2020

The gamma-globin Repressors KLF1, BCL11A, and MBD2: Knowledge Gaps in the Literature

Lauren Nufer

Virginia Commonwealth University

THE γ-GLOBIN REPRESSORS KLF1, BCL11A, AND MBD2:

KNOWLEDGE GAPS IN THE LITERATURE

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

By

LAUREN NUFER

Bachelor of Science, North Carolina State University, 2017

Director: Dr. Joyce A. Lloyd

Professor, Vice Chair of Education and Faculty Affairs

Department of Human and Molecular Genetics

School of Medicine

Virginia Commonwealth University

Richmond, Virginia

2020

ACKNOWLEDGEMENTS

This work would not have been possible without the support and motivation provided by my parents and sisters. They have always supported me in my decisions and have encouraged me to achieve my goals. I would not be where I am today without them. I am also grateful to Nate Thomas for always being there to support me and for being a sounding board for my ideas. I am fortunate to have been a student of Dr. Lloyd. She was always available and eager to discuss my research project with me and to help troubleshoot any problems that arose. Even when the global pandemic of COVID-19 shut everything down, she was always available to meet over Zoom to discuss my project. She was extremely helpful in aiding me restructure the scope of my thesis as a result of the shutdown. Dr. Lloyd has helped me become a better critical thinker, has helped me improve my writing skills, and has helped me become a better problem solver.

I would like to thank Dr. Xia Li for being my mentor in the lab. I am grateful to her for teaching me cell culture and the techniques I needed to complete my project. She was always willing to help teach me a new method or to discuss any questions I had regarding my research. I would also like to thank my committee members: Dr. Gordon Ginder from the department of Internal Medicine and Dr. James Lister from the department of Human and Molecular Genetics. They were helpful in answering any questions I had regarding my thesis. Additionally, I would like to thank Dr. Alexander Azzo, Dr. Javeria Aijaz, and Dr. Shengzhe Shang of the Ginder Lab for always being receptive of my questions and willing to help me problem solve. I would also like to thank the Ginder Lab for openly welcoming me into their lab, collaborating with, and sharing their resources with me. I am grateful for the fun conversations we had and for Xia and Javeria always knowing where the free food giveaways were. Finally, I would like to thank my friends Hayley Walston and Katlyn Myers for always being eager to have some fun and explore Richmond.

DEDICATION

I would like to dedicate this thesis to my parents.

Rex Nufer and Letha Nufer

Table of Contents:

List of Figures:

List of Tables:

List of Abbreviations:

ACH – Active chromatin hub

βYAC – Entire β-globin locus yeast artificial chromosome

BCL11A – B-Cell lymphoma/leukemia 11a

BFU-E – Burst-forming unit erythroid

BMT – Bone marrow transplant

CDA – Congenital dyserythropoietic anemia

cDNA – Complementary DNA

CFU-E – Colony-forming unit erythroid

CH – Chromatin hub

CHD4 – Chromodomain-helicase-DNA-binding protein 4

ChIP – Chromatin immunoprecipitation

ChIP-Chip – microarray chromatin IP

CML – Chronic myeloid leukemia

DMEM- Dulbecco's modified eagle medium

DMSO – Dimethyl sulfoxide

E – Embryonic day

EKLF – former name for KLF1

EryP-CFC – Erythroid progenitor colony forming cells

FBS – Fetal bovine serum

FOG1 – Friend of GATA1

GFP – Green fluorescent protein

GVHD – Graft vs. host disease

GWAS – Genome wide association study

HDCC – Histone deacetylase core complex

HEL – Human erythroleukemia

HbA – Adult hemoglobin

HbF – Fetal hemoglobin

HbS – Sickle hemoglobin

HPFH – Hereditary persistence of fetal hemoglobin

HS – DNase I hypersensitive site

HSC – Hematopoietic stem cell

HSPC – Hematopoietic stem and progenitor cell

HU – Hydroxyurea

HUDEP – Human umbilical cord blood-derived erythroid progenitor

IMDM – Iscove's modified Dulbecco's medium

IDR – Intrinsically disordered region

IP – Immunoprecipitation

KLF1 – Krüppel-like factor 1

KO – Knockout

LCR – Locus control region

M/B – MBD2/BCL11A dual lentivirus

MBD2 – Methyl binding domain 2

MEL – Mouse erythroleukemia

NuRD – Nucleosome remodeling deacetylase

PBS – Phosphate buffered saline

P/S – penicillin/streptomycin

PCR – Polymerase chain reaction

qPCR – Quantitative PCR

qRT-PCR – Quantitative reverse transcriptase PCR

QTL – Quantitative trait locus

RBC – red blood cell

SCA – Sickle cell anemia

SCD – Sickle cell disease

SCR – Scramble

SCT – Sickle cell trait

SFEM – Serum free expansion medium

shRNA – Short hairpin RNA

SNP – Single nucleotide polymorphism

TAD - Transactivation domain

TDT – transfusion dependent thalassemia

XLT – X-linked thrombocytopenia

XLTT – X-linked thrombocytopenia with thalassemia

YAC – Yeast artificial chromosome

ZFN – zinc finger nuclease

ABSTRACT

β-hemoglobinopathies affect millions of people around the world. Research into treatments for these conditions has focused on methods to increase γ-globin expression, because increased levels of γ-globin ameliorate or reduce the severity of symptoms. As more and more studies have been done, a few proteins have emerged as having crucial roles in γ-globin repression and have been established as key genes to study. These are Krüppel-like factor 1 (KLF1), B cell CLL/lymphoma 11A (BCL11A), and methyl binding domain 2 (MBD2). The roles of these proteins in the switch from fetal to adult hemoglobin and in repressing γ-globin expression have been investigated extensively. However, there are still questions that remain unanswered and knowledge gaps that need to be filled. This thesis focuses on the gaps for each of these three genes and presents approaches to answer these questions. Some of the questions discussed are how variants of KLF1 affect its function, if BCL11A regulates the γ-globin gene by binding at a distance or at its promoter, and how MBD2 binds at the β-globin locus. One of the questions that is universal to the three genes is how they each interact with the β-globin locus and if they each interact with each other in order to repress γ-globin expression. We propose knocking down different combinations of two of the three genes simultaneously to determine if the combinations have an additive or synergistic effect on increasing γ-globin expression, which could indicate whether the proteins are working in the same or different pathways. We provide evidence that dual knockdown of BCL11A and MBD2 in HUDEP-2 cells results in a significantly larger increase in γ-globin mRNA than single knockdown of either gene. Our evidence supports our rationale that further experiments on combinations of the KLF1, BCL11A,

and MBD2 gene knockdowns in HUDEP-2 cells and then further verification in CD34+ cells could lead to future effective treatments for β-hemoglobinopathies.

Chapter 1 – Background and Introduction

1. Erythropoiesis

Erythropoiesis is the synthesis of erythrocytes (red blood cells). Red blood cells (RBCs) are essential for transporting oxygen and carbon dioxide throughout the body. The process of erythropoiesis is complex and consists of two phases, termed primitive and definitive erythropoiesis (1) (Figure 1).

a. Primitive Erythropoiesis

The first blood cells in mammals emerge in the yolk sac in regions called blood islands. This gives rise to the differentiation of erythroid progenitor colony forming cells (EryP-CFCs) from the mesoderm at the start of gastrulation (1) (2) (3) beginning in humans two weeks after conception (4). Primitive erythropoietic cells contain embryonic hemoglobin and begin to circulate throughout the body upon the start of cardiac contractions around the third week of development (5) (6). Despite the hemoglobin switch and the switch from primitive to definitive erythropoiesis, small numbers of EryPs are still present in the circulation up to five days after birth (7). Around the time that blood begins to circulate the primitive erythrocytes begin to mature into definitive erythroid precursors. The characteristics of this switch include the cells becoming smaller, enucleated, and beginning to accumulate hemoglobin (8) (9) (7). Studying this process has been done mainly in animal models because it is difficult to study in humans due to ethical concerns and the lack of access to early human embryos (1).

b. Definitive Erythropoiesis

Around the sixth week of human embryonic development there is a shift from primitive to definitive erythropoiesis (4). Definitive erythropoiesis occurs in the fetal liver until after birth when it switches to the bone marrow (10). This makes definitive erythropoiesis responsible for the creation of fetal hemoglobin (HbF) and adult hemoglobin (HbA). There are two stages definitive erythroid progenitors go through: burst-forming unit erythroid (BFU-E) and colony-forming unit erythroid (CFU-E) (Figure 1). Human BFU-E progenitors are the earliest progenitors and require 14 days in culture to become mature colonies consisting of more than a thousand erythroid cells. The CFU-E progenitors arise from BFU-E progenitors, so they are more mature. Human CFU-E progenitors only require 7 days to form mature colonies consisting of 16-32 cells in culture (1). Next, the BFU-E progenitors differentiate into CFU-E progenitors (11). CFU-E progenitors are intermediate in size and divide quickly (12) (13). Their divisions result in mature cells that differentiate into proerythroblasts, basophilic erythroblasts, polychromatic erythroblasts, and finally orthochromatic erythroblasts (13) (Figure 1). The maturation into the final form, orthochromatic erythroblasts, is characterized by the accumulation of hemoglobin, a decrease in cell size, the condensation of chromatin in the nucleus, and a decrease in RNA content. This mature erythroid precursor then expels its condensed chromatin to become enucleate, forming reticulocytes – immature erythrocytes (1). The final step is for the reticulocytes to mature into RBCs. This process occurs through the loss of plasma membrane surface area, reducing cell volume, and the loss of all residual cytoplasmic organelles (mitochondria and ribosomes) – resulting in the biconcave disc of a RBC (14) (15). The steady level of RBCs is maintained by a constant release of reticulocytes into circulation in

the bloodstream to balance the removal of senescent RBCs. This results in the transportation of around 2 million reticulocytes into the bloodstream every second in a human adult (16).

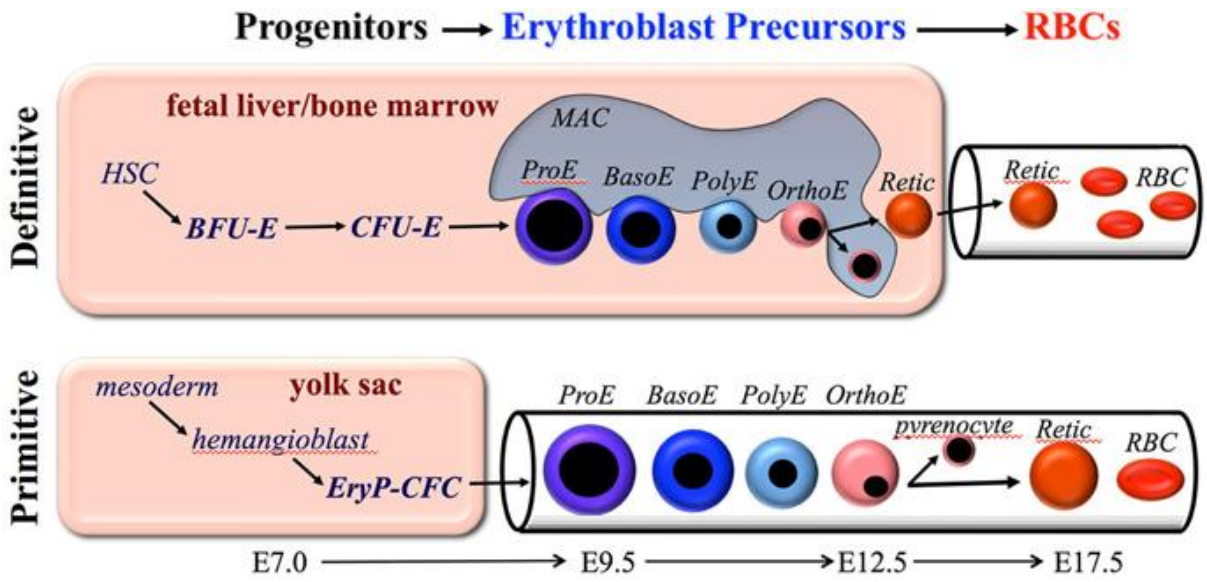

Figure 1: Primitive and Definitive Erythropoietic Pathways
The two pathways in humans that result in the generation of red blood cells.
HSC – Hematopoietic Stem Cell, BFU-E – Burst-Forming Unit Erythroid, CFU-E – Colony Forming Unit-Erythroid, ProE – Proerythroblast, BasoE – Basophilic erythroblast, PolyE – Polychromatophilic erythroblast, OrthoE – Orthochromatic erythroblast, Retic – Reticulocyte, EryP-CFC – Erythroid Progenitor Colony Forming Cells. Figure adapted from Palis J. Front Physiol. 2014 (1).

2. Hemoglobin Switching

a. Structure of the α- and β-globin genes

In humans there are two families of globin genes, the α- and β-family globin genes. The α-family globin genes are located on chromosome 16 in the order of ζ2 (embryonic), ζ1 (embryonic), α2 (fetal/adult) and α1 (fetal/adult) (17) (18) (Figure 2). The β-family globin genes are located on chromosome 11 in the order of ε (embryonic) - Gγ - Aγ (fetal) - δ - β (adult), which happens to be in the order of their developmental expression (19) (17) (20) (Figure 2).

The first hemoglobin to be expressed in humans is present during the first five weeks of fetal development and is the embryonic hemoglobin Gower 1 (21). It is comprised of two ζ- and two ε-globin chains, forming a $\zeta_2\varepsilon_2$ tetramer. The second hemoglobin to be expressed is embryonic hemoglobin Gower 2, which is comprised of two α- and two ε-globin chains to form the tetramer $\alpha_2\varepsilon_2$. The third and final embryonic hemoglobin to be expressed is Portland hemoglobin. It has two ζ- and two γ-globin chains to form a $\zeta_2\gamma_2$ tetramer (19) (17). These are the embryonic hemoglobins, so they are made in the yolk sac during primitive erythropoiesis (Figure 3). This is discussed in more detail in Chapter 1.1.a above.

After expression of the embryonic hemoglobins during primitive erythropoiesis the site of erythropoiesis switches to the fetal liver where HbF is made (Figure 3). HbF has two α- and two γ-globin chains to form the $\alpha_2\gamma_2$ tetramer. The formation of HbF replaces the embryonic hemoglobins as the predominant form of hemoglobin circulating in the blood for the remainder of fetal development (19) (17) (18) (20). Towards the end of human fetal development production of HbA begins, leading to a gradual decline in the amount of HbF as the levels of HbA increases (22) (23) (Figure 3). HbA is comprised of two α- and two β-globin chains to form a $\alpha_2\beta_2$ tetramer. At birth there is a switch where the expression of HbF is almost entirely silenced, leading to HbA being the predominant form of hemoglobin expressed (Figure 3). This switch is discussed more in the next section (Chapter 1.2.b). This results in the composition of human blood to be about 97% HbA, 2% HbA_2 ($\alpha_2\delta_2$), and 1% HbF by around 6 months of age (24) (25).

b. Hemoglobin Switching

Hemoglobin switching is the term for the developmental changes in hemoglobin expression that are described above (Figure 3). The term covers the two switches observed in humans; the first switch is from embryonic hemoglobin to HbF and the second switch is from HbF to HbA (24). Understanding how the HbF to HbA switch is controlled is a huge focus for studies aiming to develop treatments for individuals with β-hemoglobinopathies. Studies of chromosomal looping, gene interactions, protein interactions, and genetic regulation involved in this switch (all to be discussed later in this thesis) are crucial for understanding how the HbF to HbA switch occurs. If these aspects to hemoglobin switching are understood, then they can be harnessed to reactivate the expression of HbF as a treatment for β-hemoglobinopathies.

c. β-globin Locus Control Region (LCR)

The locus control region (LCR) is a DNA regulatory element located in *cis* to the human β-globin locus at a distance 50 kb upstream of the β-globin gene. It is required for the β-globin genes to be normally expressed during all stages of development (26) (27) (28) (29) (30) (31). The LCR controls the expression of the genes in the β-globin locus through its six hypersensitive sites (HS). Five of these DNase I HSs are located 5' of the ε-globin gene. HS 1, 2, 3, and 4, are erythroid specific and are distributed over about 15 kb. The final 5' DNase I HS, HS 5, is sensitive to DNase I in many tissues (32) (26) (27). The sixth HS, 3'HS 1, is located downstream of the β-globin gene. 5'HSs 2 – 4 appear to be the most important for globin gene regulation and each of these sites have a 200-300 bp highly conserved core (33) (34) (35). These cores contain binding sites for erythroid specific DNA binding proteins that are involved in transcriptional regulation of the β-globin locus, such as KLF1 and GATA1 (36). The LCR is a novel enhancer

distinct from classic enhancers due to the fact it has regulatory influence over the entire β-globin locus (31). In fact, it is characterized as a super-enhancer. It is given this classification because it is a large region that contains active chromatin features, like DNase I HSs, transcription factor binding, histone acetylation, and chromatin modifiers (37) (38).

A separate function of the LCR that is important is its role in chromatin looping. Studies supporting this role provide evidence of the novel activity of the LCR. They show it loops to bring the HSs near each other in order to bind DNA regulatory factors and then bring them in contact with the β-globin locus gene promoters (39) (36). Looping of the LCR changes through the different stages of development, binding the γ-globin promoter in fetal development and the β-globin promoter postnatally, providing a possible explanation for globin switching. Although originally believed to be important for the timing of globin switching, studies opposing this role indicate that when the LCR is deleted in βYAC mice they have reduced expression of the β-globin genes, but they are still expressed at the correct developmental time points (40) (41) (42). These findings indicate the LCR is not needed to help regulate the switches of globin gene expression but is needed for normal expression levels. A factor that is necessary to promote chromatin looping of the LCR to the β-globin promoter is lim domain binding 1 (LDB1) protein. A region of LDB1 that is required for its function is its dimerization domain (DD). LDB1 is able to complex with GATA1 and TAL1 through its LIM-interacting domain (LID) which interacts with the LIM-only protein (LMO2). This then provides LDB1 interaction with the DNA-binding proteins GATA1 and TAL1 to provide LDB1 association with chromatin (43) (44).

These studies also found in the absence of the LCR the DNase I-sensitive open chromatin conformation is still maintained, which was surprising. Prior to this finding, experiments in

transgenic mice or MEL cells had shown the 5' HSs increased expression of the β-globin genes (27) (45) (46) and had shown individual HSs were needed for forming an open chromatin state at the β-globin locus (47) (48). It was not until the entire LCR was deleted from the murine β-globin locus and was shown the locus was still active and in an open chromatin conformation that researchers begin to consider the LCR is not needed for the open chromatin conformation seen at the β-globin locus (40). Despite some uncertainty remaining about how the LCR functions, it is still well established that the LCR is a key regulatory element needed for expression of the β-globin genes and is involved in chromatin looping to maintain normal expression levels.

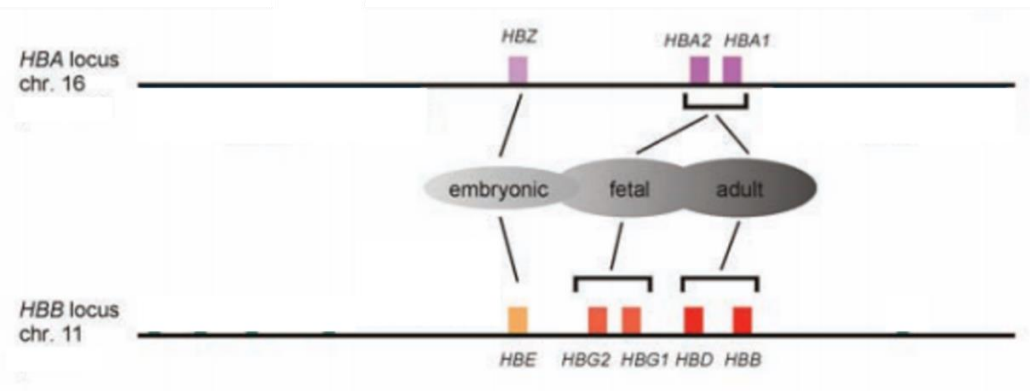

Figure 2: The Human α- and β-globin Loci
Depiction of the human globin genes and how the products from each chromosome combine to form the different globins. HBZ = ζ-, HBA2 = α2-, HBA1 = α1-, HBE = ε-, HBG2 = Gγ, HBG1 = Aγ-, HBD = δ-, and HBB = β-globin.
Top: The human α-globin genes on chromosome 16. Bottom: The human β-globin locus on chromosome 11. Middle: the ovals and their corresponding lines show how the combinations of these different genes result in the formation of the respective globin tetramers. Adapted from Philipsen S. 2013 (22).

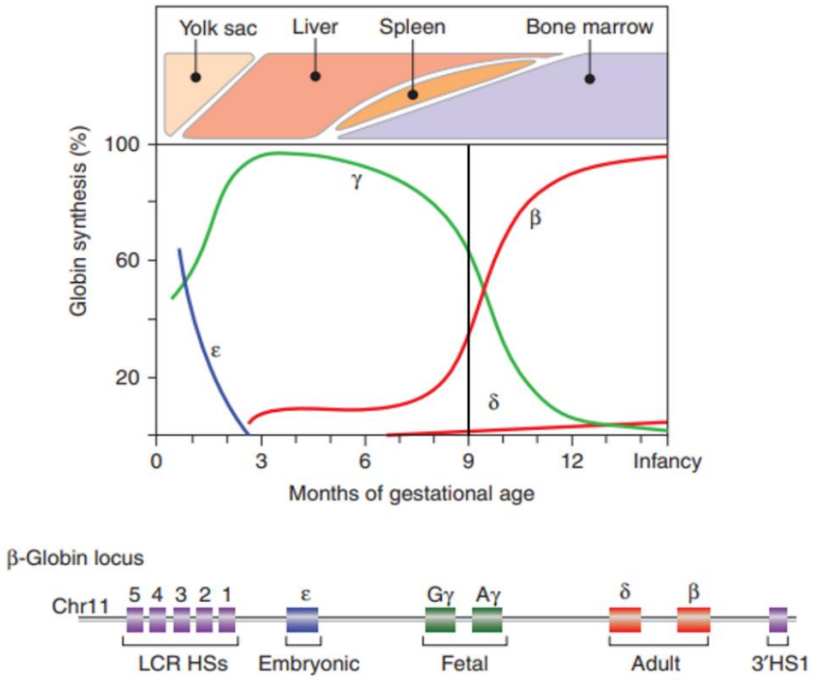

Figure 3: Hemoglobin Switching and the Human β-globin Locus
Top: The synthesis of the globins from the time of conception, to birth, to adult life and the location at which the globin synthesis is taking place. Bottom: The β-globin locus on chromosome 11. In order to better visualize when each gene is being expressed, the color of the curves correspond to the colors of the genes depicted in the schematic of the β-globin locus. Adapted from Sankaran V. and Orkin S. Cold Spring Harb Perspect Med. 2013 (23).

d. Hereditary Persistence of Fetal Hemoglobin (HPFH)

HPFH is an asymptomatic condition that occurs in some individuals, where the HbF to HbA switch is not complete. This results in individuals having abnormally high levels of HbF, ranging from 5 – 80% (49). The condition seems to be benign with no noticeable phenotype. Additionally, coinheritance of a β-hemoglobinopathy mutation with a HPFH variant can ameliorate the symptoms of the disease (50). Many variants that cause HPFH are found in the β-globin locus and the γ-globin gene (49) (50). One variant is a deletion in the 3' region of the γ-globin gene that removes a silencer element between the γ- and δ-globin genes; the result is that γ-globin expression is no longer repressed (51) (52) (53). A different deletion in the 3' region

brings enhancer elements in close proximity to the γ-globin gene; the enhancer now promotes γ-globin expression (54) (55). Point variants in the γ-globin promoter are also responsible for causing HPFH. One of these variants is a C to T transition at -158, which creates a restriction site for *XmnI* (56). Another variant is a C to G mutation at -202 (57). These variants increase HbF to 10 – 20% of the total hemoglobin. One final variant that causes HPFH is due to a non-homologous crossover between the γ- and β-globin genes and is called Hb Kenya. The crossover is in the second exons of the two genes between the codons for amino acids 80 – 87. It brings an enhancer element of the β-globin gene near the γ-globin promoters which results in an increase in expression of HbF (50). There are multiple other point variants in the promoter and in regulatory regions of the γ-globin gene that cause HPFH; the variants listed here are examples of the variety of variants that can cause HPFH (50). More recently, it has been found that at least 50% of the variants that cause HPFH are found in the HBSIL-MYB, BCL11A, and KLF1 genes (58). The HBSIL-MYB region on chromosome 6q23 is a quantitative trait locus (QTL) for HPFH identified through genome wide linkage studies (59) (60) (61) (62) (63). A genome wide association study (GWAS) performed in a population of Sardinian individuals with a high prevalence of HPFH identified a causal SNP in BCL11A (64). This variant will be discussed more in Chapter 2.3.b. Variants in KLF1 also cause HPFH (65) (66) (67) (58). These variants will be discussed more in Chapter 4.1.a, but it should be noted that the first KLF1 variant found to cause HPFH was identified in a Maltese family. The identified variant is the nonsense variant p.K288X; it ablates the DNA binding domain (58). The evidence that HPFH is benign, ameliorates β-hemoglobinopathies, and is a result of numerous variants in various locations introduces the possibility that reproducing these variants to reactivate the γ-globin gene could be a treatment for individuals with a β-hemoglobinopathy. It also raises the possibility of modifying

the factors that bind to these sites, rather than the DNA itself, as a potential treatment. This will be discussed more in Chapter 8.

3. β-Hemoglobinopathies

These diseases are the manifestation of improper forms or amounts of HbA as a result of mutations in the β-globin gene. The most common β-hemoglobinopathies are Sickle Cell Disease (SCD) and β-thalassemia.

a. Sickle Cell Disease (SCD)

SCD affects approximately 100,000 Americans, with prevalence in about 1 in 365 African-American births and 1 in 16,300 Hispanic-American births (68). The term SCD refers to all the genotypes that cause a clinical symptom. Sickle-cell anemia (SCA) is the most common form of SCD (69). Individuals who are carriers of SCD are said to have Sickle-Cell Trait (SCT). The prevalence of having SCT in the African-American population is 1 in 13. SCA is an autosomal recessive disorder caused by the substitution of a thymine for an adenine, causing for the coding of a valine instead of a glutamine in position 6 of the Hb beta chain (70). To have SCA an individual must be homozygous for this mutation. This results in the affected individual having Sickle-cell hemoglobin (HbS). Being homozygous for this substitution results in the traits associated with SCA; instead of having healthy, round RBCs, affected individuals have RBCs that are sickled. This results in the RBCs becoming hard and sticky, which causes them to get stuck in blood vessels and clog blood flow. The sickled blood cells also die early, which causes a shortage of RBCs (68). This change in morphology and reduction of RBCs is what reduces their oxygen carrying capacity.

The treatment most often used for SCA is the drug hydroxyurea (HU). Hydroxyurea was first shown to be an effective treatment for individuals with SCA in 1984 (71). A large clinical trial completed in 1995 confirmed HU as a viable treatment for ameliorating pain crises and for increasing HbF levels (72) (73). It has been proven to reduce the frequency of painful episodes in SCD patients and the need for blood transfusions by 50%. Another drug that is used to treat SCA is L-glutamine, which has only begun to be used more in the last three to five years. It has been shown to reduce the number of sickle cell crises, vaso-occlusive painful events, the need for hospitalization, and erythrocyte morphology; however, it has not been shown to increase HbF and its mechanism of action is unclear (74). Increasing HbF is important because it results in an increase in the ratio of γ- to β-globin in the individual. This is crucial because it results in the individual having a higher percentage of circulating γ-globin which can function in place of their abnormal β-globin, reducing the complications associated with SCA.

The known mechanism of action by which HU induces HbF is by inhibiting ribonucleotide reductase from transforming ribonucleosides into deoxyribonucleosides, which reduces DNA synthesis (75). This is good because inhibited DNA synthesis causes cytotoxic suppression of erythroid progenitors and cell stress signaling which leads to recruitment of erythroid progenitors to make more HbF (75). It could also work by reducing inflammation by reducing the bone marrow's production of neutrophils. Neutrophils are part of the immune system and promote inflammation. Patients with SCA tend to have higher neutrophil counts than controls and their neutrophils tend to be more active in promoting inflammation. Treatment with HU reduces neutrophil recruitment and activation, therefore reducing inflammation in patients (76).

HU treatment was initially reserved for treatment in older patients because there were concerns about giving it to children. But, clinical trials using HU treatment in children and infants have shown it is just as safe and effective as in adults (77) (78) (79). HU treatment has been shown to be the most effective available treatment for SCA, but unfortunately it does not work for every patient (80). Additionally, there have been concerns about its long-term use. To address these concerns, there have been many studies performed that show there are now patients who have been on HU treatment for 10-20 years who have had sustained benefits, and there is evidence HU treatment is associated with improved survival (81) (82) (83) (84). It is important to note that HU is a good treatment for ameliorating RBC sickling, but it does not eliminate RBC sickling. This establishes it as a partial treatment for SCA, but is certainly not a cure. While HU continues to be the best method available for treating SCA patients, the issues and questions presented here indicate the need for additional treatment methods.

b. β-thalassemia

β-thalassemia is a common genetic disorder worldwide that also reduces the production of hemoglobin. About 60,000 symptomatic individuals are born each year, and the estimated effected global population is 80 to 90 million (85). Causes of β-thalassemia are a result of single nucleotide substitutions or deletions or insertions of oligonucleotides leading to a frameshift; β-thalassemia is rarely the result of deletion of the gene (86). The type of β-thalassemia an individual has determines the severity of their disease. There are three types: β-thalassemia major (β^0/β^0), intermedia (β^0/β^+or β^+/β^+), and minor (β/β^+ or β/β^0) (85). Individuals affected with β-thalassemia major have no β-globin and need consistent medical treatment starting in the first few years of life. Individuals affected with β-thalassemia intermedia have reduced β-globin production, typically do not present until later in life, and do not need regular treatment.

Individuals with β-thalassemia minor are usually asymptomatic but occasionally have mild anemia (85).

In individuals with major or intermedia β-thalassemia, their lack or reduction of hemoglobin leads to a reduction of oxygen in the body, resulting in weakness, fatigue, and an increased risk of blood clots (87). Additionally, the ineffective erythropoiesis in these patients can cause enlarged organs, such as splenomegaly (enlarged spleen), hepatomegaly (enlarged liver), and lymphadenitis (enlarged lymph nodes), due to extramedullary erythropoiesis – hematopoiesis that occurs in organs outside of the bone marrow; and misshapen bones, such as deformities of the long bones in the legs and craniofacial changes, due to expansion of the bone marrow (86) (85). A factor contributing to the severity of anemia in β-thalassemia patients is an imbalance of α-globin. This imbalance happens due to the lack of β-globin present to form normal tetramers. The excess α-globin precipitates in erythrocytes causing damage to developing and mature red cells. This results in the premature destruction of erythrocytes in the bone marrow, which causes the observed anemia (88).

A common method used for treating β-thalassemia is blood transfusions. The objective of blood transfusions is to reduce the anemia of an individual by replenishing their RBCs so they are able to carry more oxygen (85) (89). Individuals with major β-thalassemia need transfusions every few weeks, but individuals with intermedia β-thalassemia need transfusions less frequently. A side effect to frequent blood transfusions is iron overload. This buildup of iron occurs mainly in the thyroids, liver, and heart and makes it hard for these organs to function properly (89).

c. α-thalassemia

α-thalassemia is another inherited blood disorder, except the missing/reduced globin is α-globin instead of β-globin. It mainly affects people in the Middle East, Southeast Asia, and the Mediterranean Basin (90). The normal genotype is αα/αα. The forms of α-thalassemia are expressed as α^0 and α^+, where α^0 produces no α chains from either one or both affected chromosome(s) (--/αα or --/--) and $\alpha^+$is missing one of the linked pairs of α-globin genes (-α/-α or -α/αα) (88). Erythropoiesis is more effective in α-thalassemia than in β-thalassemia, but red cell survival is still short. Like with β-thalassemia, in α-thalassemia the complications associated with it are due to the imbalance of γ and β chains that cannot form tetramers because of the reduction of α chains present. In the fetus the excess γ-globin forms tetramers (γ_4) and is referred to as Hb Bart's. After birth it is the excess β-globin that forms tetramers (β_4) and is referred to as HbH. These tetramers have a high affinity for oxygen and do not let go of their bound oxygen when necessary, which contributes complications associated with α-thalassemia. Hemolysis – the destruction of RBCs – is the main cause of anemia in α-thalassemia patients (88). Hemolysis occurs because of the β-globin tetramers of HbH; HbH is unstable and the tetramers precipitate in the RBCs, which causes them to be prematurely destroyed (91) (92). -α/αα heterozygotes have less severe symptoms and no detectable HbH. -α/-α heterozygotes have very low levels of HbH and a mild phenotype. -α/-- have a more severe phenotype and higher levels of HbH. --/-- homozygotes is the most severe form that results in stillbirth; it is referred to as Hb Bart's hydrops fetalis (93). The phenotypes most often seen with HbH are anemia, jaundice, and swelling of the liver and spleen.

d. Treatment for β-hemoglobinopathies

The only cure for both SCA and β-thalassemia is for patients to receive a bone marrow transplant (BMT). BMTs are performed more frequently in β-thalassemia patients than in SCA patients because BMT is reserved for SCA patients with severe complications where the benefits of transplant outweigh the risks, like graft-vs-host-disease (94) (95). Graft-vs-host-disease (GVHD) occurs when the donors bone marrow T-cells view the recipient's cells as foreign and attack them (96). Acute and chronic GVHD ranges from 6-50%, and results in death in 5-8% of all treated individuals (97) (85). A major limitation to receiving a BMT is that most patients do not have a matched sibling who can be a donor and unrelated donors are rarely used because it is hard to find a perfect match (85). If a donor is found, the recipient still has a risk of developing GVHD. Another factor contributing to the likelihood of a patient receiving a BMT is their location. More transplants are performed in Europe, especially in Italy, compared to elsewhere. This is because there are three centers in Italy that have focused on BMTs since the first one was performed on a β-thalassemia patient in 1982 (88). Other centers have been open around the world since then, but the Italian groups are the most experienced in this procedure. This has resulted in the development of better pre- and post-transplant procedures that contribute to patient survival (88) (98) (99).

Chapter 2 – Regulation of the Gamma and Beta Globin Genes

1. Regulation by Transcription Factors and Chromatin Modulators

The γ- and β-globin genes are regulated by elements in their promoters. Both genes have CAAT, TATA, and CACCC elements (100). γ-globin has a duplication of the CAAT box and β-globin has a duplicated CACCC box (100) (101) (Figure 4). The LCR is also important for chromatin looping. The LCRs role in gene expression was discussed in detail in Chapter 1.2.c, but it is important to note the significance of the DNAse I 5'HSs in the regulation of γ- and β-globin. The 5'HSs contain DNA binding sites for DNA transcription factors like KLF1 and GATA1 (32) (102). Many of these transcription factors will be discussed in part 2 of this chapter.

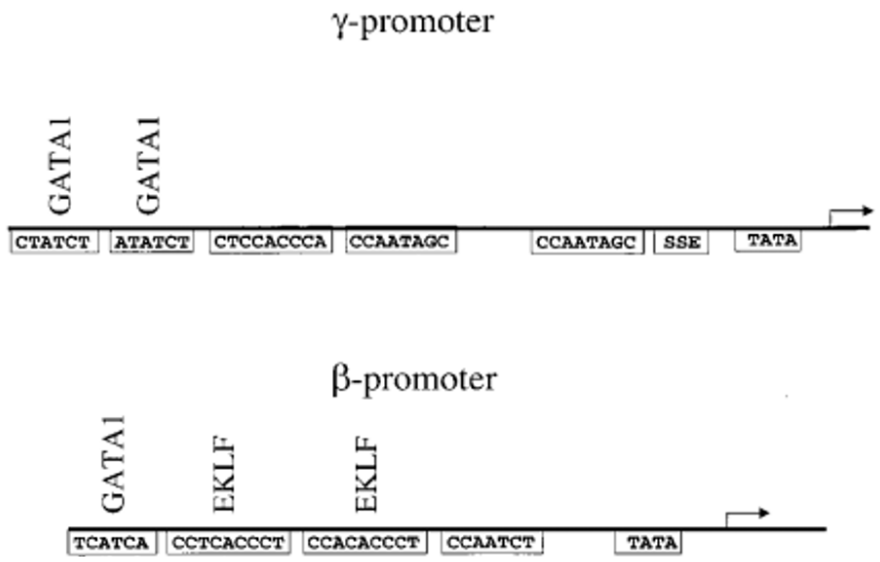

Figure 4: Regulatory Elements of the γ- and β-globin genes
The CAAT, TATA, and CACCC elements of the γ- and β-globin genes. Above the element sequences are the regulatory factors likely to bind at that site. EKLF = KLF1. Adapted from Cao A. and Moi P. Pediatric Research. 2002 (84).

2. Erythroid-specific Transcription Factors

a. GATA1

GATA1 was first discovered in 1988 through studies of the γ-globin promoter. One study identified it as a protein present only in erythroid cells at chicken globin promoters (103) and another study saw the β-globin gene enhancer contains multiple regions for binding GATA1 (104) (Figure 4). This introduced GATA1 as an important factor for transcription of the β-globin genes. The gene is located on the X chromosome (105) and the protein contains two zinc finger DNA binding domains. These domains are the carboxyl terminal finger (C-f) and the amino terminal finger (N-f). They are responsible for attaching GATA1 to its target gene and for promoting GATA1 interactions with its cofactor, friend of GATA1 (FOG1), to activate transcription (106) (107). GATA1 is highly expressed in hematopoietic cells and is required for erythropoiesis (108). GATA1 is able to perform these functions through its interaction with FOG1. They work to induce differentiation in hematopoietic cells by activating genes that promote mature cell forms and repressing genes that work to maintain an undifferentiated cell state (109). GATA1 and FOG1 are both necessary for normal development; when either is missing there is a failure to produce mature RBCs which results in the death of mice in utero due to severe anemia (110) (111) (112). Mutations can also occur in either one of these genes that result in a number of diseases, like X-linked thrombocytopenia and acute megakaryoblastic leukemia associated with Trisomy 21 (113).

One of the blood disorders that results from a GATA1 mutation is X-linked thrombocytopenia (XLT). This is the result of a missense mutation in GATA1 that reduces its affinity for binding FOG1. There are many missense mutations that have been identified to cause

this disease, but they all are a single amino acid substitution in the N-f section of GATA1 (114) (115) (116) (117) (118). These mutations reduce the platelet number and make them less effective at clotting. Therefore, individuals who have XLT often present with a tendency for hemorrhaging. Patients with XLT also present with varying forms of anemia, ranging from mild to dependence on blood transfusions (113). A second blood disorder also caused by a missense mutation is X-linked thrombocytopenia with thalassemia (XLTT). In this disorder the mutation affects the section of N-f that stabilizes GATA1 binding with palindromic DNA; it does not affect the interaction of GATA1 with FOG-1 (119). This reduces the ability for GATA1 to promote hematopoietic cell differentiation. Individuals with XLTT share many of the same clinical features as individuals with XLT, but in addition individuals with XLTT have mild β-thalassemia (120).

b. Krüppel-Like Factor 1 (KLF1)

KLF1, originally EKLF (erythroid Krüppel-like factor) because it is expressed in erythroid cells, is a transcription factor with three zinc fingers first discovered in 1992 (Figure 5). KLF1 was found by enriching for genes expressed in an murine erythroleukemia (MEL) cell line but not expressed in a mouse monocyte-macrophage cell line (121). It was named because investigators found it has three zinc finger motifs that are similar to those in the *Krüppel* gene in the fruit fly. Its name was later changed from EKLF to KLF1 when it was realized it is the first member of a 17-member family of Krüppel-like transcription factors (122). Originally KLF1 was thought to be necessary only for activation of the adult β-globin locus, but later it was found to be necessary for activating genes in many other erythroid pathways required for erythropoiesis (123) (124). One of the more frequently studied roles of KLF1 is its activation of the β-globin gene. KLF1 is able to activate the adult β-globin locus by binding its CACCC sequence (Figure

4). This allows KLF1 to transcriptionally activate the β-globin gene (121). KLF1 interacts with the proteins p300, CBP, and PCAF, which have histone acetyltransferase activities. p300 and CBP acetylate KLF1 in its transactivation domain (TAD) to enhance its transcriptional activation at the β-globin promoter. PCAF works in opposition of p300 and CBP; it inhibits KLF1 transcriptional activity by interacting with KLF1 at the same regions as CBP and p300 which inhibits their interaction with KLF1 (125). KLF2, or lung KLF (LKLF), is a member of the KLF family that has 88% homology to the zinc fingers of KLF1, but they share little homology outside of this region (126). KLF1 and KLF2 can partially compensate for each other when regulating the mouse embryonic β-globin genes (127) (128). It is predicted they are able to compensate for each other by binding to the same or similar DNA sequences due to their highly homologous zinc fingers (126).

Through its interaction with p300 and CBP, KLF1 plays an important role in maintaining an open chromatin structure at the β-globin locus. Open versus closed chromatin conformation is important for determining if a gene is transcribed or not. When chromatin is open it is more easily accessible to the transcription apparatus, including RNA polymerase, and more likely to be transcribed (129). p300 and CBP acetylate histones, which is correlated with open chromatin. Additionally, the three zinc fingers of KLF1 allow it to bind to CACCC elements of the gene promoters of the β-globin locus to help create its active chromatin structure (130) (128). It also binds to the LCR HSs. When KLF1 is present it is able to bind to CACCC elements and the LCR HSs to form the active chromatin hub (ACH), which is essential for transcription. The ACH is comprised of the upstream and downstream HSs of the LCR, the actively expressed globin genes of the β-globin locus, and KLF1 (39) (131). Without the presence of KLF1, the active globin genes, and the downstream HS, the interactions between the LCR and the remaining HSs form a

chromatin hub (CH) (132) (133) (131). The ACH cannot form when KLF1 is knocked out, changing the positioning of the LCR in relation to the genes, as discussed in the next paragraph. This in turn results in reduced transcription of the β-globin locus.

Loss of KLF1 results in a closed chromatin state through multiple mechanisms. The LCR contains HSs, which are crucial for maintaining β-globin gene expression. Loss of KLF1 affects the formation of the HSs at the β-globin gene (134). This results in the loss of the LCRs ability to regulate β-globin locus expression. Without KLF1, chromatin looping to position the LCR and globin genes in close proximity can no longer occur (131). KLF1 expression is correlated with the presence of histone H3K9Ac and H3K4me3 at the globin locus, which are associated with open chromatin and active transcription, respectively. When KLF1 is ablated, these marks are no longer enriched and the open chromatin state at the β-globin locus is diminished (128). The effect that loss of KLF1 has on these processes results in the loss of transcription at the β-globin locus.

To prove KLF1 is necessary for transcription of the β-globin gene it was knocked out in mice. This resulted in embryonic death by day 16 (E16) due to the development of β-thalassemia (123) (135). The time point at which the embryos die is significant; it is when the switch from embryonic to adult globin is complete. KLF1 null embryos appear normal during the yolk sac stage, but between E11 and E16 the embryos become more and more anemic until it results in the loss of the embryo. The progressive increase in anemia starting at E11 matches the beginning of the switch from embryonic to adult globin.

There are a number of variants associated with KLF1 that affect transcription of the gene, resulting in RBC disorders. One of these disorders is β-thalassemia. As discussed in the

paragraph above, loss of KLF1 in mice is lethal due to the severe anemia caused from the loss of expression of β-globin. Additionally, mutations in the CACCC box of the β-globin gene result in reduced affinity for KLF1. This results in a lack of binding by KLF1 and reduced expression of the β-globin gene, resulting in β-thalassemia (136). A condition that is caused by KLF1 variants is HPFH. HPFH was described in Chapter 1.2.d. A genome-wide SNP scan performed in 2010 on members of a Maltese family, in which 10 of 27 members have HPFH, identified a nonsense variant, p.K288X, in the KLF1 gene (58). This mutation ablates the zinc finger domain which prevents KLF1 from binding to DNA, therefore reducing its activation of the β-globin gene and allowing the γ-globin gene to be expressed. This results in an increase in HbF and causes HPFH. Only individuals that are heterozygous for the p.K288X mutation experience increased HbF levels, ranging from 3.3 – 19.5%.

There are two anemia condition caused by KLF1 variants. KLF1 variants cause nonspherocytic hemolytic anemia (NSHA). NHSA results in short survival of RBCs that have abnormal morphology and hemolysis. Heterozygotes for a small in-frame deletion that interferes with KLF1 function (class 2 variant) and a stop codon or frameshift mutation that causes a shortened KLF1 protein that lacks the DNA binding domain (class 3 variant) usually have this disease (137). A final disorder caused by KLF1 variants is congenital dyserythropoietic anemia (CDA). Patients with CDA have poor erythropoiesis, hemolysis, anemia, and erythroblasts with abnormal morphology. There are four types of CDA, but only type IV is caused by a KLF1 variant. There is a single base transition in exon 3 of KLF1 that results in an amino acid change in the second zinc finger. This area of the zinc finger is essential for KLF1 binding to its DNA motif, so the variant results in improper expression of the β-globin gene (138) (139) (140). Additionally, this variant has a dominant negative effect (140). Listed here are just some of the

known disorders KLF1 variants can cause, and there are likely many more disorders caused by KLF1 variants researchers have yet to discover. This shows just how involved and important KLF1 is in erythropoiesis and globin gene switching.

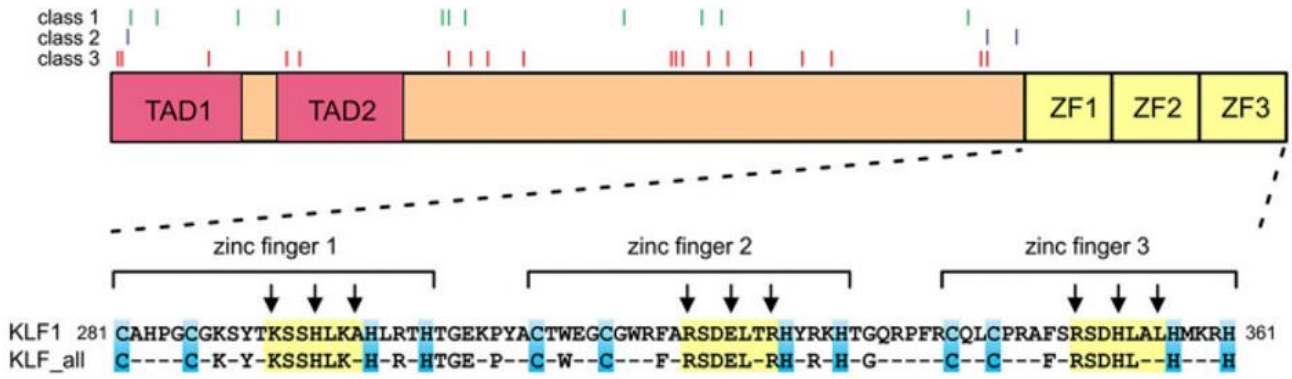

Figure 5: KLF1 Domains and Variants
Two transactivation domains (TAD) are located at the N-terminus and three zinc fingers (ZF) located at the C-terminus. Class 1 variants are green; class 2 and class 3 variants (variants that can be associated with CDA IV) are in blue and red, respectively. The blue highlights are histidine and cysteine residues involved in zinc coordination, yellow highlights are KLF1 residues that contact DNA. Adapted from Perkins AC. Blood. 2016 (106).

3. Non-erythroid specific Transcription Factor

B cell CLL/lymphoma 11A (BCL11A)

BCL11A was initially identified as the murine homologue *Evi9* in 1999. In mice the locus was discovered to be a site of retroviral integration which induces myeloid leukemias (141) (142). One study performed in 2000 identified the human homologue by isolation of isoforms expressed in CD34+ myeloid precursor cells (143). Another study in 2000 also found the BCL11A gene. This study found BCL11A to be involved in a translocation involving the immunoglobulin heavy chain (IGH) locus that causes aggressive B cell lymphocytic leukemias

and that the region in which BCL11A is located is often overexpressed in B cell lymphomas and Hodgkin disease (144). All of these studies introduced BCL11A as a zinc finger transcriptional repressor and established it to be involved in B-cell malignancies as an oncogene (144). Then, a GWAS performed in 2007 identified BCL11A as a key repressor of γ-globin (64).

This GWAS had 4,305 participants and tested association with 362,129 common SNPs. The participants chosen for this study were Sardinians; this population was selected because β-thalassemia occurs at a high frequency with co-inheritance of HPFH, so they are a good population for identifying alleles associated with HPFH (64). The GWAS identified a SNP in the BCL11A gene to be strongly associated with the HbF phenotype. This indicates that when a variant is present in the BCL11A gene it allows for higher production of γ-globin because the gene cannot function as normal to repress the γ-globin gene. Like KLF1, knockout of BCL11A in mouse models is lethal, however BCL11A null mice die postnatally (145) (146). Erythroid-specific knockout of BCL11A in transgenic mouse models results in reduced silencing of human γ-globin and does not result in anemia or embryonic death (147) (146). One study showed erythroid-specific knockout of BCL11A in transgenic mice was able to rescue the phenotype of a SCD mouse model. These compound transgenic mice, having the human sickle β-globin locus, have normal RBC counts, normal hemoglobin content, no sickled cells, normal red cell survival, and reduced damage to the spleen (146). These findings prove erythroid-specific targeting of BCL11A is a valid target as an approach to increase γ-globin levels as a potential treatment for patients with β-hemoglobinopathies.

These studies showing that reduction of BCL11A expression increases γ-globin expression and does not have severe negative effects in mouse models led to the development of

clinical trials. One clinical trial, initiated in 2018, is testing the use of ST-400 as a treatment for individuals with transfusion-dependent β-thalassemia (TDT) (148). ST-400 is a cell therapy that uses gene editing to modify a patient's own hematopoietic stem cells (HSCs) to produce more HbF and then infusing these gene edited cells back into the patient. In this trial, the patients' cells are enriched for CD34+ cells, then the cells are transfected with mRNA encoding the zinc finger nucleases (ZFNs) that have binding sites flanking a precise location in the BCL11A gene, in order to disrupt an enhancer and decrease gene expression. This location that is disrupted is the GATA-binding region of the intronic erythroid-specific enhancer (BCL11A ESE). These edited cells are then infused back into the patient (149) (150). Preliminary results from this clinical trial have shown increased HbF levels (remaining elevated at 6 months post-infusion) and on-target indels present in peripheral blood mononuclear cells (PBMCs) in three of the six patients who have undergone treatment. All three of these patients still require intermittent blood transfusions though.

The second clinical trial, also initiated in 2018, is focused on a treatment for SCD. The aim of this trial is also to perform gene therapy by harvesting the patients' bone marrow and enriching it for CD34+ cells, but they use a different gene therapy method. They transduce the cells with a lentiviral vector containing a shRNA that targets BCL11A. These edited cells are then infused back into the patient (151). Preliminary results have shown two of the three patients (9, 10, and 18 months post transfusion) who received treatment have stopped blood transfusions, with the third having to resume them because they had progressive narrowing of the carotid artery. The patients now produce normal hemoglobin levels and HbF levels are high enough to prevent sickling (152).

Both of these clinical trials are still in the early stages, but they both show promising results for their treatments. As they enroll more patients and have studied them for a longer period of time, it will be interesting to see if these methods will be viable for effective treatment of individuals with a β-hemoglobinopathy. A downside to developing gene editing treatments is they will not be feasible treatments for individuals affected with a β-hemoglobinopathy in third world countries, as they lack many of the resources needed to perform the treatments. Gene therapy will also be expensive, which will greatly limit its availability. These limitations will be discussed more in Chapter 8.

KLF1 regulates the BCL11A gene (153) (154). When KLF1 is knocked down the level of BCL11A is thereby reduced. KLF1 directly activates β-globin production, as discussed in Chapter 2.2.b, and indirectly represses γ-globin production by promoting BCL11A to repress the γ-globin gene. These genetic interactions are responsible for reducing HbF (Figure 6). KLF1 is able to regulate BCL11A expression by binding to the BCL11A promoter, which has several CACC boxes that are KLF1 binding sites (58) (153).

However, it is unclear exactly how BCL11A interacts with the γ-globin gene to repress transcription. There was debate about whether BCL11A binds DNA directly and what sequence(s) it recognizes (155). Previous work suggested BCL11A acts within the β-globin locus at a distance to the γ-globin promoter in order to repress it; the two proposed locations were at the 5'HS 3 and between the γ- and δ-globin genes (156) (157) (155). The current proposed mechanism provides evidence that BCL11A binds to the γ-globin promoter at a TGACCA motif from nucleotides -118 to -113 which prevents the occurrence of chromosomal looping between the LCR and the γ-globin gene (155). This encourages the LCR to bind to the β-

globin gene and promote its transcription, therefore increasing the ratio of HbA to HbF. When BCL11A is not present it cannot interfere with looping of the LCR to the γ-globin gene. This causes the LCR interaction to shift to the γ-globin gene to promote its transcription which increases the ratio of HbF to HbA. These theories will be discussed more in Chapter 4.2.

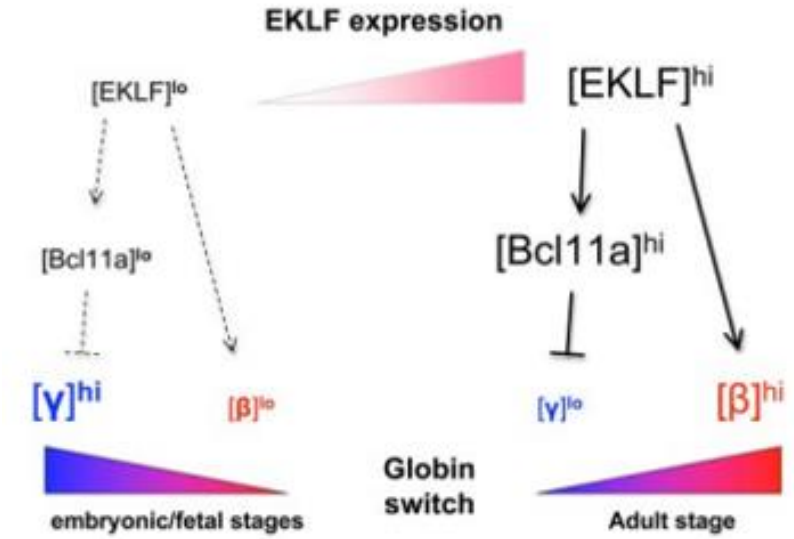

Figure 6: KLF1 Promotes BCL11A to Repress γ-globin Gene Expression
The left shows the proportion of γ- to β-globin when EKLF1 (KLF1) levels are low. Low levels of KLF1 cannot promote BCL11A to repress γ-globin gene expression and cannot promote β-globin gene expression. This results in predominantly HbF. The right shows the proportion of γ- to β-globin when KLF1 levels are high. The high levels of KLF1 promote BCL11A to repress γ-globin gene expression and promotes β-globin gene expression. This results in predominantly HbA. Adapted from Siatecka M. and Bieker J. Blood. 2011 (202).

4. Non-Erythroid Specific Chromatin Modifier

Methyl Binding Domain 2 (MBD2)

MBD2 is a member of the MBD family of proteins that also includes MeCP2, MBD1, MBD3, and MBD4 (158) (159). These proteins were all discovered as a result of the discovery of

MeCP1 in 1989 (158). MeCP1 is a methyl-CpG binding protein that complexes with other methylated DNA sequences. Further study led to the discovery of MeCP2 in 1992. Subsequently, MBD1 was identified in 1997 (160) and then the other MBDs in 1998 (159). All of these proteins, except MBD3, bind to methylated DNA. MBD2 is able to bind to densely methylated CpG islands through recruitment of the Nucleosome Remodeling and Deacetylase (NuRD) co-repressor complex. This allows MBD2 to repress transcription of multiple genes (161).

The NuRD complex has seven subunits. The histone deacetylase core proteins HDAC1 and HDAC2 and the histone binding proteins RbAp48 and RbAp46 make up the core complex. The remaining three subunits are comprised of either MBD2 or MBD3, chromodomain helicase DNA binding protein CHD3 (Mi2α) or CHD4 (Mi2β), and the metastasis-associated proteins MTA1 or MTA2 or MTA3 (162). This group of proteins provides both histone deacetylase and chromatin remodeling ATPase activities that are used for repressing transcription (163)(164). Through the interaction of MBD2 with NuRD it is able to silence the γ-globin gene.

It was shown in 2006 that MBD2 is a regulator of γ-globin gene silencing in adult human βYAC transgenic mice (165). This study showed adult βYAC transgenic mice with an MBD2 knockout have higher γ-globin expression than wild type (WT) mice, indicating that MBD2 causes transcriptional repression of the γ-globin gene. MBD2 null mice have only minor phenotypic effects – lower body weight than average and changes in B-cell differentiation. The knockout of MBD2 does not alter the viability of embryos or their fertility (166) (167). Increased γ-globin expression is also seen when human erythroid cells are depleted of MBD2 (168) (169). The proposed mechanism by which MBD2 represses γ-globin has to do with the structure of MBD2 and all of its interactions. MBD2 is made up of an N-terminal glycine-arginine repeat

(G/R domain), MBD, an intrinsically disordered region (IDR), and a coiled-coil domain (170). The coiled-coil domain binds to p66α of NuRD, forming a MBD2/NuRD complex; this recruits CHD4 which results in gene silencing (171). The IDR is also involved in this process by recruiting the histone deacetylase core components (HDCC), comprised of MTA2, HDAC2, and RbAp48, as well as recruiting the coiled-coil domain which interacts with p66α and then recruits CHD4 (170). Disruption of the IDR prevents formation of the MBD2/NuRD complex and results in increased γ-globin expression (168). These interactions will be discussed more in Chapter 4.3.

Chapter 3 – Erythroid Model Systems

1. Cell lines

Cell lines are cells that have been isolated from an organism that continue to proliferate in culture outside of their host environment. They are important because they are easier to maintain than animal models and they provide a system that can be manipulated to allow for the study of different disorders, diseases, and biological processes. Cell lines have proved a useful tool for studying globin gene regulation, erythropoiesis, and hemoglobin disorders. Many cell lines have been used in the course of studying these topics; some of those cell lines will be discussed here.

a. Murine Erythroleukemia (MEL) cells

A virus that is able to induce murine leukemia was discovered in 1957 and is called the Friend virus (172). The MEL cell line was established in 1966 from leukemic spleens induced by the Friend virus (173). MEL cells can be differentiated and synthesize hemoglobin when treated with dimethyl sulfoxide (DMSO) (174) (175). The creation of this cell line was important because it was the first erythroid cell line created that was able to be cultured and passaged long term. This, and the ability to induce the cell line to produce hemoglobin, introduced the possibility to study globin gene regulation.

The availability of MEL cells allowed for the discovery that the accumulation of hemoglobin and globin mRNA after differentiation is a result of the transcriptional activation of the globin genes. This study also determined that termination of transcription occurs in a specific region and the length of the globin transcription region is approximately 2800 nucleotides (176).

Studies using MEL cells also led to the findings that the α-globin gene is located on human chromosome 16 (177) and the β-globin gene locus is located on human chromosome 11 (178). Another study showed that even though the γ- and β-globin genes are located close to one another on chromosome 11, only β-globin is expressed in MEL hybrid cells. This led them to conclude that there is different regulation between the γ- and β-globin genes and it must involve globin specific regulation factors (179).

b. K562 cells

K562 is a cell line that was established in 1973 from a chronic myeloid leukemia patient (CML) (180). K562 was the first established human cell line in which the erythroid lineage could be studied. In 1979 it was shown K562 cells can be differentiated into RBCs *in vitro* (181). In 1980 it was confirmed that K562 cells lack several properties of adult and possess multiple characteristics of embryonic or fetal erythroid cells (182). From these observations they concluded K562 cells represent early fetal and embryonic erythroid progenitor cells (182) (183) (184). They have an intact β-globin gene but it is not expressed (185) (186) (187). The status of K562 cells provided the opportunity to study the mechanisms of embryonic and fetal globin development. They also provided a tool to study globin switching and the regulatory mechanisms of the β-globin gene. Because β-globin is not made in these cells, they provided a model to study negative control of the β-globin gene.

K562 cells were also used to identify important regulators of the globin genes. In 1990 K562 cells were used to study the function of the γ-globin promoter. This study examined three mutations in the γ-globin promoter associated with HPFH. They found none of these mutations are present in K562 cells, indicating that this is not the reason K562 cells have a fetal globin

phenotype (188). In 1997 K562 cells were used to show guanosine 5'-triphosphate (GTP) can be used to induce differentiation and proliferation (189). A study done in 2000 used this knowledge to show that GTP differentiation in K562 cells increases fetal globin mRNA expression, increases the binding capacity of the transcription factor GATA1 to the γ-globin promoter, increases GATA1 mRNA and protein expression, and GTP treatment can stimulate the γ-globin promoter (190). These findings provided an approach that could be studied as a method to increase γ-globin expression as a treatment for β-hemoglobinopathies. While K562 cells led to some discoveries regarding the globin genes, they are not an optimal model system. K562 cells are derived from cancer cells, so they have abnormal genomes which can influence the expression of the globin genes and give an inaccurate understanding of their function.

c. Human Erythroleukemia (HEL) cells

The HEL cell line was established in 1982 from the peripheral blood of a Hodgkin's disease patient who developed erythroleukemia (191). This same study established that HEL cells can be induced to express γ-globin but not β-globin.

Studies involving HEL cells helped characterize multiple HPFH mutations and helped to uncover how the altered activity due to these mutations increases HbF (49). The same study detected multiple binding sites for nuclear proteins in the promoter region of the γ-globin gene. For example, γCAC1 and γCAC2 are important to transcription because mutations in their sequences reduce binding to the γ-globin CACCC (49). The CACCC box is a regulatory element upstream of the β-globin gene first discovered to be important for transcription in 1984 from studies in individuals with β-thalassemia (192). HEL cells also helped uncover a mechanism by which HU treatment can increase the production of HbF. HU was first discovered as a treatment

for individuals with β-hemoglobinopathies in 1984 (71). A 2001 study shed light on the mechanism by which HU treatment works. It found HEL cells can be induced to express the β-globin gene via HU treatment and that GATA1 plays a role in inducing the β-globin gene in the HU treated cells and promotes cell differentiation. They also showed that GATA2 inhibits the differentiation of HEL cells (193). HEL cells have been useful for studying important regulatory regions of the globin genes and the role of transcription factors, but like K562 they are not a good model for studying hemoglobin switching because they were derived from cancer cells.

d. Human Umbilical Cord Blood Derived Erythroid Progenitor (HUDEP-2) cells

The HUDEP cell lines were established in 2012 and, as the name suggests, were derived from human umbilical cord blood. They are immortalized human erythroid progenitor cells that are able to produce RBCs upon differentiation (194). They were immortalized through a process whereby CD34+ cells were collected from umbilical cord blood (UCB) and cultured with SCF, TPO, and FLT3-L cytokines for 1 day. These cells were then infected with a lentiviral vector that contained the Tet-inducible HPV16-E6/E7 system. One day after infection the cells were cultured with SCF, EPO, and DEX cytokines for 4 days. Then, non-adherent cells were collected and cultured with SCF, EPO, DEX, and DOX for over 100 days with regular media changes (194). These proliferated cells established the immortalized HUDEP cell lines.

There are three HUDEP lines: HUDEP-1, HUDEP-2, and HUDEP-3. The HUDEP-1 line has very little β-globin mRNA before and after differentiation but has a 12-fold increase in γ-globin mRNA after differentiation. The HUDEP-2 line has very little γ-globin mRNA before and after differentiation but has a 6-fold increase in β-globin mRNA after differentiation. The HUDEP-3 line expresses very little β-globin cDNA before and after differentiation but has a 2-

fold increase in γ-globin mRNA after expression (Figure 7). Additionally, the HUDEP-2 line is the only line of the three that expresses BCL11A. The HUDEP-2 line is used in studies more often than the other two HUDEP lines. HUDEP-2 cells are a useful model that is used in many studies discussed throughout this thesis. The establishment of the HUDEP-2 line was important to the study of erythropoiesis, hemoglobin, and β-hemoglobinopathies; it was the first line created that displays adult erythropoiesis because they express mostly β-globin once differentiated (194).

HUDEP-2 cells have been useful for studying the variants involved in causing SCA, HPFH, and β-thalassemia and applying this knowledge to develop treatments for these diseases. In 2016, CRISPR-Cas9 was used to introduce an HPFH variant in HUDEP-2 cells and this resulted in an increase in HbF that was sufficient to ameliorate morphology of SCD (195). More recently, Cas9 sgRNA and sgRNA RNP electroporation were used to create HPFH variants in the human γ-globin promoter and resulted in an increase in γ-globin expression (196). HUDEP-2 cells have also been used to generate sickle cell anemia and β-thalassemia cell lines via CRISPR-Cas9 editing (197) (198). These cell lines exhibit the phenotypes for sickle cell anemia and β-thalassemia and therefore provide an *in vitro* system in which trials can be performed testing new treatments. All of these studies are introducing potential new targets for treating β-hemoglobinopathies.

HUDEP-2 cells have also been useful for studying many of the transcription factors involved in globin gene regulation. Studies using these cells have contributed to the knowledge of how KLF1, BCL11A, and MBD2 – as well as other transcription factors – mediate γ and β-globin expression, which is discussed in detail in the following chapters.

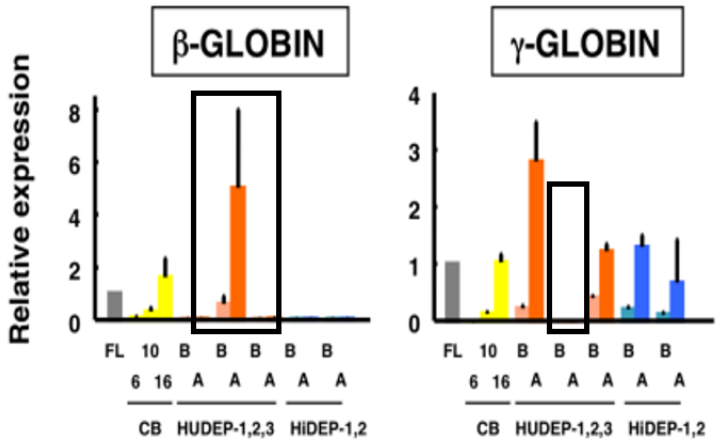

Figure 7: γ- and β-globin Gene Expression in HUDEP-2 cells
qRT-PCR results for the expression of γ- and β-globin in the different HUDEP-2 cells lines. There is about a 5-fold increase in β-globin gene expression after differentiation, but very little γ-globin gene expression before and after differentiation. Expression for the HUDEP-2 cell line is in the black box. B = before differentiation; A = after differentiation. FL = human fetal liver. CB = cord blood (CD34+ cells). Adapted from Kurita R. et.al. PLOS ONE. 2013 (165).

In Summary

Cell lines have proved a useful tool for studying all kinds of diseases, disorders, and biological functions, including those involving hemoglobin. MEL and HUDEP-2 cells are adult cell models, and K562 and HEL cells are embryonic and fetal cell models, respectively. This provides researchers the opportunity to study the expression of the globin genes at every stage of development. Some drawbacks associated with these models are that MEL cells are of mouse origin, so there are differences in some globin expression and in timing of hemoglobin switching, and that neither K562 nor HEL cells express the β-globin gene. While K562 and HEL cells addressed the issue that MEL cells are not of human origin, they cannot be used as models to study β-globin gene expression. The creation of the HUDEP-2 cell line addressed both of these issues – they represent adult erythropoiesis and are of human origin. The benefits of each line are

that they fill a hole the other lines possess: K562 cells were the first established human cell line and are useful for studying embryonic globins, HEL cells are useful for studying fetal globins, MEL cells are useful for studying adult globins, and HUDEP-2 cells are useful for studying adult globins in human cells. Studies involving these four cells lines have led to numerous findings contributing to the field of knowledge on this subject. A few of these discoveries include determining the human chromosomal location of the α- and β-globin genes, determining the location and function of the γ-globin promoter, providing a starting point for understanding how treatment with HU works, and delving deeper into understanding the genetics behind β-hemoglobinopathies so better treatments can be developed.

2. CD34+ cells

CD34 is a transmembrane phosphoglycoprotein that was discovered on HSC in 1984 (199). Expression of CD34 on hematopoietic progenitors (CD34+ cells) is used to evaluate engraftment in BMTs (200). CD34+ compose a small fraction of cells in the bone marrow; they are found at a frequency of 1- 4% (201). CD34+ HSCs have a high proliferative ability and can differentiate into all cell types of hematopoietic lineages, including erythroid cells (202). CD34+ cells were not able to be used in research until a method for collecting them was developed in 1988 (203). There remain drawbacks to their use because CD34+ cells are difficult to isolate in large quantities and cannot be kept in culture for long (202) (204). Despite these difficulties, CD34+ cells continue to be widely used for research as they are the most accurate human cell model available.

For example, CD34+ cells have been an important tool for developing new methods of gene therapy for β-hemoglobinopathy patients. In addition to the clinical trials discussed in

Chapter 2.3.a, *in vitro* studies performed in 2015 used lentiviruses to target either the γ- or β-globin gene to increase expression of HbF or inhibit HbS. This study showed transduction of cells from SCD patients with these lentiviruses resulted in healthy cells that had increased levels of HbF, reduced HbS, and therefore reduced sickling of RBCs (205). In 2016 two different groups implemented CRISPR/Cas9 technology as a tool for treating SCA. One group was successfully able to target and correct the point mutation in the β-globin gene of SCA patient cells which resulted in production of normal hemoglobin in culture (206). The other group used CRISPR/Cas9 to introduce HPFH variants that resulted in an increase in HbF in culture (207). CD34+ cells have also contributed to the understanding that BCL11A is a transcriptional repressor of the γ-globin gene. A 2011 study found lentivirus infection containing either the γ-globin gene, a transcription factor designed to interact with the γ-globin promoter, or a shRNA targeting BCL11A were all able to increase HbF in CD34+ cells taken from β-thalassemia patients (208). Another study done in 2017 used zinc finger nucleases to make a targeted disruption of the GATA motif in the BCL11A gene of SCA cells. This study showed an increase in HbF and the production of hematopoietic stem and progenitor cells (HSPCs) that are capable of long-term engraftment (209). These developments are important to the field because, as mentioned in Chapter 1, methods for gene therapy of the β-hemoglobinopathies are lacking. These studies are a step towards developing autologous gene therapy treatments for patients.

3. Transgenic Mice

Mouse models have long been used to study the mechanisms associated with disease and as models to test treatments. The physiologic similarity mice have to humans is why mice were originally chosen and continue to be used as a model system. Transgenic mice are mice that have had foreign DNA integrated into their genome through experimentation. The most common

method is by microinjecting recombinant DNA into the pronuclei of fertilized eggs (210). This method allows for almost any sequence of DNA to be introduced into mice which is then integrated into their genome. This provides the possibility to study any disease and phenotype desired, with focus on studying globin gene switching and erythropoiesis. The major drawback of earlier studies was that individual genes rather than the entire β-globin locus was inserted (211). In 1987 yeast artificial chromosome (YAC) libraries were created, which allow for large segments of DNA to be cloned into yeast (212). With this discovery came the possibility of being able to insert even larger segments of DNA into mice, including the entire β-globin locus.

This led to the creation of βYAC mice, which provided researchers with a better model for studying globin gene switching and β-hemoglobinopathies (213) (214). As discussed in Chapter 2, the human globin genes are expressed in a particular order, with a major switch occurring at birth when the γ-globin gene is silenced and the β-globin gene begins to be expressed. Like humans, mice express all genes in the order in which the genes are arranged. The difference is mice do not have a γ-globin gene and switch to expression of β-globin in the fetal liver, which means their switch to adult globin occurs earlier than in humans (215) (216) (217). This is a problem with the βYAC transgenic model because the expression of the globin genes does not occur with the same timing as in humans (215). Despite this drawback, βYAC transgenic mice are still a good model for studying hemoglobin and remain to be used often. They have contributed to the extensive studies of human hemoglobin switching, as hemoglobin switching and development can only be studied in animal models (218) (213) (219).

Studies using transgenic mice have contributed to the study of the LCR (discussed in Chapter 1.2.c) (220) (221) (215). Understanding how the LCR functions is important because it is essential for expression of all the genes at the β-globin locus (26) (27). βYAC transgenic mice

have also been used to study HPFH, sickle cell anemia, β-thalassemia, and the transcription factors discussed in Chapter 2. The discoveries made in these studies have contributed to the knowledge on these topics discussed throughout this thesis.

Chapter 4: γ-globin Regulation by KLF1, BCL11A, and MBD2: Knowledge Gaps

1. KLF1

a. Gap #1: What is the association between KLF1 and chromatin looping?

The role of KLF1 in maintaining chromatin conformation was discussed in Chapter 2.2.b. However, it is still not completely understood how KLF1 affects chromatin, and thereby globin gene expression. It remains unknown how many of the interactions KLF1 has are important for chromatin looping and how chromatin accessibility changes throughout the process of differentiation.

A paper published in 2010 performed a ChIP-seq study in mouse E14.5 fetal livers and presented the finding that KLF1 occupies 945 sites in erythroid cells (222). They saw many of these sites are in erythroid gene promoters, but most of the binding sites are at distances of 10 kb or more from a transcription start site (TSS) of any known gene. They hypothesize these distal KLF1 binding sites could be promoting chromatin looping to interact with the gene promoter or that these sites are novel erythroid-specific gene promoters (222) (223). In 2018 a study performed by a different group paired RNA-seq and ATAC-seq using HSPCs to analyze genes implicated throughout erythropoietic differentiation (224). Their analysis found chromatin accessibility at the KLF1 locus increases throughout terminal differentiation. This is associated with an increase in expression of its target genes. These findings suggest KLF1 is more active during the later stages of erythropoiesis. This finding is interesting because it sheds light on the chromatin conformation of KLF1 at different points during erythropoiesis and how this likely contributes to its different functions at various stages of differentiation. While the findings of

both of these studies are important for understanding KLF1, they still leave the question of how KLF1 is changing chromatin conformation to promote interactions. The role KLF1 plays at many of the identified binding sites is unknown. Could these sites be important for the promotion of chromatin looping and formation of the ACH? Chromatin Conformation Capture (3C) is a technique that analyzes the spatial organization of DNA regulatory elements and the target gene in a specific region or of the whole genome in order to provide information on 3D chromatin structure (225). 3C works by performing proximity ligation, which uses formaldehyde to crosslink proteins and DNA that are interacting. The cross-linked DNA is then digested with restriction enzymes and qPCR is performed to identify and quantify the products, as well as to determine spatial proximity of the linked chromatin (226) (227). To know if chromatin looping is occurring at any of these sites, but particularly within the β-globin locus, with KLF1 binding at these sites a 3C technique could be used.

b. Gap #2: Does the binding of KLF1 to the γ-globin promoter change from embryonic to adult stage, and how can this be better studied in a human erythroid cell system?

The evidence supporting the role of KLF1 in indirectly repressing γ-globin gene expression in adults is strong. In Chapter 2.2.b and 2.3.b it is discussed how KLF1 works to repress γ-globin gene expression in adults indirectly through its interaction with BCL11A (153) (154). KLF1 null embryos are not viable because the switch from embryonic to adult globin is blocked (123) (135). The question that remains is if KLF1 positively regulates embryonic and/or fetal globin expression in fetal erythroid cells. This topic has been studied well, but the majority of the studies have been performed in βYAC transgenic mice and not in human erythroid cells. The lack of studies performed in human erythroid cells is a limitation to understanding the

interactions KLF1 has with the different β-globin locus genes at the different stages of development.

A study performed using human βYAC transgenic mice showed KLF1 positively regulates γ-globin expression in embryonic and fetal erythroid cells (128). This study also found KLF1 binds directly to the human ε- and γ-globin gene promoters as well as the β-globin LCR, suggesting that KLF1 is positively regulating these genes directly through its interactions with the promoters. This study provides evidence that KLF1 is not needed only for the γ- to β-globin gene switch but also promotes gene expression during the embryonic to fetal switch. Other studies had shown KLF1 has many roles during primitive and definitive erythropoiesis, including its role in the γ- to β-globin switch (228) (124) (229). One study in CD34 UCB derived erythroblasts that represent the fetal environment saw significant KLF1 binding at the β-globin promoter and detected some binding at the γ-globin promoter (230). Additionally, there have been mouse studies that show a decrease in γ-globin expression during the embryonic stage in KLF1 knockouts (128) (231). This provides more evidence that KLF1 is involved in regulating embryonic and fetal gene expression.

In all but one of the studies discussed in the previous paragraph the model system used was βYAC transgenic mice; the single study used CD34+ cells. The studies in βYAC transgenic mice support the role of KLF1 binding to embryonic and fetal globin promoters during erythropoiesis. The study in CD34+ cells does as well but their evidence is not as strong as they only saw low levels of binding. It is known KLF1 is expressed throughout erythropoiesis during the embryonic and fetal stages from the research mentioned above. It is also established that in transgenic mouse models KLF1 interacts with human ε- and γ-globin promoters – whether or not

it is interacting with the ε- and γ-globin promoters in human cells lines needs further investigation. There is the possibility that differences in the model systems used could result in an improper understanding about when KLF1 is regulating embryonic and fetal globin during development. To address this issue more studies could be conducted in human cell systems. Using human cells systems has a drawback that none behave as an embryonic/fetal system and then switch to an adult system. Despite this drawback, it could be worth investigating this role in HUDEP cells. HUDEP-1 cells could be a useful tool for studying this. These cells are similar to HUDEP-2 cells (discussed in detail in Chapter 3.4), but they differ in that HUDEP-1 cells express predominantly γ-globin after differentiation instead of β-globin like HUDEP-2 cells (194). Studies could be done in the HUDEP-1 cell line to see if KLF1 is binding to the γ-globin promoter before and after differentiation to determine if KLF1 is regulating the γ-globin gene in human derived cells. This could potentially be done in HUDEP-2 cells as well. Since they can be manipulated to express γ-globin it is possible the interaction of KLF1 at the γ-globin promoter could then be studied. A gene that represses γ-globin expression, like MBD2, could be knocked out to restore the expression of the γ-globin gene. Then it can be investigated to see if KLF1 is binding to the promoter of the γ-globin gene. These two studies could potentially show if KLF1 is regulating the γ-globin gene in human derived cells. Performing this study would still leave a gap in the understanding of how KLF1 is binding the ε-globin gene, as there are no good human cell models that portray this stage.

c. Gap #3: Why are some KLF1 variants detrimental and others benign/beneficial? Is it related to how the variant affects KLF1 DNA binding?

As discussed in Chapter 2.2.b, there are many diseases and disorders caused by variants in the KLF1 gene. Interestingly, some KLF1 variants rescue β-thalassemia phenotypes. Different KLF1 variants have been found in individuals who should have β-thalassemia based on their genotypes, but have very mild to no dependence on blood transfusions (232) (65). This suggests that these KLF1 variants increase HbF enough to ameliorate the effects of β-thalassemia. The differences in the phenotypes caused by KLF1 variants raises the question of why some cause detrimental diseases while others are beneficial? The answer to this question could be found by analyzing where each of the variants is located within the KLF1 gene. This analysis will focus on the thalassemias and HPFH. Detailed descriptions of these and other KLF1 disorders can be found in Chapters 1.2.d and 2.2.b

β- and α-thalassemia were discussed in detail in Chapter 1.3.b and the mechanism by which loss of binding of KLF1 to the β-globin gene can cause β-thalassemia was discussed in Chapter 2.2.b. The loss of KLF1 binding to the β-globin gene can cause β-thalassemia because it can no longer activate β-globin gene expression (123). It was also discussed in Chapter 2.2.b how mutations of the β-globin gene CACCC element can result in loss of KLF1 binding. There are four known mutations in the CACCC box that cause β-thalassemia. Two mutations, identified in 1982, are a C to T transition at position -87 or -88 of the proximal CACCC box (233) (192) (234). In 1989 a C to T transition was also found at position -101 in the distal CACCC element (235) (236). In 2004 a fourth mutation, a C to G transition at position -101, was identified in the distal CACCC element (237). These mutations are not in the KLF1 gene but are

still important to mention because they affect the function of KLF1 at the β-globin gene. These mutations prevent the KLF1 zinc fingers from being able to recognize and bind to the CACCC elements, which results in β-thalassemia because KLF1 is no longer promoting expression of the β-globin gene.

KLF1 mutations can also cause α-thalassemia; there is a report of a KLF1 null human with hydrops fetalis, a form of α-thalassemia, who was able to survive past birth (238). It was found this individual is compound heterozygous for KLF1 null mutations: W30X and R319EfsX34. These genes were inherited from asymptomatic parents. The R319EfsX34 mutation is a result of a frameshift in exon 3 and the W30X mutation occurs in exon 2 and is a novel finding. RNA-seq indicated no functional protein is made from the W30X allele. The R319EfsX34 mutation results in a truncated protein that also is not functional. It is surprising that a KLF1 null human can survive past birth with no functional protein given that KLF1 null mice is lethal in utero. It is speculated that no other cases have been seen before because they are lost in utero and it is not realized the loss is due to KLF1 deficiency. It would be interesting to know if these mutations are occurring in cases of fetal loss because they could account for cases of fetal loss that are currently not explained.

In individuals with β-thalassemia, KLF1 variants that increase HbF are also sometimes found. A study performed in 2014 that had a population of 1,190 individuals from northern China, where thalassemia is not common, and 3,839 individuals from southern China, where thalassemia occurs often, identified KLF1 variants in 1.25% of the population from southern China and only 0.08% of the population from northern China. Their findings indicate KLF1 variants occur at a significantly higher prevalence in an area where thalassemia is endemic because elevated HbF is beneficial (232). As mentioned above, there have been cases reported

where an individual's genotype indicates they should have β-thalassemia, but they do not present any phenotypic symptoms. Upon investigation, it was found these individuals have KLF1 variants that increase their HbF levels seemingly ameliorating their symptoms. This study also identified three novel variants in 12 β-thalassemia patients with increased HbF; one is a missense variant in exon 2 and disrupts zinc finger 1, one is a missense variant in exon 3 and disrupts zinc finger 3, and the third is another missense variant in exon 1 that is located in the proline rich TAD (232). It is interesting to note that this final variant is not located in the DNA binding domain like the other variants. Its location in the TAD could disrupt the interaction of KLF1 with other proteins. If this is the case it could be an explanation for the phenotype. Another study identified a novel variant, p.Ser323Leu, in two brothers that have the IVS1-110; HBB: c.93–21 G > A mutation indicating they should exhibit severe β-thalassemia, but instead they are fine and do not need any transfusions. The KLF1 variant is a missense variant in exon 2 and zinc finger 2 (65). This study performed in silico structural analyses of the zinc finger 2 and found the serine at position 323 is pointed towards the DNA. The substitution of leucine does not alter the structure of the zinc finger but increases the distance of residue 323 from the DNA backbone. This is suggested by the fact that serine is polar and hydrophilic, while leucine is nonpolar and hydrophobic; this change from a polar to a nonpolar residue could affect the formation of hydrogen bonds that are important for the stability of KLF1 binding.

At least fifteen additional different variants have been identified in KLF1 by investigating variants that cause HPFH. A study performed in 2012 on 131 individuals with HPFH identified KLF1 variants in 8.4% of their subjects (239). Not all of the identified HPFH variants will be discussed here, but instead the focus will be on the most common ones, a KLF1 promoter variant, and the commonalities seen amongst them. HPFH variants are more frequently seen in

individuals of African, Indian, and Southeast Asian descent. One of the more commonly reported variants is the nonsense variant p.K288X located in exon 2 that ablates the zinc finger domain and therefore DNA binding (58). Another variant that is more commonly reported is the nonsense variant S270X located in exon 2. This variant is also predicted to eliminate KLF1 interaction with DNA by ablating the zinc finger domain. Unlike those with the p.K228X variant, individuals with the S270X variant are compound heterozygotes who also have a K332Q missense variant in exon 3 in zinc finger 2, with the two variants in trans (66). One last variant that has been seen in multiple studies is the frameshift variant G176AfsX179 in exon 2. This variant is the result of a 7 bp insertion that produces a premature stop codon (239) (240). The location of this variant in exon 2 before the zinc fingers likely explains why this variant causes HPFH – KLF1 will not have binding function. This study identified eight other missense variants with one located in exon 1 and the others distributed within the zinc fingers. The final variant they identified is another frameshift variant that occurs in exon 2 due to an 11 bp deletion that results in a premature stop codon (239). A different frameshift mutation as a result of a 4 bp insertion in exon 2 has also been identified (240). The final variant that will be discussed here is the g.-148G>A variant in the promoter of the KLF1 gene, which is the first variant to have been identified in the promoter of KLF1 (67). In silico analysis performed in this study predicts the variant occurs in a Sp1 binding site. It is predicted the loss of this binding site for Sp1 decreases transcription of the KLF1 gene.

What causes some KLF1 variants to be detrimental and some to be beneficial? Looking at the data presented above the noticeable difference between variants that cause thalassemia and HPFH is that the variants in individuals with thalassemia result in no interaction with or production of functional KLF1, while the variants in individuals with HPFH have some

functional KLF1. This can be explained by the finding from a separate study that shows when KLF1 is knocked down in CD34+ cells by more than 70% there is not a significant increase in HbF but when it is knocked down in the range of 40-70% an increase in HbF is seen (230). This shows there is a threshold amount of KLF1 needed to see the beneficial effects of the knockdown of KLF1.

Looking at the variants causing HPFH the only noticeable pattern is a lot of the variants occur within or between the three zinc fingers (Figure 8), even though the zinc fingers only make up about 20% of KLF1 (241). Zinc fingers are small but important for the interaction of KLF1 with DNA as they are what recognize and bind CACCC elements. It makes sense that disruption in these domains results in a phenotype because KLF1 is no longer able to bind to other genes to regulate them. It is also not surprising that there only needs to be a variant in one of the three zinc fingers in order to see a phenotype because KLF1 needs all three of the zinc fingers for proper binding (242). This can also explain why there does not seem to be one zinc finger that variants occur in more often than others. Other commonalities among the HPFH variants are: over half of the variants seen are due to missense variants, and the majority of variants occur in exons 2 or 3 – only 2 variants were seen in exon 1. This disparity is likely because exon 1 is very small and exon 2 is the largest. Some of these studies looked into the mechanism by which the variants cause their phenotypes, but most did not. Of the studies that did investigate the mechanism of the variant, most of them provided evidence that the variations they found have an effect because they ablate binding by the zinc fingers. The loss of zinc finger binding would result in a HPFH phenotype because KLF1 cannot bind to the β-globin gene to promote its expression, and it cannot bind BCL11A to promote its repression of the γ-globin gene. A

reduction in expression of β-globin and a decrease in repression of γ-globin would result in an increase in the HbF percentage.

To better understand how these variants cause HPFH it would be beneficial to investigate the mechanism of function of every variant. It is understood the variants in the zinc fingers cause HPFH because they reduce KLF1 DNA binding efficiency. More investigation of these variants could reveal there is a change in the DNA sequence recognized by some KLF1 variants, like what is seen in the *Nan* mutant mouse model. The *Nan* mouse is a model for neonatal anemia; homozygotes die at E10 due to a severe lack of hemopoiesis and heterozygotes survive but have severe hemolytic anemia throughout their life (231). The causal mutation for this has been found to be a single missense mutation E339D in the second zinc finger of KLF1 (231) (243). This mutation is similar to the human mutation E325K that causes CDA IV (discussed in Chapter 2.1.b), which has a similar phenotype to the *Nan* mouse model (231) (244) (245). This *Nan* mutation causes KLF1 to bind to its target sequences with less affinity and changes the specificity of its binding from what is seen in WT or KLF1 null RBCs (231). This finding introduces the idea that this could also be happening in the case of some of the variants discussed here. More studies need to be done to uncover if this is the case.

It would also be useful to understand how the variants outside of the zinc fingers result in a phenotype. Many of the variants outside of the zinc fingers are frameshift variants that result in a truncated protein that is lacking the zinc fingers, which has the same result as a mutation in the zinc fingers. But what about the missense variant in exon 1, or the variant in the TAD, or the variant in the promoter? What interactions could KLF1 be losing because of the variant in the TAD? And how does the missense variant in exon 1 cause HPFH? Does it cause protein degradation or is it also affecting an interaction of KLF1? Having the answers to these questions

could provide researchers with an understanding as to why some variants cause HPFH and some do not. Another factor that would help clarify this difference would be to know more about KLF1 variants that do not cause any disorder. The UCSC Genome Browser (https://genome.ucsc.edu/) compiles reported KLF1 variants, but for the majority of the variants the functional consequence is not available. If the phenotype of each KLF1 variant was reported the location of those variants within KLF1 could be compared to disease causing mutations. Comparing these variants to disease causing mutations could help uncover critical regions or locations in the KLF1 gene and could lead to better diagnostics and treatments.

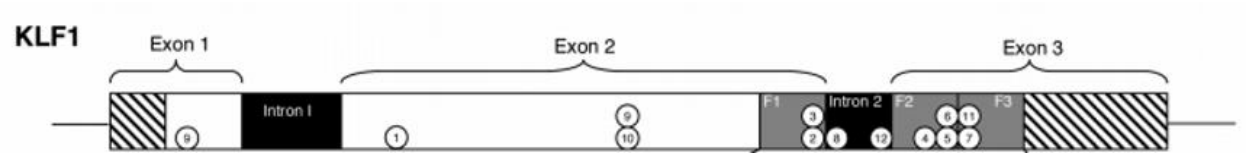

Figure 8: KLF1 Locus with Mutations Clustered in the Zinc Fingers
Depiction of the KLF1 locus with the location of some of the mutations discussed in this chapter indicated. Mutations are indicated by white circles. Most of the mutations are located in the zinc fingers. Grey boxes = zinc fingers. White boxes = exons. Black boxes = introns. Hatched = untranslated regions. Adapted from Gallienne A. et.al. Haematologica. 2012 (220).

2. BCL11A

What interactions does BCL11A have that influence its repressive function on the γ-globin gene?

When a GWAS identified BCL11A as a repressor of γ-globin expression in 2008 (64) it introduced the question of how or if BCL11A binds to and interacts with the γ-globin gene to repress it. The details of this GWAS were presented in Chapter 2.3.a. The literature has provided conflicting evidence about whether BCL11A acts on the γ-globin gene at its promoter or at a

distance. The most recent studies provide compelling evidence that BCL11A is binding at sites in both of these locations.

The first theories about how BCL11A binding represses the γ-globin gene proposed BCL11A acts outside the γ-globin promoter but within the β-globin gene cluster. This theory arose because ChIP experiments showed binding at HS 3 of the LCR and between the γ- and δ-globin genes (246) (247). This introduced the idea that BCL11A acts from a distance. To test this, a 3C assay was performed. The study used mouse fetal liver cells derived from βYAC mice that were either WT or null for BCL11A. The study used the LCR as their anchor location and found that in the WT mouse cells the LCR had the highest cross-linking efficiency with the β-globin gene; in the null cells the γ-globin gene has a high cross-linking efficiency with the LCR. They also used the 3'HS and saw the same results. The proposed mechanism that came from this was BCL11A binds to these distal control regions – the LCR HS3 and in between γ and δ – and reconfigures the β-globin locus by modifying chromosomal looping to allow for the LCR and the 3'HS elements to interact with the β-globin gene; when BCL11A is not present the chromosomal looping changes so that the LCR and the 3'HS elements favor interaction with the γ-globin gene (247).

Studies were performed to determine if chromatin occupancy of BCL11A coincides with common histone modifications – H3K9ac, H3K4me2, H3K4me3, and H3K27me3 (247). A ChIP-Chip experiment using CD34+ WT cells was performed and it was observed that there were peaks of H3K9ac, H3K4me2, and H3K4me3 upstream of the LCR, at the δ- and β-globin genes, and in the γ- δ-globin intergenic region; weak peaks of the active histone modifications were seen at the γ-globin genes; and peaks of H3K27me3 were seen at the proximal promoters of

the γ-globin genes. These findings are to be expected because H3K9ac, H3K4me2, and H3K4me3 are associated with transcriptionally active genes, while H3K27me3 is associated with transcriptional repression and the locations at which peaks of these histone modifications are seen agree with the gene expression that should be seen. With the establishment of the location of the histone modification peaks, they then looked at the peaks where BCL11A binds in the β-globin gene locus and saw there is very little overlap between BCL11A binding sites and peaks of the histone modifications mentioned above. Even at HS 3 upstream of the LCR and in the γ-δ-globin intergenic region where it is thought BCL11A binds and histone modification peaks were seen, the sites where BCL11A binds happen to be in between the peaks where the histone modifications occur. Another study looked to see how knockdown of BCL11A affects the histone marks (248). They saw that with knockdown of BCL11A the histone mark H3K4me2 – which is associated with active transcription – is increased at the γ-globin promoter and in the γ-δ-globin intergenic region. Additionally, both of these studies saw an increase of RNA polymerase II at these regions as a result of the loss of BCL11A, which promotes transcription. It is not known why there is a lack of the histone modifications discussed above at BCL11A binding sites. Some theories are: once BCL11A binds it blocks the nucleosomes from histone modifications, that the sites BCL11A binds possess specific chromatin features that protect them from histone modifications, or simply that BCL11A binds to nucleosome free regions (247). It is possible this feature of BCL11A is what enables its repressive function. Further experiments need to be done in order to answer these questions.

Additional studies into how BCL11A represses the γ-globin gene led to findings that it binds with many other regulatory factors. One study performed a survey in human primary erythroid cells using ChIP. Their findings showed BCL11A regulates and is regulated by

GATA1 (157). This interaction will be discussed in the next paragraph. Another study performed a knockdown of BCL11A in adult erythroid cells and saw a subsequent reduction of HDAC1, particularly in the region between the γ- and δ-globin genes (156). This suggests that BCL11A recruits HDAC1 to this region to establish repressive chromatin modifications, so when BCL11A is knocked down HDAC1 is no longer recruited to this site which results in less repression of the γ-globin gene. A different study affinity tagged BCL11A in MEL cells and performed mass spectrometric peptide sequencing. They identified a strong association between BCL11A and the NuRD complex (246). The relationship between BCL11A and the NuRD complex will be discussed in the following paragraph. Another study performed gel filtration chromatography on CD34+ nuclear extracts and saw a comigration of SOX6 with BCL11A. They also performed a co-IP and saw BCL11A coimmunoprecipitates SOX6 (247). One final study performed a proteomic screen in MEL cells to identify multiprotein complexes that contain BCL11A. They identified interactions with all of the previously mentioned proteins, as well as interactions with IKAROS, SWI/SNF, LSD1/CoREST, NCoR/SMRT, BCOR, and DNMT1 (248). The interaction of BCL11A with these proteins was identified using liquid chromatography tandem mass spectrometry (LC-MS/MS). The copurified proteins were extracted from SDS/PAGE and then used to perform LC-MS/MS. LC-MS/MS is a technique that first uses LC to separate peptides based on their hydrophobicity, then the eluted peptides are passed through the MS/MS and are detected based on their mass-to-charge ratio (249). Based on the data generated for the peptides eluted it can be determined which proteins interact with BCL11A. All of these identified interactions with BCL11A led into the second proposed mechanism that BCL11A acts at the LCR and intergenic sites with multiprotein complexes made up of chromatin modifying agents

and transcriptional repressors in order to repress the γ-globin gene from a distance (157) (156) (248).

It is speculated that BCL11A repressive function on the γ-globin gene is dependent on its formation of multiprotein complexes made up of chromatin modifying agents and transcriptional repressors. There have been many studies reporting its interactions with multiple different proteins (246) (247) (157) (156) (248); one of these reported interactions is with GATA1/FOG1 (246) (157). GATA1 and FOG1 were discussed in detail in Chapter 2.2.a. The first study showed via immunoprecipitation (IP) that BCL11A and GATA1/FOG1 bind to each other in MEL erythroid cells. They propose that since FOG1 binds to the NuRD complex (109) then the interaction of BCL11A with GATA1/FOG1 suggests BCL11A may interact with NuRD as well. This study supported their idea by performing size fractionations on MEL nuclear extracts. They saw overlap between NuRD and BCL11A (246). It is interesting to note the observed connection between BCL11A and the NuRD complex. Since this interaction was first shown in 2008 (246) it has been modified due to new findings. The finding in the 2008 study was identified using MBD3-NuRD. It has since been found that MBD2-NuRD is important for repressing the γ-globin gene, not MBD3-NuRD (250) (168). Researchers now believe BCL11A is not interacting with MBD3-NuRD but that it does interact with MBD2-NuRD, though there is no direct evidence to support this. The available evidence shows that disruption of MBD2-NuRD, but not MBD3-NuRD, increases γ-globin expression, supporting the idea that BCL11A interacts with MBD2-NuRD and not MBD3-NuRD.The second study performed a survey of BCL11A and GATA1 binding in human erythroid K562 cells by using microarray-based chromatin IP (ChIP-Chip) studies, which identifies protein-DNA interactions *in vivo*, in order to better understand their relationship. They also presented results that further support the interaction between

BCL11A and GATA1/FOG1. They investigated the binding at the BCL11A and GATA1 genes individually. They looked at the BCL11A gene and saw GATA1 binding at the promoter of BCL11A as well as in intron 2, which has conserved GATA1 binding motifs. They then looked at the GATA1 gene and saw binding of BCL11A at a positive auto-regulatory promoter of the GATA1 gene (157).

Another demonstrated interaction of BCL11A is with SOX6, which also generates the theory for how BCL11A promotes chromatin looping. SOX6 is a member of the Sry-related high-mobility group (HMG) box transcription factors. This interaction was identified by gel filtration chromatography of human erythroid nuclear extracts, which show BCL11A and SOX6 comigrate. Then, a co-IP was done with epitope-tagged proteins expressed in COS7 cells to determine where BCL11A and SOX6 are interacting. It was found the C-terminal of BCL11A, amino acids 732 – 835, contains three zinc fingers sufficient for SOX6 binding. Additionally, the first 90 amino acids of the N terminal have been found to be important for SOX6 binding. ChIP-chip has shown some overlap between BCL11A and SOX6 at the HS 3 site upstream of the LCR. It also showed SOX6 binds the γ-globin proximal promoters but BCL11A does not (247). This all shows BCL11A and SOX6 interact to repress γ-globin expression. It is proposed this interaction is what allows for chromatin looping to bring BCL11A to the γ-globin promoter. SOX6 contains a HMG domain that binds to the minor groove of DNA and induces drastic DNA bending, which allows for long-range enhancers to function on the DNA by bridging these distant regions (251) (252) (253) (247). When SOX6 is bound at the γ-globin promoter it could induce this chromatin bending to loop to then interact with BCL11A that is bound at the 5'HS 3. Or when BCL11A and SOX6 are both bound at the 5'HS 3 SOX6 could promote chromatin looping to bring the complex in contact with the γ-globin gene promoter.

More recently, a DNA sequence to which BCL11A binds and a new method for how it represses the γ-globin gene have been proposed. Two separate 2018 studies identified that BCL11A binds the TGACC sequence and they each adapt some of the aspects of the previous studies to form their theory. The first study identified this sequence by performing the universal protein-binding microarray (PBM) platform (155). They verified these findings using CUT&RUN – a method that maps protein binding sites and their long-range genomic interactions – in HUDEP-2 and CD34+ cells. Of the 12 motifs in the locus the ones within the γ-globin promoter were predicted to have the highest binding probability. Analyses of these motifs revealed BCL11A prefers binding the distal TGACCA motif from nucleotides -118 to -113 in the γ-globin promoter. To prove BCL11A is binding this distal TGACCA motif, they created a HUDEP-2 cell line where the distal motifs of the γ-globin promoters were mutated. qRT-PCR showed these cells had a γ/(γ+β) ratio of 77%, compared to less than 1% in WT cells, and 3C showed a switch in interaction from the LCR at the β-globin gene to the LCR at the γ-globin gene. The second study identified this binding site by using an epitope tag system that contains a tamoxifen-responsive estrogen receptor moiety (ER) and a V5 tag. They used CRISPR-Cas9 to introduce this onto the C terminus of BCL11A in HUDEP-2 cells. They then performed ChIP-qPCR and ChIP-seq. They identified a peak at -115 bp which the TGACC site encompasses and which agrees with the other study's findings (254). The first study uses this evidence to adapt the chromosomal looping theory to propose the binding of BCL11A at this distal motif in the γ-globin promoter prevents a stable loop from being able to be formed and shifts the interaction of the LCR to the β-globin gene to promote its expression, and that when BCL11A is not bound to this motif chromosomal looping is able to occur in a way that allows for the LCR to bind to the γ-globin gene and promote its expression (155) (Figure 9). The second study does not provide a

theory for how they think BCL11A is functioning from this site, but they do propose that because they see BCL11A peaks in regions that do not contain TGACC motifs, BCL11A could be acting at these distal regions to promote chromatin looping as well as at the motif in the γ-globin promoter in order to repress its expression (254). These are the most recent theories and are the first to give definitive evidence of a sequence BCL11A binds, but it is still unclear as to if this is the actual mechanism by which BCL11A functions to repress γ-globin expression. The binding of BCL11A at the -115 site could inhibit transcription of the β-globin gene because its presence there interferes with the ability of LDB1 to form a stable complex at the β-globin promoter. This could weaken the ability of LDB1 to promote chromatin looping of the LCR to stably complex with the β-globin promoter, or it could prevent LDB1 from promoting chromatin looping with the LCR at the β-globin promoter altogether. These studies also saw TGACCA motifs in between the γ and δ genes and at the LCR. These are the sites at which the previous studies showed BCL11A binds to the DNA, which could explain why the previous studies identified these regions.

Findings from these more recent studies could provide a framework to unify the theories generated from the earlier studies. BCL11A could be acting at multiple sites at once. If BCL11A binds at all of the proposed sites, then each of these studies has presented only a piece of the mechanism by which BCL11A represses γ-globin. The more recent studies present evidence of BCL11A binding at TGACCA motifs. Binding peaks are observed in between the γ and δ genes and at the LCR, in addition to the strong binding at the -115 motif in the γ-globin promoter (155) (254). This suggests that all of the theories discussed in the studies presented here have a piece of the mechanism correct, and together form a more complete description of the mechanism by which BCL11A functions. The actual mechanism could be BCL11A binds at the motif in the

distal promoter of γ-globin where it interacts with SOX6 (247), as discussed above, to promote chromatin looping. This chromatin looping brings the BCL11A that is bound to the HS3 at the LCR into proximity of the β-globin locus. The way that BCL11A causes chromatin looping is a function that needs to be studied further. The best theory presented this far is its interaction with SOX6. BCL11A could be interacting with any number of the other proteins discussed in this section in order to promote chromatin looping.

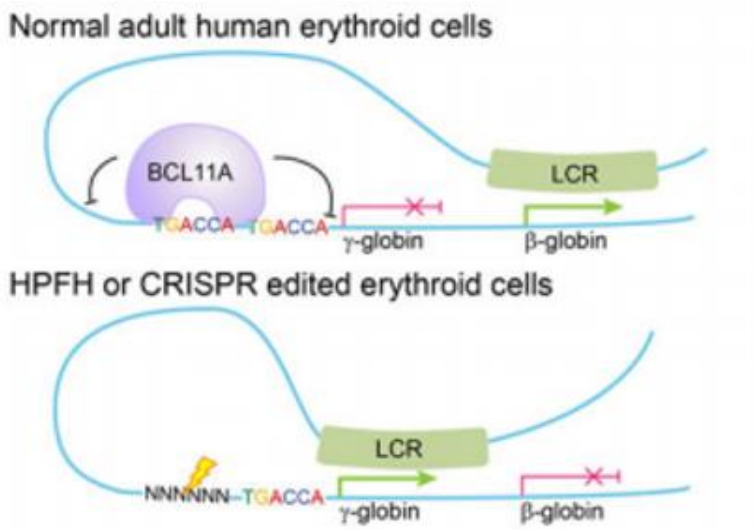

Figure 9: Proposed Mechanism of BCL11A Binding
Top: when BCL11A is present it binds to the TGACCA sequence and prevents the LCR from being able to loop and bind to the γ-globin gene; the LCR instead loops and binds at the β-globin gene. When BCL11A is not bound at the TGACCA sequence the LCR is able to loop and bind to the γ-globin gene. Adapted from Liu N. et.al. Cell. 2018 (105).

3. MBD2

How does MBD2 repress γ-globin gene expression? If it is not binding the gene directly, what proteins is MBD2 interacting with to exert its repressive function?

MBD2 typically binds to methylated CpG rich regions and recruits the NuRD complex to help regulate gene expression (162) (167). It is well established that MBD2 represses the γ-

globin gene because when it is ablated, expression of γ-globin increases (165) (168). However, there are no CpG islands within 6 kb of the γ-globin gene (165). Thus, despite knowing MBD2 represses the γ-globin gene, the mechanisms by which it works are still unknown. If MBD2 is not binding at the γ-globin promoter via a CpG island, then it must have other interactions with proteins or transcription factors, or other functions that allow it to exert its repressive effects on the γ-globin gene.

There have been a number of proposed interactions provided to explain how MBD2 binds the γ-globin gene in order to repress it. One interaction is the MBD2 coiled-coil domain with p66α. p66α is a component of the NuRD complex that is responsible for the association of MBD2 with the NuRD complex. Forced expression of the p66α coiled-coil domain in CID β-YAC bone marrow progenitor cells or MEL cells reverses MBD2 silencing of the γ-globin gene (171). This indicates the repressive function of MBD2 is dependent, at least in part, on its binding to the p66α coiled-coil domain. An explanation for how forced interaction of the p66α coiled-coil domain with MBD2 results in loss of transcriptional repression is that this interaction blocks MBD2 from binding to the NuRD complex (171).

MBD2 also interacts with CHD4. CHD4 is a member of the NuRD complex, and a study of how CHD4 functions to bring the NuRD complex to the B-cell-specific gene *mb-1* (*Cd79a*) found CHD4 is responsible for performing the chromatin remodeling activities of NuRD (255). It found MBD2 binding of CHD4 is necessary for silencing of the gene. Additionally, they found certain CHD4 mutations result in reduced function of the NuRD complex, preventing it from binding MBD2 and from exerting its repressive function (255). This association was shown in a system unrelated to the β-globin gene locus, but it is likely MBD2 is reliant on CHD4 for its

interaction with the NuRD complex. Another study found CHD4 is important for the function of MBD2 at the γ-globin gene because when enforced expression of the p66α coiled-coil domain is introduced it competitively binds to MBD2, which prevents the recruitment of CHD4 and p66α to the MBD2-NuRD complex (171). The disruption of these interactions results in an increase in γ-globin expression, which indicates both p66α and CHD4 are required for MBD2 repressive activity.

Another feature of MBD2 that is important for repression of the γ-globin is the IDR of MBD2. The IDR is required for MBD2 to recruit the HDCC of NuRD, which contains MTA2, HDAC2, and RbAp48. A study performed in 293T cells identified two important locations in the MBD2-IDR that are critical for its binding of this complex. Point mutations of these two residues weaken the interaction of MBD2 with the HDCC (170). Forced expression of these two MBD2-IDR mutants in HUDEP-2 cells that lack MBD2 results in failure of MBD2 to repress γ-globin expression (168). It is thought these mutations cause this result because they disrupt the structure of the IDR which affects its ability to bind to the HDCC.

The evaluations of each of these interactions individually indicates MBD2 requires NuRD to repress the γ-globin gene. Two of the interactions MBD2 needs to bring it in contact with NuRD are between its coiled-coil and its IDR domain with different components of the NuRD complex. To see if these domains are necessary for the repressive function of MBD2 on the γ-globin gene, forced expression of WT MBD2, a coiled-coil mutant, or an IDR mutant in MBD2 knockout HUDEP-2 cells was performed. The addback of WT MBD2 showed a significant reduction in expression of the γ-globin gene compared to control, rescuing the phenotype. Neither the IDR mutant nor the coiled-coil domain mutant could rescue the WT

phenotype; they failed to repress the γ-globin gene (168). These results show MBD2 needs these domains to function and that both need to be functional for MBD2 to repress γ-globin transcription. If MBD2 needed only one functional domain to repress γ-globin gene expression then abolishing the function of one of the domains would not affect its repressive activity, but that is not the case.

These mechanisms were suggested under the assumption that MBD2 was binding to methylated CpG rich regions, but it is now believed that there are no CpG islands in the β-globin locus. A CpG island is a cluster of CpGs in GC rich regions of the genome (256); they are about 1,000 bp long and have a GC content greater than 50% (257) (256). About 70% of gene promoters are associated with a CpG island, which is important for gene expression (258) (257). When CpGs are methylated; when they become methylated proteins, MBD2 can recognize the site and bind it, resulting in silencing of the associated gene (256) (257) (259) (260). All of these proposed mechanisms can still be valid for how MBD2 is functioning to silence the γ-globin gene, but there must be some other factor involved that allows MBD2 to bind to methylated but CpG-poor regions of the β-globin locus. It could be that there is another protein that binds to the γ-globin gene and then to MBD2 to anchor it there. Or, it could be that there is a region of MBD2 yet to be studied that gives it the ability to bind to CpG-poor DNA.

MBD2 could be binding to the γ-globin gene through an interaction with BCL11A. As discussed in the previous section, it is believed BCL11A and NuRD interact. It is thought BCL11A interacts with the NuRD complex through RbAp48 and RbAp46 and through this interaction BCL11A is able to interact with other members of the NuRD complex, like CHD4 (261) (169). If BCL11A binds CHD4, and MBD2 binds CHD4, then it is possible that their

interactions through CHD4 brings them together. The interaction of MBD2 with CHD4 could be bringing it to where BCL11A is bound on the γ-globin gene and then MBD2 could bind with BCL11A to repress the gene. This is one possible explanation for how MBD2 could bind the γ-globin gene without methylated CpG islands. Another explanation could be MBD2 does not need to bind to other factors in order to repress the γ-globin gene because it has a region that allows it to bind to CpG-poor DNA. MBD2 can bind to CpG-poor DNA (262), but it is still unknown how it does this and how it represses transcription from an unmethylated CpG-poor region. In order to determine if MBD2 has a unique region that gives it this function it should be compared to MBD3. MBD2 and MBD3 are similar in structure, but MBD3 does not bind to methylated DNA. Comparing the differences in the domains of these two could uncover a region that is not shared between the two and is responsible for how MBD2 binds CpG-poor DNA and represses the γ-globin gene. This would also give researchers an idea of where to focus their efforts in trying to uncover how MBD2 is binding.

Conclusion

One of the main gaps each of these DNA binding proteins have in common is the mechanism by which they bind to and regulate the γ-globin gene. For each of the three DNA binding proteins there is debate as to how they repress transcription of the γ-globin gene. This shows that while much is known about how these factors work, there is still a lot to be uncovered. To help answer these question some experiments that could be done are to investigate how KLF1 binds the γ-globin gene in human cells and to compare the domains of MBD2 and MBD3 to elucidate if the differences account for what gives MBD2 the ability to bind to CpG-poor DNA. Continuing to study these interactions and performing the experiments discussed

above will hopefully provide some clarity as to how the γ-globin gene is being repressed by these transcription factors.

Another gap these DNA binding proteins share is uncertainty about how they are interacting with each other to repress the γ-globin gene. There is evidence showing some of these transcription factors work together, but there is much left to be studied about how these interactions occur. Understanding how the transcription factors regulate and interact with each other would also provide more clarity about to how the γ-globin gene is being regulated, and would give a better idea of which transcription factors to target as a therapeutic treatment. It is possible that targeting factors that are not regulating each other would result in a greater increase in γ-globin gene expression than knockdown of either transcription factor alone. This idea will be expanded upon in the following chapter.

Chapter 5: Simultaneous Targeting of Multiple γ-globin Repressors: Literature on Dual Knockdowns

As discussed in Chapters 2 and 4, the transcription factors KLF1, BCL11A, and MBD2 are repressors of γ-globin gene expression. It is well established that knockdown of any one of these genes results in a significant increase in γ-globin expression. This introduces the question of whether knockdown of more than one simultaneously would increase γ-globin expression more than a single knockdown. There has been limited investigation into this question. Some of these studies focus on the genes presented here, and some investigate combinations of knockdowns of other γ-globin repressors.

In a 2013 study, a proteomic screen was done in MEL, and verified in K562 cells, to identify BCL11A-containing multiprotein complexes. The co-purified proteins were fractionated in SDS/PAGE, followed by liquid chromatography, and then subjected to tandem MS peptide sequencing. They then performed co-IP assays to confirm their result. Two of the BCL11A-interacting proteins they identified are LSD1 and DNMT1. They first investigated the combinations of BCL11A and LSD1, and then BCL11A and DNMT1, in repressing γ-globin expression (248). LSD1 is the lysine-specific demethylase 1 and is a histone demethylase. Two models were used to measure changes in γ-globin expression. The EpoR-Cre system was used to create an erythroid-specific compound BCL11A homozygous – LSD1 heterozygous knockout in human βYAC transgenic mice. Mice homozygous for LSD1 loss results in embryonic lethality due to severe anemia, so LSD1 heterozygotes were used. Using this system, they measured the mRNA concentrations via qRT-PCR in peripheral blood (PB) in 8-12 week-old mice. They saw a statistically significant increase of γ-globin mRNA from 8.4% to 13.5%, about a 1.6-fold

increase, in BCL11A -/- compared to compound BCL11A -/- : LSD1 +/- mice. The BCL11A -/- was increased 383-fold from the WT βYAC control mice, which express 0.02% γ-globin mRNA, and the compound BCL11A -/- : LSD1 +/- was increased 615-fold compared to the control mice. Both of these increases are statistically significant, with the compound -/- : +/- resulting in a statistically significant increase in γ-globin mRNA compared to the single knockdowns (Table 1). Their second model used the hematopoietic selective and IFN inducible Mx1-Cre allele to generate complete loss of BCL11A and LSD1 (263) (264). They used this model to create BCL11A -/- : LSD1 -/- (dual knockout) mice. Again, mRNA concentrations were measured via qRT-PCR in PB of 8-12 week-old mice. They first showed the effect single knockouts of the genes had in comparison to WT-βYAC control mice. LSD1 knockout resulted in no increase in γ-globin mRNA expression. BCL11A knockout resulted in a statistically significant increase in γ-globin mRNA to 12.9% – a 515-fold increase compared to control. The dual knockout resulted in a statistically significant increase of γ-globin mRNA to 18.2% – a 728-fold increase compared to control, a statistically significant 1.4-fold increase compared to BCL11A knockout, and a significant 455-fold increase compared to LSD1 knockout. The dual knockout using this model also resulted in a statistically significant increase in γ-globin mRNA compared to each of the single gene knockouts.

The difference between these two models is that the EpoR-Cre system is erythroid specific while the Mx1-Cre system is not erythroid specific. Additionally, Mx1-Cre is an inducible system that acts during the adult stage while EpoR-Cre causes a homozygous deletion at the embryonic stage, which is what makes Epor-Cre lethal and Mx1-Cre not. The Mx1-Cre system allows for the creation of LSD1 null mice, but the EpoR-Cre system can only be used to create LSD1 heterozygous mice because loss of LSD1 in erythroid cell is lethal. The BCL11A -/-

: LSD1 +/- model resulted in a smaller increase in γ-globin mRNA compared to the BCL11A -/- : LSD1 -/- system, 615-fold compared to 728-fold. This is about a 1.2-fold difference between the two models. It is not stated if this is a significant difference between the two models. It seems that the two models work about the same in increasing γ-globin mRNA. These findings indicate knockout of BCL11A and LSD1 have an additive effect on increasing γ-globin mRNA, but targeting LSD1 is probably not a feasible treatment because the EpoR-Cre generated compound KO mice exhibited mild anemia as well as a decrease in white blood cells. Additionally, they discuss that LSD1-depleted CD34+ cells exhibit proerythroblast and basophilic erythroblast morphology – the mature polychromatophilic and orthochromatic erythroblast are absent at day 9 of differentiation in culture. These findings indicate LSD1 is important for normal erythroid development and terminal differentiation.

Xu et.al. also investigated the combination of DNMT1 and BCL11A (248). Another BCL11A-containing protein they identified is DNA methyltransferase 1 (DNMT1). DNMT1 is a DNA methyltransferase, so it adds methyl groups to DNA, which plays an important role in repressing globin gene expression. They found DNMT1 has enzymatic activity with BCL11A, so they decided to investigate its role in γ-globin expression. To investigate this, they used the EpoR-Cre system in human βYAC transgenic mice to create erythroid specific BCL11A/DNMT1 double knockouts. They observed embryonic lethality with the DNMT1 knockouts, so they used erythroid specific BCL11A -/-: DNMT1 +/- mice. They saw a statistically significant increase from 11.2% γ-globin mRNA in BCL11A -/- mice to 23.6% γ-globin mRNA in BCL11A -/- : DNMT1 +/- – a 2.1-fold increase in expression (Table 1). Their results show BCL11A and DNMT1 also have an additive effect in increasing γ-globin mRNA expression. While some of the genes presented in this paper are likely not good therapeutic

targets because knockout of these genes impairs erythropoiesis and terminal differentiation (248) (265) (266), the results are still interesting because they show promise for using dual knockouts to increase γ-globin gene expression.

Another study, performed in 2016, did double knockouts of BCL11A and LRF in HUDEP-2 cells using CRISPR/Cas9. First, they used HPLC to show single knockouts of BCL11A or LRF increase the ratio of HbF to HbA from about 2% in control cells to about 60% in each of the single knockouts. Then they show the double knockouts increase the ratio to about 90% HbF, indicating the double knockout of the two genes has an additive effect on HbF expression (Table 1) (267). However, there are some issues with this paper. First, FACS of fetal liver cells from E14.5 LRF null mice indicates that disrupting LRF results in ineffective differentiation (268). This makes it impractical to use LRF as a target to increase HbF. This drawback makes it questionable how this study achieved these results with LRF knockouts. The study mentioned ineffective differentiation of CD34+ but not HUDEP-2 cells in which LRF is knocked out. CD34+ cells provide a more normal representation of human erythropoiesis than HUDEP-2 cells. Another issue with this paper is how they display their data. All of their data is depicted as the ratio of HbF – the percentage of γ-globin HbF to total globin HbF. They do not show any data on the levels of γ-globin or β-globin expression alone; this makes it hard to know if the increased ratios are due to an increase in γ-globin or a decrease in β-globin expression.

One more study, performed in 2013, investigated the effect of dual knockout of BCL11A and KLF1 on γ-globin expression (154). This study used the mouse line PAC8.1, which carries a human β-globin locus transgene. They crossed PAC8.1 mice carrying either a KLF1 KO allele or a floxed allele of BCL11A and used a knockin EpoR-Cre allele in order to create erythroid-specific BCL11A -/- : KLF1 +/- mice. KLF1 -/- results in embryonic lethality, as discussed in

Chapter 2.2.b. They first studied how γ- and β-globin expression changed in these mice at embryonic time points E14.5 and E18.5. They used fetal liver RNA to perform qRT-PCR and E18.5 blood samples to perform western blots. They saw a significant increase in γ-globin mRNA and HbF ratios in the KLF1 +/-, BCL11A -/-, and BCL11A -/- : KLF1 +/- mice when compared to PAC8.1 BCL11A floxed control mice. The largest increase was for the BCL11A -/- : KLF1 +/- mice and the smallest increase was in the KLF1 +/- mice. At both time points the BCL11A KO resulted in a statistically significant increase in the γ/(γ+β) ratio compared to the KLF1 +/-, and the BCL11A -/- : KLF1 +/- resulted in a statistically significant increase in the γ/(γ+β) ratio compared to the single knockouts (Table 1). They also studied the effect on γ-globin expression postnatally in young 4-5 week-old mice and adult >10 weeks old mice. They performed qRT-PCR from bone marrow RNA. They saw the same results as in the embryonic mice. The increase in γ-globin mRNA was reduced in adult compared to embryonic mice but still significant. This is not unexpected because it is difficult to reactivate the silenced γ-globin gene in mice. It is also worth noting that this study saw a smaller fold increase of γ-globin mRNA than in the other studies. An explanation for this could be because KLF1 and BCL11A are in the same pathway. KLF1 normally works to activate the role of BCL11A in repressing γ-globin expression, so knocking down or knocking out two genes that are in the same pathway could give less potential for a greater increase in γ-globin expression.

Target gene #1	Target gene #2	Model System	Genotype	Increase in γ-globin mRNA in BCL11A -/- compared to control	Increase in γ-globin mRNA in target gene #2 compared to control	Increase in γ-globin mRNA in dual KD compared to control	Increase in ratio of HbF to HbA for single targets compared to control	Increase in HbF to HbA for dual compared to control	Fold change between single and dual KD
BCL11A	LSD1	EpoR-Cre βYAC 8-12 week-old transgenic mice	BCL11A -/- : LSD1 +/-	0.02% to 8.4% (383-fold)	X	0.02% to 13.5% (615-fold)	X	X	1.6-fold
BCL11A	DNMT1	EpoR-Cre βYAC 8-12 week-old transgenic mice	BCL11A -/- : DNMT1 +/-	0.02% to 11.2% (560-fold)	X	0.02% to 23.6% (1,180-fold)	X	X	2.1-fold
BCL11A	KLF1	PAC8.1 mice at E18.5	BCL11A -/- : KLF1 +/-	5.7% to 37% (6.5-fold)	5.7% to 14% (2.4-fold)	5.7% to 49% (8.6-fold)	X	X	BCL11A: 1.3-fold KLF1: 3.5-fold
BCL11A	LRF	HUDEP-2 cells	BCL11A KO/LRF KO	X	X	X	2% to 60% (30-fold)	2% to 90% (45-fold)	1.5-fold

Table 1: Fold Changes in γ-globin Expression in Dual Knockdowns Previously Seen in the Literature
Data from some of the dual knockdown/out experiments discussed above. An X indicates data was not reported for that measure.

Chapter 6: Materials and Methods

1. 293T Culture

293T cells are cultured in T75 flasks at 37°C with 5% CO_2 in Dulbecco's Modified eagle Medium (DMEM). A 500 mL bottle of DMEM is made by adding of 10% fetal bovine serum (FBS), 1% of 100x Penicillin/Streptomycin (P/S), 5 mL of 1M HEPES buffer, 5 mL of 100x non-essential amino acids, and 5 mL of 200 mM L-alanyl-L-glutamine. When the cells reach approximately 70% confluency they are split. To split the cells the DMEM media was vacuumed off and the attached cells are washed with 5 mL of PBS. The PBS is vacuumed off and then 2 mL of 0.05% trypsin is added to the cells to detach them. The flask with the trypsin is placed in the incubator for 1 minute until the cells start to detach. Then 7 mL of DMEM complete media + P/S is added to the flask and all of the media is transferred to a centrifuge tube. The tube is centrifuged in an Eppendorf Centrifuge 5810 R (15 amp version) at 800 rpm for 5 minutes. After centrifuging, the supernatant is vacuumed off and the cell pellet is resuspended with 1 mL of DMEM complete media +P/S. Then the cells are either split between 3 T75 flasks at volumes of 20 mL each or a fraction of the cell pellet is seeded in one T75 flask at a volume of 20 mL and a density of 4 x 10^6; the rest of the cells are discarded.

2. Lentivirus Packaging

293T cells are seeded at 4.5 x 10^6 cells per 10 cm dish (typically 10 dishes per desired virus). The dishes are placed in the incubator at 37°C with 5% CO_2 overnight to allow the cells to attach. The next day the attached cells are used to package the lentivirus. A mixture is made for each virus plasmid (shBCL11A, shMBD2, shMBD2 and shBCL11A dual vector (shM/B dual),

or shSCR): 250 µL opti-MEM, the volume of desired virus plasmid to have 8 µg, the volume of psPAX2 to have 6 µg, pMD2.G to have 4 µg. psPAX2 is the packaging vector, which is an empty lentiviral backbone used to package the cDNA into virus. pMD2.G is envelope vector that coats the virus. A separate mixture of 196 µL of opti-MEM and 54 µL of PEI (1 µg/µL) is made and then added to the mixture with the plasmids. PEI is polyethylenimine. It is responsible for condensing DNA into positively charged particles that allow for the DNA:PEI complex to be more easily endocytosed by the cells and the DNA released into the cytoplasm (269). The mixture is incubated at room temperature. The media on the plates is changed to 5 mL of DMEM complete media with 10% FBS and with no P/S. After 8-10 minutes of incubation, 500 µL of the mixture is pipetted drop by drop in a spiral pattern onto the plates. The plates are placed at 3% CO_2 overnight. The next morning the cells are checked for GFP using the Olympus Ix70 fluorescent microscope at 200x magnification and the media is changed to 5 mL DMEM complete media with 5% FBS and 1% P/S. Then the dishes are placed at 10% CO_2. On the third day this media is collected by pipetting it off and transferring into a 50 mL tube. 5 mL of new DMEM complete media with 5% FBS is pipetted onto the dishes. The tube of collected virus is centrifuged at 3,000 g at 4°C for 5 minutes. The supernatant containing the virus is poured into a 50 mL tube and stored at 4°C. The following day, the fourth day, the supernatant collection process is repeated and then the culture dishes are discarded.

3. Ultracentrifugation to Concentrate Virus

The 30 mL of supernatant collected from lentivirus packaging is filtered using a 0.45 µM PES filter, a 10 mL syringe, 21G x 1.5 (0.8 mm x 40mm) gauge needles, and ultracentrifuge tubes are. A 10 mL syringe is used to obtain the virus supernatant. Then the syringe is attached the filter and placed onto an ultracentrifuge tube and the virus is then be filtered through the PES

membrane and into the ultracentrifuge tube. This process is repeated until all of the virus is filtered. A needle is used when the virus volume gets low and the syringe cannot reach the bottom of the tube; the needle can be attached to the syringe to reach to the bottom. The ultracentrifuge tubes are balanced and parafilm is used to seal them. A Beckman Coulter Optima L-80 XP Ultracentrifuge with the SW32Ti rotor to centrifuge the samples at 4°C at 28,000 rpm for 90 minutes. The buckets are then removed and opened in the hood to remove the ultracentrifuge tubes. The parafilm is removed and the supernatant is vacuumed off. Then, the pellet containing the virus is resuspended in 400 µL of SFEM II complete media – P/S. The tubes containing the virus are then covered with parafilm and placed at 4°C overnight. The following day the media is collected from the ultracentrifuge tubes and transferred to 1.5 mL centrifuge tubes. The tubes are centrifuged at maximum speed at 4°C for 15 minutes to get rid of any sediment. Then, the supernatant containing virus is transferred to a 1.5 mL centrifuge tube in 100 µL aliquots. The aliquots are stored at -80°C. The titer of the frozen aliquots is reduced by approximately 50% compared to fresh virus.

4. HUDEP-2 Cell Culture

HUDEP-2 cells are grown at a density of 2.5 x 10^5 per T25 flask. The media used is SFEM II. Complete media is made by taking 45 mL aliquots and adding 45 µL each of DOX (1 µg/mL) and EPO (3 IU/mL), 22.5 µL of SCF (100 µg/mL), 9 µL of DEX (10^{-6}M), and 450 µL of L-alanyl-L-glutamine (200 mM). 5 mL of media is used on the first day of passage. Two days later, 2 more mL of media are added to the flasks. On the third day the cells are split by collecting the media in a 15 mL centrifuge tube and centrifuging at 300 g for 5 minutes. After centrifuging the media is vacuumed off and the pellet is resuspended in 1 mL of SFEM II complete media. Then a fraction of the cells are pipetted into a T25 flask with 5 mL of fresh

SFEM II complete media to achieve the density of 2.5 x 10^5cells per flask. Cells are incubated at 37°C at 5% CO_2.

5. Creation of BCL11A Vector

All viral vectors used in this thesis were in the pRRL expression vector. The MBD2 shRNA vector was gifted from the Ginder lab. The map of the shMBD2 vector is in Figure 10. The shBCL11A vector was created by using the Broad Institute's website to search for human BCL11A sequences. One of the oligo sequences, oligo 1 or S1, was selected from a study that had designed it and had shown success using it for knockdowns (246). The map for the shBCL11A vector can be found in Figure 11. Four were selected to design four different pairs of oligos (Tables 2 and 3). The forward and reverse oligos for each pair were designed and ordered to be synthesized using IDT. Once the oligos arrived they were centrifuged at 12,000g for 1 minute to make sure all of the powder was at the bottom of the tube. Then the samples were reconstituted in water to bring their concentration to 100 μM. Next, the forward and reverse oligos were phosphorylated and annealed. To do this 1 μL of the forward oligo, 1 μL of the reverse oligo, 1 μL of 10x T4 ligation buffer, 6.5 μL Hyclone H_2O, and 0.5 μL of T4 PNK (10 μL total) were combined in a tube. This was done for each of the four sets of oligos. To anneal the oligos, they were incubated in the Bio Rad C1000 Touch Thermal Cycler with the following parameters: 37°C for 30 minutes, 95°C for 5 minutes, and then ramp down to 25°C at 5°C/min. An annealed a diluted stock was made: 1 μL of the annealed primers was added to 199 μL hyclone water to make a 1:200 dilution. Next, the plasmid the oligos were to be inserted into had to be digested.

The shBCL11A oligos were cloned into the vector containing shMBD2. This shMBD2 plasmid was double digested with XhoI and MluI, with recognition sites located on either side of the shMBD2, to remove the shMBD2. All enzymes were obtained from Thermo Scientific and 1 μL of enzyme can digest 1 μg of plasmid DNA in 5 minutes. Single digests with one restriction enzyme were performed to ensure the restriction enzymes were cutting. A 50 μL digestion was performed for the double digest and 20 μL digests were used for the single digestions. To perform the double digest a mixture was made containing: 2.5 μg of shMBD2 plasmid, 5 μL of FD (10x) green buffer, 2.5 μL MluI restriction enzyme, 2.5 μL XhoI restriction enzyme, and H_2O to 50 μL. For the single digests a mixture was made containing: 1.0 μg of shMBD2 plasmid, 2 μL of FD (10x) green buffer, either 1.0 μL MluI restriction enzyme or 1.0 μL XhoI restriction enzyme, and H_2O to 20 μL. The restriction digests were sealed with parafilm and incubated in a water bath at 37°C for at least 2 hours. The reaction was terminated by placing the tubes on a heat block at 80°C for 5 minutes. Then the products were separated on a 0.7% agarose gel. The samples were loaded into respective lanes on the agarose gel and then run at 130V for 30 minutes. A UV light was used to visualize the bands. The double digested band was excised and placed in a 1.5 mL centrifuge tube that had previously been weighed. DNA purification was performed to obtain the vector that no longer contains shMBD2, using the Omega E.Z.N.A Gel Extraction Kit.

The annealed oligos were then ligated into the plasmids. A mixture was prepared containing 50 ng double digested plasmid, 1 μL diluted oligo, 5 μL 2x quick ligase buffer, 1 μL Quick Ligase, and hyclone H_2O to 11 μL. A mixture was made for each oligo, as well as one that contains water in place of the oligo as a negative control. The mixtures were incubated at room temperature for 1 hour then they were transformed into Stbl3 competent cells.

Transformation into Stbl3 competent cells: 11 μL of each sample was pipetted into respective tubes containing 50 μL of Stbl3 and pipetted up and down to mix. The mixtures were incubated on ice for 30 minutes. During this time agar plates were taken out of the fridge and placed at room temperature to warm up. Agar plates are made with LB media containing ampicillin (100 μg/mL) for colony selection. The samples were then heat shocked at 42°C for 45 seconds, and then placed on ice for 2 minutes. Next, the mouth of the bottle containing SOC media was passed through a flame to sterilize it, then 200 μL of SOC media was pipetted into each sample. The tubes were then placed in the shaking incubator at 32°C at 180 rpm for 90 minutes. The tubes were retrieved from the incubator and 200 μL of each sample was pipetted onto its respective agar plate and 6-8 glass beads are added to each plate. The plates were shaken horizontally by hand for 10 minutes and then placed in an incubator at 32°C for 10 minutes. The plates were retrieved from the incubator and the glass beads were removed. The plates were placed back in the incubator, upside down, at 32°C overnight. The next day the plates were checked to confirm that no or very few colonies grew on the negative control plate compared to the plates containing samples. If the negative control plate had growth it would indicate the plasmid was not digested and the growth seen on the plates containing samples likely does not contain the product with the oligos ligated in. A pipette tip was used to select 4 single colonies. Then the pipette tip was dropped into a test tube containing 3 mL of LB broth. Additional colonies from each sample were selected to spread on fresh agar plates. The plates were incubated at 32°C overnight and the tubes were capped and placed in the shaking incubator at 220 rpm at 32°C overnight. The next day the plasmid purification is performed on each sample using the Omega E.Z.N.A Plasmid DNA Mini Kit II. The product obtained is nanodropped using the NanoDrop ND-1000 Spectrophotometer and based on the concentrations, 260/280 and

260/230 measurements two plasmid preparations with the best measurements – highest concentration, 1.9 – 2.1 for 260/280, and 2.0 – 2.4 for 260/230 – from each set of oligos were selected to send to Eurofins Genomics for sequencing. When the sequencing results were obtained, they were copied into NCBI Blast global align tool along with the designed oligo sequences to check that the plasmid products were what we expected them to be. All sequences were fully aligned, indicating they were complete and contained our intended oligos.

Once it was confirmed that the oligos contained the intended shBCL11A, they were used to package lentivirus following the protocol above. The lentivirus was not concentrated; fresh virus was used to infect 293T cells to the test efficiency of the shBCL11A plasmids. 8 x 10^5 293T cells were seeded per well in a 6 well plate with DMEM complete media – P/S one day before infection. To infect the cells the media was vacuumed off and 2 mL of virus was pipetted onto the cells as well as 4 μL of 4 μg/μL polybrene and then placed at 37°C at 5% CO_2 for 7 hours. Then the media was vacuumed off and replaced with 2 mL of DMEM complete media + P/S. After 3 days, the 293T cells were collected by scraping them off of the plate and then centrifuged at 800 rpm. The cells were then used to extract RNA and make cDNA to run a qRT-PCR following the protocol below. The knockdown efficiency of each shRNA was measured. shRNA #1 (S1) was selected to be used for BCL11A knockdown experiments because it gave the best knockdown efficiency, about 70%, of the 4 shRNAs tested. This was the shRNA previously designed by another group (246). We used shBCL11A S1 to perform the experiments in this thesis.

shBCL11A	Oligo Target
Oligo 1	ACAGAACACTCATGGATTAAG
Oligo 2	ACCTCTCCATGGGATTCATAT
Oligo 3	GCCAACCTAATTACCTGTATT
Oligo 4	GCATAGACGATGGCACTGTTA

Table 2: BCL11A Target Sequences
These are the sequences selected from the Broad Institute website to be used to design oligos.

Oligo Name	Forward Oligo	Reverse Oligo
Set 1	5 – CGCGTCCCCACAGAACACTCATGGATTAAG TCAAGAGCTTAATCCATGAGTGTTCTGTTTTTTGG AAC – 3	5 – TCGAGTTCCAAAAAACAGAACACTCATGGA TTAAGCTCTTGACTTAATCCATGAGTGTTCTGTGG GGA – 3
Set 2	5 – CGCGTCCCCACCTCTCCATGGGATTCATAT TCAAGAGATATGAATCCCATGGAGAGGTTTTTTG GAAC – 3	5 – TCGAGTTCCAAAAAACCTCTCCATGGGATTC ATATCTCTTGAATATGAATCCCATGGAGAGGTGG GGA – 3
Set 3	5 – CGCGTCCCCGCCAACCTAATTACCTGTATT TCAAGAGAATACAGGTAATTAGGTTGGCTTTTTGG AAC – 3	5 – TCGAGTTCCAAAAAGCCAACCTAATTACCT GTATTCTCTTGAAATACAGGTAATTAGGTTGGCGG GGA – 3
Set 4	5 – CGCGTCCCCGCATAGACGATGGCACTGTTA TCAAGAGTAACAGTGCCATCGTCTATGCTTTTTGG AAC – 3	5 – TCGAGTTCCAAAAAGCATAGACGATGGCAC TGTTACTCTTGATAACAGTGCCATCGTCTATGCGG GGA – 3

Table 3: shBCL11A Oligo Sequences
These are the oligo sequences designed and ordered from IDT. Red = MluI restriction enzyme sequence. Blue = sense strand. Green = loop. Purple = anti-sense strand. Yellow = XhoI restriction enzyme sequence.

6. Creation of the M/B Dual Viral Vector containing both the MBD2 and BCL11A shRNAs

To create the dual vector, the shBCL11A was amplified from its PRRL vector and ligated into the PRRL vector already containing shMBD2. The map for the shM/B dual vector is in Figure 12. The site for inserting shBCL11A was chosen directly 3' to shMBD2, between the

XhoI and EcoRV restriction sites. Primers were designed to target the H1 promoter and the shBCL11A of the shBCL11A vector. The forward primer was designed to have an XhoI site on the 5' end so that the product can be ligated into the cut plasmid. The reverse primer was designed to have a one base pair mismatch to XhoI so that the final product would not have two XhoI restriction sites as it would cause problems during digestion. The primer sequences are found in Table 4. The primers were used to amplify the region of interest from the shBCL11A PRRL vector. The Q5 Hot Start High-Fidelity 2x Master Mix protocol was followed: 12.5 μL Q5 high fidelity 2x master mix, 10 μM forward primer, 10 μM reverse primer, 0.1 ng shBCL11A PRRL plasmid, and DEPC H_2O to 25 μL. The mix was briefly vortexed and centrifuged. Then it was placed in the Bio Rad C1000 Touch Thermal Cycler using the following parameters: Initial denaturation: 98°C for 30 seconds; 35 cycles of: 98°C for 10 seconds, 65°C for 20 seconds, and 72°C for 10 seconds; the final extension step is 72°C for 2 minutes. The product was then run on a 2% agarose gel at 90V for 50 minutes. The product size is 308 bp. The resulting band was cut out of the gel and purified using the Omega E.Z.N.A gel extraction kit.

Next, a double digest of the shMBD2 PRRL vector was performed. The same protocol was followed as described in the section for designing the shBCL11A vector, except the restriction enzymes XhoI and EcoRV were used in this digest. The restriction enzymes were also used for digestion one at a time because they are located close to each other in the plasmid; 2.5 μL of EcoRV was added first and the mixture was incubated at 37°C for 2 hours. After 2 hours the XhoI enzyme was added and the mixture was incubated at 37°C for another 2 hours. The sample was gel purified following the protocol described above. The purified shBCL11A was also digested with XhoI and EcoRV to create compatible ends for ligation using the protocol

described above. Then, the sample was cleaned up using the Monarch PCR and DNA cleanup kit.

The BCL11A PCR and shMBD2 plasmid digested products were then ligated together. A mixture was made with 2 µL T4 10x ligation buffer, 50 ng of the shMBD2 PRRL double digested vector, 5.94 ng of the shBCL11A PCR insert (this concentration was determined using the online NEBiocalculator tool), 1 µL T4 ligase, 2 µL 10 mM ATP, and DI H_2O to 20 µL. This was incubated at room temperature for 1 hour and then was transformed into Stbl3 cells following the protocol outlined in the previous section. The protocols for streaking agar plates and inoculating LB broth and the plasmid preparations were also followed. The sequencing showed both of the shRNAs were present in the final product, so virus packaging and transfections were performed following the protocol laid out above.

Forward Primer	5' – CGTATATTCTCGAGTTCGAACGCTGACGTCATC – 3'
Reverse Primer	5' – GGAATTCGATATCAAGCTTCTCGcGTTCCAAAAAAC – 3'

Table 4: Primers Targeting the shBCL11A
These are the primers sequences designed and ordered from IDT. Red = XhoI restriction enzyme sequence. Blue = EcoRV. Green = shBCL11A. Yellow = H1 promoter. The reverse primer was designed to have a one base pair mismatch (the small c) so that the final product would not have two XhoI restriction sites as it would cause problems during digestion.

7. Lentiviral Infection

2 x 10^5 HUDEP-2 cells are plated per well of a 12 well plate. The number HUDEP-2 cells needed is determined, then they are centrifuged 300 g for 5 minutes. The supernatant is vacuumed off and the pellet is resuspended with 1 mL of SFEM II complete media – P/S. Then the remaining volume of SFEM II complete media – P/S needed to pipette 600 µL of cells at a density of 2 x 10^5 into the desired number of wells of the 12 well plates is added to the

resuspended cell pellet. Next, 100 μL of appropriate virus from frozen stock is pipetted into well(s) to achieve a dilution of 1:20. If using fresh virus the volume used would be 50 μL of virus to achieve the 1:20 dilution, but because frozen virus is used 100 μL of virus has to be used to achieve the same dilution to account for the approximately 50% loss of virus due to being frozen. The total volume of each well of the 12 well plate should be 1 mL, so 300 μL of SFEM II complete media – P/S is added to make up the remaining volume of the well. The plate is sealed and centrifuged in an Eppendorf Centrifuge 5810 R (15 amp version) at 2,800 rpm at 32°C for 90 minutes. After centrifuging the plate is incubated at 37°C and 5% CO_2 for 3 hours. Then 1.5 mL of SFEM II complete media – P/S is added to each well and then the plate is placed in the incubator. The following day the cells are resuspended using a bulb pipette and transferred into labeled 15 mL centrifuge tubes. The tubes are centrifuged at 300 g for 5 minutes, and the supernatants are vacuumed off. Then, each pellet is resuspended with 1 mL of SFEM II complete media + P/S and transferred into its respective well in a 6-well plate. Then 1 mL more of SFEM II complete media + P/S is added to each well, bringing the total volume to 2 mL and the plates are placed in the incubator. After 72 hours the cells are checked for GFP using the Olympus Ix70 fluorescent microscope at 200x magnification and 2 mL more of SFEM II complete media + P/S is added to each well of cells. The following day each sample is transferred into a T25 flask and SFEM II complete media + P/S media is added to bring the volume of each flask to 5 mL. The following day 2 mL more of media is added to each flask. 5 days after infection the cells are collected to be differentiated. For quality control, every time an infection is done 2 wells are infected with SCR lentivirus and 1 well is also prepared as an uninfected control. The SCR control is what the results are compared to, but the uninfected control is included to make sure no GFP is seen and to compare the SCR to. The GFP check is important because the lentiviral

plasmid contains a GFP element, so GFP is used to assess the efficiency of infection. At least 80% GFP expressing cells indicates the infection worked well. If there is GFP in the uninfected control well it indicates there was a contamination and the experiment needs to be thrown. The uninfected control is also used for qRT-PCR to compare the SCR results to make sure the SCR lentivirus did not knock down any of the genes or affect γ- and β-globin gene expression.

8. Cell Differentiation

1 million cells are differentiated per well of a 6-well plate. Cell number is counted for each sample and it is determined how many wells will need to be differentiated per sample, i.e. if the cell number for sample X is 5.6 x 10^6 then 6 wells will be used to differentiate this sample. Per well, 3 mL of differentiation media will be needed; 2 mL when initially plating the cells and 1 mL more 2 days later. When differentiating the cells, Iscove Modified Dulbecco Media (IMDM) media is used. To make the media, P/S is added to IMDM to 2% of the volume. The volume of IMDM + P/S needed for the differentiation experiment is calculated. Then Human AB serum is added to 5% of the volume. The IMDM plus Human AB serum media is filtered using a 0.45 μm PES filter. Once the media is filtered DOX (1 μg/mL), EPO (3 IU/mL), Insulin (10 μg/mL), Heparin (3 U/mL), 100x Holo-Transferrin (500 μg/mL), and 1x L-alanyl-L-glutamine (200 mM) is added to the media. The cell pellets are resuspended in full media to have 1 x 10^6 cells per 2 mL of media for each well, i.e. if sample X needs to be differentiated in 6 wells then the cell pellet will be resuspended in 6 mL of IMDM media. 1 mL of cells are plated in 6 well plates and then 1 mL more of IMDM is added to each well to bring the total volume per well to 2 mL. Then the plate(s) are placed in the incubator at 37°C with 5% CO_2. Two days later 1 mL of differentiation media is added to each well. Exactly 72 hours after differentiation is initiated the cells are collected in 15 mL centrifuge tubes and centrifuged at 1,500 rpm for 5 minutes. The

supernatant is vacuumed off, the cell pellet is resuspended with 5 mL of PBS, and centrifuged again at 1,500 rpm. Finally, the supernatant is carefully removed and then the cell pellet is stored at -80°C.

9. RNA Extraction

Cell pellets are resuspended in 750 µL of Trizol and incubated at room temperature for 5 minutes. Then, 200 µL of Chloroform is added to each sample, they are vortexed for 15 seconds, and incubated at room temperature for 15 minutes. The samples are centrifuged in an Eppendorf Centrifuge 5424 R at 12,000 g at 4°C for 15 minutes. The upper, clear aqueous phase is transferred to a new tube, taking care not to transfer any of the lower, pink layer. Then, 5 µL of glycogen (final 5 mg/mL) is pipetted into each sample. Next, an equal volume of 2-proponol is added to each sample. The samples are inverted 6-8 times (do not shake) and placed at -20°C overnight. The samples are centrifuged at 12,000 g at 4°C for 15 minutes, and the supernatant is discarded, leaving the RNA pellet. The pellet is washed with 1 mL of chilled 70% RNA ethanol and then centrifuged at 7,500 g at 4°C for 5 minutes. The supernatant is removed and the tube with pellet is put on ice and left to dry. The pellet is resuspended in 10 µL of DEPC water. The NanoDrop ND-1000 Spectrophotometer is used to determine the purity and concentration of the resuspended RNA pellet.

10. cDNA Preparation

A 20 µL reaction is used to make the cDNA, where 1 µL is Oligo dT (0.5 µg/µl), 4 µL is 5x Reaction Buffer, 0.5 µL is RiboLock RNase inhibitor (40 U/µL), 2 µL is dTNP mix (10 mM), 1 µL is RevertAid M-MuLV Reverse Transcriptase (200 U/µl), and the remaining 11.5 µL is comprised of the volume of RNA needed for 2,000 ng and DEPC water. The final concentration

of mRNA in the reaction is 100 ng/µL. Once the samples are prepared, they are loaded into Bio Rad C1000 Touch Thermal Cycler set to incubate at 42°C for 60 minutes and then terminate the reaction by heating to 70°C for 10 minutes. As a quality control one of the samples is used for an additional reaction where no RevertAid M-MuLV Reverse Transcriptase is added. cDNA samples are stored at -20°C.

11. qRT-PCR Protocol

A 20 µL reaction is prepared with 10 µL of Quantabio PerfeCta MultiPlex qPCR ToughMix Low ROX mix, 1 µL of Thermo Fisher predesigned TaqMan probe (human γ- globin, human β-globin, hMBD2, hBCL11A, or hPPIA), 8 µL of DEPC water, and 1 µL of cDNA (equal to 100 ng of RNA). The Thermo Fisher TaqMan probes are predesigned assays that contain the TaqMan probe and the PCR primer set specific for the product to be amplified. TaqMan works by annealing probes that contain a fluorophore to the sample and then when the Taq Polymerase synthesizes a new strand it cleaves the probe resulting in the release of florescence that is measured by the instrument. A master mix is made with the qPCR ToughMix, TaqMan probe, and DEPC water. 19 µL of the master mix is pipetted into the corresponding wells in the 96 well plate. Then 1 µL of cDNA is pipetted into respective wells. Each sample is run in triplicate in a 96 well plate. The plate is then sealed and wrapped in foil to protect it from light exposure while it is transported to the instrument. The foil wrapped plate is briefly vortexed and then spun down. The plate is then loaded into the QuantStudio 3 applied biosystems instrument by Thermo Fisher Scientific. The ΔΔCT method is used to calculate relative amounts of RNA and the running parameters used were 95°C for 3 minutes for initial denaturation, 95°C for 15 seconds and then 60°C for 1 minute. The last two steps are repeated 40 times. On every plate two quality control measures are included. One is a well that contains the master mix but 1

μL of DEPC water is added instead of cDNA. This is done to check that none of the reagents used in the experiment were contaminated and there should be no amplification for this well. The other quality control is to include one well that uses 1 μL of the cDNA sample prepared without Reverse Transcriptase. This well should also not be amplified and checks that none of the reagents used to make cDNA were contaminated and that genomic DNA is not being amplified. If amplification is seen in either of these wells then it indicates there is an issue with the water or reagents and the results cannot be used.

12. Statistical Analysis

One-way ANOVA with a Tukey-test for all pairwise comparisons was run using SigmaPlot 14. $p < 0.05$ was considered significant.

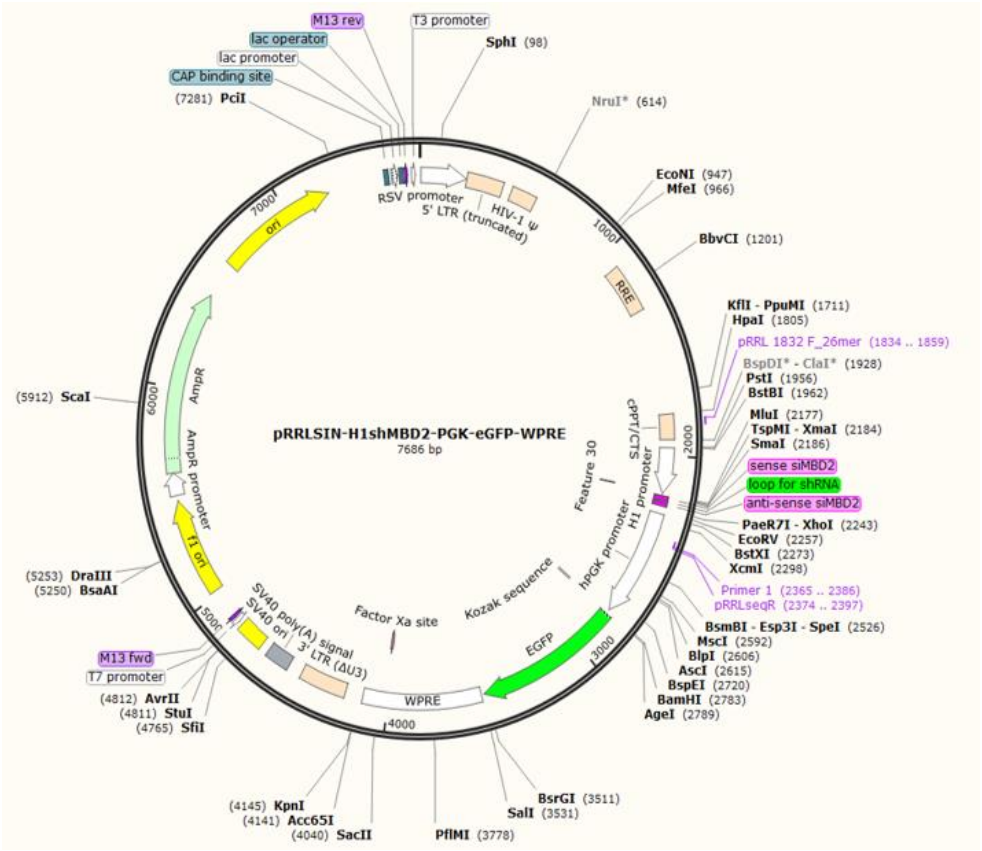

Figure 10: shMBD2 Map
The mapping for the PRRL vector containing the shMBD2. The location of the shRNA is indicated by the pink highlighting. The XhoI, MluI, and EcoRV restriction sites can be seen on either side of the pink highlighting. The H1 promoter is identified by the white arrow directly before the pink highlighting. The region that provides GFP is indicated by the bright green arrow. The region that provides ampicillin resistance is indicated by the mint arrow. This plasmid and mapping was gifted from the Ginder Lab.

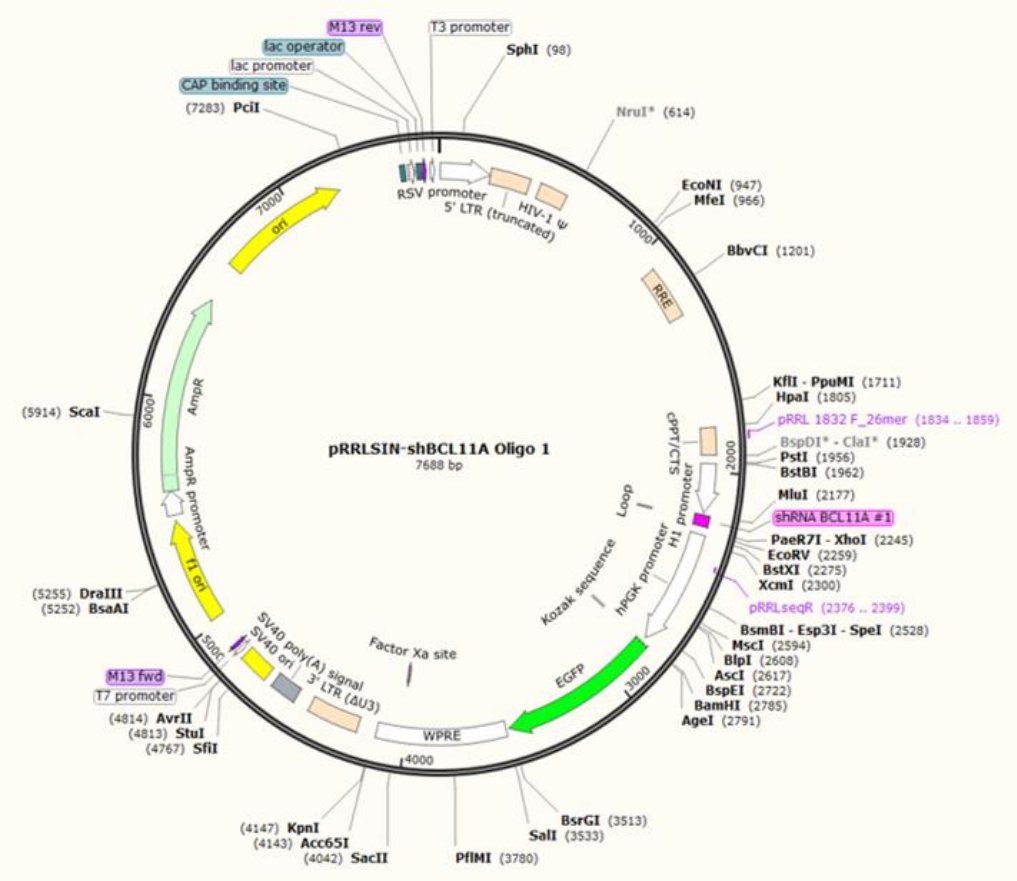

Figure 11: shBCL11A Map
The mapping for the PRRL vector containing the shBCL11A. The location of the shRNA is indicated by the pink highlighting. The XhoI and EcoRV restriction sites can be seen on either side of the pink highlighting. The H1 promoter is identified by the white arrow directly before the pink highlighting. The region that provides GFP is indicated by the bright green arrow. The region that provides ampicillin resistance is indicated by the mint arrow.

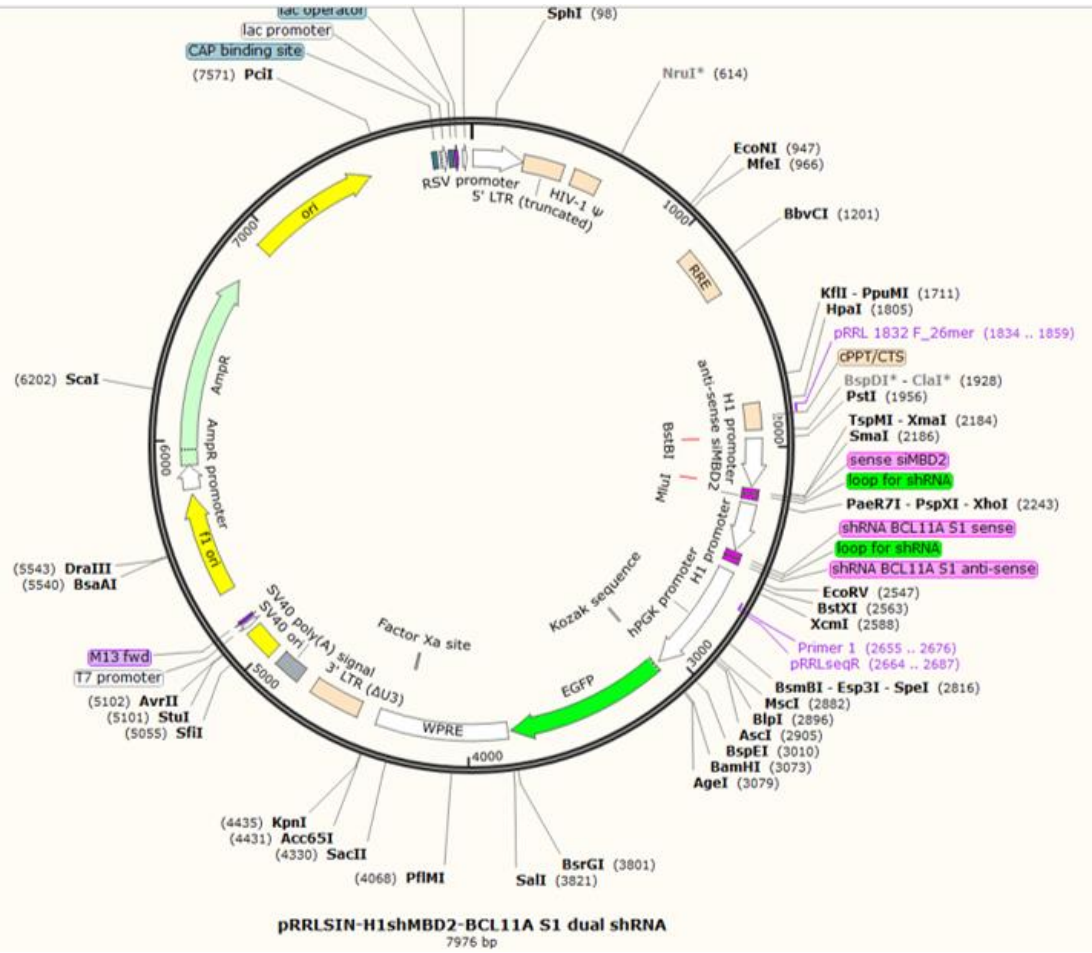

Figure 12: shMBD2/shBCL11A Dual Expression Vector Map
The mapping for the PRRL vector containing the shMBD2 and shBCL11A. The location of the shRNAs is indicated by the pink highlighting. The H1 promoter is identified by the white arrow directly before the pink highlighting. The region that provides GFP is indicated by the bright green arrow. The region that provides ampicillin resistance is indicated by the mint arrow.

Chapter 7: Rationale and supporting evidence for dual knockdowns of γ-globin repressors

Introduction

As discussed in Chapter 5, there is good evidence that targeting two different genes that repress the γ-globin gene is a feasible method for increasing γ-globin mRNA expression more than single knockdowns. This thesis proposes to perform a study where BCL11A and MBD2 are knocked down simultaneously. These genes were chosen because they are the most commonly studied; there is good evidence single knockdowns or knockouts of these genes gives a significant increase in γ-globin mRNA expression (153) (168) (171) (246) (247) (248). This thesis proposes performing knockdown of the genes rather than knockout. There are several reasons for this. One is that knockout could increase γ-globin expression so much that it would not allow us to detect an additive effect of the knockdown of a second gene. Another is that it is not feasible to knockout an entire gene in patients as there could be deleterious effects. We propose using the HUDEP-2 cell system because it portrays adult erythropoiesis. When HUDEP-2 cells are differentiated they express mostly β-globin. This is a useful model because it mimics human adult erythroid cells that have very little expression of γ-globin. After performing the experiments in HUDEP-2 cells, the results should be verified using erythroid-differentiated CD34+ cells. We think the dual knockdown combination will work better than the single knockdowns at increasing γ-globin mRNA expression. This thesis will present preliminary data to support our hypothesis that dual knockdowns increase γ-globin mRNA expression more than single knockdowns due to an additive effect. Our goal is the data presented here will encourage the analysis of other dual combinations of KLF1 and MBD2, and KLF1 with BCL11A. These

findings provide better insight about combining repressor genes to target as a therapeutic approach.

Results

First, the four oligos designed to target BCL11A were tested for their knockdown efficiency. Lentivirus was made for each of the four BCL11A shRNAs following the protocol described in Chapter 6. To check knockdown efficiency of the shRNAs, 293T cells were infected with the lentivirus. The results showed that shRNA #1 (S1) knocked down BCL11A by about 60% and shRNA #3 (S3) knocked down BCL11A by about 30% compared to the scramble (SCR) control shRNA. shRNAs #2 and #4 resulted in no knockdown of BCL11A compared to the SCR control (Figure 13). S1 and S3 lentivirus was then used to infect HUDEP-2 cells to verify the findings seen in 293T cells. S1 lentivirus resulted in about 60% knockdown of BCL11A compared to SCR, but S3 lentivirus showed very little knockdown of BCL11A in HUDEP-2 (Figure 14). From these results, the S1 shRNA was chosen for future experiments.

Single, dual, or SCR control expression vectors were used to package lentivirus in 293T cells, which was then used to infect HUDEP-2 cells. The single vectors are targeting either BCL11A or the MBD2; the dual vector contains both shRNAs and targets BCL11A and MBD2 simultaneously. From this point forward it will be referred to as the M/B dual vector; the control vector is a scrambled sequence. Three days after infection the HUDEP-2 cells were checked for GFP. GFP is used to assess the efficiency of infection because the lentiviral plasmid contains the GFP element. Only infections that resulted in GFP in $\geq$ 80% of cells were used to continue the experiments. A representative microscope image for GFP expression is in Figure 15 and percentages of cells with GFP for the four replicates used in this study are found in Table 5.

Five days after infection the cells were differentiated for three days to become more erythroid, and RNA was prepared for qRT-PCR. Refer to methods in Chapter 7 for more detailed explanations. The qRT-PCR results show 80% knockdown of BCL11A with the single viral construct and 70% knockdown of BCL11A mRNA with the dual viral construct compared to SCR, which are statistically significant knockdowns of BCL11A mRNA (Figure 16A). The results for MBD2 with the single viral construct show 90% knockdown and 70% knockdown for the dual viral construct compared to SCR, which are also statistically significant knockdowns of MBD2 mRNA (Figure 16B). All three of the viral constructs significantly increase γ-globin mRNA compared to the SCR control, with the M/B dual knockdown having higher expression than single knockdowns (Figure 17A, black lines). The knockdown of BCL11A alone results in a mean 22-fold increase of γ-globin mRNA and the knockdown of MBD2 alone results in a 17.5-fold increase in γ-globin mRNA; there is no statistically significant difference between these two knockdowns. The M/B dual knockdown results in a 45-fold increase in γ-globin mRNA. This is a statistically significant increase compared to each of the single knockdowns (Figure 17A, red lines). Cells for all three of the viral constructs exhibit significantly increased β-globin mRNA by about 1.4-fold compared to control, but there is no significant difference in β-globin mRNA between the three viral constructs (Figure 17B). The increase in β-globin is not as large as the increase in γ-globin mRNA expression.

The γ/(γ+β)% for all three of the knockdowns is significantly increased compared to control, as seen in Figure 17C. The mean baseline for γ/(γ+β)% for SCR is 0.21%. Knockdown of BCL11A alone results in a statistically significant increase to 3.2%, a 14.7-fold increase. Knockdown of MBD2 alone results in a statistically significant increase to 2.9%, a 14-fold increase. The M/B dual knockdown results in a statistically significant increase to 6.7%, a 31-

fold increase. There is no significant difference for γ/(γ+β)% between the BCL11A and MBD2 single knockdowns (Figure 17C, black lines), but there are statistically significant differences in the dual knockdown compared to each of the singles (Figure 17C, red lines). The difference between the BCL11A single and the M/B dual knockdown is a 2.1-fold increase. The difference between the MBD2 single and the M/B dual knockdown is a 2.3-fold increase. These findings show that simultaneous knockdown of BCL11A and MBD2 increases γ-globin mRNA more than single knockdowns, and that M/B dual knockdown has an additive effect on increasing γ-globin mRNA expression. The action of the M/B dual knockdown is additive rather than synergistic because the effect it has on increasing γ-globin mRNA and the γ/(γ+β)% is approximately the same as the effect of the single knockdowns if their results were combined. If the action was synergistic the increase in γ-globin mRNA and the γ/(γ+β)% caused by the M/B dual knockdown would be much larger than the effect caused by the single knockdowns.

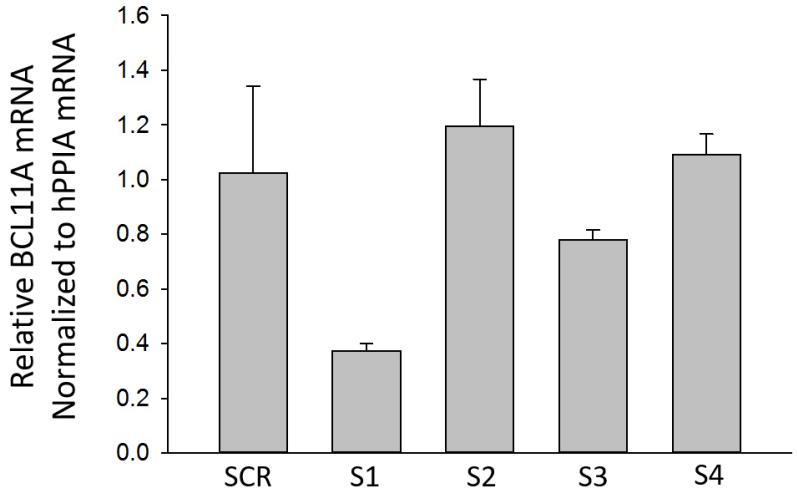

Figure 13: Knockdown Efficiency of the Four BCL11A shRNAs in 293T Cells
qRT-PCR data showing how each shRNA worked at knocking down BCL11A in 293T cells. SCR = scramble control, S1 = shRNA #1, S2 = shRNA #2, S3 = shRNA #3, and S4 = shRNA #4. n = 2. 293T cells were infected for 3 days before collected for qRT-PCR.

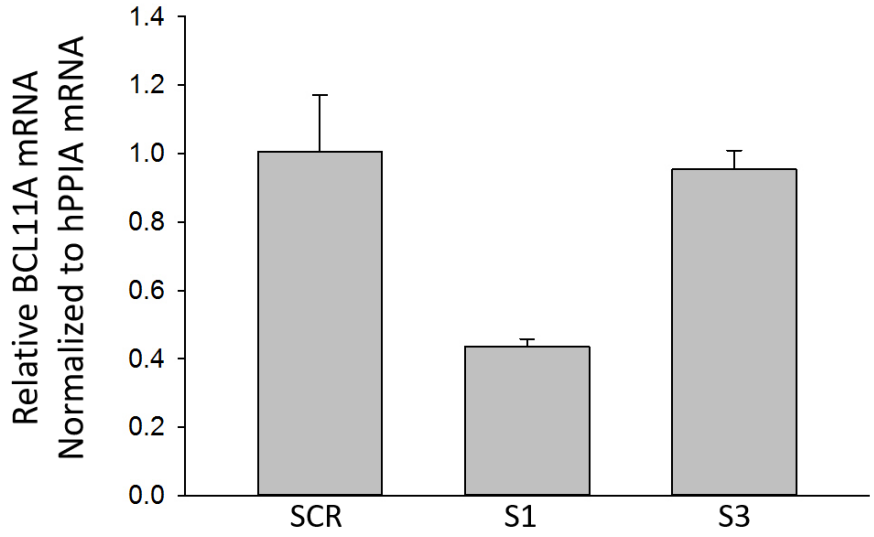

Figure 14: Knockdown Efficiency of S1 and S3 BCL11A shRNAs in HUDEP-2 Cells
qRT-PCR data showing efficiency of BCL11A S1 and BCL11A S3 shRNA knockdown of BCL11A mRNA. SCR is set to 1. SCR = scramble control, S1 = shRNA #1 and S3 = shRNA #3. n = 3. HUDEP-2 cells were infected for 5 days before collected for qRT-PCR.

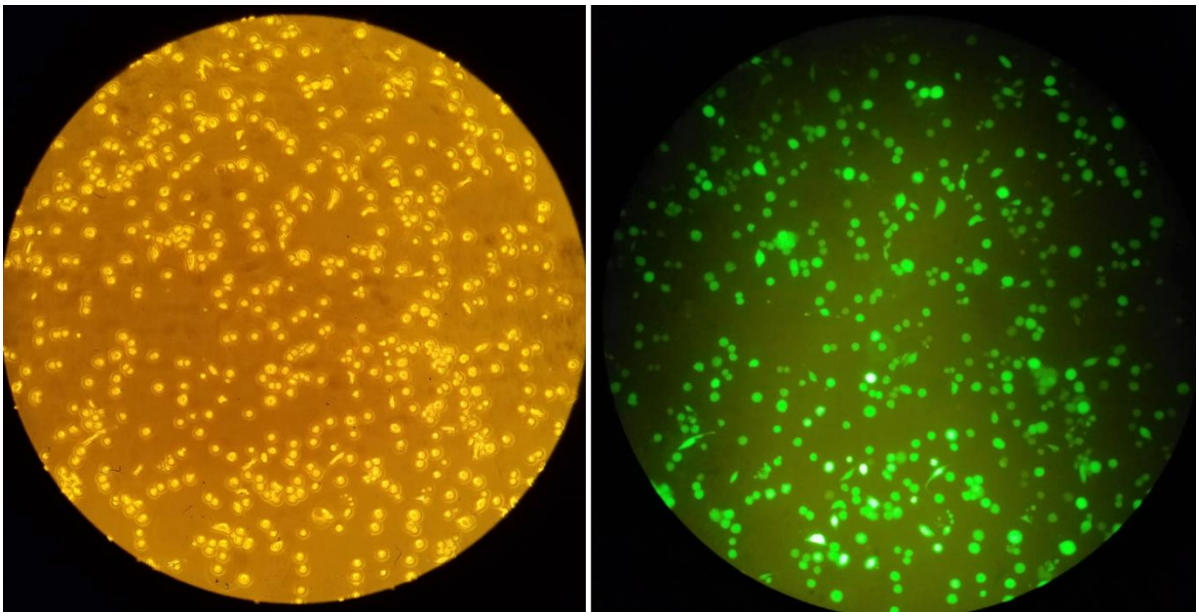

Figure 15: Microscope Images of Lentiviral Infected Cells
HUDEP-2 cells infected with shMBD2 lentivirus. Left: brightfield image to show how many cells are in the field. Right: The same field of cells showing fluorescent cells with GFP to test infection efficiency. 200x magnification was used. Pictures were taken 72 hours after infection.

Trial #	Virus	Number of cells in brightfield	Number of GFP cells	Percentage of cells infected
1	SCR	264	214	81.1
	BCL11A	173	171	98.8
	MBD2	200	200	100
	M/B dual	282	274	97.2
2	SCR	370	360	97.3
	BCL11A	127	121	95.3
	MBD2	132	132	100
	M/B dual	243	202	83.1
3	SCR	200	200	100
	BCL11A	141	131	92.91
	MBD2	136	132	97.0
	M/B dual	178	155	87.1
4	SCR	109	103	94.4
	BCL11A	111	111	100
	MBD2	140	138	98.6
	M/B dual	129	122	94.6

Table 5: Cell Counts for Lentiviral Infected Cells used in this Study
Hand counts of each sample of infected cells under brightfield or green fluorescent light. The shBCL11A used was the S1 construct. M/B dual = the shMBD2 & shBCL11A dual vector. These counts were taken 72 hours after infection.

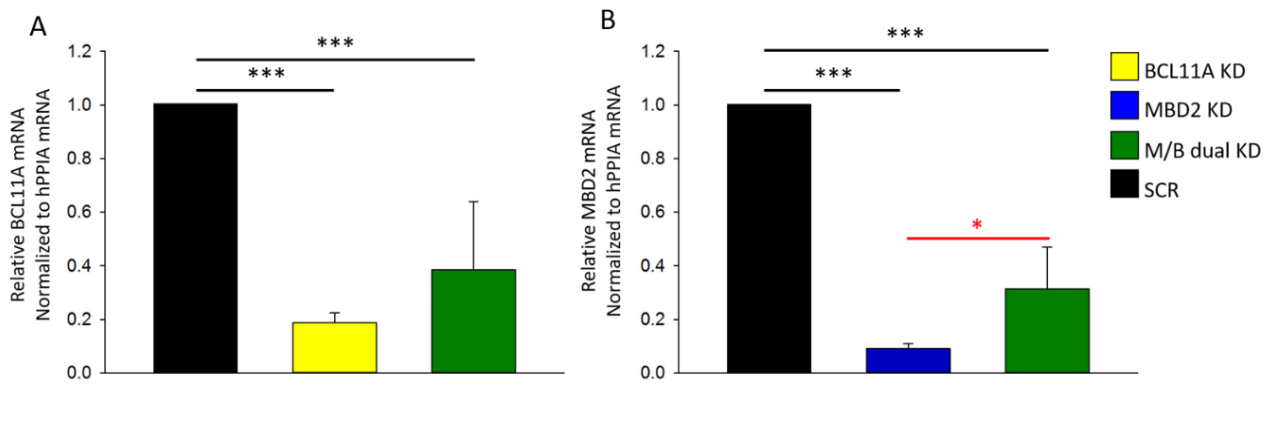

Figure 16: Knockdown of BCL11A and MBD2
qRT-PCR data. Panels A-B are compared to the SCR control set to 1. A: Knockdown of BCL11A using the shBCL11A exclusive lentiviral vector or the M/B dual lentiviral vector. B: Knockdown of MBD2 using the shMBD2 exclusive lentiviral vector or the M/B dual lentiviral vector. Sample size of 4. Statistical test: one-way ANOVA with Tukey correction. Black lines are statistical comparisons to SCR. Red lines are statistical comparisons to the M/B dual results. Cells were infected for 5 days then differentiated for 3 days before collected for qRT-PCR. Significance $p < 0.05$. *<0.05, **<0.01, ***<0.001

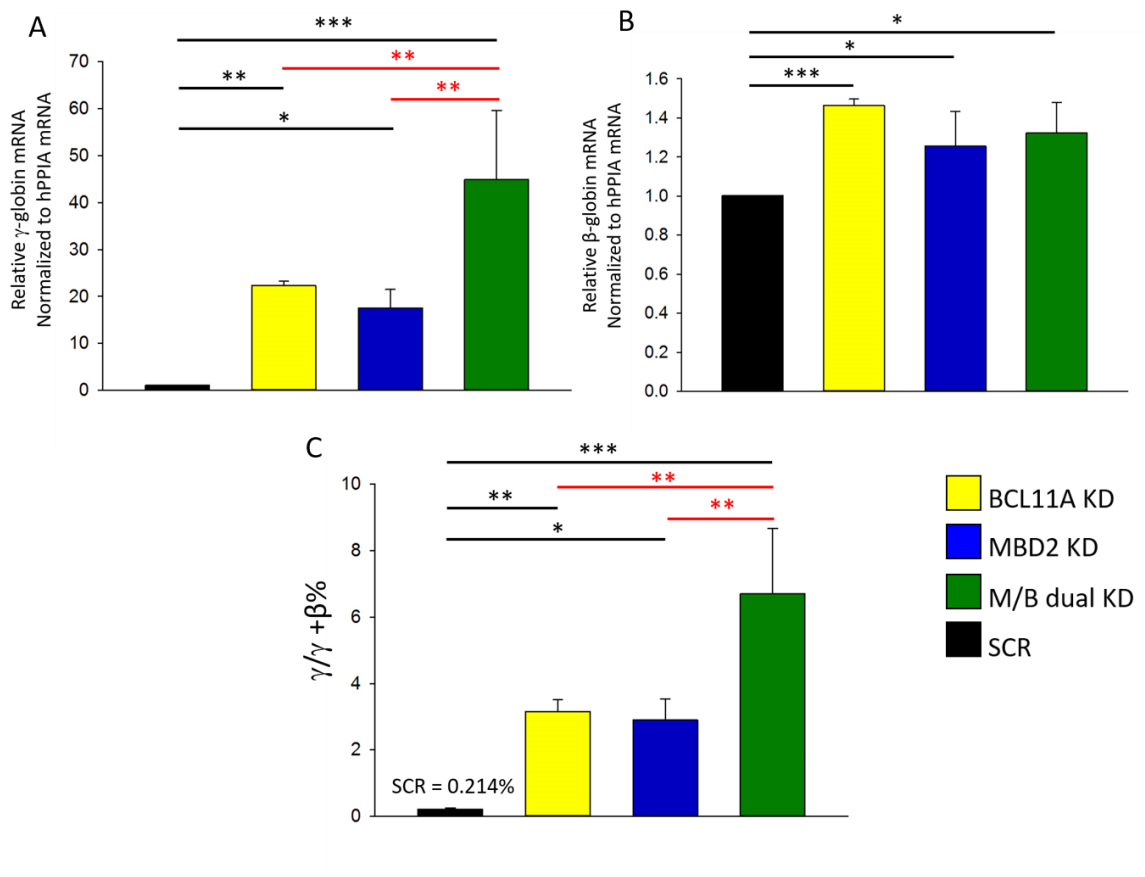

Figure 17: γ- and β-globin Expression in HUDEP-2 Cells with MBD2 and BCL11A Knockdown

qRT-PCR data. Panels A-B are compared to the SCR control set to 1. Panel C is compared to SCR which is a percentage. A: γ-globin gene expression for each of the single knockdowns and the dual knockdown. B: β-globin gene expression for each of the single knockdowns and the dual knockdown. C: Total globin ratios (γ/(γ+β)%) for each of the single knockdowns and the dual knockdown. Sample size of 4. Statistical test: one-way ANOVA with Tukey correction. Black lines are statistical comparisons to SCR. Red lines are statistical comparisons to the M/B dual results. Cells were infected for 5 days then differentiated for 3 days before collected for qRT-PCR. Significance $p < 0.05$. *<0.05, **<0.01, ***<0.001

Discussion

The data presented here is evidence that dual knockdown of MBD2 and BCL11A increases γ-globin mRNA significantly more than either of the single knockdowns. Comparing our results to the literature, studies measuring the effects of knockdown of BCL11A in CD34+ cells report 97% - 99% knockdown, where we saw 80% knockdown. These studies report a 3 to

6.5 fold increase in γ-globin mRNA, with the average being a 4.2-fold increase (153) (246) (247) (248). A study measuring the effects of BCL11A knockout in HUDEP-2 cells saw a 30-fold increase in γ-globin mRNA (267). Our knockdown of BCL11A showed a 22-fold increase in γ-globin mRNA. Our increase in γ-globin mRNA is larger than what the studies in CD34+ cells saw, but similar to the study done in HUDEP-2 cells. The differences between results in HUDEP-2 compared to CD34+ cells suggest that in a clinical setting γ-globin mRNA increases may be smaller than in HUDEP-2 cells. One study reported the change in β-globin mRNA in CD34+ cells with BCL11A knocked down 99% to have a 1.5-fold increase (248). Our findings for how BCL11A influences β-globin mRNA agree with what has been shown in the literature; we saw a 1.4-fold increase in mRNA.

A study of how MBD2 knockdown affects γ-globin mRNA in CD34+ cells showed it results in a 5-fold increase as a result of 90% knockdown of MBD2 (171). Another study in HUDEP-2 cells saw 98% knockdown of MBD2 results in about a 10-fold increase in γ/γ+β ratio compared to control. This study also saw knockout of MBD2 in HUDEP-2 cells resulted in at least a 300-fold increase in γ-globin mRNA, no significant change in β-globin mRNA, and a γ/γ+β of about 50% (168). We saw 90% knockdown of MBD2 with a resulting 17.5-fold increase γ-globin mRNA and a 14-fold increase in the γ/γ+β ratio. Our findings agree with the experiments done in HUDEP-2 cells; the increase in the γ/γ+β ratio is similar. Again, the more clinically relevant CD34+ cells appear to show less increase in γ-globin, further emphasizing the probable need for double knockdown. Additionally, the one study that looked at β-globin mRNA did not report a change, but we saw a significant 1.3-fold increase.

To our knowledge, this is the first study to explore single knockdowns of BCL11A and MBD2 at the same time in the same cell system. This allows us to compare the results for each gene to see if knockdown of one works better than knockdown of the other. We show our shRNAs knockdown these genes by about the same amount, 80 – 90%, and both significantly increase γ-globin mRNA compared to the SCR control. However, there is no significant difference in the amount of the increase seen in γ-globin mRNA, 22-fold for BCL11A vs. 17.5-fold for MBD2. There are no significant differences in the γ/γ+β ratio with the two different knockdowns either. Knockdown of BCL11A resulted in a 15-fold increase and knockdown of MBD2 resulted in a 14-fold increase, 3.2% vs. 2.9%. These findings show that knockdown of BCL11A or MBD2 significantly increases the amount of γ-globin mRNA by about the same amount and that there is not one gene that is better to target than the other with regard to increasing γ-globin expression. Our findings could mean BCL11A and MBD2 are not working in the same pathway, because there is a significant difference in the increase in γ-globin mRNA in the single knockdowns compared to the dual knockdown. In contrast, our results could be due to the dual knockdown depleting the interactions of BCL11A and MBD2 with the NuRD complex to a greater extent than single knockdown of either gene. The greater depletion of the NuRD complex would result in less silencing of the γ-globin gene and a larger increase in γ-globin mRNA even if BCL11A and MBD2 are in the same pathway.

Our findings present strong evidence that it would be promising to investigate combinations of the other DNA binding proteins discussed in this thesis – combinations of KLF1 and MBD2, and KLF1 and BCL11A. Investigating each of these combinations will hopefully determine which of the three combinations is the most effective at increasing γ-globin gene expression. This knowledge could then be used for subsequent experiments to develop a

therapeutic treatment for individuals with β-hemoglobinopathies. Future directions this thesis proposes are to perform studies where the different combinations of knockdowns of these DNA binding proteins are tested. These three genes are of interest because they are the most commonly studied; there is good evidence single knockdowns or knockouts of these genes gives a significant increase in γ-globin gene expression (153) (168) (171) (246) (247) (248). We also propose that each of the dual knockouts should be done in the same system. The studies described in Chapters 5 and 6 were done in various different model systems. Some were performed in βYAC transgenic mice, some in CD34+ cells, and some in HUDEP-2 cells. We believe that in order to compare how well each of the dual knockouts works they all need to be done in the same system. We hypothesize there will be one dual knockout combination that works better than the others at increasing γ-globin gene expression. We also hypothesize the combination of BCL11A and KLF1 will be the least effective combination because of the role KLF1 has in regulating the BCL11A gene, that is, they are in the same pathway. We think this combination should still be tested, however. In conclusion, we present strong evidence that dual knockdown of BCL11A and MBD2 simultaneously works significantly better at increasing γ-globin mRNA than single knockdown of either gene. Western blots need to be done to confirm the knockdown of BCL11A and MBD2, the decrease in β-globin mRNA, and the increase in γ-globin mRNA at the protein level. To have a more quantitative measure for γ- and β-globin expression, HPLC could be performed. We use this evidence to support future studies on dual knockdowns of other DNA binding proteins to determine which combination would be the best to target as a therapeutic treatment for individuals with β-hemoglobinopathies.

Chapter 8: Conclusions

Erythropoiesis is the synthesis of RBCs, which are essential for transporting oxygen throughout the body. RBCs are created in two phases during development. The first of these is primitive erythropoiesis, which is responsible for the creation of embryonic hemoglobin (1). Then there is a switch from primitive to definitive erythropoiesis, which correlates with expression of the γ- and β-globin genes (4). The switches in globin gene expression are referred to as hemoglobin switching, and in most humans the γ- to β-globin switch is completed by approximately 6 months of life (24) (25). In some humans this switch never or incompletely occurs, resulting in a condition called HPFH. Individuals with this condition have elevated levels of HbF compared to normal (49). The occurrence of a β-hemoglobinopathy and a HPFH variant in an individual can ameliorate their symptoms (50). This led researchers to seek ways to manipulate genes in order to increase HbF to treat patients with β-hemoglobinopathies. The two β-hemoglobinopathies most often focused on for this treatment are SCD and β-thalassemia, as they affect millions of people worldwide (68) (85). Both disorders have treatments available, but patients still have morbidity and early death compared to healthy individuals (89). The only cure available is BMT but this treatment can cause GVHD which can lead to death (96). The lack of a safe treatment method available to all patients highlights the need to develop a treatment that increases HbF. The strategies being worked on to achieve this will be discussed in this chapter.

In this thesis, the focus on increasing HbF has been on three DNA binding proteins implicated in regulating γ-globin expression: KLF1, BCL11A, and MBD2. KLF1 is an erythroid specific transcription factor that is well known for its role in activating the β-globin gene (121)

(123) (124). In fact, it is crucial for transcription of the β-globin gene, because knockouts of KLF1 in mouse models have shown it is embryonic lethal due to the subsequent development of β-thalassemia (123) (135). BCL11A is a transcription factor that regulates genes in many different processes, one of those being repression of the γ-globin gene (64) (144). BCL11A represses the γ-globin gene because when it is knocked out in mouse models this results in an increase in HbF, without embryonic lethality (147) (146). Subsequent studies led to the incidental finding that KLF1 regulates BCL11A; when KLF1 is knocked down the level of BCL11A is also reduced (153) (154). This indicates that KLF1 indirectly negatively regulates γ-globin gene expression by promoting the expression of BCL11A to repress the γ-globin gene. MBD2 is a DNA binding protein that represses γ-globin gene transcription through its recruitment of the NuRD complex (161). NuRD is a complex of proteins that provide histone deacetylase and chromatin remodeling activities that allow MBD2 to repress transcription (163) (164). MBD2 was also verified to regulate the γ-globin gene through knockouts in mouse and cell models (165) (166) (167).

Although investigation of these DNA binding proteins progresses, there are still gaps in knowledge. The gaps discussed in this thesis for KLF1 include the function of KLF1 at the γ-globin gene in human fetal erythroid cells and how the type and location of KLF1 mutations affect the severity of the condition the mutation causes. There is evidence that KLF1 binds the γ-globin gene promoter during fetal development, but the majority of the studies on this topic have been performed in βYAC transgenic mice (124) (128) (228) (229) (230) (231). We propose there could be differences in the timing of KLF1 function depending on the model system used. In an attempt to resolve this gap, I propose experiments should be done in a human cell model that

expresses HbF or HbA, like the HUDEP-1 and HUDEP-2 cell lines (194). When analyzing numerous mutations of KLF1 that cause various disorders we found the majority of the mutations occur in the zinc fingers. It is not surprising that variants in the zinc fingers cause an increase in HbF because they facilitate binding of KLF1 to the DNA. I propose that these variants could be causing KLF1 to change the specificity of its target sites. I also propose that more sequencing needs to be done to compare KLF1 mutations that do not cause a phenotype to mutations that do to see if there are critical regions in the KLF1 gene. Understanding the critical regions of the gene that cause disease or HPFH would provide researchers with the knowledge of where in the gene they should target for treatment of the β-hemoglobinopathies.

The gaps presented in this thesis for BCL11A are the methods by which it binds to the γ-globin gene in order to repress it and what other proteins or complexes it is interacting with. There was uncertainty about where BCL11A binds in the β-globin locus to repress the γ-globin gene. Previous theories suggested it acts at a distance to the gene promoters (246) (247), but convincing more recent evidence shows it binds to a sequence at the -115 distal motif of the γ-globin promoter (155) (254). Researchers are still unsure about the exact mechanism by which BCL11A functions to repress the γ-globin gene. The current theory is that BCL11A is binding at the -115 promoter site as well as at the distal sites in the β-globin locus in order to repress the γ-globin gene. There are many reported interactions of BCL11A with other proteins, for example its interactions with- RbAp48/46 and CHD4 of the NuRD complex (157) (246). The understanding of if/how BCL11A is interacting with the NuRD complex is still unclear. The answer to this question could help uncover the mechanism by which MBD2 binds to the β-globin locus.

This is the predominant gap in knowledge for MBD2. It is known MBD2 represses the γ-globin gene, but there are no methylated CpG islands within the γ-globin gene for which MBD2 can bind to (165) – its normal mechanism of action (162) (167). MBD2 is binding to CpG-poor DNA somehow, but it is not known how. We propose the differences between MBD2 and MBD3 should be studied to determine if regions not shared by the two account for the ability of MBD2 to bind to CpG-poor DNA.

We focused on the gap that knockdown of combinations of the KLF1, BCL11A, and MBD2 DNA binding proteins could be more effective at increasing γ-globin than single knockdowns. We propose knocking down all three possible combinations of two of the three factors simultaneously. We presented previous studies that have done dual knockdowns to increase γ-globin expression, to show what is available in the literature (154) (248) (267). We want to test combinations of these three genes because they are the most studied. We performed single and dual knockdown of BCL11A and MBD2 and show the dual knockdown results in a significantly greater increase in γ-globin gene expression and γ/(γ+β)% than either of the single knockdowns. These findings indicate dual knockdowns have an additive effect on increasing HbF and would likely be a more effective treatment for β-hemoglobinopathies than single knockdowns. We use these results to support our hypothesis that the other combinations of genes should be tested to determine which combination would be the best target for a treatment.

Using shRNAs that target these genes is one viable option for a treatment and has been shown to be possible in clinical trials testing gene therapy of BCL11A (149) (151). While this is an option it is not the best solution for large scale distribution. Gene therapy would not be easily accessible to most people because it would be expensive and because the technology is not

available in third world countries where a large proportion of β-hemoglobinopathy cases occur. A better treatment that could be developed is to create small molecule inhibitors that could be taken orally. Potential targets of a small molecule inhibitor for KLF1 and BCL11A could be their zinc fingers. They are what allow these proteins to bind DNA; by preventing this function they would no longer be able to promote β-globin or repress γ-globin expression. Targeting the KLF1 zinc fingers is a strong candidate as demonstrated by the cases of increased HbF as a result of missense variants in the KLF1 zinc fingers, discussed in Chapter 4.1.c. Small molecule inhibitors could also be used to target the interaction of KLF1 with p300 and CBP. KLF1 needs these interactions to transcriptionally activate the β-globin gene, so blocking these interactions would repress the expression of β-globin. For MBD2, studies have shown its IDR or coiled-coil domain would be promising targets for a small molecule inhibitor (170) (171) (168). Gene therapy and small molecule inhibitors are both promising treatment options being considered for β-hemoglobinopathies. While clinical trials to develop gene therapy that targets BCL11A are underway, more data needs to be collected on how this treatment is working before it can be decided if it is a better treatment option than what is currently available. Research needs to be done to create small molecule inhibitors that target these regions as this is the more feasible treatment option in order to reach the largest number of people. As previously mentioned, small molecule inhibitors would be a better treatment option than gene therapy as the majority of individuals with β-hemoglobinopathies are located in underdeveloped nations where access to the technology for gene editing is scarce. It will take time to develop these treatment methods, but both will likely be better treatment options than what is available now.

References

1. Palis J. Primitive and definitive erythropoiesis in mammals. Front Physiol [Internet]. 2014 Jan 28 [cited 2020 Jun 12];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3904103/

2. Wong PM, Chung SW, Chui DH, Eaves CJ. Properties of the earliest clonogenic hemopoietic precursors to appear in the developing murine yolk sac. PNAS. 1986 Jun 1;83(11):3851–4.

3. Palis J, Robertson S, Kennedy M, Wall C, Keller G. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development. 1999 Nov 15;126(22):5073–84.

4. Qiu C, Olivier EN, Velho M, Bouhassira EE. Globin switches in yolk sac–like primitive and fetal-like definitive red blood cells produced from human embryonic stem cells. Blood. 2008 Feb 15;111(4):2400–8.

5. McGrath KE, Koniski AD, Malik J, Palis J. Circulation is established in a stepwise pattern in the mammalian embryo. Blood. 2003 Mar 1;101(5):1669–75.

6. Zambidis ET, Peault B, Park TS, Bunz F, Civin CI. Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematoendothelial, primitive, and definitive stages resembling human yolk sac development. Blood. 2005 Aug 1;106(3):860–70.

7. Kingsley PD, Malik J, Fantauzzo KA, Palis J. Yolk sac–derived primitive erythroblasts enucleate during mammalian embryogenesis. Blood. 2004 Jul 1;104(1):19–25.

8. Steiner R, Vogel H. On the kinetics of erythroid cell differentiation in fetal mice. I. Microspectrophotometric determination of the hemoglobin content in erythroid cells during gestation. Journal of Cellular Physiology. 1973;81(3):323–37.

9. Morioka K, Minamikawa-Tachino R. Temporal Characteristics of the Differentiation of Embryonic Erythroid Cells in Fetal Peripheral Blood of the Syrian Hamster. Development, Growth & Differentiation. 1993;35(5):569–82.

10. Dzierzak E, Speck NA. Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nature Immunology. 2008 Feb;9(2):129–36.

11. Migliaccio G, Migliaccio AR, Petti S, Mavilio F, Russo G, Lazzaro D, et al. Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac----liver transition. J Clin Invest. 1986 Jul;78(1):51–60.

12. Chui DH, Clarke BJ. Abnormal erythroid progenitor cells in human preleukemia. Blood. 1982 Aug 1;60(2):362–7.

13. Nijhof W, Wierenga PK. Isolation and characterization of the erythroid progenitor cell: CFU-E. J Cell Biol. 1983 Feb 1;96(2):386–92.

14. Johnstone RM. Maturation of reticulocytes: formation of exosomes as a mechanism for shedding membrane proteins. Biochem Cell Biol. 1992 Mar 1;70(3–4):179–90.

15. Waugh RE, Mantalaris A, Bauserman RG, Hwang WC, Wu JHD. Membrane instability in late-stage erythropoiesis. Blood. 2001 Mar 15;97(6):1869–75.

16. Bennett GD, Kay MM. Homeostatic removal of senescent murine erythrocytes by splenic macrophages. Exp Hematol. 1981 Mar;9(3):297–307.

17. Weatherall DJ. Phenotype-genotype relationships in monogenic disease: Lessons from the thalassaemias. Nature Reviews: Genetics. 2001;2(4):245–55.

18. Hua-bing Z, De-Pei L, Chih-Chuan L. The Control of Expression of the α-Globin Gene Cluster. Int J Hematol. 2002 Dec 1;76(5):420–6.

19. Bank A, Mears JG, Ramirez F. Disorders of Human Hemoglobin. Science. 1980;207(4430):486–93.

20. Philipsen S. Molecular control of hemoglobin switching. 2013;9.

21. Manning LR, Russell JE, Padovan JC, Chait BT, Popowicz A, Manning RS, et al. Human embryonic, fetal, and adult hemoglobins have different subunit interface strengths. Correlation with lifespan in the red cell. Protein Sci. 2007 Aug;16(8):1641–58.

22. Sankaran VG, Orkin SH. The Switch from Fetal to Adult Hemoglobin. Cold Spring Harb Perspect Med [Internet]. 2013 Jan [cited 2020 Jun 18];3(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530042/

23. Stamatoyannopoulos George, Grosveld F. Hemoglobin switching. In: The Molecular basis of blood diseases [Internet]. 3rd ed. Philadelphia: W.B. Saunders Publishing Co.; 2001. 1028 p. Available from: https://hdl.handle.net/2027/mdp.39015050252892

24. Sankaran VG, Xu J, Orkin SH. Advances in the understanding of haemoglobin switching. British Journal of Haematology. 2010;149(2):181–94.

25. Bauer DE, Orkin SH. Hemoglobin switching's surprise: the versatile transcription factor BCL11A is a master repressor of fetal hemoglobin. Current Opinion in Genetics & Development. 2015 Aug;33:62–70.

26. Forrester WC, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, Groudine M. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucleic Acids Res. 1987 Dec 23;15(24):10159–77.

27. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell. 1987 Dec 24;51(6):975–85.

28. Enver T, Raich N, Ebens AJ, Papayannopoulou T, Costantini F, Stamatoyannopoulos G. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. Nature; London. 1990 Mar 22;344(6264):309–13.

29. Dillon N, Grosveld F. Human (gamma)-Globin Genes Silenced Independently of Other Genes in the (beta)-Locus. Nature. 1991 Mar 21;350(6315):252–252.

30. Hanscombe O, Whyatt D, Fraser P, Yannoutsos N, Greaves D, Dillon N, et al. Importance of globin gene order for correct developmental expression. Genes Dev. 1991 Aug 1;5(8):1387–94.

31. Epner E, Kim CG, Groudine M. What does the locus control region control? Current Biology. 1992 May 1;2(5):262–4.

32. Tuan D, Solomon W, Li Q, London IM. The ``β -Like-Globin'' Gene Domain in Human Erythroid Cells. Proceedings of the National Academy of Sciences of the United States of America. 1985;82(19):6384–8.

33. Philipsen S, Talbot D, Fraser P, Grosveld F. The beta-globin dominant control region: hypersensitive site 2. EMBO J. 1990 Jul;9(7):2159–67.

34. Pruzina S, Hanscombe O, Whyatt D, Grosveld F, Philipsen S. Hypersensitive site 4 of the human beta globin locus control region. Nucleic Acids Res. 1991 Apr 11;19(7):1413–9.

35. Wijgerde M, Grosveld F, Fraser P. Transcription complex stability and chromatin dynamics in vivo. Nature. 1995 Sep 21;377(6546):209–13.

36. Peterson KR, Fedosyuk H, Harju-Baker S. LCR 5′ hypersensitive site specificity for globin gene activation within the active chromatin hub. Nucleic Acids Res. 2012 Dec;40(22):11256–69.

37. Hnisz D, Schuijers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, et al. Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. Molecular Cell. 2015 Apr 16;58(2):362–70.

38. Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. Molecular Cell. 2017 Apr 20;66(2):285-299.e5.

39. Tolhuis B, Palstra R-J, Splinter E, Grosveld F, de Laat W. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. Molecular Cell. 2002 Dec 1;10(6):1453–65.

40. Epner E, Reik A, Cimbora D, Telling A, Bender MA, Fiering S, et al. The β-Globin LCR Is Not Necessary for an Open Chromatin Structure or Developmentally Regulated Transcription of the Native Mouse β-Globin Locus. Molecular Cell. 1998 Oct 1;2(4):447–55.

41. Bender MA, Bulger M, Close J, Groudine M. β-globin Gene Switching and DNase I Sensitivity of the Endogenous β-globin Locus in Mice Do Not Require the Locus Control Region. Molecular Cell. 2000 Feb 1;5(2):387–93.

42. Bender MA, Byron R, Ragoczy T, Telling A, Bulger M, Groudine M. Flanking HS-62.5 and 3′ HS1, and regions upstream of the LCR, are not required for β-globin transcription. Blood. 2006 Aug 15;108(4):1395–401.

43. Krivega I, Dale RK, Dean A. Role of LDB1 in the transition from chromatin looping to transcription activation. Genes Dev. 2014 Jun 15;28(12):1278–90.

44. Liu G, Dean A. Enhancer long-range contacts: The multi-adaptor protein LDB1 is the tie that binds. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2019 Jun 1;1862(6):625–33.

45. Forrester WC, Novak U, Gelinas R, Groudine M. Molecular analysis of the human beta-globin locus activation region. PNAS. 1989 Jul 1;86(14):5439–43.

46. Ryan TM, Behringer RR, Martin NC, Townes TM, Palmiter RD, Brinster RL. A single erythroid-specific DNase I super-hypersensitive site activates high levels of human beta-globin gene expression in transgenic mice. Genes Dev. 1989 Mar 1;3(3):314–23.

47. Martin DI, Fiering S, Groudine M. Regulation of β-globin gene expression: straightening out the locus. Current Opinion in Genetics & Development. 1996 Aug 1;6(4):488–95.

48. Geyer PK. The role of insulator elements in defining domains of gene expression. Current Opinion in Genetics & Development. 1997 Apr 1;7(2):242–8.

49. Gumucio DL, Rood KL, Gray TA, Riordan MF, Sartor CI, Collins FS. Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. Molecular and Cellular Biology. 1988 Dec 1;8(12):5310–22.

50. Forget BG. Molecular basis of hereditary persistence of fetal hemoglobin. Ann N Y Acad Sci. 1998 Jun 30;850:38–44.

51. Calzolari R, McMorrow T, Yannoutsos N, Langeveld A, Grosveld F. Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and deltabeta-thalassemia affects beta- but not gamma-globin gene expression. EMBO J. 1999 Feb 15;18(4):949–58.

52. Bhardwaj U, McCabe ERB. Multiplex-PCR Assay for the Deletions Causing Hereditary Persistence of Fetal Hemoglobin. Molecular Diagnosis 2005. 2005;9(3):151–6.

53. Gazouli M, Katsantoni E, Kosteas T, Anagnou NP. Persistent Fetal γ-Globin Expression in Adult Transgenic Mice following Deletion of Two Silencer Elements Located 3′ to the Human Aγ-Globin Gene. Mol Med. 2009 Nov;15(11–12):415–24.

54. Feingold EA, Forget BG. The breakpoint of a large deletion causing hereditary persistence of fetal hemoglobin occurs within an erythroid DNA domain remote from the beta-globin gene cluster. Blood. 1989 Nov 1;74(6):2178–86.

55. Anagnou NP, Perez-Stable C, Gelinas G, Costantini C, Liapaki L, Constantopoulou C, et al. Sequences Located 3′ to the Breakpoint of the Hereditary Persistence of Fetal Hemoglobin-3 Deletion Exhibit Enhancer Activity and Can Modify the Developmental Expression of the Human Fetal Aγ-Globin Gene in Transgenic Mice. J Biol Chem. 1995 Apr 28;270(17):10256–63.

56. Gilman JG, Huisman TH. DNA sequence variation associated with elevated fetal G gamma globin production. Blood. 1985 Oct 1;66(4):783–7.

57. Collins FS, Stoeckert CJ, Serjeant GR, Forget BG, Weissman SM. G gamma beta+ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the G gamma gene. PNAS. 1984 Aug 1;81(15):4894–8.

58. Borg J, Papadopoulos P, Georgitsi M, Gutiérrez L, Grech G, Fanis P, et al. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nature Genetics. 2010 Sep;42(9):801–5.

59. Craig JE, Rochette J, Fisher CA, Weatherall DJ, Marc S, Lathrop GM, et al. Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. Nature Genetics. 1996 Jan;12(1):58–64.

60. Garner C, Mitchell J, Hatzis T, Reittie J, Farrall M, Thein SL. Haplotype Mapping of a Major Quantitative-Trait Locus for Fetal Hemoglobin Production, on Chromosome 6q23. The American Journal of Human Genetics. 1998 Jun 1;62(6):1468–74.

61. Close J, Game L, Clark B, Bergounioux J, Gerovassili A, Thein SL. Genome annotation of a 1.5 Mb region of human chromosome 6q23 encompassing a quantitative trait locus for fetal hemoglobin expression in adults. BMC Genomics. 2004 May 31;5:33.

62. Thein SL, Menzel S, Peng X, Best S, Jiang J, Close J, et al. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc Natl Acad Sci U S A. 2007 Jul 3;104(27):11346–51.

63. Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, Heath S, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nature Genetics. 2007 Oct;39(10):1197–9.

64. Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of β-thalassemia. PNAS. 2008 Feb 5;105(5):1620–5.

65. Fanis P, Kousiappa I, Phylactides M, Kyrri A, Hadjigavriel M, Christou S, et al. A novel mutation in the erythroid transcription factor KLF1 is likely responsible for ameliorating β-thalassemia major. Human Mutation. 2019;40(10):1768–80.

66. Satta S, Perseu L, Moi P, Asunis I, Cabriolu A, Maccioni L, et al. Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. Haematologica. 2011 May 1;96(5):767–70.

67. Radmilovic M, Zukic B, Petrovic MS, Bartsakoulia M, Stankovic B, Kotur N, et al. Functional analysis of a novel KLF1 gene promoter variation associated with hereditary persistence of fetal hemoglobin. Ann Hematol. 2013 Jan 1;92(1):53–8.

68. CDC. Data & Statistics on Sickle Cell Disease | CDC [Internet]. Centers for Disease Control and Prevention. 2016 [cited 2020 Mar 11]. Available from: https://www.cdc.gov/ncbddd/sicklecell/data.html

69. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. The Lancet. 2010 Dec 11;376(9757):2018–31.

70. Sickle Cell Anemia: Practice Essentials, Background, Genetics. 2020 Jan 31 [cited 2020 Mar 11]; Available from: https://emedicine.medscape.com/article/205926-overview#a4

71. Platt OS, Orkin SH, Dover G, Beardsley GP, Miller B, Nathan DG. Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. J Clin Invest. 1984 Aug;74(2):652–6.

72. Charache S, Terrin ML, Moore RD, Dover GJ, Barton FB, Eckert SV, et al. Effect of Hydroxyurea on the Frequency of Painful Crises in Sickle Cell Anemia. New England Journal of Medicine. 1995 May 18;332(20):1317–22.

73. Charache S, Dover GJ, Moore RD, Eckert S, Ballas SK, Koshy M, et al. Hydroxyurea: effects on hemoglobin F production in patients with sickle cell anemia [see comments]. Blood. 1992 May 15;79(10):2555–65.

74. Quinn CT. l-Glutamine for sickle cell anemia: more questions than answers. Blood. 2018 Aug 16;132(7):689–93.

75. Agrawal RK, Patel RK, shah V, Nainiwal L, Trivedi B. Hydroxyurea in Sickle Cell Disease: Drug Review. Indian J Hematol Blood Transfus. 2014 Jun;30(2):91–6.

76. Zhang D, Xu C, Manwani D, Frenette PS. Neutrophils, platelets, and inflammatory pathways at the nexus of sickle cell disease pathophysiology. Blood. 2016 Feb 18;127(7):801–9.

77. Ferster A, Vermylen C, Cornu G, Buyse M, Corazza F, Devalck C, et al. Hydroxyurea for treatment of severe sickle cell anemia: a pediatric clinical trial. Blood. 1996 Sep 15;88(6):1960–4.

78. Kinney TR, Helms RW, O'Branski EE, Ohene-Frempong K, Wang W, Daeschner C, et al. Safety of Hydroxyurea in Children With Sickle Cell Anemia: Results of the HUG-KIDS Study, a Phase I/II Trial. Blood. 1999 Sep 1;94(5):1550–4.

79. Thompson BW, Miller ST, Rogers ZR, Rees RC, Ware RE, Waclawiw MA, et al. The pediatric hydroxyurea phase III clinical trial (BABY HUG): Challenges of study design. Pediatric Blood & Cancer. 2010;54(2):250–5.

80. Charache S, Barton FB, Moore RD, Terrin ML, Steinberg MH, Dover GJ, et al. Hydroxyurea and Sickle Cell Anemia Clinical Utility of a Myelosuppressive "Switching" Agent. Medicine. 1996 Nov;75(6):300–26.

81. Steinberg MH, Barton F, Castro O, Pegelow CH, Ballas SK, Kutlar A, et al. Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment. JAMA. 2003 Apr 2;289(13):1645–51.

82. Voskaridou E, Christoulas D, Bilalis A, Plata E, Varvagiannis K, Stamatopoulos G, et al. The effect of prolonged administration of hydroxyurea on morbidity and mortality in adult patients with sickle cell syndromes: results of a 17-year, single-center trial (LaSHS). Blood. 2010 Mar 25;115(12):2354–63.

83. Steinberg M, McCarthy W, Castro O, Ballas S, Armstrong F, Smith W, et al. The Risks and Benefits of Long-term Use of Hydroxyurea in Sickle Cell Anemia: A 17.5 Year Follow-Up. Am J Hematol. 2010 Jun;85(6):403–8.

84. Hankins JS, Aygun B, Nottage K, Thornburg C, Smeltzer MP, Ware RE, et al. From Infancy to Adolescence: Fifteen Years of Continuous Treatment With Hydroxyurea in Sickle Cell Anemia. Medicine (Baltimore) [Internet]. 2014 Dec 2 [cited 2020 Apr 29];93(28). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4603125/

85. Galanello R, Origa R. Beta-thalassemia. Orphanet J Rare Dis. 2010 May 21;5:11.

86. Cao A, Galanello R. Beta-thalassemia. Genetics in Medicine. 2010 Feb;12(2):61–76.

87. Beta thalassemia - NIH [Internet]. Genetics Home Reference. [cited 2020 Mar 18]. Available from: https://ghr.nlm.nih.gov/condition/beta-thalassemia

88. Weatherall DJ, Clegg JB. The Thalassaemia Syndromes [Internet]. 4th ed. Oxford; Malden (MA): Blackwell Science; 2001 [cited 2020 Apr 30]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470696705

89. CDC. Thalassemia: Complications and Treatment | CDC [Internet]. Centers for Disease Control and Prevention. 2016 [cited 2020 Mar 18]. Available from: https://www.cdc.gov/ncbddd/thalassemia/treatment.html

90. Alpha thalassemia [Internet]. Genetics Home Reference. [cited 2020 Apr 30]. Available from: https://ghr.nlm.nih.gov/condition/alpha-thalassemia

91. Galanello R, Cao A. Alpha-thalassemia. Genetics in Medicine. 2011 Feb;13(2):83–8.

92. Farashi S, Harteveld CL. Molecular basis of α-thalassemia. Blood Cells, Molecules, and Diseases. 2018 May 1;70:43–53.

93. Higgs DR, Weatherall DJ. The Alpha Thalassaemias. Cell Mol Life Sci. 2009 Apr 1;66(7):1154–62.

94. Ashorobi D, Bhatt R. Bone Marrow Transplantation In Sickle Cell Disease. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 [cited 2020 May 12]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK538515/

95. Alonso L, González-Vicent M, Belendez C, Badell I, Sastre A, Rodríguez-Villa A, et al. Hematopoietic stem cell transplantation in pediatric patients with β-thalassemia and sickle cell disease: An experience of the Spanish Working Group for Bone Marrow Transplantation in Children (GETMON). Medicina Clínica (English Edition). 2019 Feb 15;152(4):135–40.

96. Leukemia and Lymphoma Society. Graft-Versus-Host Disease [Internet]. 2015 [cited 2020 Mar 18]. Available from: https://www.lls.org/treatment/types-of-treatment/stem-cell-transplantation/graft-versus-host-disease

97. King A, Shenoy S. Evidence-based focused review of the status of hematopoietic stem cell transplantation as treatment of sickle cell disease and thalassemia. Blood. 2014 May 15;123(20):3089–94.

98. Hoppe CC, Walters MC. Bone marrow transplantation in sickle cell anemia. Current Opinion in Oncology. 2001 Mar;13(2):85–90.

99. Akif Yesilipek M. Stem Cell Transplantation in Hemoglobinopathies. Hemoglobin. 2007 May;31(2):251–6.

100. Myers RM, Tilly K, Maniatis T. Fine Structure Genetic Analysis of a β-Globin Promoter. Science. 1986;232(4750):613–8.

101. Cao A, Moi P. Regulation of the Globin Genes. [Review]. Pediatric Research. 2002 Apr;51(4):415–21.

102. Noordermeer D, Laat W de. Joining the loops: β-Globin gene regulation. IUBMB Life. 2008;60(12):824–33.

103. Evans T, Reitman M, Felsenfeld G. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. PNAS. 1988 Aug 1;85(16):5976–80.

104. Wall L, deBoer E, Grosveld F. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev. 1988 Sep 1;2(9):1089–100.

105. Zon LI, Tsai SF, Burgess S, Matsudaira P, Bruns GA, Orkin SH. The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. PNAS. 1990 Jan 1;87(2):668–72.

106. Tsai SF, Wong G, Orkin SH. Cloning of eDNA for the major DNA·binding protein of the erythroid lineage through expression in mammalian cells. 1989;339:6.

107. Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, et al. FOG, a Multitype Zinc Finger Protein, Acts as a Cofactor for Transcription Factor GATA-1 in Erythroid and Megakaryocytic Differentiation. Cell. 1997 Jul 11;90(1):109–19.

108. Orkin SH. GATA-binding transcription factors in hematopoietic cells. Blood. 1992 Aug 1;80(3):575–81.

109. Hong W, Nakazawa M, Chen Y-Y, Kori R, Vakoc CR, Rakowski C, et al. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. EMBO Journal. 2005;24(13):2367–78.

110. Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature. 1991;349(6306):257–60.

111. Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. Arrested Development of Embryonic Red Cell Precursors in Mouse Embryos Lacking Transcription Factor GATA-1. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(22):12355–8.

112. Tsang AP, Fujiwara Y, Hom DB, Orkin SH. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. Genes Dev. 1998 Apr 15;12(8):1176–88.

113. Ciovacco WA, Raskind WH, Kacena MA. Human phenotypes associated with GATA-1 mutations. Gene. 2008 Dec 31;427(1):1–6.

114. Nichols KE, Crispino JD, Poncz M, White JG, Orkin SH, Maris JM, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. Nature Genetics. 2000;24(3):266–70.

115. Mehaffey MG, Newton AL, Gandhi MJ, Crossley M, Drachman JG. X-linked thrombocytopenia caused by a novel mutation ofGATA-1. Blood. 2001 Nov 1;98(9):2681–8.

116. Freson K, Devriendt K, Matthijs G, Van Hoof A, De Vos R, Thys C, et al. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1mutation. Blood. 2001 Jul 1;98(1):85–92.

117. Freson K, Matthijs G, Thys C, Mariën P, Hoylaerts MF, Vermylen J, et al. Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation. Hum Mol Genet. 2002 Jan 15;11(2):147–52.

118. Vecchio GCD, Giordani L, Santis AD, Mattia DD. Dyserythropoietic Anemia and Thrombocytopenia due to a Novel Mutation in GATA-1. AHA. 2005;114(2):113–6.

119. Yu C, Niakan KK, Matsushita M, Stamatoyannopoulos G, Orkin SH, Raskind WH. X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. Blood. 2002 Sep 15;100(6):2040–5.

120. Balduini C, Pecci A, Loffredo G, Izzo P, Noris P, Grosso M, et al. Effects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis. Thromb Haemost. 2004;91(01):129–40.

121. Miller IJ, Bieker JJ. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. Molecular and Cellular Biology. 1993 May 1;13(5):2776–86.

122. Perkins A, Xu X, Higgs DR, Patrinos GP, Arnaud L, Bieker JJ, et al. Krüppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. Blood. 2016 Apr 14;127(15):1856–62.

123. Perkins AC, Sharpe AH, Orkin SH. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature. 1995 May 25;375(6529):318–22.

124. Hodge D, Coghill E, Keys J, Maguire T, Hartmann B, McDowall A, et al. A global role for EKLF in definitive and primitive erythropoiesis. Blood. 2006 Apr 15;107(8):3359–70.

125. Zhang W, Bieker JJ. Acetylation and modulation of erythroid Krüppel-like factor (EKLF) activity by interaction with histone acetyltransferases. Proc Natl Acad Sci U S A. 1998 Aug 18;95(17):9855–60.

126. Anderson KP, Kern CB, Crable SC, Lingrel JB. Isolation of a gene encoding a functional zinc finger protein homologous to erythroid Krüppel-like factor: identification of a new multigene family. Molecular and Cellular Biology. 1995 Nov 1;15(11):5957–65.

127. Basu P, Morris PE, Haar JL, Wani MA, Lingrel JB, Gaensler KML, et al. KLF2 is essential for primitive erythropoiesis and regulates the human and murine embryonic β-like globin genes in vivo. Blood. 2005 Oct 1;106(7):2566–71.

128. Alhashem YN, Vinjamur DS, Basu M, Klingmüller U, Gaensler KML, Lloyd JA. Transcription Factors KLF1 and KLF2 Positively Regulate Embryonic and Fetal β-Globin Genes through Direct Promoter Binding. J Biol Chem. 2011 Jul 15;286(28):24819–27.

129. Felsenfeld G, Groudine M. Controlling the double helix. Nature. 2003 Jan;421(6921):448–53.

130. Zhou D, Pawlik KM, Ren J, Sun C-W, Townes TM. Differential Binding of Erythroid Krupple-like Factor to Embryonic/Fetal Globin Gene Promoters during Development. J Biol Chem. 2006 Jun 9;281(23):16052–7.

131. Drissen R, Palstra R-J, Gillemans N, Splinter E, Grosveld F, Philipsen S, et al. The active spatial organization of the β-globin locus requires the transcription factor EKLF. Genes Dev. 2004 Oct 15;18(20):2485–90.

132. Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AMA, Strouboulis J, et al. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. Genes Dev. 2004 Jun 15;18(12):1495–509.

133. Palstra R-J, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The β-globin nuclear compartment in development and erythroid differentiation. Nature Genetics. 2003 Oct;35(2):190–4.

134. Yien YY, Bieker JJ. EKLF/KLF1, a Tissue-Restricted Integrator of Transcriptional Control, Chromatin Remodeling, and Lineage Determination. Molecular and Cellular Biology. 2013 Jan 1;33(1):4–13.

135. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature; London. 1995 May 25;375(6529):316–8.

136. Feng WC, Southwood CM, Bieker JJ. Analyses of beta-thalassemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor. J Biol Chem. 1994 Jan 14;269(2):1493–500.

137. Viprakasit V, Ekwattanakit S, Riolueang S, Chalaow N, Fisher C, Lower K, et al. Mutations in Krüppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. Blood. 2014 Mar 6;123(10):1586–95.

138. Singleton BK, Fairweather VS, Lau W, Parsons SF, Burton NM, Frayne J, et al. A Novel EKLF Mutation in a Patient with Dyserythropoietic Anemia: The First Association of EKLF with Disease in Man. Blood. 2009 Nov 20;114(22):162–162.

139. Arnaud L, Saison C, Helias V, Lucien N, Steschenko D, Giarratana M-C, et al. A Dominant Mutation in the Gene Encoding the Erythroid Transcription Factor KLF1 Causes a Congenital Dyserythropoietic Anemia. The American Journal of Human Genetics. 2010 Nov 12;87(5):721–7.

140. Jaffray JA, Mitchell WB, Gnanapragasam MN, Seshan SV, Guo X, Westhoff CM, et al. Erythroid transcription factor EKLF/KLF1 mutation causing congenital dyserythropoietic anemia type IV in a patient of Taiwanese origin: Review of all reported cases and development of a clinical diagnostic paradigm. Blood Cells, Molecules, and Diseases. 2013 Aug 1;51(2):71–5.

141. Li J, Shen H, Himmel KL, Dupuy AJ, Largaespada DA, Nakamura T, et al. Leukaemia disease genes: large-scale cloning and pathway predictions. Nature Genetics. 1999 Nov;23(3):348–53.

142. Nakamura T, Yamazaki Y, Saiki Y, Moriyama M, Largaespada DA, Jenkins NA, et al. Evi9 Encodes a Novel Zinc Finger Protein That Physically Interacts with BCL6, a Known Human B-Cell Proto-Oncogene Product. Molecular and Cellular Biology. 2000 May 1;20(9):3178–86.

143. Saiki Y, Yamazaki Y, Yoshida M, Katoh O, Nakamura T. Human EVI9, a Homologue of the Mouse Myeloid Leukemia Gene, Is Expressed in the Hematopoietic Progenitors and Down-Regulated during Myeloid Differentiation of HL60 Cells. Genomics. 2000 Dec 15;70(3):387–91.

144. Satterwhite E, Sonoki T, Willis TG, Harder L, Nowak R, Arriola EL, et al. The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. Blood. 2001 Dec 1;98(12):3413–20.

145. Liu P, Keller JR, Ortiz M, Tessarollo L, Rachel RA, Nakamura T, et al. Bcl11a is essential for normal lymphoid development. Nature immunology. 2003;4(6):525–32.

146. Xu J, Peng C, Sankaran VG, Shao Z, Esrick EB, Chong BG, et al. Correction of Sickle Cell Disease in Adult Mice by Interference with Fetal Hemoglobin Silencing. Science. 2011;334(6058):993–6.

147. Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, et al. Developmental and species-divergent globin switching are driven by BCL11A. Nature. 2009 Aug;460(7259):1093–7.

148. A Study to Assess the Safety, Tolerability, and Efficacy of ST-400 for Treatment of Transfusion-Dependent Beta-thalassemia (TDT) - Full Text View - ClinicalTrials.gov [Internet]. [cited 2020 Jun 25]. Available from: https://clinicaltrials.gov/ct2/show/NCT03432364

149. Smith AR, Schiller GJ, Vercellotti GM, Kwiatkowski JL, Krishnamurti L, Esrick EB, et al. Preliminary Results of a Phase 1/2 Clinical Study of Zinc Finger Nuclease-Mediated Editing of BCL11A in Autologous Hematopoietic Stem Cells for Transfusion-Dependent Beta Thalassemia. Blood. 2019 Nov 13;134(Supplement_1):3544–3544.

150. Sangamo Announces Preliminary Results From the First Three Patients in a Phase 1/2 Study Evaluating ST-400 Ex Vivo Gene-edited Cell Therapy in Beta Thalassemia | Sangamo Therapeutics, Inc. [Internet]. [cited 2020 Jun 25]. Available from: https://investor.sangamo.com/news-releases/news-release-details/sangamo-announces-preliminary-results-first-three-patients-phase

151. Williams D. Pilot and Feasibility Study of Hematopoietic Stem Cell Gene Transfer for Sickle Cell Disease [Internet]. clinicaltrials.gov; 2020 Feb [cited 2020 Jun 25]. Report No.: study/NCT03282656. Available from: https://clinicaltrials.gov/ct2/show/study/NCT03282656

152. Revolutionary Treatment for Sickle Cell Disease Shows Promise in Clinical Trial [Internet]. [cited 2020 Jun 25]. Available from: https://www.ajmc.com/focus-of-the-week/revolutionary-treatment-for-sickle-cell-disease-shows-promise-in-clinical-trial

153. Zhou D, Liu K, Sun C-W, Pawlik KM, Townes TM. KLF1 regulates BCL11A expression and γ- to β-globin gene switching. Nature Genetics. 2010 Sep;42(9):742–4.

154. Esteghamat F, Gillemans N, Bilic I, van den Akker E, Cantù I, van Gent T, et al. Erythropoiesis and globin switching in compound Klf1::Bcl11a mutant mice. Blood. 2013 Mar 28;121(13):2553–62.

155. Liu N, Hargreaves VV, Zhu Q, Kurland JV, Hong J, Kim W, et al. Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell. 2018 Apr 5;173(2):430-442.e17.

156. Sankaran VG, Xu J, Byron R, Greisman HA, Fisher C, Weatherall DJ, et al. A functional element necessary for fetal hemoglobin silencing. The New England journal of medicine. 2011;365(9):807–14.

157. Jawaid K, Wahlberg K, Thein SL, Best S. Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus. Blood Cells, Molecules, and Diseases. 2010 Aug 15;45(2):140–6.

158. Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell. 1989 Aug 11;58(3):499–507.

159. Hendrich B, Bird A. Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. MOL CELL BIOL. 1998;18:10.

160. Cross SH, Meehan RR, Nan X, Bird A. A component of the transcriptional represser MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nature Genetics. 1997 Jul;16(3):256–9.

161. Feng Q, Zhang Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev. 2001 Apr 1;15(7):827–32.

162. Zhang Y, Ng H-H, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999 Aug 1;13(15):1924–35.

163. Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W. NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. Molecular Cell. 1998 Dec 1;2(6):851–61.

164. Denslow SA, Wade PA. The human Mi-2/NuRD complex and gene regulation. Oncogene. 2007;26(37):5433–8.

165. Rupon JW, Wang SZ, Gaensler K, Lloyd J, Ginder GD. Methyl binding domain protein 2 mediates γ-globin gene silencing in adult human βYAC transgenic mice. Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6617–22.

166. Wood KH, Zhou Z. Emerging Molecular and Biological Functions of MBD2, a Reader of DNA Methylation. Front Genet [Internet]. 2016 May 26 [cited 2020 May 11];7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4880565/

167. Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 2001 Mar 15;15(6):710–23.

168. Yu X, Azzo A, Bilinovich SM, Li X, Dozmorov M, Kurita R, et al. Disruption of the MBD2-NuRD complex but not MBD3-NuRD induces high level HbF expression in human erythroid cells. Haematologica. 2019 Apr 19;haematol.2018.210963.

169. Sher F, Hossain M, Seruggia D, Schoonenberg VAC, Yao Q, Cifani P, et al. Rational targeting of a NuRD subcomplex guided by comprehensive in situ mutagenesis. Nature Genetics. 2019 Jul;51(7):1149–59.

170. Desai MA, Webb HD, Sinanan LM, Scarsdale JN, Walavalkar NM, Ginder GD, et al. An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. Nucleic Acids Res. 2015 Mar 31;43(6):3100–13.

171. Gnanapragasam MN, Scarsdale JN, Amaya ML, Webb HD, Desai MA, Walavalkar NM, et al. p66α–MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2–NuRD complex. PNAS. 2011 May 3;108(18):7487–92.

172. Friend C. CELL-FREE TRANSMISSION IN ADULT SWISS MICE OF A DISEASE HAVING THE CHARACTER OF A LEUKEMIA. J Exp Med. 1957 Mar 31;105(4):307–18.

173. Patuleia MC, Friend C. Tissue Culture Studies on Murine Virus-induced Leukemia Cells: Isolation of Single Cells in Agar-Liquid Medium. Cancer Res. 1967 Apr 1;27(4):726–30.

174. Friend C, Scher W, Holland JG, Sato T. Hemoglobin Synthesis in Murine Virus-Induced Leukemic Cells in Vitro: Stimulation of Erythroid Differentiation by Dimethyl Sulfoxide. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(2):378–82.

175. Singer D, Cooper M, Maniatis GM, Marks PA, Rifkind RA. Erythropoietic Differentiation in Colonies of Cells Transformed by Friend Virus. PNAS. 1974 Jul 1;71(7):2668–70.

176. Hofer E, Hofer-Warbinek R, Darnell JE. Globin RNA transcription: A possible termination site and demonstration of transcriptional control correlated with altered chromatin structure. Cell. 1982 Jul 1;29(3):887–93.

177. Deisseroth A, Hendrick D. Human α-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. Cell. 1978 Sep 1;15(1):55–63.

178. Pyati J, Kucherlapati RS, Skoultchi AI. Activation of human beta-globin genes from nonerythroid cells by fusion with murine erythroleukemia cells. PNAS. 1980 Jun 1;77(6):3435–9.

179. Willing MC, Nienhuis AW, Anderson WF. Selective activation of human β - but not γ -globin gene in human fibroblast × mouse erythroleukaemia cell hybrids. Nature. 1979 Feb;277(5697):534–8.

180. Lozzio CB, Lozzio BB. Cytotoxicity of a Factor Isolated From Human Spleen. J Natl Cancer Inst. 1973 Feb 1;50(2):535–8.

181. Andersson LC, Jokinen M, Gahmberg CG. Induction of erythroid differentiation in the human leukaemia cell line K562. Nature. 1979 Mar;278(5702):364–5.

182. Benz EJ, Murnane MJ, Tonkonow BL, Berman BW, Mazur EM, Cavallesco C, et al. Embryonic-fetal erythroid characteristics of a human leukemic cell line. PNAS. 1980 Jun 1;77(6):3509–13.

183. Horton MA, Cedar SH, Edwards P a. W. Expression of Red Cell Specific Determinants during Differentiation in the K562 Erythroleukaemia Cell Line. Scandinavian Journal of Haematology. 1981;27(4):231–40.

184. Villeval JL, Pelicci PG, Tabilio A, Titeux M, Henri A, Houesche F, et al. Erythroid properties of K562 cells: Effect of hemin, butyrate and TPA induction. Experimental Cell Research. 1983 Jul 1;146(2):428–35.

185. Rutherford T, Clegg JB, Higgs DR, Jones RW, Thompson J, Weatherall DJ. Embryonic erythroid differentiation in the human leukemic cell line K562. PNAS. 1981 Jan 1;78(1):348–52.

186. Fordis CM, Anagnou NP, Dean A, Nienhuis AW, Schechter AN. A beta-globin gene, inactive in the K562 leukemic cell, functions normally in a heterologous expression system. PNAS. 1984 Jul 1;81(14):4485–9.

187. Donovan-Peluso M, Young K, Dobkin C, Bank A. Erythroleukemia (K562) cells contain a functional beta-globin gene. Molecular and Cellular Biology. 1984 Nov 1;4(11):2553–5.

188. Ulrich MJ, Ley TJ. Function of normal and mutated gamma-globin gene promoters in electroporated K562 erythroleukemia cells. Blood. 1990 Feb 15;75(4):990–9.

189. Osti F, Corradini FG, Hanau S, Matteuzzi M, Gambari R. Human leukemia K562 cells: induction to erythroid differentiation by guanine, guanosine and guanine nucleotides. Haematologica. 1997 Jan 1;82(4):395–401.

190. Morceau F, Dupont C, Palissot V, Borde-Chiché P, Trentesaux C, Dicato M, et al. GTP-mediated differentiation of the human K562 cell line: transient overexpression of GATA-1 and stabilization of the γ-globin mRNA. Leukemia. 2000 Sep;14(9):1589–97.

191. Martin P, Papayannopoulou T. HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. Science. 1982 Jun 11;216(4551):1233–5.

192. Orkin SH, Antonarakis SE, Kazazian HH. Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. J Biol Chem. 1984 Jul 25;259(14):8679–81.

193. Zhang SB, He QY, Zhao H, Gui CY, Jiang C, Qian RL. Function of GATA transcription factors in hydroxyurea-induced HEL cells. Cell Research. 2001 Dec;11(4):301–10.

194. Kurita R, Suda N, Sudo K, Miharada K, Hiroyama T, Miyoshi H, et al. Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. PLOS ONE. 2013 Mar 22;8(3):e59890.

195. Traxler EA, Yao Y, Wang Y-D, Woodard KJ, Kurita R, Nakamura Y, et al. A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. Nature Medicine. 2016 Sep;22(9):987–90.

196. Wang L, Li L, Ma Y, Hu H, Li Q, Yang Y, et al. Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. Cell Research. 2020 Mar;30(3):276–8.

197. Park SH, Lee CM, Zhang Y, Chang A, Sheehan VA, Bao G. Engineered Human Umbilical Cord Derived Erythroid Progenitor Cells (HUDEP2) with Sickle or β-Thalassemia Mutation: An in-Vitro System for Testing Pharmacological Induction of Fetal Hemoglobin. Blood. 2018 Nov 29;132(Supplement 1):3478–3478.

198. Gupta D, Sturtevant S, Vieira B, Nakamura Y, Krishnamoorthy S, Demers M. Characterization of a Genetically Engineered HUDEP2 Cell Line Harboring a Sickle Cell Disease Mutation As a Potential Research Tool for Preclinical Sickle Cell Disease Drug Discovery. Blood. 2019 Nov 13;134(Supplement_1):3559–3559.

199. Civin CI, Strauss LC, Brovall C, Fackler MJ, Schwartz JF, Shaper JH. Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. The Journal of Immunology. 1984 Jul 1;133(1):157–65.

200. Berenson RJ, Bensinger WI, Hill RS, Andrews RG, Garcia-Lopez J, Kalamasz DF, et al. Engraftment after infusion of CD34+ marrow cells in patients with breast cancer or neuroblastoma. Blood. 1991 Apr 15;77(8):1717–22.

201. Andrews RG, Singer JW, Bernstein ID. Monoclonal antibody 12-8 recognizes a 115-kd molecule present on both unipotent and multipotent hematopoietic colony-forming cells and their precursors. Blood. 1986 Mar 1;67(3):842–5.

202. Servida F, Soligo D, Caneva L, Bertolini F, Harven E de, Campiglio S, et al. Functional and Morphological Characterization of Immunomagnetically Selected CD34+ Hematopoietic Progenitor Cells. STEM CELLS. 1996;14(4):430–8.

203. Berenson RJ, Andrews RG, Bensinger WI, Kalamasz D, Knitter G, Buckner CD, et al. Antigen CD34+ marrow cells engraft lethally irradiated baboons. J Clin Invest. 1988 Mar;81(3):951–5.

204. Sidney LE, Branch MJ, Dunphy SE, Dua HS, Hopkinson A. Concise Review: Evidence for CD34 as a Common Marker for Diverse Progenitors. Stem Cells. 2014 Jun;32(6):1380–9.

205. Urbinati F, Hargrove PW, Geiger S, Romero Z, Wherley J, Kaufman ML, et al. Potentially therapeutic levels of anti-sickling globin gene expression following lentivirus-mediated gene transfer in sickle cell disease bone marrow CD34+ cells. Experimental Hematology. 2015 May 1;43(5):346–51.

206. Hoban MD, Lumaquin D, Kuo CY, Romero Z, Long J, Ho M, et al. CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ cells. Molecular Therapy. 2016 Sep 1;24(9):1561–9.

207. Mishra B, Chou S, Lin MI, Paik EJ, Zhang Y, Liang R, et al. Crispr/Cas9- Mediated Genome Editing of Human CD34+ Cells Upregulate Fetal Hemoglobin to Clinically Relevant Levels in Single Cell-Derived Erythroid Colonies. Blood. 2016 Dec 2;128(22):3623–3623.

208. Wilber A, Hargrove PW, Kim Y-S, Riberdy JM, Sankaran VG, Papanikolaou E, et al. Therapeutic levels of fetal hemoglobin in erythroid progeny of β-thalassemic CD34+ cells after lentiviral vector-mediated gene transfer. Blood. 2011 Mar 10;117(10):2817–26.

209. Chang K-H, Smith SE, Sullivan T, Chen K, Zhou Q, West JA, et al. Long-Term Engraftment and Fetal Globin Induction upon BCL11A Gene Editing in Bone-Marrow-Derived CD34+ Hematopoietic Stem and Progenitor Cells. Molecular Therapy - Methods & Clinical Development. 2017 Mar 17;4:137–48.

210. Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH. Genetic transformation of mouse embryos by microinjection of purified DNA. PNAS. 1980 Dec 1;77(12):7380–4.

211. Parker MP, Peterson KR. Mouse Models of Erythropoiesis and Associated Diseases. In: Lloyd JA, editor. Erythropoiesis: Methods and Protocols [Internet]. New York, NY: Springer; 2018 [cited 2020 Apr 15]. p. 37–65. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-4939-7428-3_3

212. Burke DT, Carle GF, Olson MV. Cloning of Large Segments of Exogenous DNA into Yeast by means of Artificial Chromosome Vectors. Science. 1987;236(4803):806–12.

213. Peterson KR, Clegg CH, Huxley C, Josephson BM, Haugen HS, Furukawa T, et al. Transgenic mice containing a 248-kb yeast artificial chromosome carrying the human beta-globin locus display proper developmental control of human globin genes. PNAS. 1993 Aug 15;90(16):7593–7.

214. Gaensler KM, Kitamura M, Kan YW. Germ-line transmission and developmental regulation of a 150-kb yeast artificial chromosome containing the human beta-globin locus in transgenic mice. PNAS. 1993 Dec 1;90(23):11381–5.

215. Hu X, Bulger M, Roach JN, Eszterhas SK, Olivier E, Bouhassira EE, et al. Promoters of the murine embryonic β-like globin genes Ey and βh1 do not compete for interaction with the β-globin locus control region. PNAS. 2003 Feb 4;100(3):1111–5.

216. Fantoni A, Bank A, Marks PA. Globin Composition and Synthesis of Hemoglobins in Developing Fetal Mice Erythroid Cells. Science. 1967;157(3794):1327–9.

217. Hansen JN, Konkel DA, Leder P. The sequence of a mouse embryonic beta-globin gene. Evolution of the gene and its signal region. J Biol Chem. 1982 Jan 25;257(2):1048–52.

218. Behringer RR, Ryan TM, Palmiter RD, Brinster RL, Townes TM. Human gamma- to beta-globin gene switching in transgenic mice. Genes Dev. 1990 Mar 1;4(3):380–9.

219. Peterson KR, Li QL, Clegg CH, Furukawa T, Navas PA, Norton EJ, et al. Use of yeast artificial chromosomes (YACs) in studies of mammalian development: production of beta-globin locus YAC mice carrying human globin developmental mutants. PNAS. 1995 Jun 6;92(12):5655–9.

220. Peterson KR, Clegg CH, Navas PA, Norton EJ, Kimbrough TG, Stamatoyannopoulos G. Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. PNAS. 1996 Jun 25;93(13):6605–9.

221. Ryan TM, Sun C-W, Ren J, Townes TM. Human γ-globin gene promoter element regulates human β-globin gene developmental specificity. Nucleic Acids Res. 2000 Jul 15;28(14):2736–40.

222. Tallack MR, Whitington T, Shan Yuen W, Wainwright EN, Keys JR, Gardiner BB, et al. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Res. 2010 Aug;20(8):1052–63.

223. Tallack MR, Magor GW, Dartigues B, Sun L, Huang S, Fittock JM, et al. Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. Genome Res. 2012 Dec 1;22(12):2385–98.

224. Ludwig LS, Lareau CA, Bao EL, Nandakumar SK, Muus C, Ulirsch JC, et al. Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. Cell Reports. 2019 Jun 11;27(11):3228-3240.e7.

225. Han J, Zhang Z, Wang K. 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. Mol Cytogenet [Internet]. 2018 Mar 9 [cited 2020 Aug 1];11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5845197/

226. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing Chromosome Conformation. Science (Washington). 2002;295(5558):1306–11.

227. Fullwood MJ, Ruan Y. ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem. 2009 May 1;107(1):30–9.

228. Guy L-G, Mei Q, Perkins AC, Orkin SH, Wall L. Erythroid Krüppel-Like Factor Is Essential for β-Globin Gene Expression Even in Absence of Gene Competition, But Is Not Sufficient to Induce the Switch From γ-Globin to β-Globin Gene Expression. Blood. 1998 Apr 1;91(7):2259–63.

229. Pang CJ, Lemsaddek W, Alhashem YN, Bondzi C, Redmond LC, Ah-Son N, et al. Krüppel-Like Factor 1 (KLF1), KLF2, and Myc Control a Regulatory Network Essential for Embryonic Erythropoiesis. Molecular and Cellular Biology. 2012 Jul 1;32(13):2628–44.

230. Vinjamur DS, Alhashem YN, Mohamad SF, Amin P, Williams DC, Lloyd JA. Krüppel-Like Transcription Factor KLF1 Is Required for Optimal γ- and β-Globin Expression in Human Fetal

Erythroblasts. PLoS One [Internet]. 2016 Feb 3 [cited 2020 May 17];11(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4739742/

231. Siatecka M, Sahr KE, Andersen SG, Mezei M, Bieker JJ, Peters LL. Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Krüppel-like factor. PNAS. 2010 Aug 24;107(34):15151–6.

232. Liu D, Zhang X, Yu L, Cai R, Ma X, Zheng C, et al. KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of β-thalassemia. Blood. 2014 Jul 31;124(5):803–11.

233. Orkin SH, Kazazian HH, Antonarakis SE, Goff SC, Boehm CD, Sexton JP, et al. Linkage of β -thalassaemia mutations and β -globin gene polymorphisms with DNA polymorphisms in human β -globin gene cluster. Nature. 1982 Apr;296(5858):627–31.

234. Kulozik AE, Bellan-Koch A, Bail S, Kohne E, Kleihauer E. Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. Blood. 1991 May 1;77(9):2054–8.

235. Gonzalez-Redondo JM, Stoming TA, Kutlar A, Kutlar F, Lanclos KD, Howard EF, et al. A C----T substitution at nt--101 in a conserved DNA sequence of the promotor region of the beta-globin gene is associated with “silent” beta-thalassemia. Blood. 1989 May 1;73(6):1705–11.

236. Ristaldi MS, Murru S, Loudianos G, Casula L, Porcu S, Pigheddu D, et al. The C-T substitution in the distal CACCC box of the beta-globin gene promoter is a common cause of silent beta thalassemia in the Italian population. British Journal of Haematology. 1990;74:480–6.

237. Moi P, Faà V, Marini MG, Asunis I, Ibba G, Cao A, et al. A novel silent β-thalassemia mutation in the distal CACCC box affects the binding and responsiveness to EKLF. British Journal of Haematology. 2004;126(6):881–4.

238. Magor GW, Tallack MR, Gillinder KR, Bell CC, McCallum N, Williams B, et al. KLF1-null neonates display hydrops fetalis and a deranged erythroid transcriptome. Blood. 2015 Apr 9;125(15):2405–17.

239. Gallienne AE, Dréau HMP, Schuh A, Old JM, Henderson S. Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. Haematologica. 2012 Mar 1;97(3):340–3.

240. Wang T, He Y, Zhou J-Y, Xie X-M, Li J, Li R, et al. KLF1 Gene Mutations in Chinese Adults with Increased Fetal Hemoglobin. Hemoglobin. 2013 Oct;37(5):501–6.

241. Waye JS, Eng B. Krüppel-like factor 1: hematologic phenotypes associated with KLF1 gene mutations. International Journal of Laboratory Hematology. 2015;37(S1):78–84.

242. Pandya K, Townes TM. Basic Residues within the Kruppel Zinc Finger DNA Binding Domains Are the Critical Nuclear Localization Determinants of EKLF/KLF-1. J Biol Chem. 2002 May 3;277(18):16304–12.

243. Heruth DP, Hawkins T, Logsdon DP, Gibson MI, Sokolovsky IV, Nsumu NN, et al. Mutation in erythroid specific transcription factor KLF1 causes Hereditary Spherocytosis in the Nan hemolytic anemia mouse model. Genomics. 2010 Nov 1;96(5):303–7.

244. Nébor D, Graber JH, Ciciotte SL, Robledo RF, Papoin J, Hartman E, et al. Mutant KLF1 in Adult Anemic Nan Mice Leads to Profound Transcriptome Changes and Disordered Erythropoiesis. Scientific Reports. 2018 Aug 24;8(1):12793.

245. Ilsley MD, Huang S, Magor GW, Landsberg MJ, Gillinder KR, Perkins AC. Corrupted DNA-binding specificity and ectopic transcription underpin dominant neomorphic mutations in KLF/SP transcription factors. BMC Genomics. 2019 May 24;20(1):417.

246. Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Handel BV, et al. Human Fetal Hemoglobin Expression Is Regulated by the Developmental Stage-Specific Repressor BCL11A. Science. 2008 Dec 19;322(5909):1839–42.

247. Xu J, Sankaran VG, Ni M, Menne TF, Puram RV, Kim W, et al. Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. 2010 Apr 15;24(8):783–98.

248. Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu Y-J, et al. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proceedings of the National Academy of Sciences. 2013 Apr 16;110(16):6518–23.

249. Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. Ann Appl Stat. 2010;4(4):1797–823.

250. Amaya M, Desai M, Gnanapragasam MN, Wang SZ, Zhu SZ, Williams DC, et al. Mi2β-mediated silencing of the fetal γ-globin gene in adult erythroid cells. Blood. 2013 Apr 25;121(17):3493–501.

251. Grosschedl R, Giese K, Pagel J. HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends in Genetics. 1994 Mar 1;10(3):94–100.

252. Connor F, Cary PD, Read CM, Preston NS, Driscoll PC, Denny P, et al. DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. Nucleic Acids Res. 1994 Aug 25;22(16):3339–46.

253. Thomas JO, Travers AA. HMG1 and 2, and related 'architectural' DNA-binding proteins. Trends in Biochemical Sciences. 2001 Mar 1;26(3):167–74.

254. Martyn GE, Wienert B, Yang L, Shah M, Norton LJ, Burdach J, et al. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. Nat Genet. 2018 Apr;50(4):498–503.

255. Ramírez J, Dege C, Kutateladze TG, Hagman J. MBD2 and Multiple Domains of CHD4 Are Required for Transcriptional Repression by Mi-2/NuRD Complexes. Molecular and Cellular Biology. 2012 Dec 15;32(24):5078–88.

256. Takai D, Jones PA. The CpG Island Searcher: A New WWW Resource. In Silico Biology. 2003 Jun;3(3):235.

257. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011 May 15;25(10):1010–22.

258. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A. 2006 Jan 31;103(5):1412–7.

259. Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Laboratory Investigation. 2005 Sep;85(9):1172–80.

260. Illingworth RS, Bird AP. CpG islands – 'A rough guide.' FEBS Letters. 2009 Jun 5;583(11):1713–20.

261. Moody RR, Lo M-C, Meagher JL, Lin C-C, Stevers NO, Tinsley SL, et al. Probing the interaction between the histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. J Biol Chem. 2018 Feb 9;293(6):2125–36.

262. Pan H, Bilinovich SM, Kaur P, Riehn R, Wang H, Williams DC. CpG and methylation-dependent DNA binding and dynamics of the methylcytosine binding domain 2 protein at the single-molecule level. Nucleic Acids Res. 2017 Sep 6;45(15):9164–77.

263. Bahr C, von Paleske L, Uslu VV, Remeseiro S, Takayama N, Ng SW, et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. Nature. 2018 25;553(7689):515–20.

264. Kühn R, Schwenk F, Aguet M, Rajewsky K. Inducible Gene Targeting in Mice. Science. 1995;269(5229):1427–9.

265. Trowbridge JJ, Snow JW, Kim J, Orkin SH. DNA Methyltransferase 1 Is Essential for and Uniquely Regulates Hematopoietic Stem and Progenitor Cells. Blood. 2009 Nov 20;114(22):392–392.

266. Sprüssel A, Schulte JH, Weber S, Necke M, Händschke K, Thor T, et al. Lysine-specific demethylase 1 restricts hematopoietic progenitor proliferation and is essential for terminal differentiation. Leukemia. 2012 Sep;26(9):2039–51.

267. Masuda T, Wang X, Maeda M, Canver MC, Sher F, Funnell APW, et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. Science. 2016 Jan 15;351(6270):285–9.

268. Maeda T, Ito K, Merghoub T, Poliseno L, Hobbs RM, Wang G, et al. LRF Is an Essential Downstream Target of GATA1 in Erythroid Development and Regulates BIM-Dependent Apoptosis. Developmental Cell. 2009 Oct 20;17(4):527–40.

269. Longo PA, Kavran JM, Kim M-S, Leahy DJ. Transient Mammalian Cell Transfection with Polyethylenimine (PEI). Methods Enzymol. 2013;529:227–40.