** Editing efficiency and surface marker expression of sorted cells submitted for sequencing.

To enable identification and characterization of the progenitor populations produced through this co-culture system, I have recently established a single-cell RNA sequencing experiment (Figure 2.14). This will enable me to further elucidate

the pathways and processes that are impacted by ETV6, as well as more specifically characterize the cells of interest previously identified by flow cytometry.

While the analysis of this sequencing data is currently ongoing, I hope it will provide a substantial amount of data in service of highlighting genes and pathways which are implicated as a result of ETV6 perturbation.

2.4 – Summary of results

In summary, here I present an *in vitro* model that enables functional studies of how genetic perturbation of a transcription factor affects human hematopoiesis. By utilizing CRISPR-mediated gene editing, a familiar technique in our laboratory, variable but sufficient knockdown of *ETV6* in both an early exon and the ETS domain were achieved. This perturbation of ETV6 did not significantly alter cell proliferation or ETV6 transcript levels, but resulted in reduced ETV6 protein expression. Edited HSPCs driven towards B cell differentiation exhibited an altered trajectory in early lymphoid development. Specifically, I observed a reproducible flow cytometry phenotype of reduced $CD34^{+}CD90^{-}CD45RA^{+}CD33^{-}$ cells during early differentiation. Ultimately no significant change in terminal B cell differentiation was seen, although some data suggests a transient increase in CD19 expression in early development. These results not only confirm a role for ETV6 in regulating lymphoid development, but also support the use of *in vitro* co-culture to assess factors regulating B cell differentiation.

Chapter 3
3.1 – Discussion

In this thesis, I set out to further investigate the role ETV6 might have in clinically predisposing individuals to B-ALL by observing its impact on B cell development. To this end, I replicated the heterozygous loss of function seen in germline ETV6 variants and induced B cell differentiation *in vitro* to measure the effects of editing and identify errors in this process.

It is evident that ETV6 plays a functional role in the development and/or proliferation of B cells and their progenitors. It is likely that this affects an early progenitor population and did not demonstrate any significant change in overall terminal IgM output. While not consistent enough to warrant official publication in this thesis, more recent experiments provided evidence pointing towards an increase in CD19-expressing cells when *ETV6* is edited.

How can one reconcile the reduction of the $CD34^{+}CD90^{-}CD45RA^{+}CD33^{-}$ population with a potential increase in $CD19^{+}$ expression in early weeks of differentiation? It might be reasonable to hypothesize that ETV6 perturbation causes an aberrant acceleration in the differentiation process. At one week in co-culture, these cells might have progressed past the early MPP stage in which we would expect to observe them, and express in higher quantities markers of more committed B cell lineage. Such an effect would result in improper development of a B cell progenitor that is able to proliferate and produce preleukemic clones. While

these cells are still able to fully develop into B cells, indicative of a clinically silent mutation, they are perhaps more susceptible to additional somatic mutations that trigger leukemic progression.

While the population identified through flow cytometry was highly reproducible and consistent based on *ETV6* editing, the exact identity of this progenitor population is not fully understood. The panel of surface markers used to characterize this compartment broadly classifies an early progenitor pool that has lost certain lineage potential, but does not necessarily denote a fully committed lymphoid cell. To this end I have been working towards analyzing single cell RNA sequencing data, which was unfortunately not ready at the time of writing this thesis. As mentioned previously, I hope that the sequencing analysis will help to not only better characterize the progenitor population identified through flow cytometry, but also identify genes and pathways highly correlated with ETV6 perturbation.

Beyond the identification of changes in surface markers and generating B cells *in vitro*, the broader goal of this project was to identify potential mechanisms ETV6 might have in predisposing individuals to B-ALL. In my introduction, I highlighted several functional characteristics of ETV6 mutants and how they might contribute to disease progression.

Understandably, genes more directly involved in B cell development specifically have had more studies describing their involvement in leukemia risk. In some

genes, for example PAX5, these mechanisms have been linked to infection exposure or other immune processes[55]. Given the clear role of functional B cells for fighting infections and regulating the immune system, there is a growing appreciation for immunological mechanisms of predisposition to B-ALL. Indeed, it will be interesting to investigate how ETV6 perturbation in conjunction with immune challenge affects lymphoid development.

Through this project I have developed a system that combines CRISPR editing, in vitro culturing, and sequencing platforms to model the effects of genetic perturbation on hematopoietic processes. To study this system, I utilized three separate models of co-culture (OP9, BM-MSCs, and WJ-MSCs). These cell lines were quite variable in their proliferation and stimulus tolerance, which made consistent comparisons quite difficult. Over time, reanalyzing and reevaluating early experiments, I have found benefits and challenges to using each model, as well as identified technical adjustments to increase confidence in the data. Nevertheless, each co-culture model has contributed to the generation of data in this thesis. Ultimately there is still much work to be done to more comprehensively understand the role ETV6 plays in regulating lymphoid differentiation.

3.2 – Challenges and shortcomings

CRISPR editing

I have spent much of the past year troubleshooting and working to achieve high indel efficiency in my editing. The challenges are both technical and biological. The efficacy of Cas9 reagents had some variability between batches, prompting cancellations of a number of experiments due to low editing efficiency. The experimental process of introducing Cas9 and sgRNAs into the cells is a somewhat technical protocol, which can always introduce human error or variability. Given that the ETS domain presented significantly more challenges in achieving high editing efficiency, I am confident that the obstacles are more rooted in biology, given the critical role ETV6 plays in HSC homeostasis.

In addition, there was some concern whether the levels of editing in *ETV6* would accurately reflect the haploinsufficiency seen in patients. There were numerous discussions regarding the optimal level of editing and whether subsequent effects seen would be as a direct result of ETV6 or other factors, such as a dominant negative effect. In future studies I would also hope to utilize base editor platforms to introduce specific point mutations. Nevertheless, I am confident that the models used for this thesis faithfully emulate the biological effects of *ETV6* mutations.

Modeling in vitro B cell development

Similar to editing, I spent a significant amount of timing working to troubleshoot and optimize the co-culture models to reduce variability and increase confidence

in the data. While the use of OP9 and MSC layers have been previously described, they have had mixed results in my experiments. The proliferation and survival rate of the cell lines are quite variable. The confluence of the feeder layer, in addition to the concentrations of exogenous cytokines and culture reagents, seemed to significantly impact the growth and differentiation potential of the edited cells.

Flow cytometry markers corresponding to discrete developmental stages

Canonically, the stages of B cell development can be described by the expression of surface markers. Experimentally, it is clear that these markers can only be observed transiently and transitionally rather than discretely. A number of surface markers in my panels did not express at expected levels, and showed high variability across machines and antibody lots. Different fluorophores for the same marker would yield differing levels of expression in the same cell population, casting doubt on many of the flow experiments. Surface expression of CD19 was influenced heavily by compensating CD45RA in the same flow cytometry panel. Given that Cy7 and H7 have emission peaks at 779 nm and 782 nm respectively, there is certainly spectral overlap between the two channels. Given the growing body of literature suggesting that the hematopoietic process is less linear and binary, it is important to appreciate that the trajectories of lymphoid differentiation, especially *in vitro*, may not be fully observable by traditional flow cytometry.

Setting up single cell RNA sequencing

Technically, attaining all the conditions necessary to submit a sample for single cell RNA sequencing was difficult. The edited $CD34^+$ cells and BM-MSCs must survive consistently for more than two weeks, especially after electroporation, and proliferate enough to sort. The nucleofection must not yield an editing efficiency too low in order to confirm editing or too high to mimic haploinsufficiency.

In the first run, the AAVS1 control samples were not successfully edited. In the second run, the flow sort did not reliably capture the desired $CD34^+$ and $CD19^+$ populations based on compensation issues. The fragmentation step during scRNA-seq library prep had to be rerun to yield cleaner traces. Each of these setbacks delayed the final sequencing on the order of weeks.

As a note of caution, previous studies have suggested that within cases with damaging germline *ETV6* mutations, the transcriptional profile was largely driven by leukemia ploidy rather than the nature of the variant.

3.3 – Future of the field and future studies

Utilizing single cell multi-omics to understand hematopoiesis/lymphopoiesis

The development of a functional hematopoietic system is shaped by complex processes regulating survival, proliferation, and migration of HSCs. Even today, our understanding of the hematopoietic system is constantly evolving.

To this end, we can couple *in vitro* differentiation of edited HSPCs with novel sequencing platforms and algorithms to further dissect causal variants to their relevant cellular context at single-cell resolution[56]. Given the current and rapidly expanding field of single cell multi-omics to capture high resolution and large magnitude data, many questions about function and interaction partners can be investigated through these platforms. In particularly, I am interested in learning more about the direct targets of ETV6 transcriptional repressor activity and how abolishing this repression plays a role in lymphopoiesis. In addition, I hope further studies can be conducted to examine the interactions between ETV6 and PAX5, given the breadth of research in their involvement in leukemia development and its direct role in B cell development.

Ultimately the use of this model with single cell multi-omics is not limited to specifically studying ETV6. This system can analyze targets identified from larger scale studies such as GWAS.

Interplay between B-ALL and infection

Studies have proposed the hypothesis that exposure to infections can be a driver of preleukemic clones towards overt leukemia[37]. An increase in childhood leukemia incidence has been linked to modern lifestyles, and thus more attention has been given to understanding the role of infectious agents and immune function to childhood B-ALL risk.

On the other hand, epidemiological studies have demonstrated that early exposure to infection can confer protection against B-ALL development. These studies support a “delayed infection” model in which infections promote secondary genetic events, but only in the context of previous insufficient exposure.

In vitro data has suggested that pro-inflammatory cytokines (IL-6, IL-1B, TNF) can predispose preleukemic B cells to malignant transformation[57]. Following infection, the release of Th2 cytokines as a homeostatic response can also place a proliferative pressure on immature B cells, including preleukemic cells[58]. Infection-induced RAG1/2- and AID-dependent genomic alterations, in the context of the cytokine and innate immune environment, contributed to progression towards overt leukemia[48,57,59–61]. In mouse models, *etv6* mutants exhibit a proinflammatory transcriptional signature, with a disrupted balance in myeloid and lymphoid differentiation and altered cytokine levels in the bone marrow.

Overall, there is data supporting that development of B-ALL results from immune stress after infection exposure in early life, and there is more work to be done around that.

In vivo studies

In order to gain more clinical and therapeutic relevance, it is important to understand the role of ETV6 in leukemia predisposition and lymphoid development *in vivo*. However, it is not clear which approach to mouse models would most closely recapitulate the processes of hematopoiesis, given that the differences between human and murine hematopoiesis impact the study of this gene. In addition, previous studies have shown that heterozygous disruption of *ETV6* in mice is not associated with any clear phenotype, including thrombocytopenia.

BCR sequencing/Immunoglobulin sequencing

Given the lack of depth offered by flow cytometric analysis of surface markers, it would be useful to investigate the effect of ETV6 perturbation on Ig rearrangement and BCR formation. A potential role for ETV6 in impacting formation of the BCR has not yet been explored. Examining the clonal dynamics and diversity of these cells in both normal and perturbed states will provide further clarity in the role of *ETV6* mutants in lymphoid-specific diseases that require clonal expansion.

Clinical prognosis for B-ALL using genetic screening

While the current survival rate for B-ALL is quite high, prognosis and treatment could be improved through genetic screening of germline mutations including ETV6. Identifying and functionally characterizing germline variants associated with ALL has clear implications for developing preventative measures and improving patient outcomes through personalized medicine. A better understanding of these ETV6 mutations is required before we are able to directly target them therapeutically or adjust treatment, but the end goal of characterizing this mechanism of predisposition is to inform future clinical practice.

3.4 Conclusion

The *ETV6* gene was first identified nearly 30 years ago contributing to the multi-step pathogenesis of leukemia. Throughout the years, ETV6 has been shown to be a critical transcription factor in hematopoiesis and HSC survival, and subsequently its mutation has been implicated in blood diseases of all non-erythroid lineages. Only recently have germline *ETV6* mutations been identified in patients with inherited thrombocytopenia, curiously also significantly predisposing patients to B-ALL. It is this intersection of biological uncertainty and clinical relevance that drew me to this project as an immunology student.

I established co-culture techniques combined with CRISPR editing to investigate the effects of ETV6 perturbation. I identified an early progenitor population that was significantly reduced as a result of ETV6 loss. While the nature of this shift, as well as the exact population, is not yet fully understood, it confirms a clear role for ETV6 specifically in lymphoid hematopoiesis, and enables future studies. The developmental alteration seen could hypothetically generate a clone of B cell progenitors that render individuals more susceptible to infection and subsequent oncogenic mutation. The potential acceleration and increased proliferation of differentiated cells could by itself increase the likelihood of acquired somatic mutations that directly cause leukemic transformation.

Ultimately, I hope this thesis can contribute to a greater body of work detailing the role of ETV6, and more broadly disruption of lymphoid development, in predisposing individuals to leukemia, allowing for more personalized diagnoses and treatments and improved clinical outcomes.

Bibliography

1. Liggett, L. A. & Sankaran, V. G. Unraveling Hematopoiesis through the Lens of Genomics. *Cell* vol. 182 1384–1400 (2020).
2. Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* vol. 132 631–644 (2008).
3. Fischer, U. *et al.* Cell Fate Decisions: The Role of Transcription Factors in Early B-cell Development and Leukemia. *Blood Cancer Discov.* **1**, 224–233 (2020).
4. Hock, H. *et al.* Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* **18**, 2336–2341 (2004).
5. Hock, H. & Shimamura, A. ETV6 in hematopoiesis and leukemia predisposition. *Seminars in Hematology* vol. 54 98–104 (2017).
6. Feurstein, S. & Godley, L. A. Germline ETV6 mutations and predisposition to hematological malignancies. *International Journal of Hematology* vol. 106 189–195 (2017).
7. Park, H., Seo, Y., Kim, J. Il, Kim, W. jae & Choe, S. Y. Identification of the nuclear localization motif in the ETV6 (TEL) protein. *Cancer Genet. Cytogenet.* **167**, 117–121 (2006).
8. Golub, ', T. R., Barker, G. F., Love&s, M. & Gilliland', D. G. Fusion of PDGF Receptor p to a Novel e&-like Gene, tel, in Chronic Myelomonocytic Leukemia with t(5;12) Chromosomal Translocation. *Cell* **77**, 307–316 (1994).
9. Wang, L. C. *et al.* The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev.* **12**, 2392–2402 (1998).
10. Wang, L. C. *et al.* Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J.* **16**, 4374–4383 (1997).
11. Fenrick, R. *et al.* TEL, a Putative Tumor Suppressor, Modulates Cell Growth and Cell Morphology of Ras-Transformed Cells While Repressing the Transcription of stromelysin-1 . *Mol. Cell. Biol.* **20**, 5828–5839 (2000).
12. Irvin, B. J. *et al.* TEL, a Putative Tumor Suppressor, Induces Apoptosis and Represses Transcription of Bcl-XL. *J. Biol. Chem.* **278**, 46378–46386 (2003).
13. Neveu, B. *et al.* CLIC5: A novel ETV6 target gene in childhood acute lymphoblastic leukemia. *Haematologica* **101**, 1534–1543 (2016).
14. Wang, L. & Hiebert, S. W. TEL contacts multiple co-repressors and specifically associates with histone deacetylase-3. *Oncogene* **20**, 3716–3725 (2001).
15. Fisher, M. H. *et al.* ETV6 germline mutations cause HDAC3/NCOR2 mislocalization and upregulation of interferon response genes. *JCI Insight* **5**, (2020).
16. Van Waalwijk Van Doorn-Khosrovani, S. B. *et al.* Somatic heterozygous mutations in ETV6 (TEL) and frequent absence of ETV6 protein in acute myeloid leukemia. *Oncogene* **24**, 4129–4137 (2005).
17. Pui, C. H., Nichols, K. E. & Yang, J. J. Somatic and germline genomics in

paediatric acute lymphoblastic leukaemia. *Nature Reviews Clinical Oncology* vol. 16 227–240 (2019).
18. Noetzli, L. *et al.* Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat. Genet.* **47**, 535–538 (2015).
19. 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* **96**, 2557–2561 (2000).
20. Sun, C., Chang, L. & Zhu, X. Pathogenesis of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia and mechanisms underlying its relapse. *Oncotarget* vol. 8 35445–35459 (2017).
21. Hu, Y., Yoshida, T. & Georgopoulos, K. Transcriptional circuits in B cell transformation. *Curr. Opin. Hematol.* **24**, 345 (2017).
22. Moriyama, T. *et al.* Germline Genetic Variation in ETV6 and Risk of Childhood Acute Lymphoblastic Leukemia: a Systematic Genetic Study HHS Public Access. *Lancet Oncol* **16**, 1659–1666 (2015).
23. Papaemmanuil, E. *et al.* RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat. Genet.* **46**, 116–125 (2014).
24. Melazzini, F. *et al.* Clinical and pathogenic features of ETV6-related thrombocytopenia with predisposition to acute lymphoblastic leukemia. *Haematologica* **101**, 1333–1342 (2016).
25. Topka, S. *et al.* Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet.* **11**, (2015).
26. Zhang, M. Y. *et al.* Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat. Genet.* **47**, 180 (2015).
27. Poggi, M. *et al.* Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors. *Haematologica* **102**, 282 (2017).
28. Paola, J. Di & Porter, C. C. ETV6-related thrombocytopenia and leukemia predisposition. *Blood* **134**, 663–667 (2019).
29. Hunger, S. P. & Mullighan, C. G. Redefining ALL classification: Toward detecting high-risk ALL and implementing precision medicine. *Blood* vol. 125 3977–3987 (2015).
30. Smith, M. A. *et al.* Outcomes for children and adolescents with cancer: Challenges for the twenty-first century. *J. Clin. Oncol.* **28**, 2625–2634 (2010).
31. Zhou, Y. *et al.* Advances in the molecular pathobiology of B-lymphoblastic leukemia. (2012) doi:10.1016/j.humpath.2012.02.004.
32. Terwilliger, T. & Abdul-Hay, M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J.* **7**, e577 (2017).
33. Nguyen, K. *et al.* Factors influencing survival after relapse from acute lymphoblastic leukemia: A Children's Oncology Group study. *Leukemia* **22**, 2142–2150 (2008).
34. Alexander, T. B., Bhakta, N., Kolb, E. A. & Rubnitz, J. E. Opportunities for expanding clinical trial enrollment for relapsed and refractory pediatric

acute myeloid leukemia in the United States and Canada. *Pediatr. Blood Cancer* **64**, e26632 (2017).
35. Klco, J. M. & Mullighan, C. G. Advances in germline predisposition to acute leukaemias and myeloid neoplasms. *Nature Reviews Cancer* vol. 21 122–137 (2021).
36. Greaves, M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nature Reviews Cancer* vol. 18 471–484 (2018).
37. Cobaleda, C., Vicente-Dueñas, C. & Sanchez-Garcia, I. Infectious triggers and novel therapeutic opportunities in childhood B cell leukaemia. *Nature Reviews Immunology* 1–12 (2021) doi:10.1038/s41577-021-00505-2.
38. Vijayakrishnan, J. *et al.* Genome-wide association study identifies susceptibility loci for B-cell childhood acute lymphoblastic leukemia. *Nat. Commun.* **9**, (2018).
39. Wiemels, J. L. *et al.* GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes. *Nat. Commun.* **9**, (2018).
40. Vijayakrishnan, J. *et al.* Identification of four novel associations for B-cell acute lymphoblastic leukaemia risk. *Nat. Commun.* **10**, (2019).
41. Jeon, S. *et al.* Genome-wide trans-ethnic meta-analysis identifies novel susceptibility loci for childhood acute lymphoblastic leukemia. *Leukemia* **36**, 865–868 (2022).
42. Gocho, Y. & Yang, J. J. Genetic defects in hematopoietic transcription factors and predisposition to acute lymphoblastic leukemia. *Blood* **134**, 793–797 (2019).
43. Moriyama, T., Relling, M. V. & Yang, J. J. Inherited genetic variation in childhood acute lymphoblastic leukemia. *Blood* vol. 125 3988–3995 (2015).
44. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: A human perspective. *Cell Stem Cell* vol. 10 120–136 (2012).
45. Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* **11**, 635–643 (2010).
46. Duployez, N. *et al.* Germline PAX5 mutation predisposes to familial B-cell precursor acute lymphoblastic leukemia. *Blood* vol. 137 1424–1428 (2021).
47. Wada, H. *et al.* Successful differentiation to T cells, but unsuccessful B-cell generation, from B-cell-derived induced pluripotent stem cells. *Int. Immunol.* **23**, 65–74 (2011).
48. Swaminathan, S. *et al.* Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. *Nat. Immunol.* **16**, 766–774 (2015).
49. Le Coz, C. *et al.* Constrained chromatin accessibility in PU.1-mutated agammaglobulinemia patients. *J. Exp. Med.* **218**, (2021).
50. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
51. Bao, E. L. *et al.* Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* **586**, 769–775 (2020).
52. Voit, R. A. *et al.* A genetic disorder reveals a hematopoietic stem cell regulatory network co-opted in leukemia. *bioRxiv* 2021.12.09.471942

(2022) doi:10.1101/2021.12.09.471942.
53. Tomellini, E. *et al.* Integrin-α3 Is a Functional Marker of Ex Vivo Expanded Human Long-Term Hematopoietic Stem Cells. *Cell Rep.* **28**, 1063-1073.e5 (2019).
54. Murphy, K. *Janeway's Immunobiology*. (2012).
55. Martín-Lorenzo, A. *et al.* Infection exposure is a causal factor in B-cell precursor acute lymphoblastic leukemia as a result of Pax5-inherited susceptibility. *Cancer Discov.* **5**, 1328–1343 (2015).
56. Yu, F. *et al.* Variant to function mapping at single-cell resolution through network propagation. *bioRxiv Prepr. Serv. Biol.* (2022) doi:10.1101/2022.01.23.477426.
57. Beneforti, L. *et al.* Pro-inflammatory cytokines favor the emergence of ETV6-RUNX1-positive pre-leukemic cells in a model of mesenchymal niche. *Br. J. Haematol.* **190**, 262–273 (2020).
58. Hauer, J., Fischer, U. & Borkhardt, A. Toward prevention of childhood ALL by early-life immune training. *Blood* vol. 138 1412–1428 (2021).
59. Ford, A. M. *et al.* The TEL-AML1 leukemia fusion gene dysregulates the TGF-β pathway in early B lineage progenitor cells. *J. Clin. Invest.* **119**, 826–836 (2009).
60. Fidanza, M. *et al.* IFN-γ directly inhibits murine B-cell precursor leukemia-initiating cell proliferation early in life. *Eur. J. Immunol.* **47**, 892–899 (2017).
61. Fidanza, M. *et al.* Inhibition of precursor B-cell malignancy progression by toll-like receptor ligand-induced immune responses. *Leukemia* vol. 30 2116–2119 (2016).