2017

Characterization of a transcript found within the HBS1L-MYB intergenic region and its role in hemoglobin regulation in erythroid cells

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

CHARACTERIZATION OF A TRANSCRIPT FOUND WITHIN THE *HBS1L-MYB* INTERGENIC REGION AND ITS ROLE IN HEMOGLOBIN REGULATION IN ERYTHROID CELLS

by

TASHA ALEASE MORRISON

B.A., Temple University, 2005
M.P.H., Boston University, 2013

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader _____

David H. K. Chui, M.D.
Professor of Medicine

Second Reader _____

Martin H. Steinberg, M.D.
Professor of Medicine

"I want to make sure we use all our talent, not just 25 percent. Don't let anyone rob you of your imagination, your creativity, or your curiosity. It's your place in the world; it's your life. Go on and do all you can with it, and make it the life you want to live."

—Mae C. Jemison (first African American woman in space)

DEDICATION

This doctoral work is dedicated to my nieces and nephews. From oldest to youngest: Eddie, Pearl, Shafiq, Marcel, Johnny III, Terrance Jr, Malachi, Makaia, Troy, Camille, Christian, Jahseem, Zoe, Myah, Malia and Tyler. Because of you, I strive to excel in order to pave the way for your future accomplishments. Know that there is nothing in this world you can't do. So dream big, stay focused and disciplined, don't hold on to past mistakes, love and be loved, don't be scared of change, don't be scared to be different, and NEVER EVER give up. Life can be interesting sometimes, and you may find it difficult to climb up the ladder of success. But keep climbing…keep moving…stay positive…follow your dreams, not your fears…all things will work together for the good. And know that Aunt Tasha (aka Titi) has your back and loves you so very much.

ACKNOWLEDGMENTS

First, I would like to thank my mentor Dr. David Chui for the support he has given me these past four years. I grew as a scientist because he encouraged me to think independently, facilitated collaboration with very insightful and helpful individuals, and challenged me to improve my approach in addressing the objectives of my doctoral work. I also thank the past and current members of the Chui lab—Ibi Wilcox, Hong Yuan Luo, Lance Davis, Zhihua Jiang, Heather Edward and Alawi Habara. Ibi, thanks for helping with the project in the beginning stages. Everything I know about the methodology of studying hemoglobin is because of Hong. She took the time to ensure that I understood and was able to execute my experiments successfully. Lance, thanks for collecting patient samples when I needed them; taking blood from me when I needed it; and helping me with all of the intricate lab business that was sometimes a huge pain to deal with. I also want to give a big thanks to Dr. Martin Steinberg, who was like a second mentor to me. I appreciate his guidance in helping me submit and successfully get my first grant, as well as his encouragement, guidance and insight. I want to give another big thanks to Dr. George Murphy and his lab—Kim Vanuytsel, Amy Leung, Richard Giadone and Zaw Naing. For the last 6-8 months of my doctoral work, he stepped up to ensure I stayed on track and did the best science possible. I appreciate that the Murphy lab made time for me during their lab meetings, and all the insight, comments and suggestions made to help me successfully finish my work. I also thank Dr. Murphy for his letter of support for my grant. A huge thanks to my thesis committee—George Murphy (chair), David Chui, Martin Steinberg (second reader), Rick Myers, Paola Sebastiani and Shuaiying Cui.

There is no way I could have half-stepped getting my PhD with this group of people. They kept me on my toes, and I am a better scientist for it. When we first started this project, I felt like our ideas and hypothesis was really far-fetched. However, our early conversations with Dr. Myers and Dr. Wenqian Hu made this project viable. Thanks to Dr. Myers for looking at his RNA-Seq data, which made us believe even more that the transcript was expressed. And thanks to Dr. Hu for helping me with the RACE protocols, and his letter of support for my grant. Thank you to Dr. Cui who was part of this process during the last year, and was very helpful with the experimental design. Thank you to Dr. Adam Lerner for the hematology training grant. I also thank the Molecular and Translational Medicine program of GMS for their support, specifically, the former director, Dr. Bill Cruikshank, and current director, Dr. Matthew Jones.

My interest in basic science research first began during my last semester at Temple University, in the developmental genetics course taught by Dr. Douglas Baird. He made research so relevant to the advancement of medicine that it inspired me to pursue basic science research experience after I graduated. This led me to the laboratory of Dr. Elizabeth Henske, who was studying TSC and LAM at Fox Chase Cancer Center. Dr. Henske is the reason why I pursued this PhD. During the six years I was in her laboratory, from Fox Chase to Brigham and Women's Hospital, she gave me the tools to think scientifically about disease pathophysiology. She also introduced me to the world of conferences, lab meetings, journal clubs, grant writing, oral and poster presentations and manuscripts. She has also been a great friend, and has done numerous things to ensure my success both professionally and personally, so much so that I consider her

family. She's a boss…my role model…and I aspire to be at least half of who she is as a scientist and a human being.

My final acknowledgement goes to my loving and beautiful family—my mom, dad and siblings. People like me, who grew up in North Philadelphia in impoverish conditions, do not normally go on in life to get a PhD in Molecular and Translational Medicine. But I did, because I have such a supportive and loving family. I thank my parents, Janice and Johnny Morrison, for instilling values of faith, love and confidence. I also thank my parents for giving me eight of the best brothers and sister anyone could ask for. Eddie, Damon, Terrance, Devon, Monica, Johnny, Bryant and Juwan, thanks for filling my world with love, joy and laughter. We have been through a lot together, yet we are still standing…and standing strong with purpose. When I look back at all the negative things that have happened to me since college, which could have given me a good reason to quit, I fill up with so much emotion because (I have to personalize this now) y'all stepped in to encourage me, to tell me this is my purpose in life, to ensure me that everything will work out for the good, and to make sure I kept going no matter what.

I am truly blessed to have each and every one of you in my life. THANK YOU!!!

CHARACTERIZATION OF A TRANSCRIPT FOUND WITHIN THE *HBS1L-MYB* INTERGENIC REGION AND ITS ROLE IN HEMOGLOBIN REGULATION IN ERYTHROID CELLS

TASHA ALEASE MORRISON

Boston University School of Medicine, 2017

Major Professor: David H. K. Chui, M.D., Professor of Medicine

ABSTRACT

Sickle cell disease (SCD) is one of the most common hemoglobinopathies worldwide. It is caused by a homozygous mutation in codon 6 of the beta globin gene (*HBB*), which leads to polymerization of the variant hemoglobin and sickled red blood cells that obstruct blood vessels and reduce oxygen delivery to tissues. Patients with SCD have multiple clinical problems, including pain crises, anemia and organ damage. However, not all patients with SCD display all these clinical manifestations. One major factor for reduced occurrences of symptoms is fetal hemoglobin (HbF). HbF is the main hemoglobin in the fetus, and declines one year after birth to less than one percent of total hemoglobin. Nevertheless, there are individuals who continue to have high levels of HbF into adulthood, which is beneficial for an individual with SCD because HbF reduces the amount of sickle polymer in red blood cells. There are three major quantitative trait loci (QTL) associated with high HbF. However, these QTL account for 20-45% of HbF variance. Therefore, further investigation is required to fully understand how HbF is regulated.

The *HBS1L-MYB* intergenic polymorphism (HMIP) on chromosome 6q23 is one of the major QTL associated with high HbF. This region is also known to regulate other erythroid-specific traits due to an enhancer element that promotes the expression of the downstream gene, *MYB*, which controls hemoglobin expression and erythroid proliferation and maturation. The presence of RNA polymerase II binding and a 50-bp transcript suggested that a long noncoding RNA (lncRNA) is transcribed from this region. LncRNAs are non-protein-coding transcripts greater than 200 nucleotides and are involved in gene regulation. Therefore, it was hypothesized that a lncRNA is transcribed from the enhancer of *MYB* and regulates hemoglobin expression.

I characterized a novel lncRNA, 1283 bp in length that was differentially expressed among various tissue types, among erythroid progenitor cells with different hemoglobin makeup, and also during erythroid differentiation. Furthermore, knockdown of this lncRNA, named the *HBS1L-MYB* intergenic long noncoding RNA (*HMI-LNCRNA*), significantly increased HbF. Taken together, these observations suggest that *HMI-LNCRNA* can be a possible therapeutic target to increase HbF expression in patients with SCD and β-thalassemia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3’ 3-prime

5’ 5-prime

A ..... adenine

α ..... alpha

$\alpha_2\beta_2$ ..... adult hemoglobin A

$\alpha_2\beta^S{}_2$ ..... sickle hemoglobin

$\alpha_2\beta^S\gamma$ ..... sickle-gamma hemoglobin

$\alpha_2\delta_2$ ..... adult hemoglobin A_2

$\alpha_2\varepsilon_2$ ..... Hemoglobin Gower II

$\alpha_2\gamma_2$ ..... fetal hemoglobin

AA ..... African American

AGM ..... aorta-gonad-mesonephros

alncRNA ..... antisense lncRNA

ASO ..... antisense oligonucleotide

β ..... beta

β^S ..... sickle beta globin

Baso ..... basophilic erythroblasts

Bcl11a ..... B-cell lymphoma/leukemia

bp ..... base pair

BFU-E ..... burst-forming unit erythroid

BU ..... Boston University

C ..... cytosine

cDNA ..... complementary DNA

CD ..... cluster of differentiation

CD235a ..... glycophorin A

CD71 ..... transferrin receptor

CFU-E ..... colony-forming unit erythroid

ChIP ..... chromatin immunoprecipitation

ChIRP-seq ..... chromatin isolation by RNA purification sequencing

chr ..... chromosome

CLP ..... common lymphoid progenitor

CMP ..... common myeloid progenitor

cPPT ..... central polypurine tract

CRISPR ..... clustered regularly interspaced short palindromic repeats

CSSCD ..... Cooperative Study of Sickle Cell Disease

Ct ..... cycle threshold

δ ..... delta

Δ ..... delta

$(\delta\beta)^0$-thalassemia ..... delta-beta zero thalassemia

DMEM ..... Dulbecco's Modified Eagle's Medium

DNA ..... deoxyribonucleic acid

dsRNA ..... double-stranded RNA

ε ..... epsilon

ECL ..... enhanced chemiluminescence

EDTA ..... ethylenediaminetetraacetic acid

EMP ..... erythro-myeloid progenitor

ENCODE ..... Encyclopedia of DNA Elements

EPO ..... erythropoietin

EPOR ..... erythropoietin receptor

eRNA ..... enhancer RNA

EryP-CFC ..... primitive erythroid progenitor cell

FACS ..... fluorescence-activated cell sorting

FDA ..... Food and Drug Administration

γ ..... gamma

g ..... gravitational

G ..... guanine

GAG ..... group-specific antigen

GFP ..... green fluorescent protein

GTP ..... guanosine triphosphate

GWAS ..... genome-wide association study

H_2O ..... water

H3K4Me1 ..... histone 3 lysine 4 monomethylation

HBA ..... alpha globin protein

HBA1 ..... alpha globin gene

HbA_1 ..... adult hemoglobin A

HbA_2 ..... adult hemoglobin A_2

HBB ..... beta globin protein

HBB ..... beta globin gene

HBD ..... delta globin gene

HBE ..... epsilon globin gene

HbF ..... fetal hemoglobin

HBG ..... gamma globin protein

HBG ..... gamma globin gene

HbS ..... sickle hemoglobin

HBZ ..... zeta globin gene

HIV ..... human immunodeficiency virus

HMG ..... high-mobility group

HMIP ..... *HBS1L-MYB* intergenic polymorphism

HMI-LNCRNA ..... *HBS1L-MYB* intergenic long noncoding RNA

HMIT .....*HBS1L-MYB* intergenic transcript

HPFH .....hereditary persistence of fetal hemoglobin

HS .....hypersensitive site

HSC.....hematopoietic stem cell

HUDEP .....human umbilical cord blood-derived erythroid progenitor

IL-6 ..... interleukin 6

ilncRNA .....intronic overlapping lncRNA

IMDM ..... Iscove's modified Dulbecco's medium

iPSC .....induced pluripotent stem cell

kb..... kilobase

LB ..... Luria broth

LCR.....locus control region

LD ..... linkage disequilibrium

LDS .....lithium dodecyl sulfate

lincRNA ..... intergenic lncRNA

LNA .....locked nucleic acid

lncRNA ..... long noncoding RNA

LTR ..... long terminal repeat

M ..... molar

mESC ..... murine embryonic stem cells

min ..... minute

MIRA ..... Mistral

ml ..... milliliter

MLL1 ..... mixed lineage leukemia 1

mM ..... millimolar

MOI ..... multiplicity of infection

mRNA ..... messenger RNA

n ..... sample size

NaCl ..... sodium chloride

ncRNA ..... noncoding RNA

ng ..... nanogram

NGS ..... next generation sequencing

NHLBI ..... National Heart, Lung, and Blood Institute

NIH ..... National Institutes of Health

OrthoE ..... orthochromatic erythroblasts

PBMC ..... peripheral blood mononuclear cell

PBS ..... phosphate-buffered saline

PCR ..... polymerase chain reaction

plncRNA ..... pseudogene lncRNA

Pol II ..... polymerase 2

poly(A) ..... polyadenylation

PolyE ..... polychromatophilic erythroblasts

ProE ..... proerythroblasts

PVDF ..... polyvinylidene difluoride

qPCR ..... quantitative polymerase chain reaction

QTL ..... quantitative trait loci

RACE ..... rapid amplification of cDNA ends

RBC ..... red blood cell

RISC ..... RNA-induced silencing complex

RLM ..... RNA ligase mediated

RNA ..... ribonucleic acid

RNA ChIP ..... RNA-chromatin immunoprecipitation

RNA-Seq ..... RNA sequencing

RPM ..... revolutions per minute

RPMI ..... Roswell Park Memorial Institute

rRNA ..... ribosomal RNA

rs ..... reference SNP ID

RSV ..... Rous sarcoma virus

RT-PCR ..... reverse transcription polymerase chain reaction

S ..... sample

SCD ..... sickle cell disease

SCF ..... stem cell factor

sec ..... second

SFEM ..... serum-free expansion medium

shlncRNA ..... small RNA host lncRNA

shRNA ..... short hairpin RNA

siRNA ..... small interference RNA

SNP ..... single nucleotide polymorphism

sRNA ..... small RNA

SV40 ..... simian virus 40

T ..... thymine

T87Q ..... threonine to glutamine substitution at amino acid 87

TALEN ..... transcription-activator like effector nuclease

TAP ..... tobacco acid pyrophosphatase

TOI ..... transcript-of-interest

Tris-HCL ..... tris hydrochloride

tRNA ..... transfer RNA

TSS.....transcription start site

U.....units

µg .....microgram

µl .....microliter

µM.....micromolar

US .....United States

USA .....United States of America

ζ.....zeta

$\zeta_2\varepsilon_2$ ..... Hemoglobin Gower I

ZFN.....zinc finger nuclease

< .....less than

> .....greater than

= .....equal

% .....percent

+ .....positive

°C .....degree Celsius

CHAPTER 1. INTRODUCTION

1.1 Red Blood Cells

Cellular respiration—the set of metabolic processes that convert the nutrients we consume into energy—is required to sustain all the biological functions of the human body. However, energy can only be provided if oxygen is available. The presence of oxygen is controlled by the circulatory system, in which the heart pumps blood throughout the body. Three types of cells are found in the blood—leukocytes, thrombocytes and erythrocytes. Erythrocytes, also known as red blood cells (RBCs), are red, round, biconcave cells, are the most common cell type in the blood, and carry oxygen to tissues throughout the body. The process by which this occurs begins in the lungs, where the oxygen that is inhaled through the lungs is absorbed by RBCs. Once in the RBC, oxygen binds to the heme group that is attached to globin.[1] Next, RBCs circulate through the vascular system, releasing oxygen to all vital organs. Since RBCs are so important to human life, it is essential to know and understand how they are developed—a process called erythropoiesis—and the main protein content responsible for their function—hemoglobin.

1.1.1 Erythropoiesis

The ontogeny of erythropoiesis changes while the human embryo is developing in the womb. Collectively, there are two types of erythropoiesis—primitive and definitive—and three sites of origin for RBCs—the yolk sac, liver and bone marrow.[2–4] For both primitive and definitive erythropoiesis, the basic order of these processes is to

expand or proliferate erythroid precursors from progenitor cells, promote expression of hemoglobin, decrease cell size, enucleate and function as a red blood cell in circulation. However, the type of progenitors these cells are derived from and the location of erythroid differentiation differs between these two processes (**Figure 1.1A**).

Primitive erythropoiesis begins 15 days after fertilization within the yolk sac.[2,3] Hemangioblast progenitor cells derived from the mesoderm differentiate in the yolk sac into primitive erythroid progenitor cells (EryP-CFC), which form into clusters of cells called "blood islands."[5] Next, these erythroid progenitor cells are released into blood vessels and differentiate further as four different erythroblast precursors that are named based on the morphology of the cells. In the order of maturation, they are proerythroblasts (ProE), basophilic erythroblasts (BasoE), polychromatophilic erythroblasts (PolyE) and orthochromatic erythroblasts (OrthoE), during which the erythroblasts' nuclei condense, the cell size decreases and hemoglobin production increases (**Figure 1.1B**).[2,6] As OrthoE, the erythroblasts enucleate and become reticulocytes, and then achieve terminal maturation into RBCs indicated by the biconcave disc shape of the cell membrane.[7,8]

Definitive erythropoiesis occurs in both the liver and bone marrow.[2] However, the source of progenitor cells is different in both locations. In the liver, erythroblast precursors called burst-forming units erythroid (BFU-E) are originally from the yolk sac.[9,10] In the yolk sac, at around 4 to 5 weeks of gestation, erythro-myeloid progenitors (EMPs), which are multipotent hematopoietic progenitor cells, differentiate into BFU-E. Once expansion of BFU-E occurs, they are released into circulation and migrate to the

liver. Concurrently, hematopoietic stem cells (HSCs) begin to emerge from the aorta-gonad-mesonephros (AGM) region, and also travel to and expand in the fetal liver. At 10 to 11 weeks of gestation, HSCs begin to travel from the liver to the bone marrow, where they will permanently stay. This transition continues until after birth, when the bone marrow becomes the only source of red blood cells for the duration of life. In both the fetal liver and bone marrow, HSCs give rise to BFU-E, which will give rise to another erythroid precursor called colony-forming unit erythroid (CFU-E). Both are named based on their morphology when they are cultured *in vitro* in semisolid medium. Next, erythroid maturation occurs in the same order as primitive erythropoiesis. However, maturation does not occur in the blood vessel. It occurs in the fetal liver and bone marrow within erythroblastic islands, which are made up of erythroblasts that are physically attached to macrophages.[11] Once the OrthoE enucleate, reticulocytes are formed and released into the bloodstream where they terminally maturate into RBCs. The lifespan of a RBC is 120 days, and thereafter senesce and are removed by macrophages located in the spleen.[2]

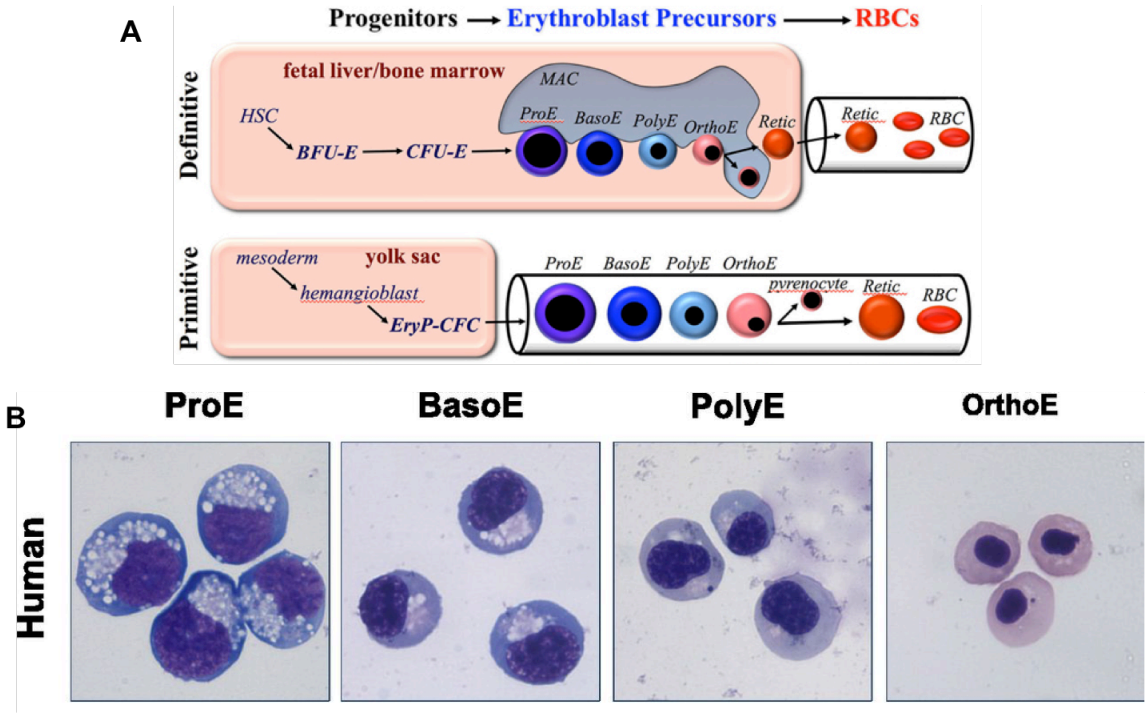


Figure 1.1. Erythropoiesis.

(**A**) Illustration of the direction of erythropoiesis from progenitor cells in the fetal liver and bone marrow (definitive), and progenitor cells in the yolk sac (primitive). During definitive erythropoiesis, erythroid differentiation and maturation occurs in the fetal liver or bone marrow until enucleation occurs, and the reticulocyte is released into the blood to undergo terminal maturation into a red blood cell. During primitive erythropoiesis, progenitor cells are released from the yolk sac and into the blood where they undergo erythroid differentiation from erythroblasts to red blood cells. Image adapted from Palis, *Frontiers in Physiology 2014*. (**B**) Morphology of human erythroblasts as proeythroblasts (ProE), basophilic erythroblasts (BasoE), polychromatophilic erythroblasts (PolyE) and orthochromatic erythroblasts (OrthoE). Image adapted from Palis, *Blood 2014*.

The most important driver of erythroid maturation is erythropoietin (EPO).[2,9,12] EPO is a cytokine that is released from the kidneys when RBC count is low, and binds to the erythropoietin receptors (EPOR) of CFU-E to promote differentiation. During erythroid maturation, two markers of maturation are expressed—transferrin receptor (CD71) and glycophorin A (CD235a).[10] CD71 is a marker for early erythroid maturation, and it is found in the plasma membrane of erythroblasts to promote uptake of iron within the cell.[13] This iron will help create the heme group that will be attached to globin to form hemoglobin that can reversibly bind oxygen. As the erythroblasts mature they also begin to express CD235a, which is another protein found in the plasma membrane of erythroblasts, as well as mature RBCs. CD235a contains sialic acid, which makes the RBCs hydrophobic. This allows the RBCs to flow through the blood without adhering to blood vessels and other cells. Once erythroid cells become fully mature RBCs, they lack expression of CD71 and only express CD235a.

1.1.2 Hemoglobin switching

Hemoglobin is a protein expressed by RBCs to bind to the oxygen that will be delivered to tissues in the entire body. It is made up of four subunits of globin chains, with each subunit bound to a heme group, which is necessary for oxygen transportation. From the development of the first RBCs derived from the yolk sac to the RBCs that are ultimately made from the bone marrow, the type of hemoglobin expressed changes, and this is known as "hemoglobin switching."[14–16] Hemoglobin switching can also occur as the erythroblasts mature into RBCs, and is called "maturational switching."[17] This is due to the regulation of gene expression on two loci—the alpha (α) globin gene cluster on

chromosome (chr) 16p13.3, and the beta (β) globin gene cluster on chr11p15. The functional genes of the α-globin gene cluster are made up of zeta (ζ) globin (*HBZ*) and two α-globin genes (*HBA2, HBA1*)[18]; and the β-globin gene cluster contains epsilon (ε) globin (*HBE)*, two gamma (γ) globin (*HBG2, HBG1*), delta (δ) globin (*HBD*), and β-globin (*HBB*) genes.[19] On both loci, the direction of globin gene expression moves from the 5' end to the 3' end. Hemoglobin is composed of two subunits from the α-globin gene cluster and two subunits from the β-globin gene cluster. During primitive erythropoiesis, erythroblasts derived from the yolk sac undergo a maturational switch that occurs between five to seven weeks of gestation.[17] The switch is only made within the α-globin gene cluster, from hemoglobin Gower I ($\zeta_2\varepsilon_2$) to hemoglobin Gower II ($\alpha_2\varepsilon_2$). At around three months of gestation, when the fetal liver begins to take over RBC production, the switch is now within the β-globin gene cluster, and the cells make fetal hemoglobin (HbF) ($\alpha_2\gamma_2$). HbF expression continues when the location of definitive erythropoiesis changes to the bone marrow. However, HbF does not remain the most abundant hemoglobin, and another switch occurs within the β-globin locus. After birth, HbF declines while adult hemoglobin (HbA_1) ($\alpha_2\beta_2$) increases as the main type of hemoglobin expressed in RBCs. Another type of adult hemoglobin (HbA_2) ($\alpha_2\delta_2$) is also made, but in small quantities. Eventually, the hemoglobin makeup in an adult becomes less than 1% HbF, 2.5% HbA_2 and 95.5% HbA_1.

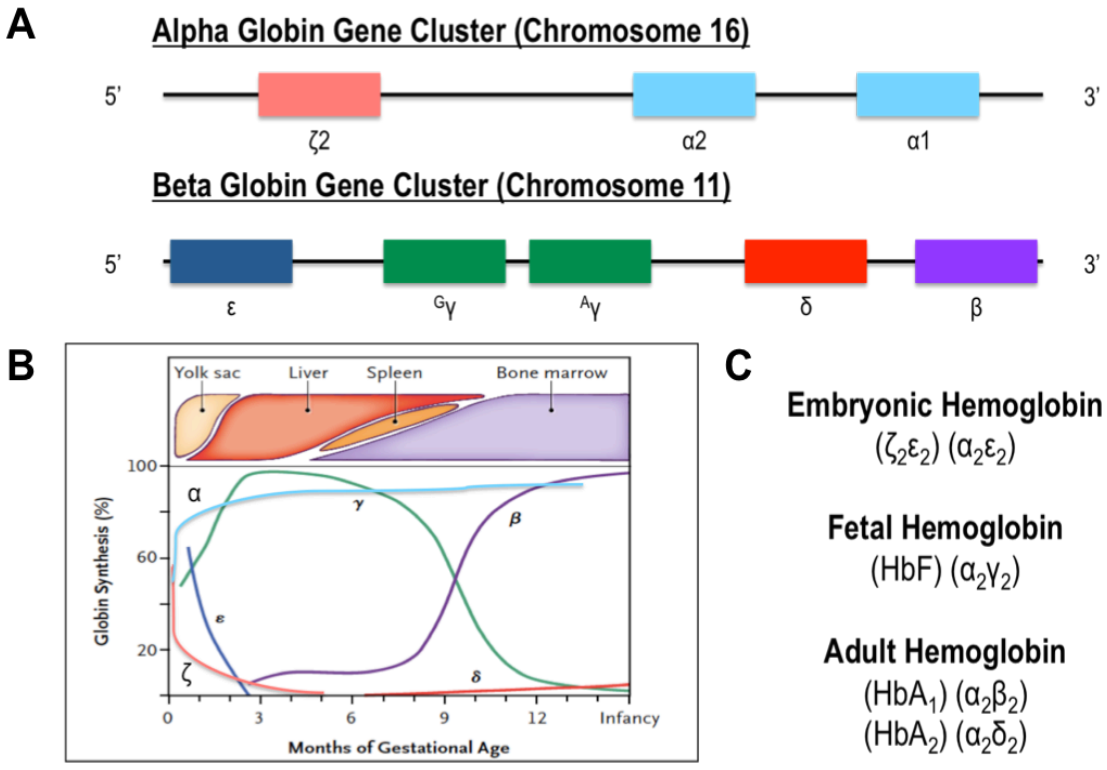


Figure 1.2. Hemoglobin Switching.

(**A**) Globin chains are transcribed from both the alpha globin gene cluster on chromosome 16 and the beta globin gene cluster on chromosome 11. (**B**) Genes on both loci are differentially expressed during gestation, and translates into individual globin subunits that form into a tetramer to make embryonic, fetal or adult hemoglobin (**C**). Chart in (B) is adapted from Forget, *N Engl J Med 2011*.

1.1.3 Transcription factors that regulate erythropoiesis and hemoglobin switching

Primitive and definitive erythropoiesis, and hemoglobin switching are regulated by the expression pattern of key transcription factors. *GATA1*, *KLF1*, *TAL1*, *LMO2* and *LDB1* are all expressed during primitive and definitive erythropoiesis.[2,4,20,21] They not only regulate hemoglobin expression but also erythroid maturation. GATA1 promotes erythropoiesis by regulating the expression of erythroid-specific genes. It does so by working in complex with TAL1, LMO2 and LDB1. Loss of GATA1 inhibits the maturation of erythroblasts.[22] KLF1 is another regulator of erythropoiesis as well as hemoglobin production.[23] It controls the expression of other erythroid-specific genes to promote expression of both embryonic and adult hemoglobin, stabilizes α-globin, ensures the biosynthesis of heme, and regulates other transcription factors.

The differences between primitive and definitive erythropoiesis are the expression of additional transcriptions factors only during the definitive stage (**Table 1.1**). These transcription factors work mostly to regulate expression of genes on the β-globin gene locus via the locus control region (LCR) that is located upstream this region. The LCR is a super enhancer with five DNase I-hypersensitive sites (HSs), which interacts with promoters of the downstream globin genes to promote their expression.[24] However, which promoter it will interact with depends on the expression of key transcription factors. c-MYB is one of these transcription factors that have been shown to promote erythropoiesis and work in complex with other transcription factors on the β-globin locus to reduce expression of γ-globin genes *HBG1* and *HBG2*.[25–29] The next transcription factor, SOX6, is a Sry-related high-mobility group (HMG) box transcription factor, and

has been shown to directly bind to the promoter of *HBE* to repress its expression.[30] The last transcription factor, BCL11a, contains zinc fingers that bind directly to the LCR and multiple sites on the β-globin gene cluster to specifically reduce expression of *HBG* and increase expression of the *HBB*.[31–34] It has been shown to do so by working with other transcription factors, including SOX6.[35]

Table 1.1. Globin genes and transcriptions factors that are expressed during primitive and definitive erythropoiesis
Table adapted from Palis, *Blood 2014*.

	Primitive Erythropoiesis	Definitive Erythropoiesis
Globin Genes	*HBZ* and *HBA1* *HBE* and *HBG*	*HBA1* *HBB*
Transcription Factors	GATA1, KLF1, TAL1, LMO2 and LDB1	GATA1, KLF1, TAL1, LMO2, LDB1, c-MYB, SOX6, and BCL11a

Taken together, RBCs are vital to human life, and require regulated processes at both the biological and molecular levels to ensure that healthy RBCs are developed and have the ability to effectively transport oxygen throughout the body.

1.2 Sickle Cell Disease

Genetic variants or mutations within both the α-globin and the β-globin loci, and transcription factors that regulate erythroid-specific traits can threaten erythropoiesis, hemoglobin expression and/or the viability of RBCs, and cause various blood disorders

or hemoglobinopathies.[18,36,37] One of these blood disorders is called sickle cell disease (SCD).

SCD was first described in 1910 by a cardiologist named James B. Herrick who saw that the appearance of RBCs from his patient was abnormal. However, Herrick could not decipher whether the patient had a primary blood disorder or whether the peculiar shaped cells occurred because of another clinical disorder.[38–40] It was not until 1945 that Linus Pauling hypothesized that sickled RBCs were due to defects within the hemoglobin, and in 1949 Pauling and others showed by gel electrophoresis that hemoglobin from sickle cells migrate differently than hemoglobin from normal cells, thus proving that his hypothesis was indeed correct.[40–42] Later that year, it was established that SCD can be inherited and as an autosomal recessive trait. However, the exact mutation of SCD was still not known. Ten years later, Vernon Ingram established that SCD is due to a mutation of codon 6 of *HBB*, which causes a change in an amino acid of adult hemoglobin (HbA_1) (glutamic acid to valine).[43]

Since then, with more understanding of erythropoiesis and hemoglobin switching, scientists were able to establish the pathophysiology of SCD. One aspect of SCD that was not quite understood during the earlier years of study is the sickled shape. The glutamic acid-to-valine change turns HbA_1 to sickle hemoglobin (HbS), where the two β-chains found in HbA_1 ($\alpha_2\beta_2$) becomes sickle β-globin (β^S).[44,45] This modifies the properties of hemoglobin produced in RBCs because valine is more hydrophobic than glutamic acid. When an RBC releases its oxygen, it is deoxygenated. In a RBC with normal hemoglobin, deoxygenation does not modify the structure of the hemoglobin

molecule. However, in a RBC with HbS, deoxygenation causes the more hydrophobic hemoglobin to polymerize. This leads to disruption of the cytoskeleton and protrusion of HbS polymer through the cell membrane, and is the reason sickle RBCs display many abnormal morphological characteristics. Damage to the cell membrane alternates the transmembrane proteins and the lipid bilayer, consequently changing many properties of the RBC. Normal RBCs can flow freely in circulation, without attaching to blood vessels and other cells. However, sickled RBCs undergo hemolysis, and attach to both the endothelium of blood vessels and other cells, and this promotes vascular occlusion (**Figure 1.3**).[46–48] Hemolysis and vaso-occlusion of sickle RBCs leads to acute pain crises, anemia and organ damage due to tissue oxygen deprivation. Other symptoms of SCD are swelling of the hands and feet, bacterial infection and acute chest syndrome. SCD is associated with premature death. Another aspect of SCD is the timeline when symptoms occur. Babies born with SCD do not show clinical evidence of the disease until sometime after the first three to six months of life. This is due to hemoglobin switching. After birth, the main type of hemoglobin a baby makes is HbF. However, during that first year of life, HbF significantly reduces while HbA_1 becomes the main type of hemoglobin expressed in RBCs. In an infant homozygous for the mutation of codon 6 in *HBB*, it is HbS, not HbA_1 that is mainly expressed. Therefore, clinical symptoms related to SCD are delayed while HbF is still high, and do not start until HbF is replaced by HbS.

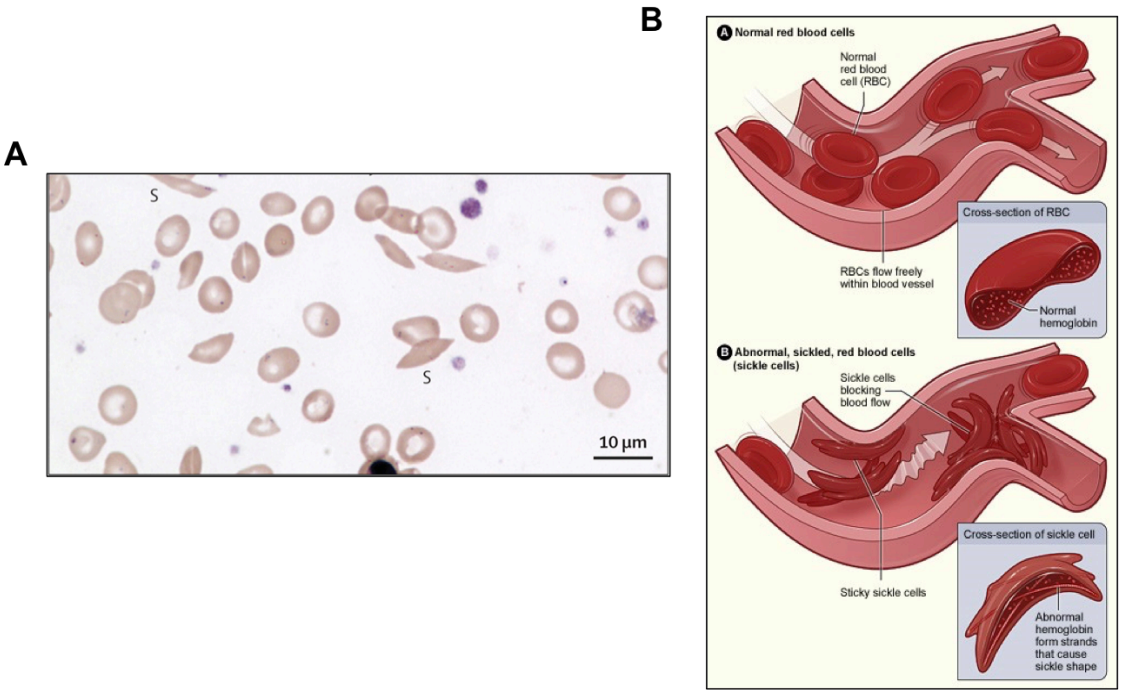


Figure 1.3. Sickle Cell Disease.

(**A**) Peripheral blood smear from a patient with sickle cell disease, stained with May-Grunwald-Giemsa. Red blood cells displaying a sickled shape (S) are distinct from normal round red blood cells. Image adapted from Rees et al, *The Lancet 2010*. (**B**) Illustration of how red blood cells with normal hemoglobin can flow freely through blood vessels, compared to red blood cells with sickle hemoglobin, which cause vaso-occlusion. Image adapted from NIH/NHLBI website (https://www.nhlbi.nih.gov/health/health-topics/topics/sca).

1.2.1 Epidemiology of SCD

SCD has become one of the most common inherited hemoglobinopathies worldwide.[44,45,49] Originally, SCD was mostly found in populations of people located in regions where malaria was endemic.[50] When individuals heterozygous for the sickle mutation—also known as having sickle cell trait—are infected with malaria, they have increased survival and are able to reproduce. Public health initiatives to decrease the incidence rate of malaria and other infectious diseases, and to promote better clinical practices within these regions have also increased the lifespan of people with SCD. Furthermore, the rise in human migration from these regions has now made SCD a global health problem. SCD is a prime example of a genetic balanced polymorphism; increased fitness (survival, reproduction) of the heterozygote is accompanied by decreased fitness of the homozygote.

SCD is mostly found in people of African, Arabian and Indian descent. Globally, about 300,000 newborns are born with SCD each year, and it has been projected that it will rise to about 400,000 newborns by 2050.[49] The highest incidence rate of SCD is found in West Africa, with up to 90,000 babies born with the disease per year in Nigeria alone (the highest rate worldwide).[51] In the United States of America (USA), approximately 100,000 people have SCD; one out of 365 African American (AA) babies are born with SCD; and one out of 13 AA babies are born with the sickle cell trait.[52]

1.2.2 HbF and SCD

The delayed appearance of sickle RBCs and the clinical manifestations of disease that hematologists observed throughout the first year after birth of infants with SCD led

them to consider HbF as a key factor for this delay. Subsequently, laboratory studies discovered that HbF inhibits the polymerization of HbS by obstructing the formation of HbS polymer in RBCs.[52] The tetramer that makes up HbS is $\alpha_2\beta^S_2$. However, when γ-globin is also expressed, the HbF molecule $\alpha_2\gamma_2$ is excluded from the HbS polymer, and the mixed hybrid tetramer $\alpha_2\beta^S\gamma$, has reduced incorporation into the polymer. Together, this reduces the polymerization potential of HbS leading to more normal RBCs.

SCD patients display a wide range of disease severity. Also different among SCD patients is the percent of HbF—some with nearly normal HbF, and others with higher levels of HbF. Epidemiological studies, which looked at the natural history of SCD, found that patients with SCD and high HbF have lower disease burden than SCD patients with lower HbF. One such study is the Cooperative Study of Sickle Cell Disease (CSSCD), which prospectively studied the natural history of about 4,000 patients in 23 health facilities around the US.[53–55] Funded by the National Heart, Lung, and Blood Institute (NHLBI) at the National Institutes of Health (NIH), the CSSCD found that various complications of SCD, including death, were reduced in patients with HbF > 8.6%. Taken together, SCD patients benefit from having high expression of HbF, and this is due to the inhibitory effects it has on the polymerization of HbS.

1.2.3 Prevention and treatment of SCD

There are preventive measures to reduce the incidence rate of SCD. Programs for premarital, antenatal and neonatal screenings determine if both individuals who are engaged to get married have the sickle cell trait, and if an unborn child or a newborn child has SCD.[44] This will allow couples to make an informed decision about whether to

continue with their nuptials or to have children if they are both carriers of the sickle cell trait; allow for antenatal diagnosis that will provide a choice to continue with the pregnancy if the fetus is found to have SCD; and will alert clinical professionals to newborns born with SCD in order to monitor and start treatment at an earlier time.

As for therapeutic treatments for SCD, there are two United States (US) Food and Drug Administration (FDA)-approved drugs to treat SCD. First approved is hydroxyurea. Hydroxyurea was initially used as a chemotherapeutic agent to treat myeloproliferative disorders. Scientists found that hydroxyurea treatment increased HbF levels in baboons[56], and therefore started clinical trials to determine if it is helpful for patients with SCD, leading to it being the standard of treatment.[57–61] However, the mechanism of how hydroxyurea upregulates HbF is not clearly understood. Furthermore, not all SCD patients respond to hydroxyurea treatment. Recently, L-glutamine, given as daily supplement, was associated with reduced vaso-occlusive crises and was FDA approved.[62] Whether or not this will change the standard of care is unclear.

Blood transfusions are another form of treatment, which reduces the percent of sickle cells in circulation by giving the patient healthy RBCs. Even though blood transfusions are effective, it is a burden because patients have to frequently have this procedure.[63–65] Also, patients who undergo frequent blood transfusions can develop antibodies against red cell antigens from transfused blood. Furthermore, only individuals living in countries with a good health care structure will be able to access this treatment.

A curative form of treatment is bone marrow transplantation, which replaces the bone marrow of SCD patients with bone marrow from healthy individuals, thus

substituting abnormal HSCs with normal HSCs.[66] The limitations with this procedure are that complications can be fatal, it is expensive and there are few suitable donors. An approach taken to address the lack of donors is gene therapy. Gene therapy takes autologous HSCs, infects them with a self-inactivating lentiviral vector that encodes a variant for β-globin chain (T87Q substitution), which inhibits the polymerization of HbS, and engraft the transduced HSCs into the bone marrow of the patient.[67] Recently, a case study was published about the 15-month follow-up of an SCD patient who underwent gene therapy.[68] Not only did high expression of the variant β-globin persist, the patient had no sickle crises. Gene therapy is a promising treatment, but the remaining two limitations of bone transplantation still is a concern.

1.3 HbF Variability

The observation that HbF is differentially expressed in patients with SCD, and that it is beneficial for patients to have higher levels of HbF, have steered researchers to determine the causes for this occurrence. What has been elucidated as the main causes for high HbF are two types of genetic variants—deletions along the β-globin gene locus, and single nucleotide polymorphisms (SNPs) found in various regions of the genome.[69] These deletions have been named based on sequence of discovery, or the geographical location or ethnic group the variant is mostly prevalent.

1.3.1 $(\delta\beta)^0$-thalassemia and hereditary persistence of fetal hemoglobin

Numerous deletions have been discovered within the β-globin gene cluster. Two types of deletions—$(\delta\beta)^0$-thalassemia and hereditary persistence of fetal hemoglobin

(HPFH)—are associated with high HbF.[70,71] Multiple deletions have been characterized for each type; they vary in size, and new deletions continue to be discovered. The deciding factors for what makes a deletion either a $(\delta\beta)^0$-thalassemia or an HPFH are the percentage of HbF that is expressed and the distribution of HbF among RBCs. Adults with $(\delta\beta)^0$-thalassemia have HbF of 5-15% and it is distributed heterocellularly among RBCs, which means that not all cells have the same amount of HbF. Conversely, adults with HPFH have higher levels of HbF (15-30%) and is distributed evenly or pancellularly among RBCs.

The location and size variability has led to discoveries of cis-acting elements essential for globin switching. The first element is a 3.5 (kilobase) kb intergenic region 5-prime (5') to the δ-globin gene (*HBD*), as has been found to be essential for the silencing of *HBG* expression via binding of Bcl11a and other transcription factors within this region.[72] Consequently, a deletion within this region will increase *HBG* expression. Next, is the 3-prime (3') HS1, which is a DNase1 hypersensitive site (HS) located downstream of *HBB* and is thought to be an insulator against enhancers for *HBG*.[73] Deletion of this region is also thought to increase *HBG* expression. Last, are the three 3'-enhancers, which are also downstream of *HBB* and have been found to enhance *HBG* expression.[73] A deletion through *HBB* to a position 5' to anyone of these enhancer regions is thought to bring the enhancer closer to *HBG* and increase its expression.

1.3.2 The 3 major quantitative trait loci

SNPs are single nucleotide variants and are found everywhere in the genome. Some SNPs do not change gene expression or epigenetic regulatory elements, while

others do. Genome-wide association study (GWAS) is a tool used to determine which SNPs are significantly associated with a specific trait. Since HbF is a trait that is beneficial for individuals with SCD, investigators have used GWAS to determine which SNPs are associated with high HbF, with the intent to determine if these SNPs could lead them to understand how HbF is regulated.[74] Using GWAS, three regions of the genome have been revealed to be associated with high HbF. These are the *HBG2* promoter region, *BCL11A*, and the *HBS1L-MYB* intergenic region, and together are the three major quantitative trait loci (QTL) for HbF.

First of these QTL to be discovered was the Xmn1-*HBG2* C-T polymorphism located within the promoter of *HBG2* 158 base pairs (bp) upstream from the *HBG2* transcription start site (TSS). This SNP creates an Xmn1 restriction site and is associated with increased expression of *HBG2*.[19,75] The mechanism of action or the functional basis of this association is currently unknown. The next region is *BCL11A* found on chr2p16, which encodes the transcription factor B-cell lymphoma/leukemia 11A (BCL11a). As previously discussed (section 1.1.3,) BCL11a is a transcription factor that is important for the γ-globin to β-globin switch.[76–79] SNPs found in intron 2 of *BCL11A* inhibit transcription of the gene, and therefore the hemoglobin switch does not occur, causing constitutive expression of *HBG*. The last QTL is the *HBS1L-MYB* intergenic polymorphisms (HMIP) found on chr6q23, which consists of three HMIP trait-associated blocks with SNPs that are in linkage disequilibrium (LD).[80,81] LD implies that SNPs are likely to be inherited with each other. This QTL will be discussed further in the next section.

1.4 The *HBS1L-MYB* Intergenic Region

The *HBS1L-MYB* intergenic region is 126 kb in length, and is in between the genes *HBS1L* and *MYB*. *HBS1L* is a member of the GTP-binding elongation factor family with no known association with erythropoiesis or hemoglobin expression. As discussed before (section 1.1.3), *MYB* encodes for the transcription factor c-MYB, and is known to regulate erythroid proliferation and maturation, and hemoglobin expression.[25–27] There are many SNPs within this intergenic region, and for a while it was not understood why these SNPs were associated with high HbF. Our lab was one of many laboratories to confirm the association of these SNPs with high HbF.[82] Among a group of non-related Chinese individuals heterozygous for β-thalassemia, HbF variability was associated with known SNPs such as rs9399137. What was novel was the discovery of a 3-bp deletion polymorphism (rs66650371), which is in LD with rs9399137 and therefore also associated with high HbF. Additionally, it was discovered by chromatin immunoprecipitation (ChIP) around the 3-bp deletion that in K562 cells, an immortalized erythroid cell line, are binding sites for four transcription factors that are involved in erythroid cell differentiation, GATA, TAL1, E47 and RUNX1. Also, Encyclopedia of DNA Elements (ENCODE) datasets revealed binding of RNA polymerase II and other factors such as Brg1 signaling, which signifies transcriptional activity, histone H3 lysine 4 monomethylation (H3K4Me1) activity (marker for enhancer regions), and expression of a 50-bp transcript. To confirm that this region was an enhancer, the fragment was cloned into a vector with an *HBG2* promoter and a luciferase reporter gene. Cells transfected with the fragment containing this region showed higher levels of luciferase

activity than cells cloned without this region. We concluded that the 3-bp deletion within HMIP was associated with increased HbF production, had signatures of transcription activity very similar to erythroid-specific transcription activity, and contained an enhancer element.

This work was later confirmed and extended by others who found that within the HMIP region was an enhancer region, which included the 3-bp polymorphism and binding sites for erythroid-specific transcription factors, and that the enhancer element actually regulated the expression of *MYB*.[83,84] They determined the mechanism of how polymorphisms in this region were associated with high HbF, and showed that when this enhancer region is intact, it has the ability to bind to the promoter of *MYB* and upregulate its expression. However, if the 3-bp polymorphism was present, the enhancer cannot bind to the promoter, resulting in a reduction in the expression of *MYB* and increase in HbF expression.

1.5 Long Noncoding RNA

The genome contains genes that are transcribed and processed into messenger RNAs (mRNAs), and these mRNAs are then translated into proteins to regulate the multiple processes in the cell. However, it was discovered that while 50-70% of the genome is transcribed, only 2% is translated into protein.[85] Except for transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), the other remaining untranslated RNAs were considered to have no biological function and were labeled as "junk RNAs." This was puzzling because many less complex species translate more of their genome, while more complex species, such as humans, translate less. As next generation sequencing (NGS)

became a popular scientific tool to look at total RNA expression, scientists discovered that noncoding RNAs (ncRNAs) are differentially expressed during development and within various disease models.[86] The discovery of changes in ncRNA expression patterns led to functional analyses to determine if modulation of ncRNA expression affect any downstream molecular or cellular activity. Their findings showed that ncRNAs actually do regulate molecular and cellular activities, and therefore changed the dogma of ncRNAs—they do have biological functions. There are various types of ncRNAs and among them are the long noncoding RNAs or lncRNAs.

LncRNAs are transcripts of 200 nucleotides or greater that do not code for protein.[87,88] They are transcribed from any region in the genome; low in abundance; mostly tissue-specific (more so than mRNA); located mostly in the nucleus of the cell; and can have many signatures of mRNA, such as the 5'-cap, 3'-polyadenylation and the ability to be spliced.[89] Furthermore, they have been shown to regulate cellular processes at various levels.[90,91] They can recruit epigenetic factors, control pluripotency, regulate transcription (in *cis* or *trans*) and translation, and be involved in post-transcriptional processing of mRNA. There are various types of lncRNAs that are named based on location relative to neighboring protein-coding genes.[92,93] Intergenic lncRNAs (lincRNAs) are located between two protein coding genes; antisense lncRNAs (alncRNAs) are transcribed complementary to a mRNA; intronic overlapping lncRNAs (ilncRNAs) are transcribed from the intronic region of mRNA; small RNA (sRNA) host lncRNAs (shlncRNAs) have sRNAs transcribed from their intronic regions; pseudogene

lncRNAs (plncRNAs) are transcribed from pseudogenes; and enhancer RNAs (eRNAs) are transcribed from the enhancer of an mRNA.

1.5.1 Enhancer RNAs

Enhancers are regions in the genome that when bound by transcription factors, are able to change the conformation of the chromatin in order to activate transcription of a gene. Active enhancers are known to have transcriptional activity. There are three reasons for enhancer transcription—noise, transcription-dependent effects, or RNA-dependent effects.[94] Noise is due to collision of Pol II with accessible chromatin, and does not have any biological function. A transcription-dependent effect is when movement of Pol II causes the conformation of chromatin to change, and epigenetic processes such as histone acetylation and methylation occur. This is eRNA-independent. eRNA-dependent effects are when transcripts derived from enhancer regions function to regulate transcription either in *cis* or *trans*.[95–99] This effect is eRNA-dependent, where eRNAs recruit or bind coactivators or transcription factors, and at act as a scaffold to ensure binding of these proteins at the promoter of neighboring protein-coding genes to promote gene expression (**Figure 1.4**).[95]

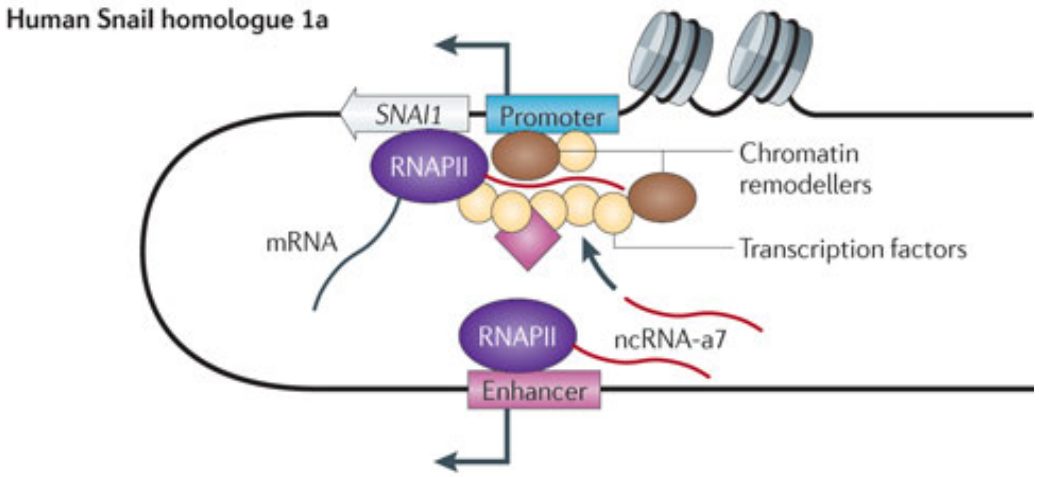


Figure 1.4. The function of an enhancer RNA (eRNA).

In this illustration, long noncoding RNA, ncRNA-a7, is transcribed from the enhancer element for *SNAI1* and acts as a scaffold to transcription factors bound to the gene's promoter to promote transcription of *SNAI1*. Image adapted from Ong and Corces, *Nature Reviews, Genetics, April 2011*.

1.5.2 Hematopoietic/Erythroid-associated long noncoding RNAs

There are multiple lncRNAs found to regulate either hematopoiesis or erythropoiesis.[93,100–111] Most of what is known has been studied mostly in murine-derived cells. Hu *et al* found that more than 400 putative mouse lncRNAs fluctuate during erythropoiesis.[112] Functional studies on one lncRNA, named long intergenic noncoding RNA (lincRNA) erythroid prosurvival (*LincRNA-EPS*), discovered that it inhibits apoptosis to promote erythroid differentiation. Other work characterized a variety of lncRNAs by using microarray analyses from the ENCODE datasets.[111] Since the *HOXA* gene cluster is differentially expressed during granulocytic differentiation and is myeloid-specific, they looked for expression of intergenic transcripts within this cluster in human cell lines that are able to undergo granulocytic differentiation. This laboratory was able to find a lncRNA, which they named *HOTAIRM1*, located between *HOXA1* and *HOXA2* genes. During granulocytic differentiation, *HOTAIRM1* is upregulated, and was dependent on the expression pattern of both *HOXA1* and *HOXA2*. Bertani et al looked for lncRNAs that are involved in epigenetic activation via Mixed lineage leukemia 1 (*MLL1*), which is known to be vital for embryonic development and hematopoiesis. RNA-chromatin immunoprecipitation (RNA ChIP) along with DNA microarray (RNA ChIP-on-chip) assay showed that the lncRNA Mistral (*MIRA*) binds and recruits Mll1 to activate transcription of homeotic genes *HOXA6* and *HOXA7*.[113] When Hoxa6 and Hoxa7 are expressed, they are able to activate expression of germ layer marker genes to induced differentiation of murine embryonic stem cells (mESC). Finally, work done by Paralker *et al*, used RNA sequencing in erythroblasts from both murine fetal liver and

human $CD34^+$ cells derived from cord blood, identified 1109 and 594 polyadenylated lncRNAs, respectively.[108] For both species, more than 50% of the lncRNAs were unannotated. Data from mouse erythroblasts showed that the expression of many of these lncRNAs are regulated by Gata1 and Tal1, which are key erythroid-specific transcription factors; 15% of lncRNAs are expressed in human erythroblasts; and knockdown of 7 out of 21 of the most abundant lncRNAs inhibited terminal erythroid differentiation (6 of the 7 were not expressed in human erythroblasts). Only 15% of the lncRNAs in human erythroblasts were expressed in mouse erythroblasts, suggesting that lncRNAs expressed in erythroid cells are highly species-specific.

1.5.3 Disease-associated lncRNAs

LncRNAs have also been implicated as key players in disease pathophysiology, including Alzheimer's disease, cancer and cardiovascular disease.[114–119] β-secretase-1 (*BACE1)-AS* is an antisense lncRNA transcribed complementary to the gene *BACE1*, which is known to be necessary for the formation of beta-amyloid and amyloid plagues found in patients with Alzheimer's disease.[120] Further investigation into the function of *BACE1-AS* discovered that this lncRNA actually interacts with and stabilizes the mRNA of *BACE1*, causing the upregulation of BACE1 protein levels. This implies that *BACE1-AS* is the key target to stop the production of amyloid plagues for treatment of Alzheimer's disease. In cancer, many lncRNAs have been found to be differentially expressed compared to their noncancerous counterparts. *MALAT1* is expressed in normal tissue; however, it is increased in multiple types of cancer.[121–123] Functional analyses for *MALAT1* show that when it is downregulated in cancer cells, they no longer have the

ability to proliferate, migrate and invade distal tissues. In cardiovascular disease, SNPs found in lncRNAs such as *ANRIL*, have been found to be associated with myocardial infarction.[124,125] Certain SNPs upregulated the expression of *ANRIL*, and promoted the production of peripheral blood mononuclear cells within the blood and buildup of atherosclerotic plagues within blood vessels, which can lead to a heart attack and stroke.

LncRNAs are now considered important regulators of gene expression and processes of the cell, including erythroid cells. Moreover, lncRNAs are implicated in disease pathophysiology that suggests the possibility that they might have some role in hemoglobinopathies such as SCD.

1.6 Rationale and Hypothesis

SCD is a genetic disease that affects the functions of both hemoglobin and RBCs. Not all individuals with SCD display the same disease phenotypes. One factor known to affect the phenotype of disease is HbF. The three known QTL only account for 20-45% of HbF variability.[77,126,127] Therefore, more is to be discovered about the regulation of HbF.

The *HBS1L-MYB* intergenic region is a key HbF-associated QTL. Within this region is an enhancer element for *MYB*, which is known to regulate erythroid-specific traits. Querying the ENCODE datasets in erythroid cells revealed the presence of a 50-bp RNA transcript from this enhancer element. Knowing that enhancers can express functional lncRNAs, we hypothesized that this 50-bp RNA might be part of an eRNA transcribed from within the *HBS1L-MYB* intergenic region, specifically from the enhancer for *MYB*, and represents a putative lncRNA that regulates *MYB* and its

downstream targets. If true, this might contribute to a better understanding of the regulation of erythropoiesis and hemoglobin expression, and this lncRNA could possibly be considered a therapeutic target for the treatment of SCD.

1.7 Specific Aims

There are two specific aims to address this hypothesis. The first aim is to characterize the transcript by determining the following: the full length, cellular location, relative expression compared to protein-coding genes, tissue-specificity, and the expression pattern among erythroid cells with different hemoglobin makeup and during erythroid differentiation. The second aim is to address the functional aspect of the transcript, and determine if the transcript regulates *MYB* and hemoglobin gene expression.

CHAPTER 2. CHARACTERIZATION OF THE TRANSCRIPT LOCATED WITHIN THE *HBS1L-MYB* INTERGENIC REGION

2.1 Introduction

Since the year 2000, when the International Human Genome Project made the assembly of the human genome available to the public, it became easier for investigators to know the location, length and sequence of genes of interest.[128,129] However, the caveat to knowing these characteristics is that these genes would have to be annotated within the human genome assembly. As scientific tools became more advanced, new genes were continuously added to the list. RNA sequencing (RNA-Seq) is a tool used to profile the transcriptome by using next generation deep-sequencing technologies.[130] The purpose of RNA-Seq is to determine which transcripts—mRNA, ncRNA and small RNA—are found, and quantify the amount of each individual transcript in a particular sample. However, there are still limitations, especially when it comes to characterizing lncRNAs due to the fact that they are expressed at very low levels compared to protein coding genes; and since lncRNAs are mostly tissue-specific, RNA has to be extracted from the right cell type.

As discussed in Chapter 1, the 50-bp transcript found in the *HBS1L-MYB* intergenic region was discovered using ENCODE datasets from RNA-Seq analyses of K562 cells. While it was plausible to consider this transcript a small noncoding RNA, we decided to perform additional analyses to determine if the outcome could possibly be different. RNA-Seq of RNA derived from primary erythroblasts showed multiple reads from this same region.

2.2 Determining the Full Length of the Transcript

The first set of experiments done to address the question of whether the 50-bp transcript was part of a lncRNA, used reverse transcription polymerase chain reactions (RT-PCRs). First, multiple primers were developed consecutively downstream and upstream the genomic region of the 50-bp transcript, at 100-bp intervals, spanning a 1400 bp region. We used for PCR complementary DNA (cDNA) (DNA synthesized from RNA) of K562 cells, which are immortalized erythroid cells that can be induced to make HbF. The controls used for PCR were PCR mix without DNA input (negative control), PCR mix with genomic DNA (positive control), and PCR mix with cDNA generated without reverse transcriptase (negative control to ensure RNA was not contaminated with DNA). The PCR reactions were done with primers starting from the 50-bp region and extended further with the next consecutive primers until there was no amplification. The longest PCR product amplified was 1180 bp (chr6: 135096355-135097534; UCSC Genome Browser assembly ID: hg38) (**Figure 2.1A**), and therefore upheld our hypothesis that a lncRNA was expressed from the *HBS1L-MYB* intergenic region.

We next determined the entire length of the lncRNA by using rapid amplification of cDNA ends (RACE). RACE is used to determine the 5'- and 3'-ends of full length RNA by ligating adapters to ends of cDNA, and then amplifying the transcript of interest with adapter-specific and gene-specific primers.[131] After confirmation of a PCR product by running the reaction on an agarose gel, the product is purified, cloned into a vector, and analyzed further to determine the exact location of the ends of the RNA (see subsection 2.8.6). Most RACE protocols use cDNA as the starting material, which can

produce non-specific amplification (especially for the 5'-end reactions) due to the lack of selectivity of specific RNAs. However, improvements have been made to produce a single band after PCR. Therefore, RNA Ligase Mediated (RLM) RACE was used, which, for the 5'-end, used RNA as the starting material; removed all degraded RNA, rRNA, tRNA and DNA; removed the 5'-cap (leaving only a monophosphate); ligated an adapter (will only ligate to RNA with the 5'-monophosphate); reverse transcribed RNA; and amplified 5'-end with adapter-specific and gene-specific primers. The RLM-RACE protocol for the 3'-end has fewer steps and higher success rate than the 5'-RACE. To determine the 3'-end, total RNA is reverse transcribed, adapter is ligated to 3'-end, and then PCR reactions are done with adapter-specific and gene-specific primers. RNA from K562 cells was initially used for both 5'- and 3'-RACE, however there was a huge DNA smear for the 5'-RACE, which could not be cloned into a vector for further analysis. The sequencing results for the 3'-RACE of K562 cells revealed that the 3'-end of the lncRNA was at location ch6: 135097644 (hg38) (**Supplemental Figure 3**), which was 110 bp downstream from what was found by RT-PCR (**Figure 2.1B**). As for the 5'-RACE, it was repeated using RNA from erythroblasts derived from cord blood $CD34^+$ cells. This time there were multiple bands, and each band was cloned and analyzed further, however, none of the sequencing results were found to be from this region. During the analysis of the expression pattern of this lncRNA among various types of tissues, it was found that the thymus expressed the lncRNA at much higher levels than both K562 cells and $CD34^+$ cells (see section 2.3). Repeating 5' RLM-RACE using RNA from the thymus showed only one band. Sequencing results of that band revealed that the 5'-end of this lncRNA

was at location ch6: 135096362 (hg38) (**Supplemental Figure 4 and Figure 2.1B**). HUDEP-2 cells, another immortalized erythroid cell line that expresses HbA, expressed more of the transcript than both K562 cells and erythroblasts derived from $CD34^+$ cells. Both the 5' and 3' RLM-RACE analyses gave the same results in these cells. The full length of this lncRNA, which we named the *HBS1L-MYB* intergenic long noncoding RNA (*HMI-LNCRNA*), is 1283 bp at chr6: 13509362-135097644 (hg38) (**Figure 2.1C**). *HMI-LNCRNA* is located 84.9 kb upstream from *MYB* and 41.5 kb upstream from *HBS1L*, and includes binding sites for erythroid transcript factors TAL1/E47, GATA1 and RUNX1, HbF-associated polymorphism rs66650371, and is transcribed from the enhancer for *MYB*. There is also no evidence of protein coding potential of *HMI-LNCRNA* based on the fact that it is absent in protein databases and contains no open reading frames longer than 300 nucleotides.

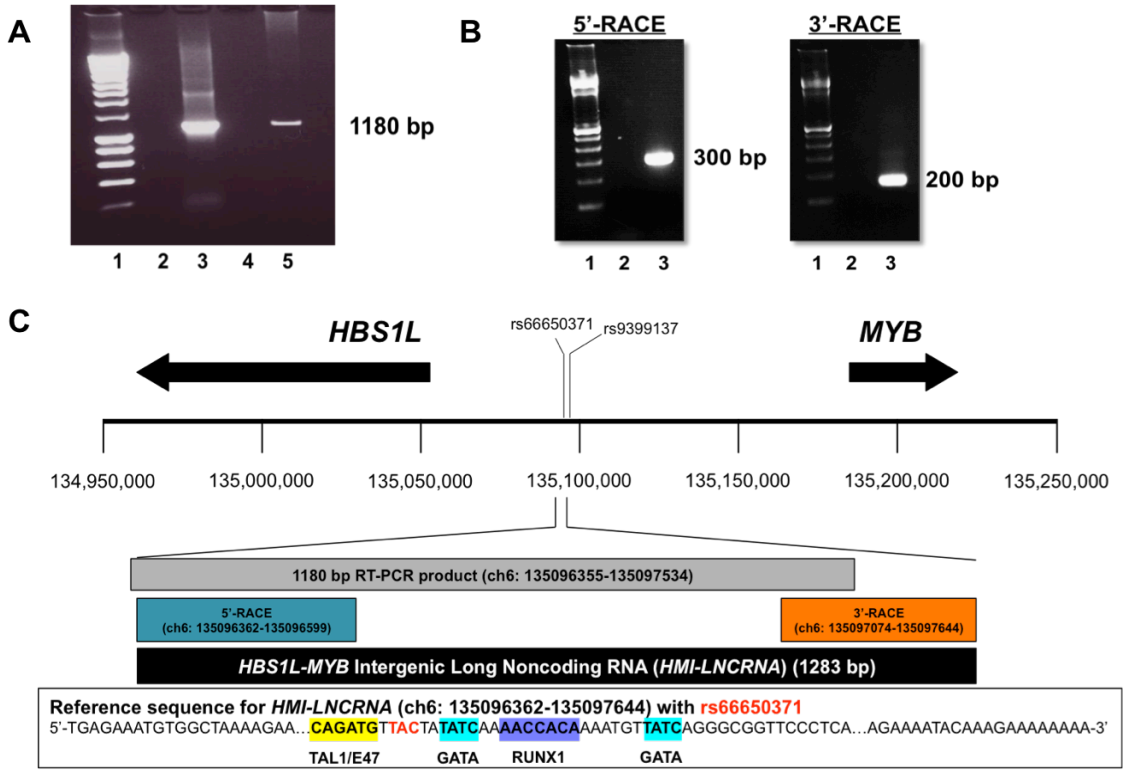


Figure 2.1. Determining the full length of the transcript found within the *HBS1L-MYB* intergenic region.

(**A**) PCR was done to amplify a 1180 bp region within the *HBS1L-MYB* intergenic region using cDNA from K652 cells. Samples were run on agarose gels and exposed under UV light. Lane 1: 100-bp DNA ladder; lane 2: negative control, PCR mix without DNA input; lane 3: positive control, PCR mix with genomic DNA; lane 4: PCR mix with cDNA generated without reverse transcriptase; and Lane 5: PCR mix with cDNA generated with reverse transcriptase. (**B**) Agarose gels show PCR products from the 5'-RACE reactions (using RNA from thymus) and 3'-RACE reactions (using RNA from K562 cells). For 5'-RACE, lane 1: 100-bp DNA ladder; lane 2: PCR mix with cDNA generated without Tobacco Acid Pyrophosphatase (TAP) treatment; and lane 3: PCR mix with cDNA generated with TAP treatment. For 3'-RACE, lane 1: 100-bp DNA ladder; lane 2: PCR mix with cDNA generated without reverse transcriptase; and lane 3: PCR mix with cDNA generated with reverse transcriptase. (**C**) Illustration of genomic region between 134950000-135250000 (hg38 coordinates) of chromosome 6, showing

approximate locations of rs66650371 and rs9399137, and *HBS1L* and *MYB* (arrows represent transcription direction and approximate length of genes). Based on DNA sequencing of RACE products, the 5'- and 3'-ends of the transcript were revealed to determine the full length of the transcript, which is 1283 bp in length and named the *HBS1L-MYB* Intergenic Long Noncoding RNA (*HMI-LNCRNA*). Located in the genomic sequence for *HMI-LNCRNA* are binding sites for erythroid-specific transcription factors TAL1/E47, GATA and RUNX1, and the 3-bp polymorphism (rs66650371). *HMI-LNCRNA* does not include rs9399137.

2.3 Tissue Specificity of *HMI-LNCRNA*

Since lncRNAs are expressed in either one or a few tissue types (more tissue-specific than mRNA), we explored the expression pattern of *HMI-LNCRNA* among 25 different types of cells and tissues. RNA samples were either purchased or extracted, and *HMI-LNCRNA* expression was analyzed by quantitative PCR (qPCR). Not all tissue expressed *HMI-LNCRNA*. The fold change of expression by qPCR in multiple samples that included erythroblasts derived from cord blood $CD34^+$ cells, bone marrow, fetal liver, brain, spleen, skeletal muscle, spinal cord, testis and thymus, had expression level of at least 1-fold above K562 cells; only three samples (erythroblasts derived from cord blood $CD34^+$ cells, testis and thymus) had a fold change above 10; and only one sample (thymus) had a fold change above 1000 (**Figure 2.2**). T-cells found in the thymus and erythroid cells both are derived from hematopoietic stem cells (HSCs), and require *MYB* to regulate differentiation. *MYB* is low in both mature erythroid cells and mature T-cells.[27] This suggests that *HMI-LNCRNA* may work together with *MYB* to regulate differentiation of both cell types. If this is so, we might expect higher levels of *HMI-LNCRNA* in HSCs, and declining levels during the maturation of each cell type. However, it is unknown if the expression of *HMI-LNCRNA* in the thymus is from the T-cells, their precursors or a non-associated cell type. *HMI-LNCRNA* in peripheral blood mononuclear cells (PBMCs), which consists of 40-70% mature T-cells, was low, suggesting that the high levels of this lncRNA maybe coming from T-cell precursors.

The expression pattern of *MYB* in these same samples was evaluated to determine if it correlated with the expression pattern of *HMI-LNCRNA*. K562 cells, which express

very low levels of *HMI-LNCRNA* compared to erythroblasts derived from cord blood $CD34^+$ cells and the thymus, had the highest expression of *MYB* among all the samples tested (**Supplemental Figure 1**). However, *MYB* has 13 isoforms, and the primer used for qPCR only recognizes one isoform. If different isoforms for *MYB* were expressed among these samples, primers specific to the isoform expressed in each sample would be needed to evaluate *MYB* expression.

2.4 Cellular Location of *HMI-LNCRNA*

Unlike mRNAs, which translocate to the cytoplasm to undergo translation, lncRNAs can either stay in the nucleus or move to the cytoplasm after transcription, depending on their function. Most lncRNAs are located in the nucleus because they regulate gene expression. Hence, it was important to know if *HMI-LNCRNA* is expressed in the nucleus or the cytoplasm. Primary erythroblasts derived from cord blood $CD34^+$ mononuclear cells were harvested and underwent nuclear and cytoplasmic fractionation. RNA was extracted from each fraction, and the relative transcript level of *HMI-LNCRNA* was analyzed by qPCR. *HMI-LNCRNA* was expressed almost entirely in the nucleus (**Figure 2.3**), suggesting that *HMI-LNCRNA* regulates gene expression. Regulation could affect *MYB*, *HBS1L* or another protein-coding gene.

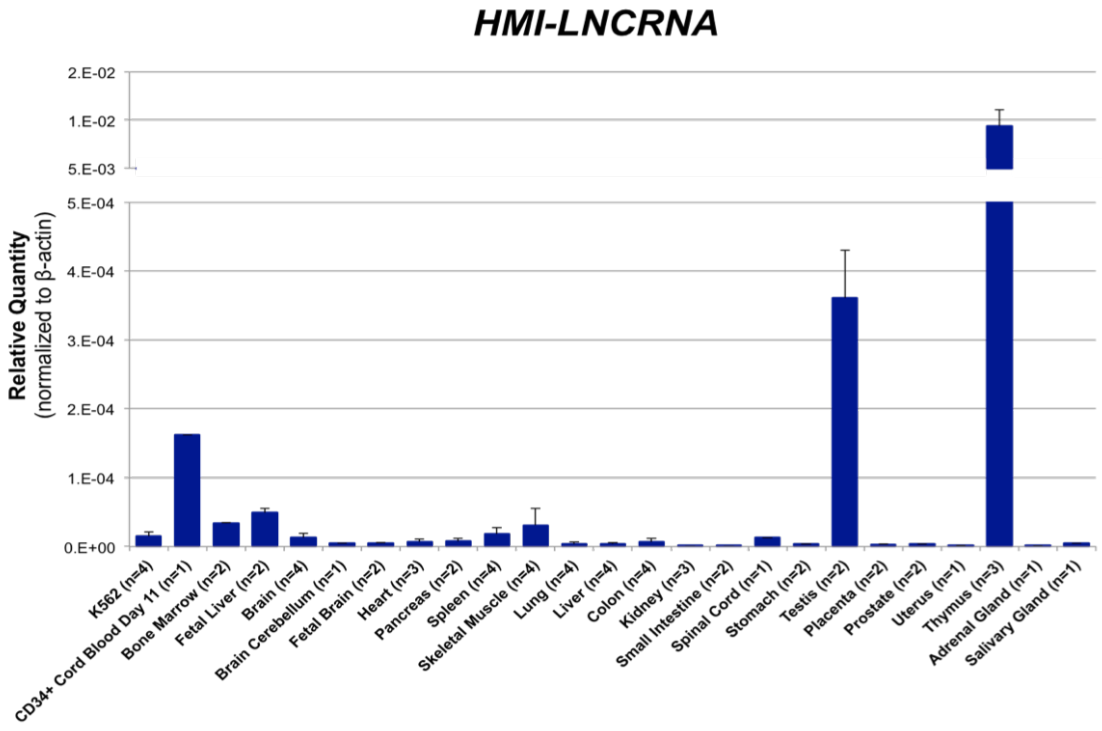


Figure 2.2. Expression pattern of *HMI-LNCRNA* among various human cells and tissue.

Relative quantity of *HMI-LNCRNA* was measured by qPCR in 25 different samples. $CD34^+$ cord blood day 11 is erythroblasts harvested at Day 11 of two-phase expansion and differentiation culture of $CD34^+$ mononuclear cells derived from cord blood. Means are shown for samples with two or more independent samples. *ACTB* was used as the endogenous control.

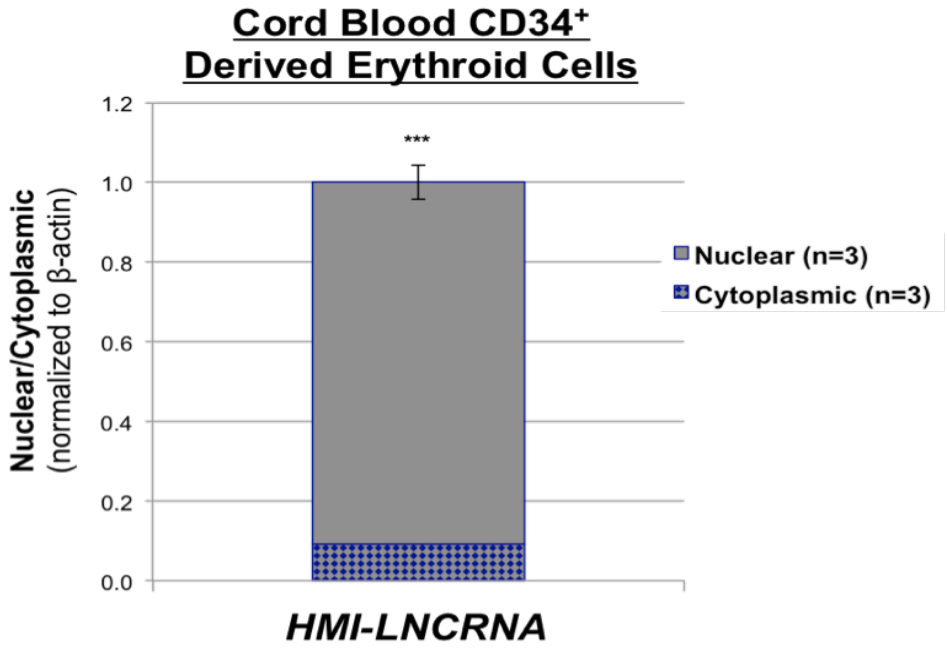


Figure 2.3. *HMI-LNCRNA* expression in nuclear and cytoplasmic fractions of primary erythroid cells derived from cord blood CD34$^+$ mononuclear cells.

Nuclear and cytoplasmic fractions were extracted from erythroid cells derived from cord blood CD34$^+$ mononuclear cells at Day 8 (n=1) and Day 12 (n=2) of differentiation. RNA extracted from each fraction was analyzed by qPCR analysis for expression of *HMI-LNCRNA*. The ratio of the relative quantity of nuclear to cytoplasmic *HMI-LNCRNA* transcripts was measured. *ACTB* was used as the endogenous control. p-value: ***<0.005. p-value obtained by Student T-test.

2.5 Comparing the Relative Expression Level of *HMI-LNCRNA* to a Protein-Coding Gene

LncRNAs have been shown to be less abundant than mRNA from protein-coding genes. To determine if *HMI-LNCRNA* shares this same characteristic, the relative expression level of *HMI-LNCRNA* was compared to the expression of *MYB* by qPCR. Using RNA extracted from HUDEP-2 cells and erythroblasts derived from peripheral blood $CD34^+$ cells, *MYB* was expressed 300 to 500 times more than *HMI-LNCRNA* (**Figure 2.4**).

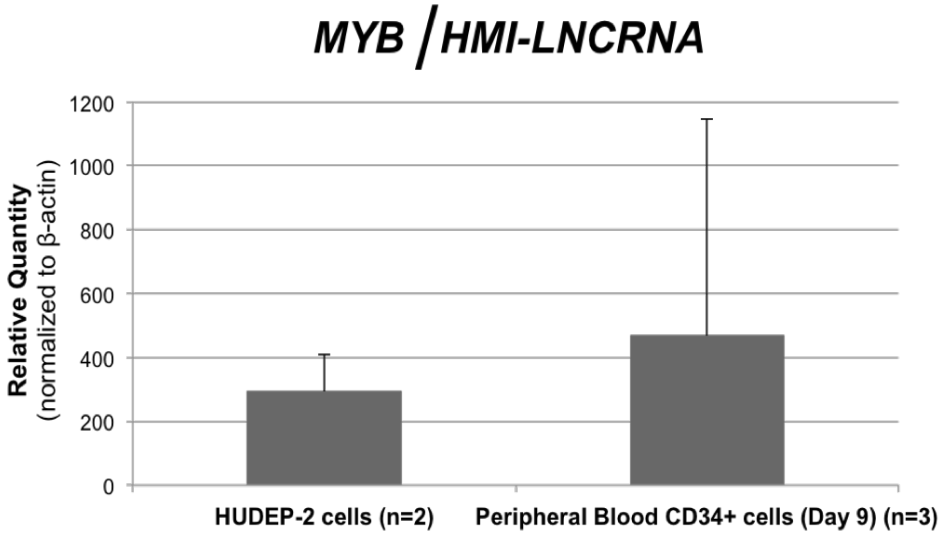


Figure 2.4. Ratio of *MYB* to *HMI-LNCRNA* expression in HUDEP-2 cells and erythroid cells derived from peripheral blood $CD34^+$ mononuclear cells.

Relative quantity ratio of *MYB* to *HMI-LNCRNA* was determined in HUDEP-2 cells (n=2) and erythroid cells derived from primary peripheral blood mononuclear cells at Day 9 of differentiation (n=3) by qPCR. *ACTB* was used as the endogenous control.

2.6 Relative Expression Pattern of *HMI-LNCRNA* Between Erythroid Cells with Different Hemoglobin Phenotypes and During Erythroid Differentiation

MYB is known to be associated with low *HBG* and high *HBB* expression. We therefore sought to determine if *HMI-LNCRNA* was associated with the expression of fetal or adult globin genes. *HMI-LNCRNA* and *MYB* mRNA expression were compared in fetal-like erythroid cells expressing *HBG* and adult-like cells expressing *HBB*. The first set of cells analyzed were HUDEP-1 cells, which express 99% *HBG* and HUDEP-2 cells, which express 99% *HBB*. *HMI-LNCRNA* was significantly higher in HUDEP-2 cells compared to HUDEP-1 cells, however, there was no difference in expression of *MYB* (**Figure 2.5A**). The results for *MYB* do not mirror protein expression of c-MYB, which was found to be higher in HUDEP-2 cells compared to HUDEP-1 cells (see Chapter 3, subsection 3.2.4). This suggested that even though there may not be a difference transcriptionally, there could be differences in the translation of *MYB* mRNA, or degradation and/or stability of c-MYB between HUDEP-1 and HUDEP-2 cells. In primary erythroblasts derived from cord blood $CD34^+$ cells that express 95% *HBG* and peripheral blood $CD34^+$ cells that express 65% *HBG*, both *HMI-LNCRNA* and *MYB* were significantly higher in erythroblasts derived from peripheral blood compared to cells derived from cord blood (**Figure 2.5B**). These data suggest that high levels of *MYB* and *HMI-LNCRNA* expression are associated with increased *HBB* expression and reduced *HBG* expression.

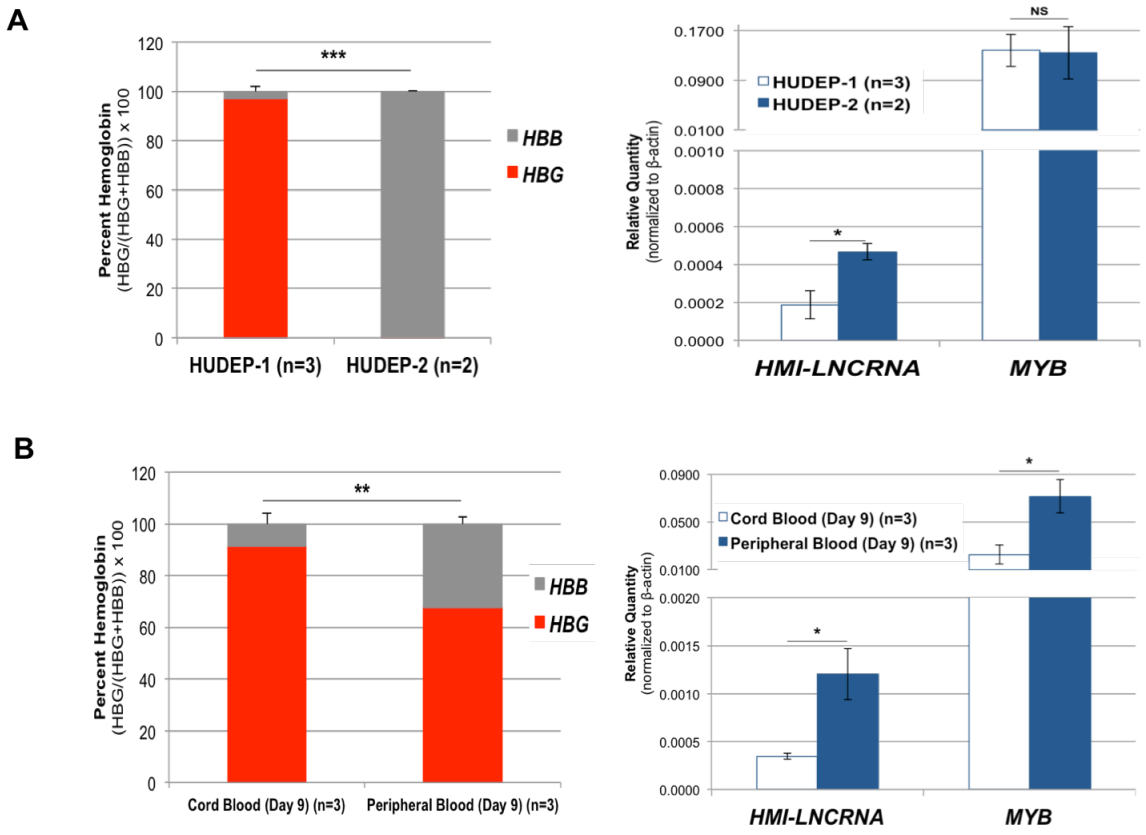


Figure 2.5. *HMI-LNCRNA* and *MYB* expression pattern in HUDEP-1 and HUDEP-2 cells, and erythroid cells derived from cord blood and peripheral blood CD34$^+$ mononuclear cells.

(**A**) Percent *HBG* to *HBB*, and the relative quantity of *HMI-LNCRNA* and *MYB* were determined by qPCR analysis in HUDEP-1 (n=3) and HUDEP-2 (n=2) cells (both maintained in expansion medium). (**B**) Primary erythroid cells derived from cord blood (n=3) and peripheral blood (n=3) CD34+ mononuclear cells (both expanded for 7 days and differentiated for 2 days) were analyzed by qPCR analysis to determine percent *HBG* to *HBB*, and the relative quantity of *MYB* and *HMI-LNCRNA*. *ACTB* was used as the endogenous control. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

Another characteristic of lncRNAs is that they are differentially expressed during development. Therefore, since *HMI-LNCRNA* is associated with erythroid cells that express more *HBB*, the expression pattern of this lncRNA was analyzed during differentiation of HUDEP-2 cells and peripheral blood $CD34^+$ mononuclear cells. During the differentiation of both cell types, *HMI-LNCRNA* and *MYB* mRNA decreased (**Figure 2.6**).

Taken together, these data show that both *HMI-LNCRNA* and *MYB* have similar expression patterns in erythroid cells, and suggest that *HMI-LNCRNA* may regulate the transcription of *MYB*.

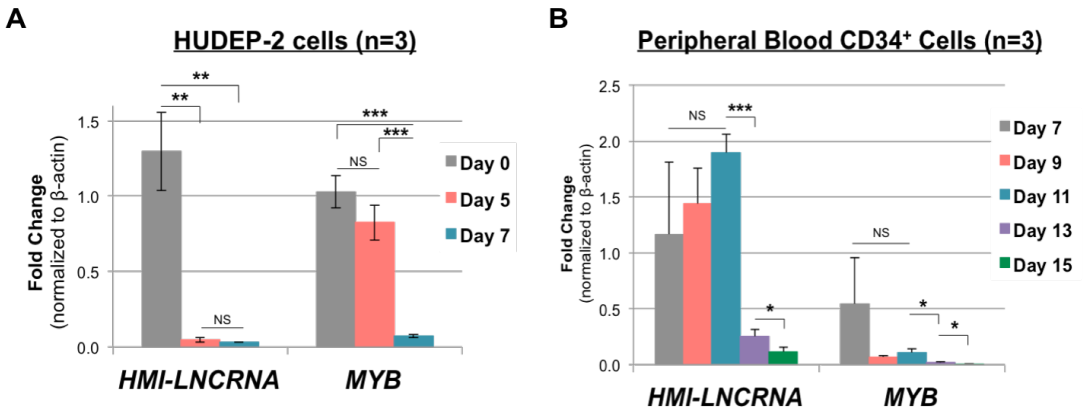


Figure 2.6. *HMI-LNCRNA* and *MYB* expression pattern during differentiation of HUDEP-2 cells and erythroid cells derived from peripheral blood $CD34^+$ mononuclear cells.

(**A**) Fold change of *HMI-LNCRNA* and *MYB* were determined in HUDEP-2 cells at Days 0, 5 and 7 in cultures with differentiation medium. (**B**) Fold change of *HMI-LNCRNA* and *MYB* were determined during differentiation of $CD34^+$ mononuclear cells derived from adult peripheral blood (n=3) at Days 7, 9, 11, 13 and 15. *ACTB* was used as the endogenous control. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

2.7 Summary and Conclusions

The 50-bp transcript annotated by ENCODE and located within the *HBS1L-MYB* intergenic region, which is known to regulate erythroid-specific traits, became an interesting element in our mission to better understand hemoglobin regulation. Therefore, additional steps were taken to fully characterize this transcript. The first step was to determine the full length of this transcript. By RT-PCR, and 5'- and 3'-RACE, we discovered that the transcript is 1283 bp (chr 6: 135096362-135097644, hg38) and is a lncRNA; it contains the binding sites for erythroid-specific transcription factors; and it is transcribed from the enhancer region of *MYB*, which brands it an enhancer RNA (eRNA). This lncRNA was named the *HBS1L-MYB* intergenic region long noncoding RNA (*HMI-LNCRNA*). *HMI-LNCRNA* is not erythroid-specific, however, it was expressed in only a few cell/tissue types and could possibly be hematopoietic-specific. The expression pattern of *HMI-LNCRNA* in erythroid cells was mostly nuclear; it was less abundant than *MYB* mRNA; higher expression was associated with erythroid cells that express more *HBB*; and it was downregulated during erythroid differentiation. *HMI-LNCRNA* and *MYB* showed the same expression pattern in erythroid cells and during erythroid differentiation (**Figure 2.7**).

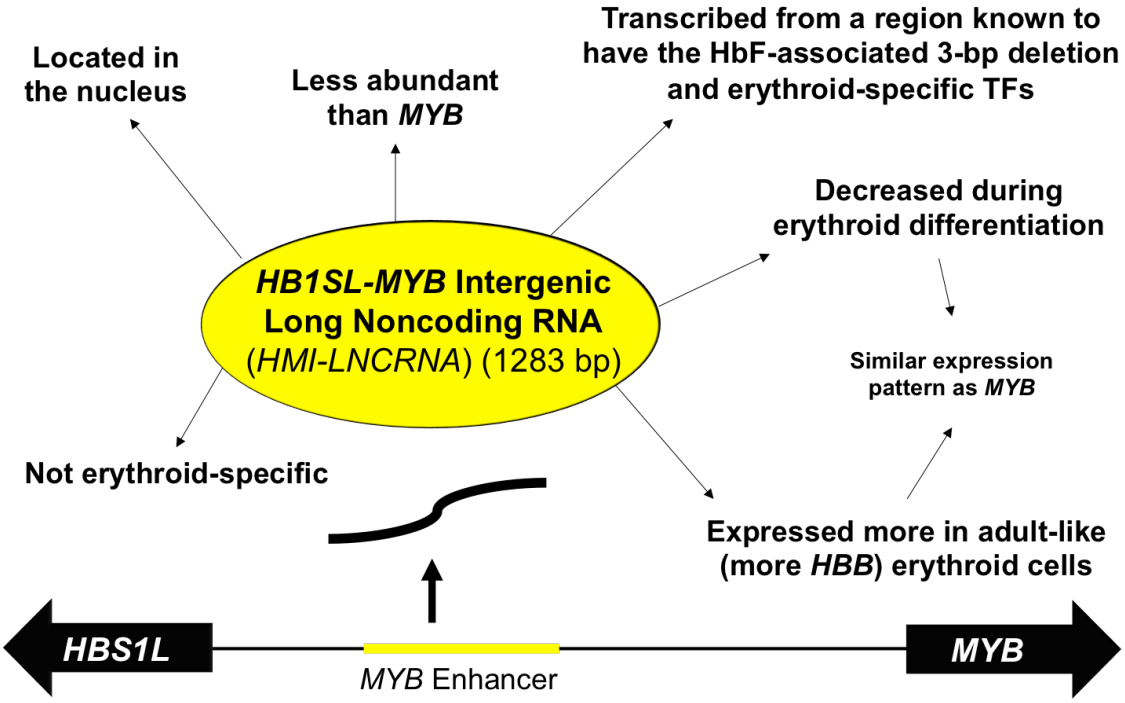


Figure 2.7. Summary of the characteristics of *HMI-LNCRNA*.

2.8 Materials and Methods

2.8.1 K562 cells

K562 cells are immortalized erythroid progenitors derived from pleural effusion of a patient with chronic myelogenous leukemia. They can be differentiated with arginine butyrate treatment toward erythroblasts that express only embryonic and fetal hemoglobin (**Supplemental Figure 2.4**).[132] They were cultured at 37 degrees Celsius (°C) in RPMI medium containing 10% FBS and 2% penicillin/streptomycin. To induce hemoglobin expression, cells are seeded at a concentration of $1x10^5$ cells/ml with 1 mM arginine butyrate, and cultured at 37°C for seven days (medium changed every two days).

2.8.2 HUDEP cells

Human umbilical cord blood-derived erythroid progenitor (HUDEP) cells are an immortalized erythroid cell lines derived from cord blood $CD34^+$ mononuclear cells.[133] HUDEP-1 and HUDEP-2 cells were cultured at 37°C maintained in StemSpan SFEM medium (StemCell Technologies) supplemented with SCF (50 ng/ml, Invitrogen), EPO (3 U/ml, Invitrogen), dexamethasone (1 μM, Sigma), doxycycline (1 μg/ml, Clontech), L-glutamine (1%, Life Technologies) and penicillin/streptomycin (2%, Life Technologies). For erythroid maturation, cells were cultured at 37°C in IMDM medium (Invitrogen) supplemented with heat inactivated human serum from human male AB plasma (5%, Sigma), EPO (3 U/ml, Invitrogen), insulin (10 μg/ml, Sigma), doxycycline (1 μg/ml, Clontech), holo-transferrin (500 μg/ml, Sigma), heparin (3 U/ml, Sigma), SCF (100 ng/ml, Invitrogen), L-glutamine (1%, Life Technologies) and penicillin/streptomycin

(2%, Life Technologies) for 5 days. For further erythroid maturation, doxycycline was removed and cells were cultured for two more days.

2.8.3 Primary $CD34^+$ mononuclear cells

Primary $CD34^+$ mononuclear cells derived from cord blood and peripheral blood (StemCell Technologies) contain hematopoietic stem/progenitor cells that can be differentiated into multiple lineage-specific cells. To generate erythroid cells, $CD34^+$ cells underwent a 2-phase culture system. The first phase is to expand the cells at 37°C in StemSpan SFEM II medium (StemCell Technologies) supplemented with StemSpan CC100 (1X, StemCell Technologies) and penicillin/streptomycin (2%, Life Technologies) for six or seven days. The second phase is to drive the cells to erythroid lineage. To induce erythroid differentiation, cells were cultured at 37°C in StemSpan SFEM II medium (StemCell Technologies) supplemented with SCF (10 ng/ml, Invitrogen), EPO (5 U/ml, Invitrogen), IL-6 (10 ng/ml, Sigma) and penicillin/streptomycin (2%, Life Technologies) for up to 10 days (**Supplemental Figure 2**).

2.8.4 RNA extraction

Cells taken from culture were centrifuged at 300 x g for 10 minutes at 4°C. Next, supernatant was removed without disturbing cell pellet, cells were resuspended in cold 1X Phosphate-buffered saline (PBS) and centrifuged again at 300 x g for 10 minutes at 4°C. After removal of supernatant, cell pellets were stored at -80°C or processed for RNA extraction. Total RNA was extracted using RNeasy Mini Kit (Qiagen), treated with

DNase (RNase-Free DNase Set, Qiagen), and then purified by RNA cleanup using RNeasy Mini Kit. For tissue-specificity experiment, multiple human organ RNA panels (Invitrogen and Clontech) were treated with DNase, followed by RNA cleanup.

DNase treatment is necessary for qPCR analysis for *HMI-LNCRNA*. In my earlier experiments, there was non-specific amplification in reactions without reverse transcriptase, which suggest that the reaction was contaminated with DNA. However, the non-specificity was only a problem when I used the primer for *HMI-LNCRNA*, and not for the other primers, such as *ACTB* and *HBG*, which had no amplification in qPCR reactions without reverse transcriptase. I have heard from other investigators that this is problem when analyzing lncRNAs by qPCR, but they do not understand why this occurs. I tried multiple DNase treatment protocols, such as incubating the column with DNase before eluting the RNA, however there was still non-specific amplification. What did work is after extracting the total RNA, incubate the entire sample in DNase for 10 minutes, and then cleanup the RNA to remove the DNase and buffer.

2.8.5 Reverse transcription polymerase chain reaction (RT-PCR)

cDNA was synthesized from DNase-treated RNA using SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). Oligo$(dT)_{20}$ was used for the annealing step, therefore, cDNA was synthesized from only poly$(A)^+$ RNAs.

PCR reactions were done using the Multiplex PCR kit (Qiagen). The following primers were used to amplify the 1180 bp product: 5'-ATCGCTCATGAGAAATGTGG-3' (forward) and 5'-GGAACCGCCCTGATAACATT-3' (reverse).

2.8.6 Rapid amplification of cDNA ends (RACE)

5'- and 3'-RACE were done using the FirstChoice RLM-RACE Kit (Ambion), following the manufacturer's instructions, using SuperTaq Plus Polymerase (Life Technologies) for PCR reactions. For 5'-RACE, the negative control was cDNA synthesized from RNA that was not treated with Tobacco Acid Pyrophosphatase (TAP). TAP removes the 5'-cap from RNA and leaves only a monophosphate. After TAP treatment, the adapter is ligated to only RNA with the monophosphate at the 5'-end. Therefore, RNAs not treated with TAP cannot ligate to the adapter, and will not amplify during the PCR reactions because the primer specific for the adapter cannot bind to them. The following gene-specific primers were used: 5'-GTCTAATGGTGTGGCTCACAAA-3' (5'-outer), 5'-CCCCAGCTTCCTTATCTGTAAA-3' (5'-inner), 5'-TTCACTCTGGACAGCAGATGTT-3' (3'-outer) and 5'-CGGTTCCCTCAGAAGACACTTA-3' (3'-inner). RACE PCR products were ligated to pCRII vector using TA Cloning Dual Promoter Kit (Invitrogen), transformed into One Shot INVαF chemically competent E. coli (Invitrogen), and grown on LB plates containing 100 μg/ml ampicillin and X-Gal. Insert-positive white colonies were picked and grown for DNA extraction. PCR reactions were done to amplify the regions where the RACE products were ligated to the vector (Forward: 5'-TGTGGAATTGTGAGCGGATA-3' and Reverse: 5'-GTTTTCCCAGTCACGACGTT-3'), and then processed for DNA sequencing.

2.8.7 DNA sequencing

PCR products were purified using AccuPrep PCR Purification Kit, and prepared for sequencing using ABI Big Dye Terminator v3.1 Cycle Sequencing Kit, which uses only one primer (forward or reverse) that was previously used for PCR. The program used for PCR was based on the DNA sample. If sequencing genomic DNA or cDNA, the following program was used: 96°C for 1 min; 15 cycles (96°C for 10 sec, 50°C for 5 sec, 60°C for 1 min 15 sec; 5 cycles (96°C for 10 sec, 50°C for 5 sec, 60°C for 1 min 30 sec); 5 cycles (96°C for 10 sec, 50°C for 5 sec, 60°C for 2 min); 10°C hold. If sequencing plasmid DNA, the following program was used: 96°C for 1min; 25 cycles (96°C for 10 sec, 50°C for 5 min, 60°C for 4 min); 4°C hold. ABI 3130xl Genetic Analyzer was used for sequencing. Sequence data was analyzed on FinchTV version 1.5.0. NCBI BLAST was used to determine location of sequence.

2.8.8 Nuclear and cytoplasmic fractionation

Cells taken from culture were centrifuged at 300 x g for 10 minutes at 4°C. Next, supernatant was removed without disturbing cell pellet, cells were resuspended in cold 1X Phosphate-buffered saline (PBS) and centrifuged again at 300 x g for 10 minutes at 4°C. After removal of supernatant, cells were processed using the Nuclear/Cytosol Fractionation Kit (BioVision), following manufacturer's instructions.

2.8.9 Quantitative PCR

qPCR—also known as real-time PCR—is a tool used to quantitate the relative transcript levels of a specific RNA. In the PCR phase, gene-specific primers are used to

amplify the target during a 40-cycle reaction period (1-40). More abundant transcripts are amplified at earlier cycle thresholds (Ct) values, while less abundant transcripts are amplified at later Ct values. To calculate the relative quantity of a transcript-of-interest (TOI), first take the difference between Ct of TOI and Ct of endogenous control, such as *ACTB* (β-actin), to get ΔCt ($Ct^{TOI} - Ct^{ACTB} = \Delta Ct^{TOI}$). Finally, the relative quantity of TOI is calculated using $2^{(-(\Delta Ct))}$. To calculate fold change between two samples—sample 1 (S1) and sample 2 (S2)—another approach is taken. For each sample, calculate ΔCt^{TOI}. Next, calculate $\Delta\Delta Ct^{TOI}$ for each sample—for S1, [(ΔCt^{TOI} for S1) – (ΔCt^{TOI} for S1)] = ($\Delta\Delta Ct^{TOI}$ for S1) = 0; and for S2, [(ΔCt^{TOI} for S2) – (ΔCt^{TOI} for S1)] = ($\Delta\Delta Ct^{TOI}$ for S2). Last, the fold change of the TOI for each sample is calculated using $2^{(-(\Delta\Delta Ct))}$, in which S1 will equal 1.

Total RNA, TaqMan primers and reagents from the TaqMan RNA-to-CT 1-Step Kit (Applied Biosystems) were the components used to make 20 µl reaction mix for each sample (done in triplicate) that was loaded onto a 96-well plate, following manufacturer's instructions. The following TaqMan gene expression assays (Applied Biosystems) were used: *HBG1/2* (Hs00361131_g1), *HBB* (Hs00758889_s1), *MYB* (Hs00920556_m1), *HMI-LNCRNA* (custom TaqMan assay designed by Applied Biosystems to target genome position chr6: 135096354-135097644, hg38; assay ID number AJI1MTQ), and *ACTB* (Hs01060665_g1). QPCR reactions were ran on a StepOne Plus qPCR machine (Applied Biosystems). *ACTB* was used as the endogenous control.

CHAPTER 3. THE EFFECTS OF MODULATING *HMI-LNCRNA* EXPRESSION IN ERYTHROID CELLS

3.1 Introduction

Characterization of *HMI-LNCRNA* has given insight into how this lncRNA possibly functions in erythroid cells, possibly as an eRNA that promotes the expression of *MYB*, thereby regulating hemoglobin expression, and erythroid proliferation and maturation. To determine if *HMI-LNCRNA* regulated *MYB* and erythroid-specific traits, we employed various systems to modulate its expression in erythroid cells.

There are multiple techniques used to either knockdown or overexpress RNAs *in vitro*. To downregulate expression of RNAs, small interference RNAs (siRNAs), locked nucleic acids (LNAs) and short hairpin RNAs (shRNAs) can be used. siRNAs are synthetically made double-stranded RNAs (dsRNAs) of 20-25 nucleotides that are complementary to the RNA of interest. Cells are transfected with siRNAs using various methods that will allow the siRNAs to travel through the cell membrane and into the cytoplasm. Once in the cytoplasm, siRNAs are processed by Dicer to separate the two strands, and then binds to the RNA-induced silencing complex (RISC). The siRNA-RISC complex binds to the RNA of interest and cleaves it, reducing the amount of RNA in the cell. LNAs are a newer system that uses RNase H to cut the RNA of interest.[134] Cells are treated with a single-stranded RNase H activating DNA flanked by antisense oligonucleotides specific to the RNA of interest, and require no transfection reagent. After the LNA is in the cell, the antisense oligonucleotides binds to the RNA of interest and the RNase H activating DNA recruits RNase H to cut the RNA. Similar to siRNAs,

shRNAs use Dicer to separate the double-stranded DNA, and RISC to cleave RNA. However, they differ in the way they are made and introduced to the cell. An shRNA is a synthetically made double-stranded DNA template containing four specific regions—sense, loop, antisense and terminator. The sense region is about 20 bp of the genomic region where the RNA of interest is transcribed from; the loop is 10 bp of nonsense nucleotides; antisense is complementary to the sense; and the terminator consists of nucleotides that will stop transcription. The shRNA template is ligated to a lentivector, which contains virus-specific genes that allow for efficient cell transduction and genome integration. Once integrated into the genome, shRNA is expressed, forms into a hairpin structure and undergoes the same processing as siRNAs.

There are also methods to completely stop transcription of RNAs. Gene editing uses nucleases such as zinc finger nucleases (ZFNs), transcription-activator like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR) to edit the genome to inhibit gene transcription. One caveat to using gene editing, especially when studying intergenic lncRNAs, is that genomic regions with epigenetic regulatory elements such as enhancers and promoters cannot be targeted. The effects observed maybe due to the lack of such elements, not the lack of transcription of the lncRNA of interest.

The decision in determining which system to use to knockdown *HMI-LNCRNA* was based on one key characteristic—its expression is almost entirely in the nucleus. siRNAs will not be effective as they stay in the cytoplasm. Therefore, shRNAs were used due to the fact that they integrate into the genome, and therefore are expressed in the

nucleus. Overexpressing *HMI-LNCRNA* also required its expression in the nucleus. Hence, full length *HMI-LNCRNA* was cloned into a lentivirus and used to transduce erythroid cells.

3.2 Downregulation of *HMI-LNCRNA* in erythroid cells

3.2.1 Determining which shRNA most effectively reduces expression of HMI-LNCRNA

Three different shRNA templates (HMIT-1, HMIT-2 and HMIT-3) were developed to target *HMI-LNCRNA*, and cloned into the pGreenPuro lentivector. pGreenPuro is an HIV-based expression lentivector, and contains multiple components for cloning, lentiviral production and transduction, and shRNA expression (see subsection 3.5.1 for details). pGreenPuro also expresses green fluorescent protein (GFP), which can be detected by flow cytometry to determine the percentage of cells infected with lentivirus. To determine which shRNA template most efficiently downregulates *HMI-LNCRNA*, each shRNA plasmid was transduced into TF-1a cells and cultured in puromycin for up to two weeks to select for only cells infected with lentivirus. At the time cells were harvested for analysis for expression of *HMI-LNCRNA*, about 95% of the cells were GFP-positive (GFP^+). *HMI-LNCRNA* was reduced with all shRNA templates, with HMIT-1 showing the greatest reduction at 70% (**Supplemental Figure 3.1**). This experiment was repeated in HUDEP-2 cells, however the cells were analyzed two days after transduction without puromycin treatment. When cells were harvested, about 30% were GFP^+, and a significant decrease in *HMI-LNCRNA* was found only in cells transduced with HMIT-1 (**Supplemental Figure 3.2**). Based on the observations from

qPCR analysis for *HMI-LNCRNA*, HMIT-1 shRNA was used for all knockdown experiments, and was labeled as HMI-lncRNA shRNA.

3.2.2 Downregulation of HMI-LNCRNA in HUDEP-1 and HUDEP-2 cells

HUDEP-1 and HUDEP-2 cells were used for knockdown of *HMI-LNCRNA* expression. Each cell line was split into three groups—naïve (non-transduced), scrambled shRNA and HMI-lncRNA shRNA. Each group had three biological replicates. Two weeks after the second transduction, cells were collected for the following analyses: flow cytometry for GFP; qPCR for *HMI-LNCRNA, HBS1L, MYB, HBG, HBB,* and *ACTB*; Western blot analysis for HBG, HBB, c-MYB; and slide preparations for immunofluorescent staining for HBG and HBB (**Figure 3.1A**). The first two experiments that were done to ensure that almost all cells were infected and that *HMI-LNCRNA* was significantly downregulated, and therefore used flow cytometry for GFP and qPCR for *HMI-LNCRNA*, respectively.

Flow cytometry results in HUDEP-2 cells showed that there were no GFP^+ cells in the naïve group, 84% GFP^+ in the scrambled shRNA group and 90% GFP^+ cells in the HMI-lncRNA shRNA group (**Supplemental Figure 3.3**). qPCR results for *HMI-LNCRNA* showed there was about 50% reduction in the HMI-lncRNA group compared to both naïve and scrambled shRNA groups (**Figure 3.1B**). Almost the entire cell population of HUDEP-2 cells transduced with either scrambled shRNA or HMI-lncRNA shRNA was infected, and there was a significant reduction in of *HMI-LNCRNA*. Transduction of HUDEP-2 cells was repeated but with the addition of allowing the cells

to mature. Both GFP-positivity and reduction of *HMI-LNCRNA* was similar to the first experiment.

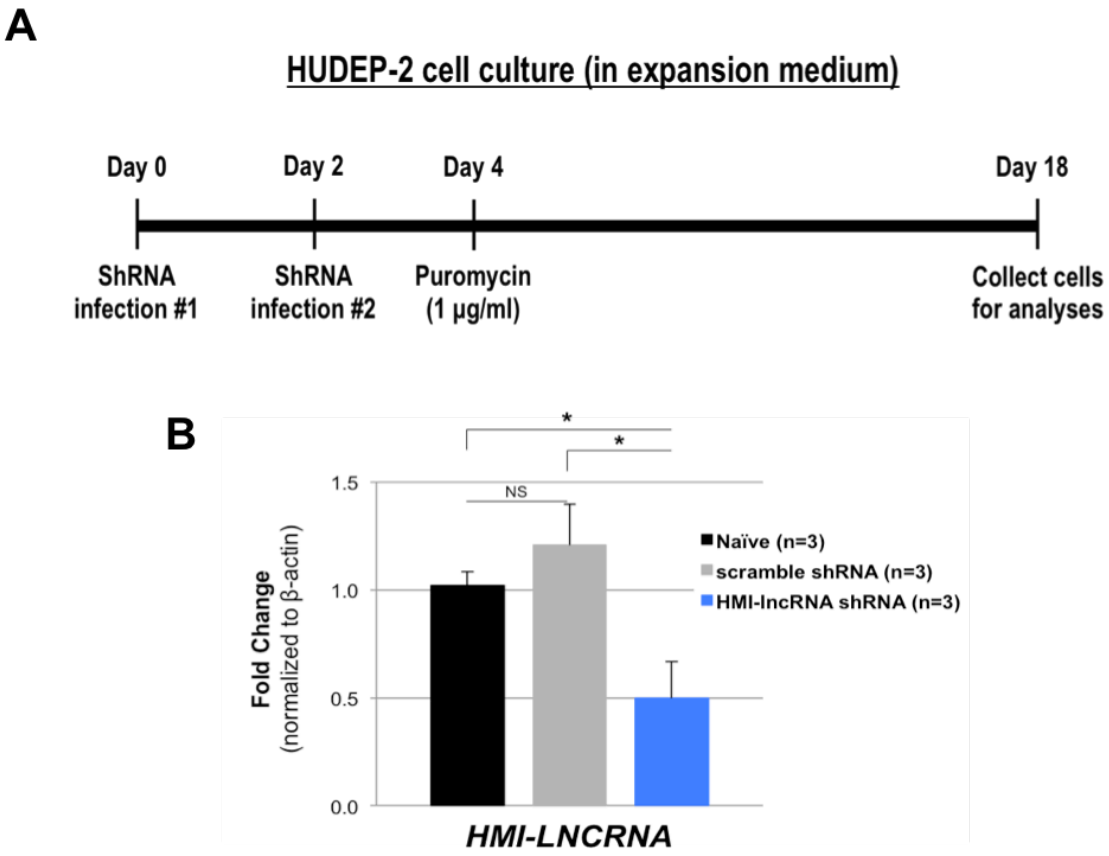


Figure 3.1. Downregulation of *HMI-LNCRNA* in HUDEP-2 cells.

(**A**) Illustration of timeline for transduction and culture of HUDEP-2 cells. Cells were maintained in expansion medium. (**B**) Expression level of *HMI-LNCRNA* was determined by qPCR in HUDEP-2 cells that were not transduced (naïve), and transduced with either scramble shRNA or HMI-lncRNA shRNA lentiviruses by qPCR. *ACTB* was used as the endogenous control. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

In HUDEP-1 cells, no GFP^+ cells were found in the naïve control group, 89% GFP^+ cells were found in the scrambled shRNA group and 92% GFP^+ cells were present in the HMI-lncRNA shRNA group (**Supplemental Figure 3.4A**). This result showed that almost all cells that were transduced expressed the shRNAs. However, there was no change in *HMI-LNCRNA* expression (**Supplemental Figure 3.4B**). This was probably due to the fact that these cells expressed relatively low levels of *HMI-LNCRNA* compared with HUDEP-2 cells, reducing any effect of shRNA knockdown. In further experiments, knockdown of *HMI-LNCRNA* was only done in HUDEP-2 cells.

3.2.3 The effect of knocking down HMI-LNCRNA on the expression of HBS1L and MYB

Since *HBS1L* and *MYB* are the two genes in proximity to *HMI-LNCRNA*, the mRNA levels of both genes in HUDEP-2 cells infected with HMI-lncRNA shRNA was assessed. qPCR analysis showed that there was no change in *HBS1L* and a significant 30% reduction in *MYB* compared with HUDEP-2 cells transduced with scrambled shRNA control, and this was also reflected in the protein level of c-MYB (**Figures 3.2A and 3.3A**). This change in *MYB* expression was reflected in one of two experiments. Pending additional studies, the regulation of *MYB* by *HMI-LNCRNA* is inconclusive.

3.2.4 Effects of knocking down HMI-LNCRNA on the expression of HBG and HBB

The expression pattern of both *HBG* and *HBB* among the naïve, scrambled shRNA and HMI-lncRNA shRNA groups were investigated by qPCR. Knocking down *HMI-LNCRNA* had no effect on *HBB* expression in one experiment and a 4-fold increase in expression in the second experiment. In contrast, there was a 200-fold increase in

HBG mRNA in HUDEP-2 cells transduced with HMI-lncRNA shRNA compared to both naïve and scrambled shRNA controls (**Figure 3.2A**). The percent *HBG* increased from less than 1% in naïve and scramble shRNA cells to more than 20% of the total *HBG* and *HBB* expression in cells with HMI-lncRNA shRNA (**Figure 3.2B**). Modulation of γ-globin expression was also seen at the protein level, with detectable HBG expression only found in cells with knockdown of *HMI-LNCRNA* (**Figure 3.3B**). These results were corroborated with immunofluorescent staining using anti-HBG antibody tagged with a secondary red fluorescent antibody. The high HBG expression appeared to be restricted to a small subpopulation of HUDEP-2 cells with knockdown of *HMI-LNCRNA*. There was no difference in protein levels of HBA and HBB between the three groups (**Supplemental Figure 3.5**). In another experiment, *HBG* mRNA was also significantly elevated 900-fold with knockdown of *HMI-LNCRNA* compared with naïve and scrambled shRNA controls, with about 20% *HBG* of total *HBG* and *HBB* mRNA. This effect was also present at the protein level.

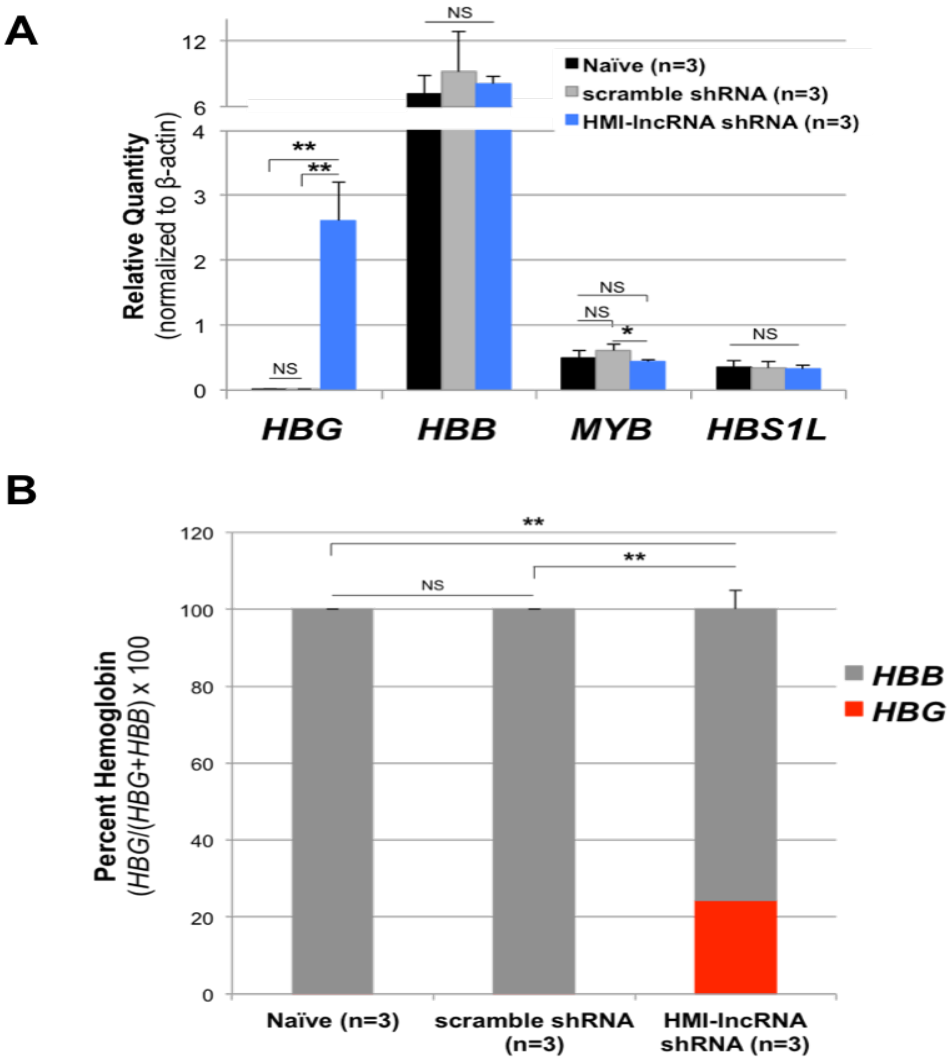


Figure 3.2. The effects of knocking down *HMI-LNCRNA* on the expression of *HBG*, *HBB*, *MYB* and *HBS1L* in HUDEP-2 cells.

(**A**) *HBG, HBB, MYB* and *HBS1L* transcript levels, and (**B**) percent *HBG* and *HBB* out of the total of both transcripts were measured in these same samples For qPCR analyses, *ACTB* was used as the endogenous control. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

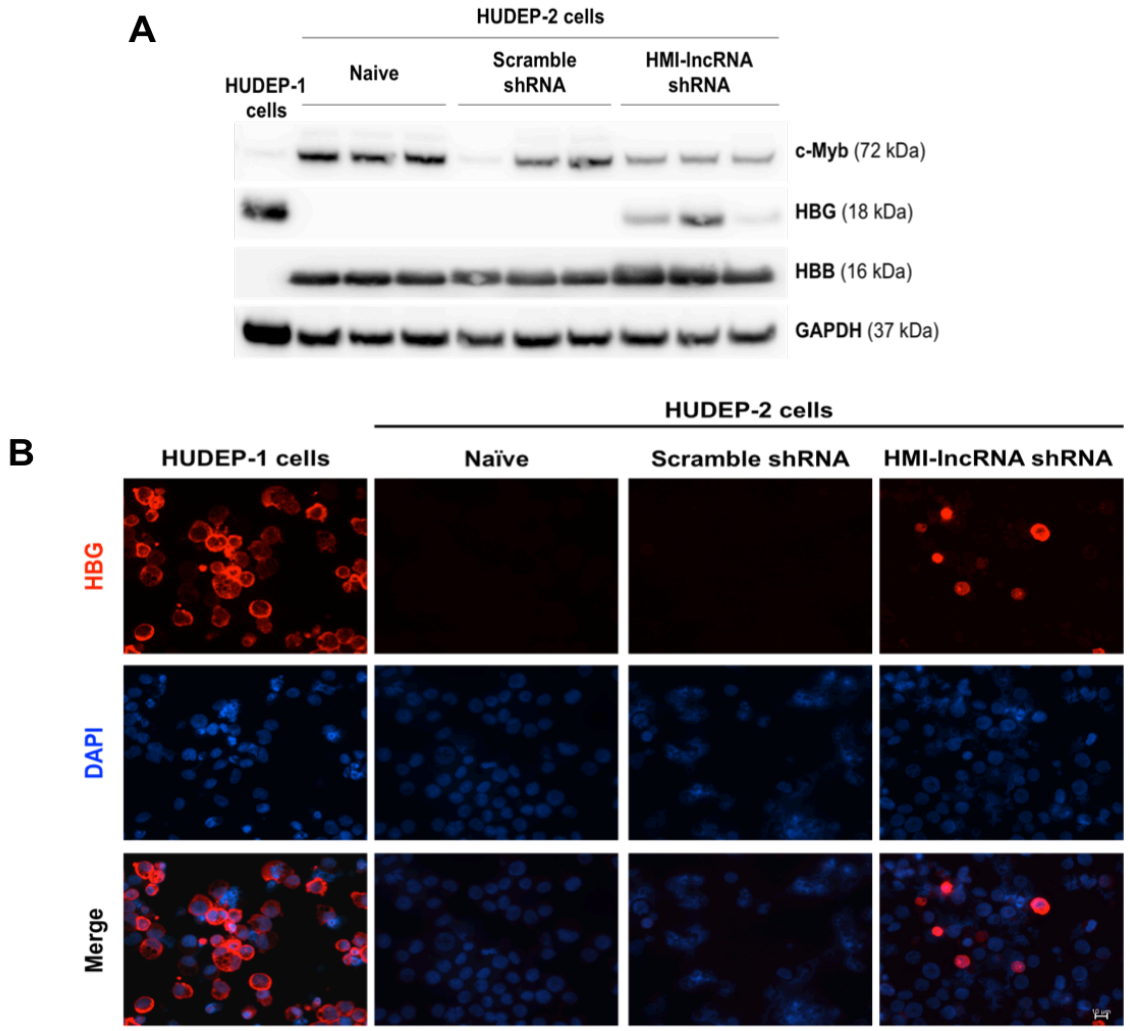


Figure 3.3. The effects of knocking down *HMI-LNCRNA* on the protein expression of c-MYB, HBG and HBB.

(**A**) Protein expression of c-MYB, HBG and HBB were analyzed by Western blot in naïve HUDEP-2 cells, and cells transduced with scrambled shRNA and HMI-lncRNA shRNA (HUDEP-1 cells were used as control). GAPDH was used as loading control. (**B**) Naïve HUDEP-2 cells, and cells transduced with scrambled and HMI-lncRNA shRNAs were stained with anti-HBG antibody, followed by secondary antibody labeled with Alexa Fluor-594 (in red) and DAPI to stain nuclei (in blue), and imaged with a fluorescent microscope at 40X magnification.

In addition to analyzing the effect of downregulating *HMI-LNCRNA* in non-differentiated HUDEP-2 cells, after the two-week selection for transduced cells with puromycin treatment, the cells were allowed to undergo erythroid maturation for up to seven days (**Figure 3.4**). *HBG* was significantly higher at Days 0, 5 and 7 in HUDEP-2 cells with *HMI-lncRNA* shRNA, and *HBB* was significantly higher at Day 0 and 7 with knockdown of *HMI-LNCRNA* (**Figure 3.5A**). *HBG* expression remained at about 20% of total *HBG* and *HBB* transcripts at each timepoint (**Figure 3.5B**). These observations were also seen at the protein level (**Supplemental Figures 3.6 and 3.7**).

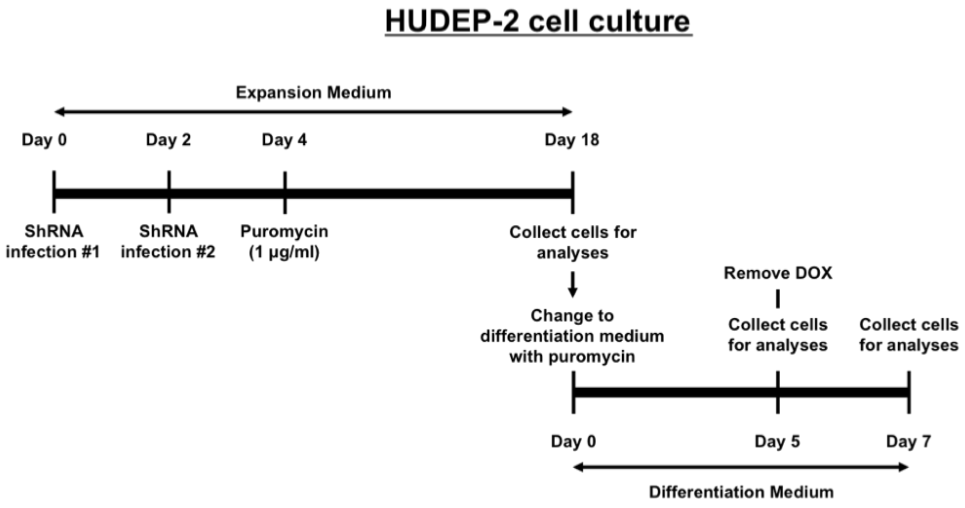


Figure 3.4. Timeline for infection and culture of HUDEP-2 cells.

Cells were maintained in expansion medium for 2 weeks after transduction, and then placed in differentiation medium for up to 7 days. Doxycycline (DOX) was removed at Day 5 to promote erythroid maturation.

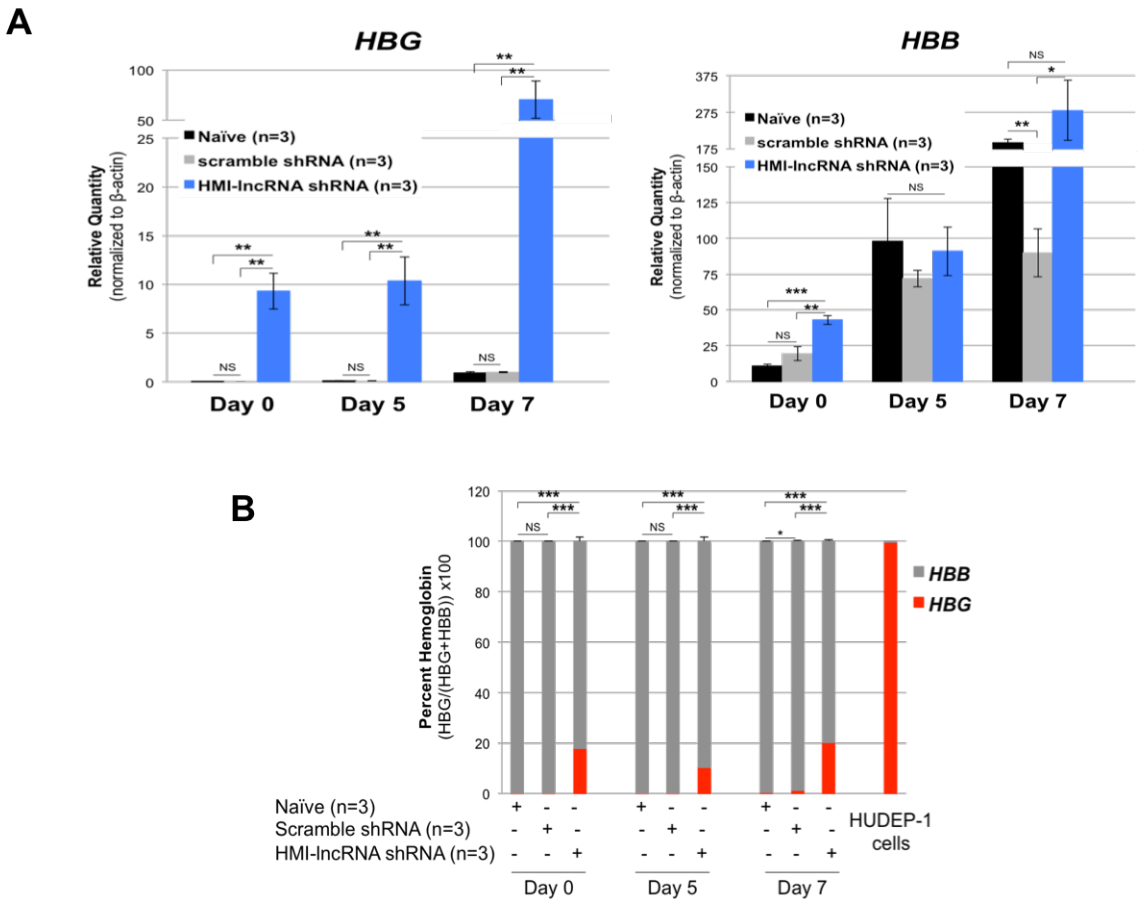


Figure 3.5. *HBG* and *HBB* expression during erythroid differentiation of HUDEP-2 cells with knockdown of *HMI-LNCRNA*.

Cells were maintained in expansion medium for 2 weeks after transduction, and then placed in differentiation medium for up to 7 days. Doxycycline (DOX) was removed at Day 5 to promote erythroid maturation. (**A**) Relative quantity for *HBG* and *HBB* transcripts were analyzed by qPCR in naïve HUDEP-2 cells (n=3), and cells transduced with scramble shRNA (n=3) and HMI-lncRNA shRNA (n=3) at Day 0, 5 and 7 of differentiation. (**B**) Percent hemoglobin of *HBG* to *HBB* was determined by qPCR. For all qPCR analyses, *ACTB* was used as the endogenous control. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

3.2.5 The effect of knocking down HMI-LNCRNA on the transcription factors that regulate the expression hemoglobin

As mentioned in Chapter 1, various transcriptions factors regulate hemoglobin gene expression. Since a significant change in γ-globin expression is observed with downregulation of *HMI-LNCRNA*, the expression pattern of a few of these transcription factors were studied in HUDEP-2 cells transduced with naïve and scrambled shRNA controls and HMI-lncRNA shRNA. The expression of *KLF1, BCL11A, ZBTB7A, CHD4, NR2C1, NR2C2* and *KDM1α*, all of which were known to downregulate *HBG*, were analyzed by qPCR. Expression of these genes was either unchanged or slightly upregulated in HUDEP-2 cells transduced with HMI-lncRNA shRNA compared to controls (**Figure 3.6**). This suggests that the significant increase in γ-globin gene expression when *HMI-LNCRNA* is reduced in HUDEP-2 cells is independent of the expression of the transcription factors analyzed by qPCR.

3.2.6 The effect of knocking down HMI-LNCRNA on erythroid maturation

HUDEP-2 cells are characterized as basophilic erythroblasts expressing both transferrin receptor (CD71), a cell surface marker for immature erythroblasts, and glycophorin A (CD235), a cell surface marker for more mature erythroblasts.[31] The percentage of cells positive for these markers were examined by flow cytometry in naïve cells, cells expressing scrambled shRNA and cells expressing HMI-lncRNA shRNA (**Figure 3.7A**). Within both control groups, approximately 55-65% of the cells were positive for both CD71 and CD235, while about 85% of cells transduced with HMI-lncRNA shRNA were positive for both markers, which was significantly higher than

control groups (**Figure 3.7B**). These observations suggest that downregulation of *HMI-LNCRNA* promotes maturation of HUDEP-2 cells.

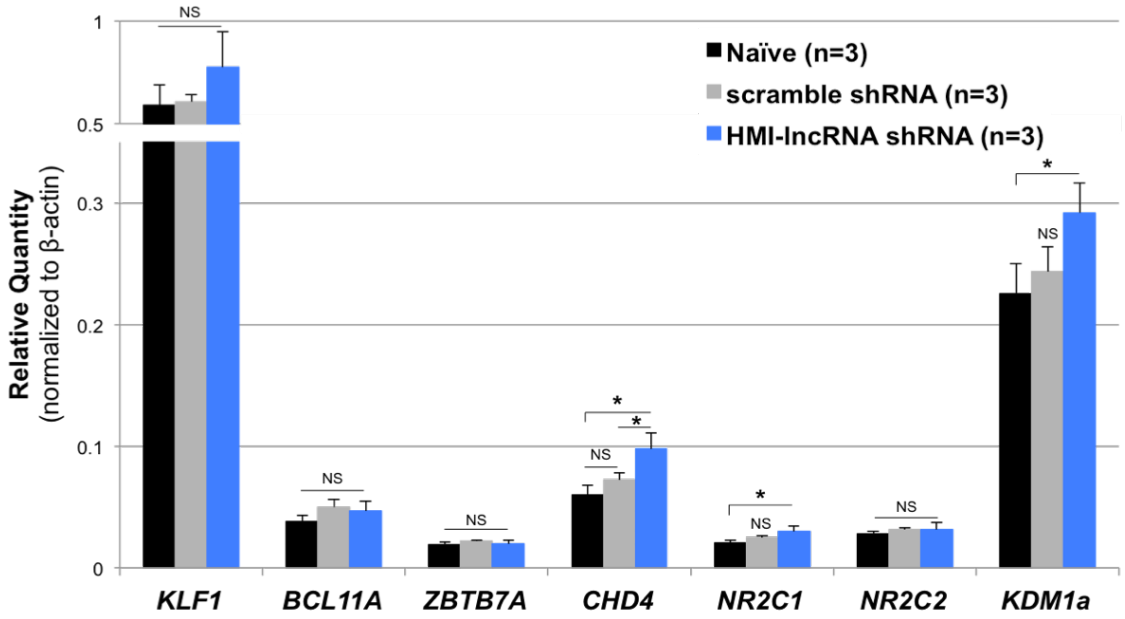


Figure 3.6. The effects of knocking down *HMI-LNCRNA* on erythroid-regulating transcription factors in HUDEP-2 cells.

Relative expression of *KLF1, BCL11A, ZBTB7A, CHD4, NR2C1, NR2C2* and *KDM1a* transcripts were analyzed by qPCR in naïve HUDEP-2 cells, and cells infected with scrambled shRNA and HMI-lncRNA shRNA. Beta actin was used as the endogenous control. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

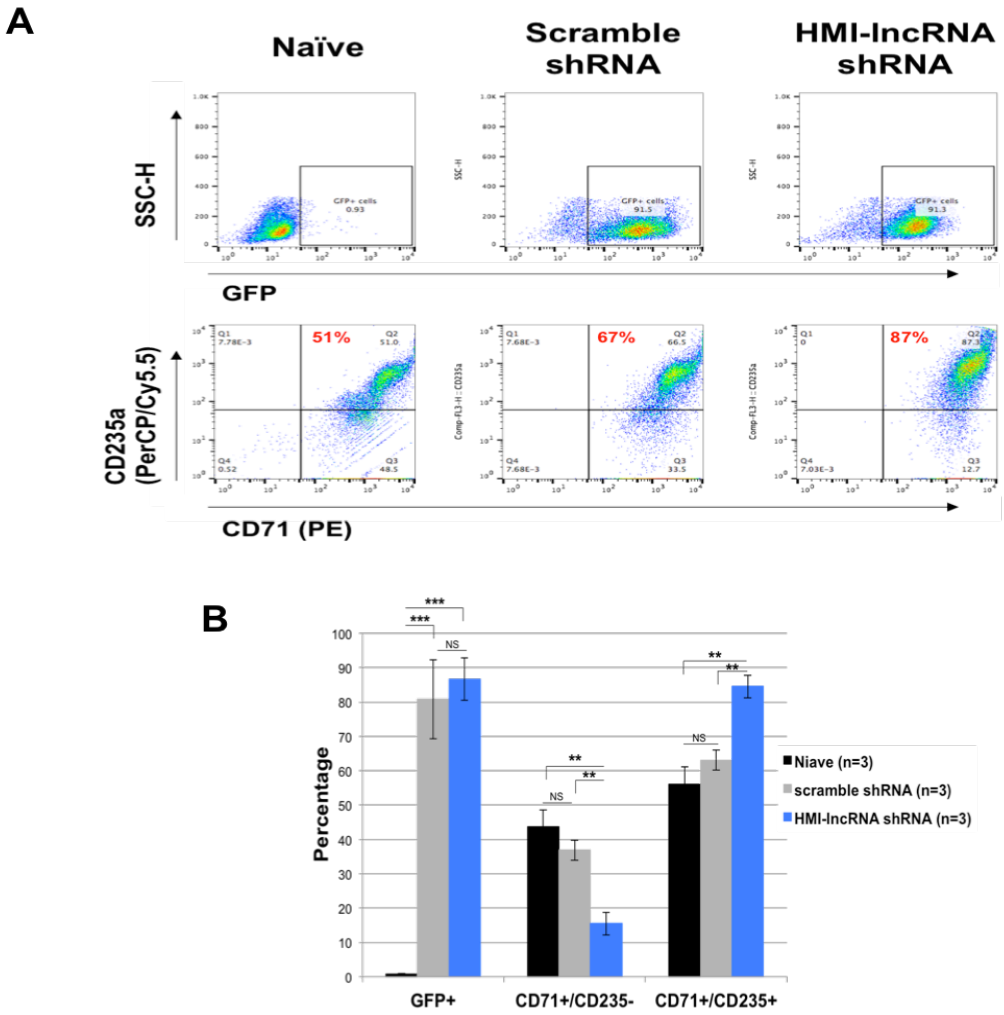


Figure 3.7. The effects of knocking down *HMI-LNCRNA* on erythroid maturation of HUDEP-2 cells.

(**A**) Naïve HUDEP-2 cells, and cells transduced with scrambled shRNA and HMI-lncRNA shRNA (cultured in expansion medium) were stained with PE-labeled transferrin receptor (CD71) and PerCP/Cy5.5-labeled glycophorin-A (CD235) antibodies, and analyzed by flow cytometry to discriminate between cells that are positive and negative for GFP, CD71 and CD235. Data was analyzed with FlowJo. (**B**) Average percent GFP+, CD71+/CD235- and CD71+/CD235+ for each group was plotted on a bar graph. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

3.3 Overexpression of *HMI-LNCRNA* in erythroid cells

Knocking down *HMI-LNCRNA* significantly increases expression of *HBG*. To ascertain the effects of overexpressing *HMI-LNCRNA*, full-length *HMI-LNCRNA* was cloned into pLVX-EF1a-IRES-ZsGreen1 expression lentivector (ZsGreen lentivector), which co-expresses ZsGreen1, a fluorescent protein that can be detected by flow cytometry. After successfully cloning *HMI-LNCRNA* in this ZsGreen lentivector (ZsGreen HMI-lncRNA), it was used to transduce HUDEP-1 cells, which, unlike HUDEP-2 cells, express mostly *HBG* and have lower expression of *HMI-LNCRNA*. Unlike pGreenPuro lentivector, ZsGreen lentivector contains no genomic cassette for antibiotic selection in cell culture. Therefore, two days after transduction, cells were sorted for ZsGreen-positive ($ZsGreen^{+}$) cells at the Flow Cytometry Core and cultured for two more weeks for analyses. Naïve cells and cells transduced with vector alone (ZsGreen) were used as controls. Similar to the knockdown experiments, the percent $ZsGreen^{+}$ cells was checked by flow cytometry and showed 0% for naïve, 84% for ZsGreen and 88% for ZsGreen HMI-lncRNA (**Supplemental Figure 3.7**). qPCR analysis for expression of *HMI-LNCRNA* showed over a 2000-fold increase in expression compared to naïve and ZsGreen cells allowing further analyses as most of the cells were transduced with either ZsGreen or ZsGreen HMI-lncRNA, and there was significant upregulation of *HMI-LNCRNA* (**Figure 3.8A**).

We hypothesized that when *HMI-LNCRNA* is overexpressed in HUDEP-1 cells, *MYB* would increase and *HBG* will decrease. However, among the three groups of cells—naïve, ZsGreen and ZsGreen HMI-lncRNA—there was no difference in *HBG,*

HBB, HBS1L or *MYB* (**Figure 3.8B**). It is not necessarily true that overexpressing a gene will give the opposite outcome to knocking down a gene. It is possible that other components that work with *HMI-LNCRNA* to regulate *HBG* expression are not expressed in HUDEP-1 cells, and therefore overexpressing *HMI-LNCRNA* will not show any effect.

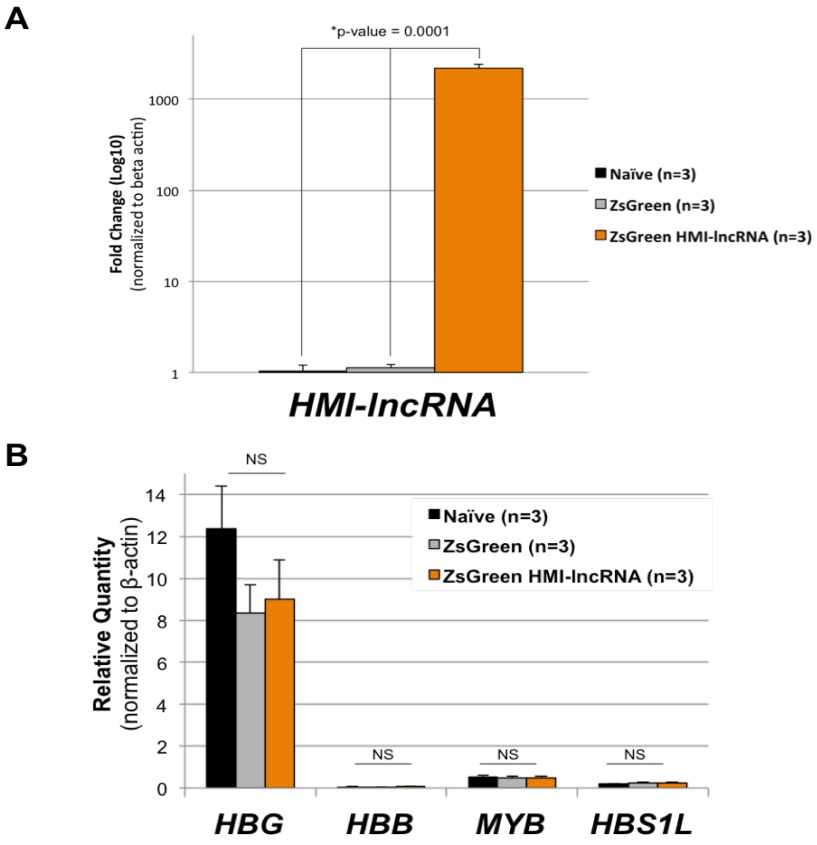


Figure 3.8. The effects of overexpressing *HMI-LNCRNA* on the expression of *HBG*, *HBB*, *MYB* and *HBS1L* in HUDEP-1 cells.

Expression levels of (**A**) *HMI-LNCRNA* and (**B**) *HBG, HBB, MYB* and *HBS1L* were determined by qPCR in HUDEP-1 cells that were not transduced (naïve), and transduced with either ZsGreen or ZsGreen HMI-lncRNA lentiviruses by qPCR. *ACTB* was used as the endogenous control. p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.

3.4 Summary and Conclusions

To determine the effect of modulating *HMI-LNCRNA* on neighboring protein-coding genes, *HBS1L* and *MYB*, and *HBG* and *HBB* in erythroid cells, we knocked down and overexpressed this lncRNA in HUDEP-2 and HUDEP-1 cells that produce primarily adult and fetal globins, respectively. To downregulate *HMI-LNCRNA*, HUDEP-2 cells were transduced with lentiviral particles that express an shRNA template targeting the lncRNA. After confirming successful transduction of cells and significant reduction of *HMI-LNCRNA*, we observed that these cells also displayed high upregulation of γ-globin at both the mRNA and protein levels, and a higher population of mature erythroid cells compared to cells in the naïve and scrambled shRNA control groups. Furthermore, there was consistent upregulation and percent *HBG* of total *HBG* and *HBB* during differentiation of these cells. Thus, these results suggest that *HMI-LNCRNA* regulates the expression of *HBG* and erythroid maturation (**Figure 3.9**). Overexpression *HMI-LNCRNA* in HUDEP-1 cells, despite a high level of cell transduction and increase in *HMI-LNCRNA*, was not accompanied by a change in the expression of *MYB* and globin genes.

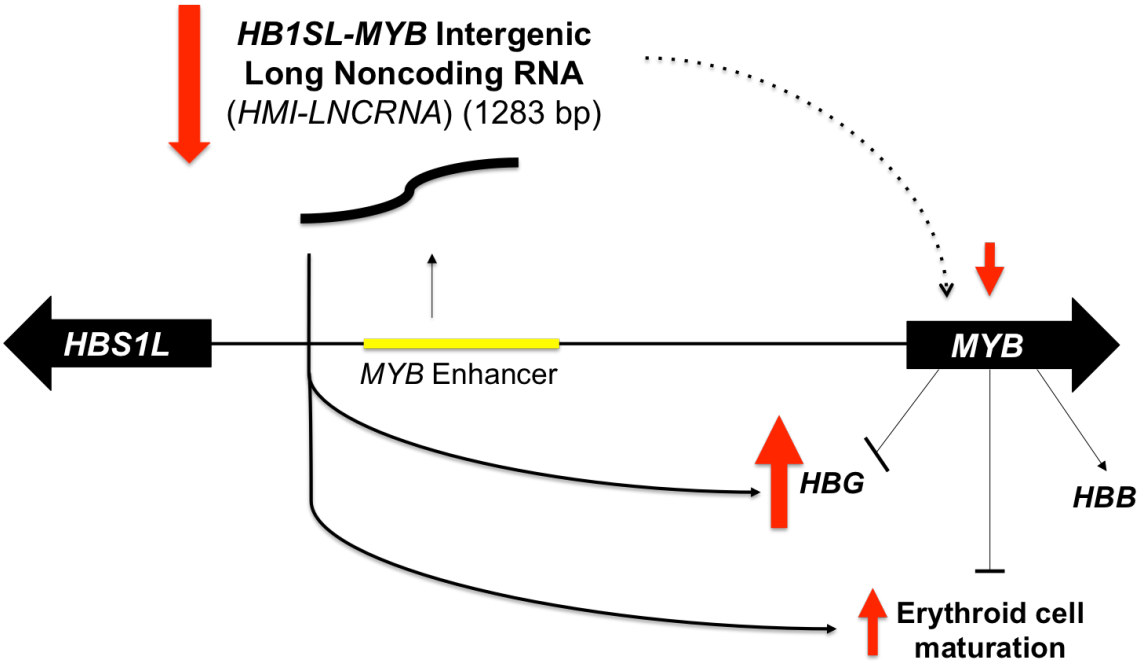


Figure 3.9. Model for the function of *HMI-LNCRNA*.

HMI-LNCRNA is expressed from the enhancer region of *MYB*. Downregulation of *HMI-LNCRNA* increased the expression of *HBG* and promoted erythroid maturation. Further investigation is required to determine if *HMI-LNCRNA* also regulates the expression of *MYB*.

3.5 Materials and Methods

3.5.1 Development of shRNA plasmids for knockdown of HMI-LNCRNA

Three different shRNA templates were made to target three different areas of *HMI-LNCRNA*. This was accomplished by using Block-iT™ RNAi Designer (Life Technologies), which analyzed the genomic region of *HMI-LNCRNA* and designed shRNA templates that are ranked based on the likelihood that it is effective. I used the top three ranking shRNA templates and labeled them *HBS1L-MYB* Intergenic Transcript 1 (HMIT-1), HMIT-2 and HMIT-3. Top and bottom strands of each shRNA were synthetically made (**Table 3.2**).

Table 3.1. shRNA templates for *HMI-LNCRNA*

shRNA Template		
HMIT-1	**Top strand**	5'- GATCCGCTAGTATGTGAAGCACTTAGCTTCCTGTC AGACTAAGTGCTTCACATACTAGCTTTTTG-3'
	Bottom Strand	5'- AATTCAAAAAGCTAGTATGTGAAGCACTTAGTCTG ACAGGAAGCTAAGTGCTTCACATACTAGCG-3'
	Location (ch6, hg38)	135096650-135096670
HMIT-2	**Top strand**	5'- GATCCGCCGCTGCCTTTAATTGATGTCTTCCTGTCA GAACATCAATTAAAGGCAGCGGCTTTTTG-3'
	Bottom Strand	5'- AATTCAAAAAGCCGCTGCCTTTAATTGATGTTCTG ACAGGAAGACATCAATTAAAGGCAGCGGCG-3'
	Location (ch6, hg38)	135096889-135096909
HMIT-3	**Top strand**	5'- GATCCGCACCATCATGGTTTCAGTGCCTTCCTGTCA GAGCACTGAAACCATGATGGTGCTTTTTG-3'
	Bottom Strand	5'- AATTCAAAAAGCACCATCATGGTTTCAGTGCTCTG ACAGGAAGGCACTGAAACCATGATGGTGCG-3'
	Location (ch6, hg38)	135097331-135097351

Each shRNA template contained the following regions: sense, loop, antisense and terminator. Also, restriction sites for BamHI and EcoRI were placed on the 5'-end of the top strand and the bottom strand, respectively, to ligate shRNA templates to the

pGreenPuro expression lentivector (System Biosciences). Next, the top strand and bottom strand of each shRNA template were annealed to make double-stranded oligonucleotides by incubating both strands at 95°C in annealing buffer (100 mM Tris-HCL, 10 mM EDTA, 1 M NaCl and H_2O) for 4 minutes. The end products are double-stranded shRNA templates with restriction sites overhangs.

We next ligated each shRNA individually to pGreenPuro shRNA expression lentivector. pGreenPuro is an HIV-based expression lentivector with the following attributes: H1 promoter that is bound by RNA polymerase III for transcription of shRNA; EF1α promoter that expresses both GFP and puromycin-N-acetyl transferase in transduced cells; RSV-5'LTR to express high levels of lentiviral particles in 293T cells; cPPT, GAG and LTRs, which are required for packaging, transduction and integration of viral components; ampicillin-resistance gene for selection in competent cells; and SV40 polyadenylation, which ensures termination of transcription. pGreenPuro lentivector was linearized by restriction digest with both BamHI and EcoRI, and then ligated to each of the double-stranded shRNA templates. After ligation, products were transformed into competent E. coli cells, grown overnight on LB plates treated with ampicillin, and colonies were selected and screened further by PCR to identify which clones had the shRNA inserts. The following primers were used for PCR: Forward (5'-AATGTCTTTGGATTTGGGAATCTTAT-3') and Reverse (5'-TGGTCTAACCAGAGAGACCCAGTA-3'), which would amplify about a 180 bp product if shRNA insert is present, and a 110 bp product is insert is not present. After positive clones were identified, each shRNA insert was confirmed by DNA sequencing.

pGreenPuro scramble shRNA (purchased from vendor, System Biosciences) was used as the negative control.

3.5.2 Development of plasmid to overexpress HMI-LNCRNA

Full-length *HMI-LNCRNA* was synthesized by IDT with an XbaI restriction site on the 5'-end and a BamHI restriction site on the 3'-end, and inserted into their pUCIDT-AMP vector. We next cloned *HMI-LNCRNA* into the pLVX-ER1a-IRES-ZsGreen1 (ZsGreen) expression lentivector. ZsGreen is an HIV-based expression lentivector that co-expresses ZsGreen, which is another green fluorescent protein. ZsGreen have the same components as pGreenPuro that are necessary to cloning, lentivirus production and integration, except there is so antibiotic cassette to select for infected cells in culture. Next, both pUCIDT-AMP with *HMI-LNCRNA* insert and the ZsGreen underwent restriction digest with XbaI and BamHI to isolate and purify *HMI-LNCRNA* insert and to linearize ZsGreen. Following digest, *HMI-LNCRNA* was ligated to ZsGreen, transformed into competent E. coli cells and cultured overnight on LB plates treated with ampicillin, and colonies were selected and screened further by restriction digest to identify which clones had the *HMI-LNCRNA* insert. DNA sequencing was utilized to confirm insert-positive clones, which was labeled as ZsGreen HMI-lncRNA. The ZsGreen vector alone was used as a negative control.

3.5.3 TF-1a cells

TF-1a cells are immortalized erythroblasts derived from the bone marrow of a patient with erythroleukemia.[135] Cells are cultured at 37°C in RPMI medium containing 10% FBS and 2% penicillin/streptomycin.

3.5.4 HUDEP cells

Human umbilical cord blood-derived erythroid progenitor (HUDEP) cells are immortalized erythroid cell lines derived from cord blood $CD34^+$ mononuclear cells.[133] HUDEP-1 and HUDEP-2 cells were cultured at 37°C maintained in StemSpan SFEM medium (StemCell Technologies) supplemented with SCF (50 ng/ml, Invitrogen), EPO (3 U/ml, Invitrogen), dexamethasone (1 μM, Sigma), doxycycline (1 μg/ml, Clontech), L-glutamine (1%, Life Technologies) and penicillin/streptomycin (2%, Life Technologies). For erythroid maturation, cells were cultured at 37°C in IMDM medium (Invitrogen) supplemented with heat inactivated human serum from human male AB plasma (5%, Sigma), EPO (3 U/ml, Invitrogen), insulin (10 μg/ml, Sigma), doxycycline (1 μg/ml, Clontech), holo-transferrin (500 μg/ml, Sigma), heparin (3 U/ml, Sigma), SCF (100 ng/ml, Invitrogen), L-glutamine (1%, Life Technologies) and penicillin/streptomycin (2%, Life Technologies) for 5 days. For further erythroid maturation, doxycycline was removed and cells were cultured for two more days.

3.5.5 Quantitative PCR

qPCR, also known as real-time PCR, is a tool used to quantitate the relative transcript levels of a specific RNA. In the PCR phase, gene-specific primers are used to

amplify the target during a 40-cycle reaction period (1-40). More abundant transcripts are amplified at earlier cycle thresholds (Ct) values, while less abundant transcripts are amplified at later Ct values. To calculate the relative quantity of a transcript-of-interest (TOI), we take difference between Ct of TOI and Ct of endogenous control, such as *ACTB* (β-actin), to get ΔCt ($Ct^{TOI} - Ct^{ACTB} = \Delta Ct^{TOI}$). Finally, the relative quantity of TOI is calculated using $2^{(-(\Delta Ct))}$. To calculate fold change between two samples—sample 1 (S1) and sample 2 (S2)—another approach is taken. For each sample, we calculate ΔCt^{TOI} followed by calculating $\Delta\Delta Ct^{TOI}$ for each sample—for S1, [(ΔCt^{TOI} for S1) – (ΔCt^{TOI} for S1)] = ($\Delta\Delta Ct^{TOI}$ for S1) = 0; and for S2, [(ΔCt^{TOI} for S2) – (ΔCt^{TOI} for S1)] = ($\Delta\Delta Ct^{TOI}$ for S2). The fold change of the TOI for each sample is calculated using $2^{(-(\Delta\Delta Ct))}$, in which S1 will equal 1.

Total RNA, TaqMan primers and reagents from the TaqMan RNA-to-CT 1-Step Kit (Applied Biosystems) were used to make 20 µl reaction mix for each sample (done in triplicate) that was loaded onto a 96-well plate, following manufacturer's instructions. The following TaqMan gene expression assays (Applied Biosystems) were used: *HBG1/2* (Hs00361131_g1), *HBB* (Hs00758889_s1), *MYB* (Hs00920556_m1), *HBS1L* (Hs04188641_g1), *HMI-LNCRNA* (custom TaqMan assay designed by Applied Biosystems to target genome position chr6: 135096354-135097644, hg38; assay ID number AJI1MTQ), *BCL11A* (Hs01093197_m1), *CHD4* (Hs00172349_m1), *KLF1* (Hs00610592_m1), *ZBTB7A* (Hs00252415_s1), *NR2C1* (Hs00915957_m1), *NR2C2* (Hs00991824_m1), *KDM1a* (Hs01002741_m1) and *ACTB* (Hs01060665_g1). QPCR

reactions were ran on a StepOne Plus qPCR machine (Applied Biosystems). *ACTB* was used as the endogenous control.

3.5.6 293T cells

293T cells are a sub-line of 293 cells, which are immortalized human embryonic kidney cells. 293T cells are competent for replicating plasmids carrying SV40 origin of replication, and are used to produce high titers of pseudoviral particles such as lentivirus. They are cultured at 37°C in DMEM supplemented with 10% FBS and 2% penicillin/streptomycin.

3.5.7 HT1080 cells

HT1080 cells are derived from human fibrosarcoma, and are cultured at 37°C in DMEM supplemented with 10% FBS and 2% penicillin/streptomycin. HT1080 cells can efficiently be transduced with lentiviruses, and therefore are used to determine viral titer.

3.5.8 Generation of lentiviral particles

For shRNA experiments, 293T cells were transfected with scrambled shRNA and HMI-lncRNA shRNA using the LentiStarter 2.0 kit (System Biosciences), which uses pPACKH1 lentivector packaging system, following manufacturer's instructions, to generate lentiviral particles. For overexpression experiments, 293 T cells were transfected with either ZsGreen vector or ZsGreen HMI-lncRNA plasmid using the Lenti-X Packaging Single Shots, following manufacturer's instructions, to generate lentiviral particles.

Viral titer was determined using HT1080 cells—$1x10^5$ cells were transduced with various amount of lentivirus using polybrene, and GFP-positivity was determined 48 hours later by flow cytometry. Transducing units per mL were determined using the following equation: (($1x10^5$ x %GFP-positive cells) x 1000)/Lentivirus volume (μL).

3.5.9 Lentivirus transduction

For shRNA experiment, HUDEP-1 and HUDEP-2 cells were transduced at an multiplicity of infection (MOI) 50 with 5 μg/mL polybrene, and centrifuged at room temperature for 30 minutes at 1250 x g, and repeated again 48 hours later. 48 hours post-transduction, cells were cultured in expansion medium with 1 ug/mL puromycin for up to two weeks to select for transduced cells. After 2 weeks, FACS for GFP expression was done to ensure most cells were GFP-positive, and therefore express the shRNA plasmids. For overexpression experiments, HUDEP-1 cells were transduced with either ZsGreen or ZsGreen HMI-lncRNA at an MOI of 50 with 5 μg/mL polybrene, and centrifuge at room temperature for 30 minutes at 1250 x g. Two days after transduction, cells were collected and sorted for GFP^+ cells using flow cytometry. After sorting, cells were cultured in expansion medium for up to two weeks to provide sufficient cells for analyses. After 2 weeks, FACS for ZsGreen expression was done to ensure most cells were infected.

3.5.10 Cell surface staining for FACS

$5x10^5$ cells were harvested and washed in FACS buffer. Cells were resuspended in 100 μl of FACS buffer with antibody, and incubated on ice for 30 minutes. Next, cells were washed twice in FACS buffer by centrifugation at 300 x g for 5 minutes each.

Finally, cells were fixed in IC Fixation buffer (eBioscience) for 30 minutes before analyzing the cells using the BD FACScan. The following antibodies were used: CD71-PE (334105, BioLegend) and CD235-PerCP/Cy5.5 (306613, BioLegend).

3.5.11 Western blot analysis

Culture cells were centrifuged at 300 x g for 10 minutes at 4°C, washed in cold 1X PBS and centrifuged again at 300 x g for 10 minutes at 4°C. Cell pellets were suspended in Roche lysis buffer (protease inhibitor, 0.3% NP40, 10% glycerine, 2 mM EDTA, 246 mM NaCl, 10% phosphatase inhibitor, PBS and water), placed on ice for 1 hour and centrifuged at 14,500 rpm for 15 minutes at 4°C to extract protein.

For Western blot analysis, protein lysates (10 μg) were denatured with lithium dodecyl sulfate (LDS; Life Technologies), reduced in sample reducing agent (Life Technologies) and incubated at 70°C for 10 minutes. Samples were ran on 4-12% Bis-Tris protein gel (Life Technologies) in 1X MED SDS running buffer and 1 ml NuPAGE antioxidant (Life Technologies) at 200 volts for 40 minutes. Protein from the Bis-Tris gel was transferred to a polyvinylidene difluoride (PVDF) membrane (Bio-Rad) in transfer buffer (2X NuPAGE transfer buffer, 10% methanol, 1 ml NuPAGE antioxidant and H_2O) at 100 volts for 1 hour in a transfer cell kept cold with an ice package and ice. After transfer, membranes were immersed in Ponceau S solution (Sigma) to determine if transfer was successful and to cut membrane evenly to probe with different antibodies that recognize proteins of different molecular weights. The following steps were taken to probe membrane with antibodies (all done on a rocker at 5-6 rpm at room temperature): wash in 1X PBS-tween 3 times for 5 minutes each; block in 1X PBS-tween with 5% milk

for 1 hour; wash in 1X PBS-tween; probe in primary antibody in 1X PBS with 5% milk for 2 hours; wash in 1X PBS-tween 3 times for 5 minutes each; probe in secondary antibody in 1X PBS-tween for 1 hour; and wash in 1X PBS-tween 3 times for 5 minutes each. Last, the membranes were treated with enhanced chemiluminescence (ECL) substrate for 5 minutes, and then exposed by chemiluminescence using a LAS-4000 Luminescent Image Analyzer at various exposure times. The following antibodies were used: c-Myb (ab109127, Abcam), hemoglobin β (16216-1-AP, Proteintech), hemoglobin γ (sc-21756, Santa Cruz Biotechnology) and GAPDH (sc-47724, Santa Cruz Biotechnology). For densitometry, ImageJ software was used to analyze western blots to determine percent intensity of each band, which was used to calculate the relative densities. Relative densities for c-Myb, hemoglobin beta and hemoglobin gamma were adjusted by GAPDH.

3.5.12 Slide preparation for microscopy

Cells were rinsed in PBS and resuspended in 2% BSA (in PBS). 200 μL of cell suspension (5-7x10^4 cells) were added to a cytology funnel in a slide holder and centrifuged in a cytospin at 800 rpm for 5 minutes. Cells were allowed to dry on slide for at least 30 minutes, and then fixed in 100% methanol for 10 minutes. Slides were stored at 4°C.

3.5.13 Wright-Giemsa staining

Fixed cells on slides were placed in diluted (1:3) Wright-Giemsa Stain (RICCA Chemical) for up to 1 hour. Slides were rinsed, dried and mounted with Permount

(Fisher). Images were taken in bright field with the Eclipse Ci (Nikon) microscope with a DS-Fi-U3 color-type (Nikon) camera, using NIS-Elements Br Imaging software (Nikon).

3.5.14 Immunofluorescence

Fixed cells were washed in PBS, blocked in 5% BSA (in PBS) and incubated in primary antibody for 1 hour. Cells were washed, blocked in 5% normal goat serum and incubated in secondary fluorescent-labeled antibody for 30 minutes. Finally, cells were washed and covered with mounting medium containing DAPI. Images were taken with the Eclipse Ci (Nikon) microscope with a DS-Qi1Mc-03 mono-type (Nikon) camera, using NIS-Elements Br Imaging software (Nikon). The following antibodies were used: Hemoglobin γ (sc-21756, Santa Cruz), Monoclonal Mouse Anti-human β-globin (a gift from IsoLab), Anti-human alpha globin (Wallac Inc.), and Alexa Fluor-594-conjugated AffiniPure F(ab)2 Fragment Goat Anti-mouse IgG (H+L) (115-586-062, Jackson ImmunoResearch).

CHAPTER 4. DISCUSSION

4.1 The *HBS1L-MYB* Intergenic Long Noncoding RNA (*HMI-LNCRNA*)

The *HBS1L-MYB* intergenic region on chr6q23 is important for regulating erythroid-specific traits. It contains an enhancer for the downstream gene *MYB* and binding sites for erythroid-specific transcription factors, which all bind to the promoter of *MYB* to increase its transcription. Furthermore, it is also known that upregulation of *MYB* increases erythroid cell proliferation, inhibits erythroid cell maturation, decreases *HBG* and increases *HBB* expression. Consequently, polymorphisms found within this region are associated with high HbF because they disrupt the enhancer-promoter interaction, thus reducing *MYB* expression and increasing expression of *HBG* and HbF. This doctoral work adds another element to this intergenic region, a novel 1283 bp lncRNA (chr6: 135096362-135097644, hg38) that is transcribed from the enhancer of *MYB*, and contains the HbF-associated polymorphic region rs66650371. Further characterization of *HMI-LNCRNA* revealed that it is not erythroid-specific but might be hematopoietic-specific; is located in the nucleus; is expressed at very low levels compared to the protein-coding gene *MYB*, and it is expressed at significantly higher levels in erythroid cells expressing *HBB* compared to cells that expressing *HBG*. The expression pattern of *HMI-LNCRNA* in these cells is similar to that of *MYB*.

Since *HMI-LNCRNA* is transcribed from an enhancer regulatory element for *MYB*, this suggests that it functions as an eRNA to support and promote binding of the erythroid-specific transcription factors to the promoter of *MYB*. However, additional experimentation will have to be done to determine if *HMI-LNCRNA* does work in

complex with these transcription factors. One method that can be used is chromatin isolation by RNA purification sequencing (ChIRP-seq).[136] ChIRP-seq is a relatively new approach to determine which genomic regions and proteins bind to a specific lncRNA. It crosslinks genomic DNA and proteins that are bound to lncRNA, adds biotinylated tilling oligonucleotides to bind to the lncRNA of interest, and then purifies out the complex by using magnetic beads to pull out the biotinylated oligonucleotides that are attached to the lncRNA. After purification, the complex can be dissociated and further analyses can be done to determine which genomic region and proteins bound to the lncRNA. If *HMI-LNCRNA* does function as an eRNA to promote the expression of *MYB*, we should find that it is associated with the promoter of *MYB*, and binds to TAL1/E47, RUNX1, GATA, LDB1 and KLF1. It is also possible for *HMI-LNCRNA* to regulate other protein-coding genes either in *cis* or *trans*, which ChIRP-seq will be able to elucidate.

We observed that the expression pattern of *HMI-LNCRNA* decreases during erythroid differentiation, as does the expression pattern of *MYB*. It will also be informative to determine the expression pattern of *HMI-LNCRNA* during the earlier phase of hematopoiesis, such as the HSCs, the common myeloid progenitors (CMPs) and the common lymphoid progenitors (CLPs) to determine if expression differs from that of erythroblasts. Since the expression of *MYB* is higher in HSCs, CMPs and CLPs, I would hypothesize that *HMI-LNCRNA* will be higher in these cells as well.[27] Additionally, *MYB* is not expressed during primitive erythropoiesis.[2] Therefore, it will also be of interest to determine if *HMI-LNCRNA* is not expressed in primitive erythroid cells.

4.2 *HMI-LNCRNA* and HbF

Increasing HbF levels in individuals with SCD is a major goal of clinical treatment. Hydroxyurea, which was the first FDA-approved drug for SCD, blood transfusions and gene therapy, are either not effective in all patients or financially and physically taxing. Therefore, it is vital to better understand how hemoglobin expression can be regulated in order to lead the field to more effective therapeutic treatments that will promote the expression of HbF. Various transcription factors such as BCL11A, KLF1 and LRF have been discovered to regulate HbF, and are considered possible therapeutic targets.[31,137,138] Once thought of as "junk" RNA, lncRNAs are now considered to be important regulators of cellular processes, biomarkers for disease prognosis and possible therapeutic targets for disease treatment. This provides a rationale for further study of lncRNAs and erythroid ontogeny. Published data implicate lncRNAs as important regulators of hematopoiesis and erythropoiesis, however, before my studies, none have been implicated as regulators of hemoglobin expression. Therefore, it was very interesting and exciting to see that when *HMI-LNCRNA* was downregulated in HUDEP-2 cells, γ-globin expression increased both at the mRNA and protein levels. HUDEP-2 cells have been used to study the regulation of hemoglobin expression and erythroid maturation.[31,137] However, additional functional analyses will be necessary in other cell culture systems, such as induced pluripotent stem cells (iPSCs)[139] and primary erythroid cells derived from $CD34^+$ mononuclear cells, to ensure these results can be replicated in cells of different genetic backgrounds and normal cells directly from human tissue. Also, we can determine if the upregulation of *HBG* provides sufficient HbF to

inhibit polymerization of sickle hemoglobin after knockdown of *HMI-LNCRNA*, which will clinically benefit patients with SCD.

As mentioned in Chapter 3, gene editing is a tool used to stop transcription by removing genomic regions where genes are transcribed. In early work, investigators used gene editing to remove lncRNAs in order to determine their function. However, it has been discovered that removal of these genomic regions, which are mostly in intergenic regions, not only inhibits transcription of the lncRNA, but also removes regulatory regions that regulate the expression of neighboring genes. Paralkar et al determined if the lncRNA, *Lockd*, actually does regulate the transcription of its neighboring gene, *Cdkn1b*, as was suggested in the literature.[140] Therefore, gene editing was used in two different experiments to delete the entire *Lockd* gene, and to stop transcription of *Lockd* by placing a polyadenylation signal at its transcription start site. When *Lockd* is deleted from the genome, there was low expression of *Cdkn1b*. However, when a polyadenylation signal was inserted at the transcription start site of *Lockd*, thus inhibiting its transcription, *Cdkn1b* was transcribed normally. They further showed that *Lockd* resided within an enhancer region for *Cdkn1b*, and when the region was deleted, it reduced expression of this gene. However, if the genomic region was left basically intact and only the transcription of the lncRNA was inhibited by inserting a polyadenylation signal, there was no change in *Cdkn1b* expression. This suggested that *Lockd* did not regulate *Cdkn1b* and that previously published work was likely to be incorrect. A similar approach would have to be taken if we stop transcription of *HMI-LNCRNA*. Since shRNA does not

entirely knockdown *HMI-LNCRNA*, and since gene editing can fully curtail its expression, without residual lncRNA, a significant downregulation of *MYB* might occur.

4.3 Generating a Mouse Model for *HMI-LNCRNA*

If downregulation of *HMI-LNCRNA* can be confirmed to significantly increase HbF in primary erythroid cells to levels that could be therapeutically useful, the next step to consider is developing a mouse model for this lncRNA. Investigators have used various methods to determine the functionality of lncRNAs via mouse models that are similar to mouse models of protein coding genes. Xenograft models have been useful in determining if neoplastic cells with a specific genetic modification have the ability to grow and/or metastasize in immunocompromised mice.[107] Many lncRNAs have been discovered to be involved in the pathophysiology of specific cancers, and therefore this approach is frequently used. Another methodology is treating mice with antisense oligonucleotides (ASOs) that will bind to and thus reduce the amount of the specific lncRNA targeted.[141] If ASOs were used for *HMI-LNCRNA*, this could be a preclinical model, which could possibly translate to the clinic as an approach for drug development for patients with SCD. Transgenic mouse models also have been useful for studying disease pathophysiology and treatment.[142]

To study *HMI-LNCRNA in vivo*, there are two approaches to consider, and both include using the sickle cell transgenic mouse model.[143] The first approach is to treat the SCD mice with ASOs for *hmi-lncrna*, and determine if HbF increases. The second approach is to make an *hmi-lncrna* knockout mouse and cross it with an SCD mouse, and

determine if HbF increases in these mice, and if it is enough to inhibit polymerization of HbS.

There are a couple factors to consider before developing a mouse model for *hmi-lncrna*, and the first requires that it be expressed in the mouse. There is a lncRNA annotated in the mouse genome (Gm33728; chr10: 21228529-21252242; mm10) that is 23.7 kb in length. Human *HMI-LNCRNA* maps to an intron of Gm33728, and therefore they do not correspond to each other. However, it does not mean they do not have similar functions. There is no published data on Gm33728, therefore it will be of interest to determine if Gm33728 can regulate the expression of *myb* and hemoglobin as well. Another factor to consider is murine hemoglobin expression. Mice do not express HbF, but they do have embryonic hemoglobin, and it will be interesting to determine if Gm33728 regulates its expression.[143]

4.4 *HMI-LNCRNA* as a Therapeutic Target

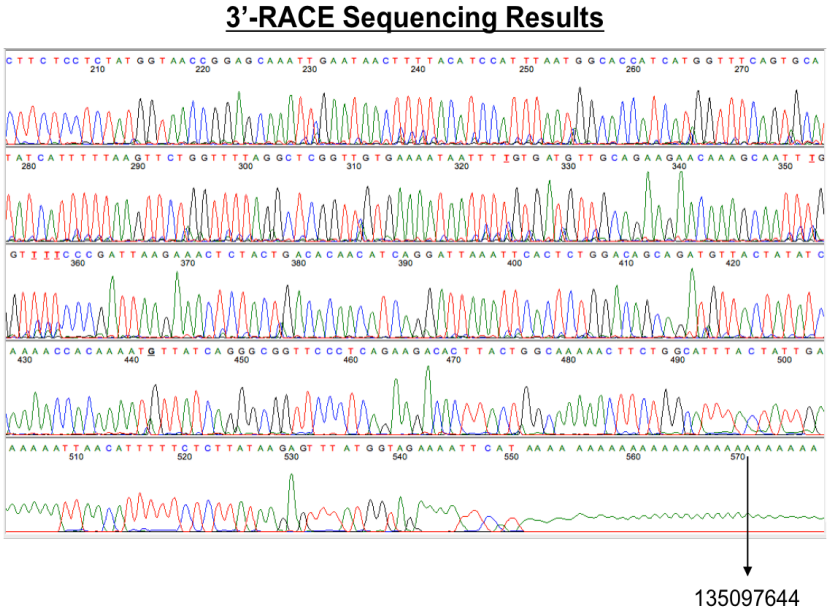
The severity of symptoms related to SCD is correlated not only to whether a patient express high levels of HbF, but also the HbF concentration in sickle erythrocytes and the distribution of HbF amongst F-cells.[144] Therefore, it is important to know if reduction of *HMI-LNCRNA* will increase HbF to levels that are beneficial to SCD patients. Knockdown of *HMI-LNCRNA* in HUDEP-2 cells resulted in an increase of *HBG* to 20% of the total *HBG* and *HBB* transcripts. Furthermore, this led to detectable protein expression of HBG. But will that ameliorate symptoms of SCD? HUDEP-2 cells also displayed a heterocellular expression pattern of HBG when *HMI-LNCRNA* was downregulated, which reflected a phenotype found in patients with $(\delta\beta)^0$-thalassemia. As

mentioned in Chapter 1 (subsection 1.3.1), $(\delta\beta)^0$-thalassemia and HPFH vary both in the concentration and distribution of HbF among RBCs.[70] These variations are the reason why patients with $(\delta\beta)^0$-thalassemia are symptomatic and patients with HPFH are not. Therefore, the heterocellular distribution of HbF found in HUDEP-2 cells when *HMI-LNCRNA* was downregulated is not ideal when considering it as a therapeutic target. However, there were factors in the experiment that would have caused this uneven expression pattern of HbF. Transduction of HUDEP-2 cells with lentiviral particles is not homogenous. Therefore, not every cell expressed the same copy number of shRNA template. This suggests that cells with more copies of HMI-lncRNA shRNA expressed more HBG than cells with fewer copies. One way to address this issue is to undertake clonal expansion from a single cell after transduction.

4.5 Summary

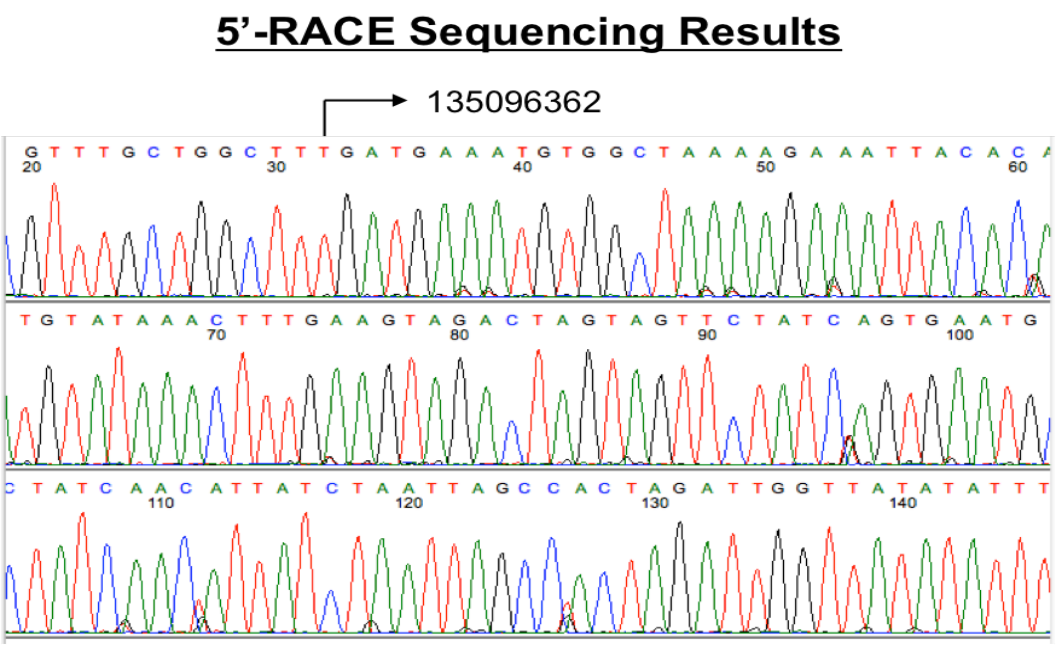
The specific aims of my doctoral thesis was to characterize a transcript found within the *HBS1L-MYB* intergenic region on chr6q23, and to determine its function as it relates to the expression of *MYB* and hemoglobin. I discovered a novel 1283 bp lncRNA, named the *HBS1L-MYB* intergenic lncRNA (*HMI-LNCRNA*), which is transcribed from an intergenic region important for the regulation of erythroid-associated traits, and has been found to significantly increase γ-globin expression at both the mRNA and protein levels when downregulated by shRNA. Furthermore, due to this remarkable increase in γ-globin expression, *HMI-LNCRNA* can be considered as a possible therapeutic target for individuals with SCD and β-thalassemia where increased HbF is clinically beneficial.

SUPPLEMENTAL DATA



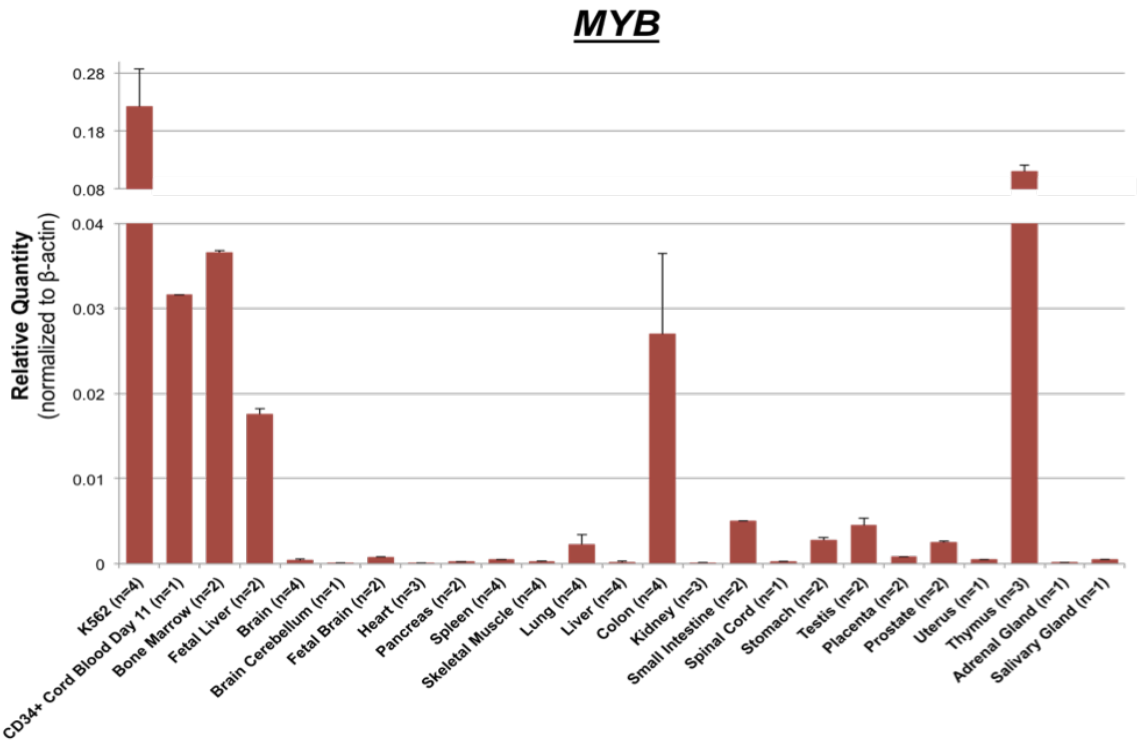
Supplemental Figure 2.1. Sequencing results for 3'-RACE PCR product.

Position 571 of sequencing results was the last nucleotide aligned on chr6q23 of NCBI Blast results, which is 135097644 (hg38).



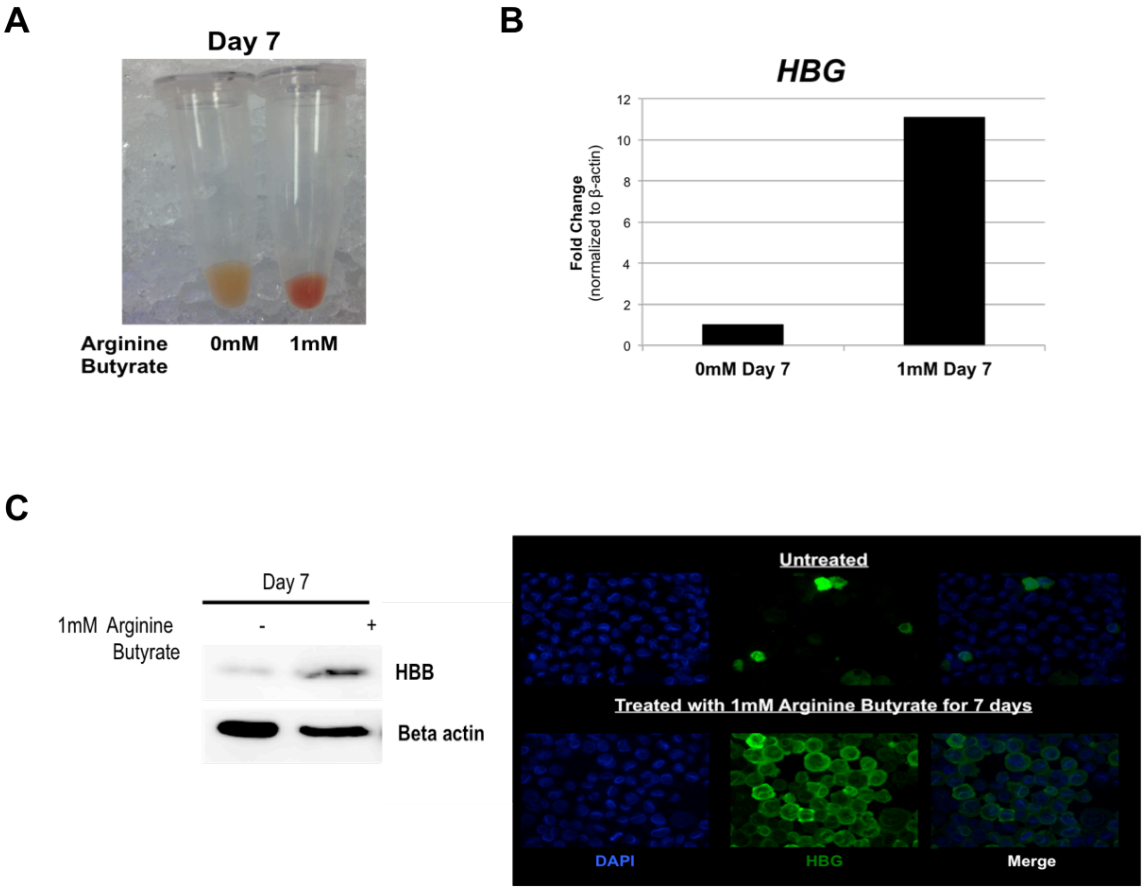
Supplemental Figure 2.2. Sequencing results for 5'-RACE PCR product.

Position 32 of sequencing results was the last nucleotide aligned on chr6q23 of NCBI Blast results, which is 135096362 (hg38).



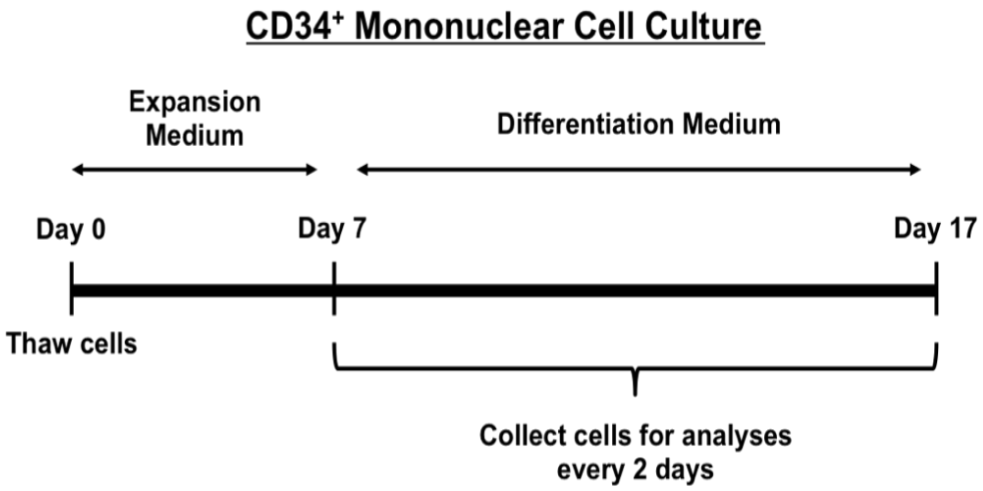
Supplemental Figure 2.3. Expression pattern of *MYB* among various human cells and tissue

Relative quantity of *MYB* was measured my qPCR in 25 different samples. Means are shown for samples with two or more independent samples. *ACTB* was used as the endogenous control.

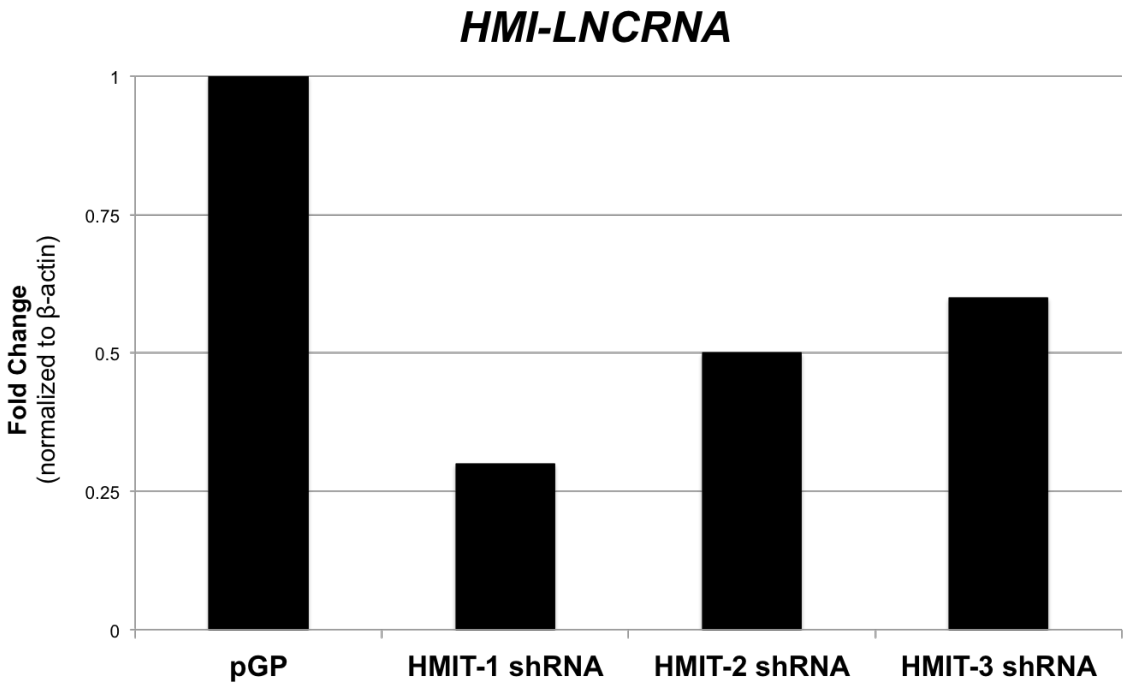


Supplemental Figure 2.4. Hemoglobin expression in K562 cells treated with arginine butyrate.

K562 cells were either untreated (n=1) or treated (n=3) with 1 mM arginine butyrate for seven days. At day 7, cells were harvested to determine (**A**) cell color, (**B**) *HBG* mRNA expression by qPCR, and (**C**) HBG protein expression by Western blot analysis (left panel) and immunofluorescence staining (blue is DAPI, green is HBG). *ACTB* was used as endogenous control for qPCR analysis.

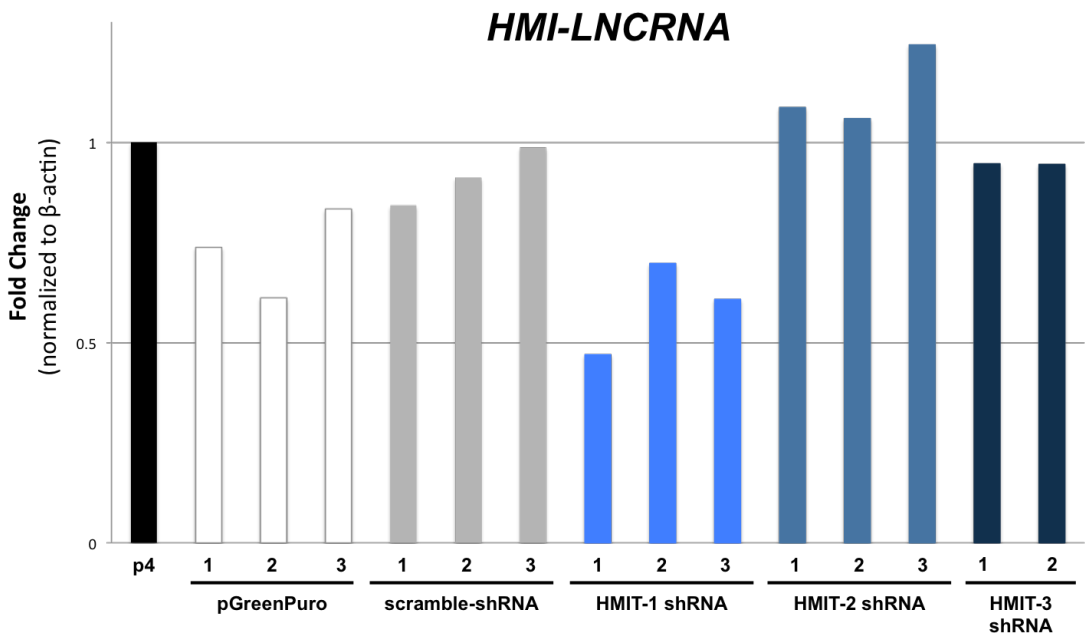


Supplemental Figure 2.5. Timeline for the expansion and differentiation of CD34+ cells derived cord blood and peripheral blood.



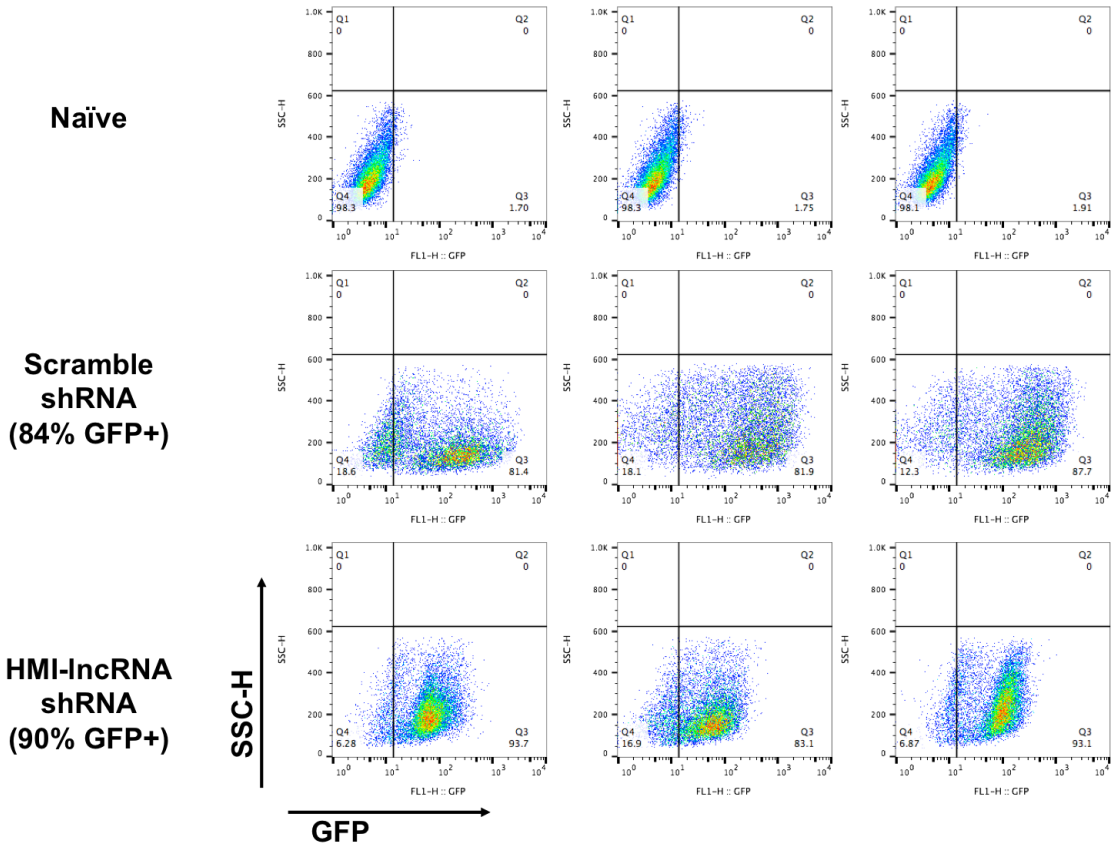
Supplemental Figure 3.1. *HMI-LNCRNA* expression in TF-1a cells transduced with three different HMI-lncRNA shRNA templates.

TF-1a cells were transduced with pGreenPuro (pGP) (n=1), and HMIT-1 (n=1), HMIT-2 (n=1) and HMIT-3 (n=1) shRNAs and cultured for two weeks in 8 μg/ml puromycin for selection. qPCR analysis for *HMI-LNCRNA* was analyzed to determine fold change for each sample. *ACTB* was used as endogenous control.



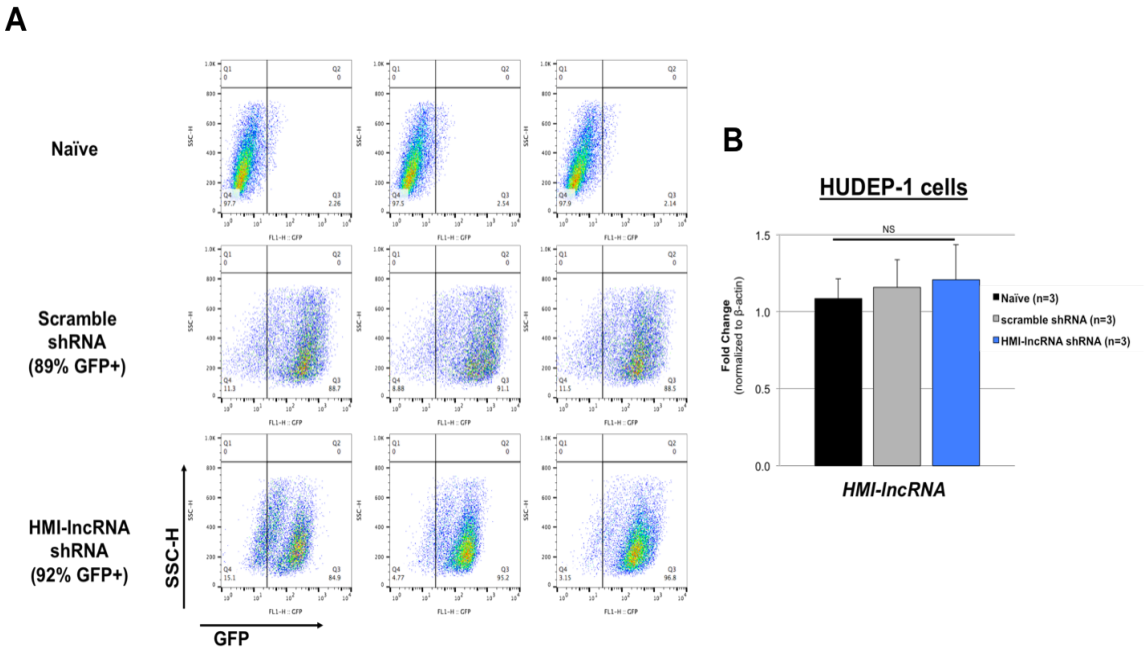
Supplemental Figure 3.2. *HMI-LNCRNA* expression in HUDEP-2 cells transduced with three different HMI-lncRNA shRNA templates.

HUDEP-2 cells were transduced with pGreenPuro (pGP) (n=3), and HMIT-1 (n=3), HMIT-2 (n=3) and HMIT-3 (n=2) shRNAs and cultured for one week in 1 μg/ml puromycin for selection. qPCR analysis for *HMI-LNCRNA* was analyzed to determine fold change for each sample. *ACTB* was used as endogenous control. p4 sample is non-transduced HUDEP-2 cells.



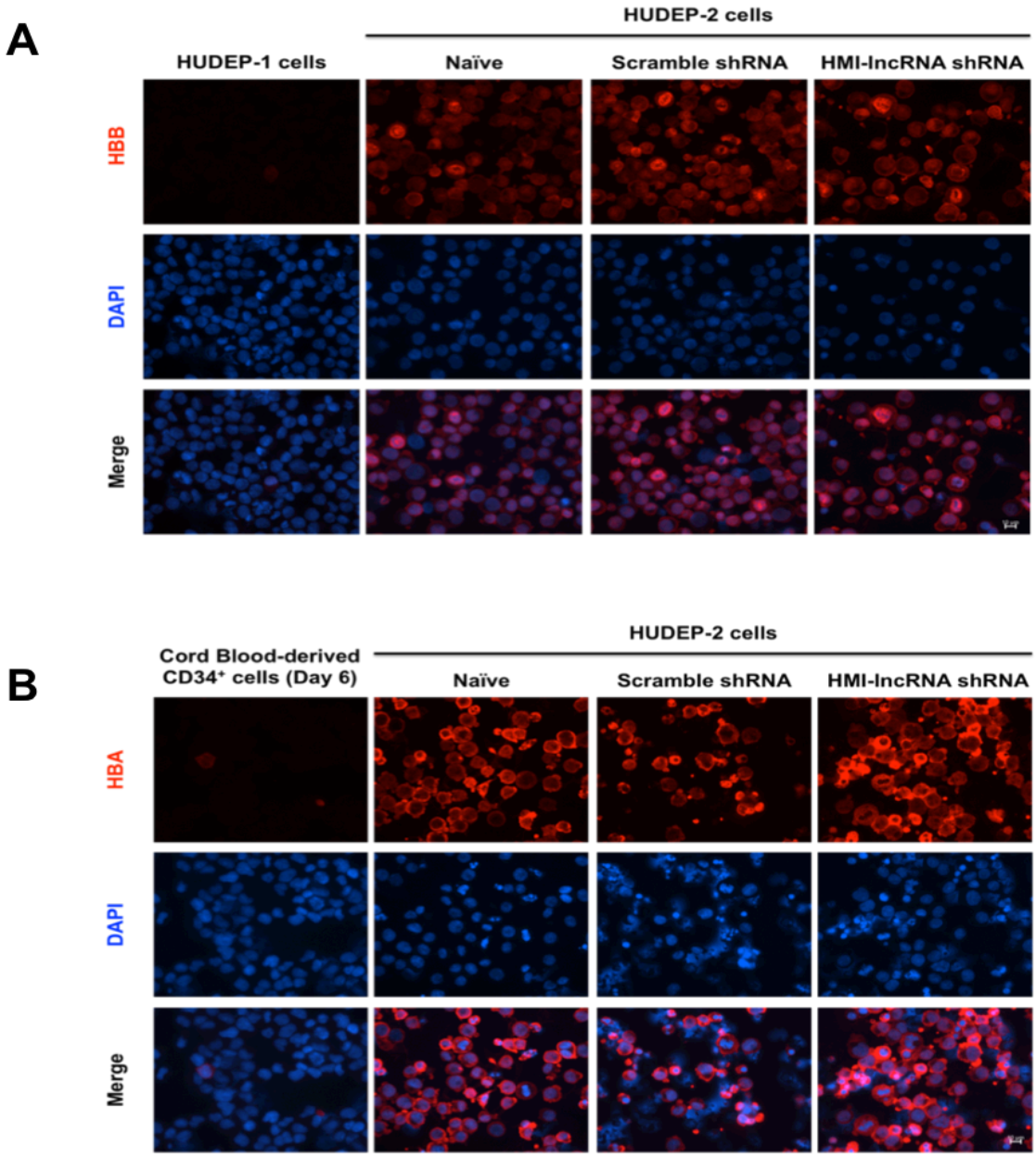
Supplemental Figure 3.3. FACS analysis for GFP expression in HUDEP-2 cells transduced with scramble shRNA and HMI-lncRNA shRNA.

Naïve HUDEP-2 cells (n=3), and cells transduced with scramble shRNA (n=3) and HMI-lncRNA shRNA (n=3) were fix, and analyzed by flow cytometry based on SSC-H (side scatter for granularity) and expression of GFP^{+}.



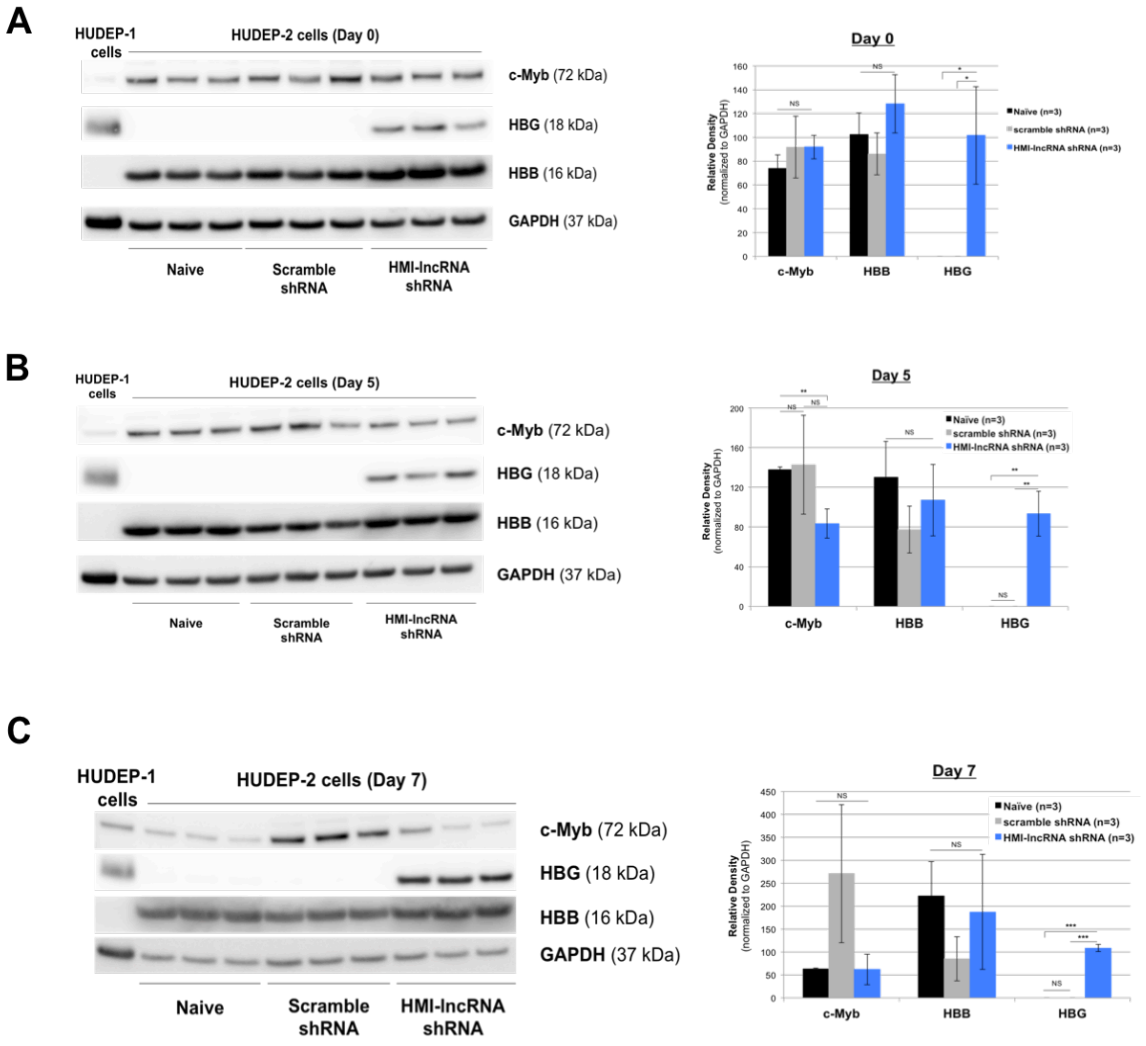
Supplemental Figure 3.4. GFP and *HMI-LNCRNA* expression in HUDEP-1 cells transduced with scramble shRNA and HMI-lncRNA shRNA.

(**A**) Naïve HUDEP-1 cells (n=3), and cells transduced with scramble shRNA (n=3) and HMI-lncRNA shRNA (n=3) were fix, and analyzed by flow cytometry based on SSC-H (side scatter for granularity) and expression of GFP^{+}. (**B**) qPCR analysis for *HMI-LNCRNA* of HUDEP-1 cells from each group. *ACTB* was used as endogenous control. NS means not significant.



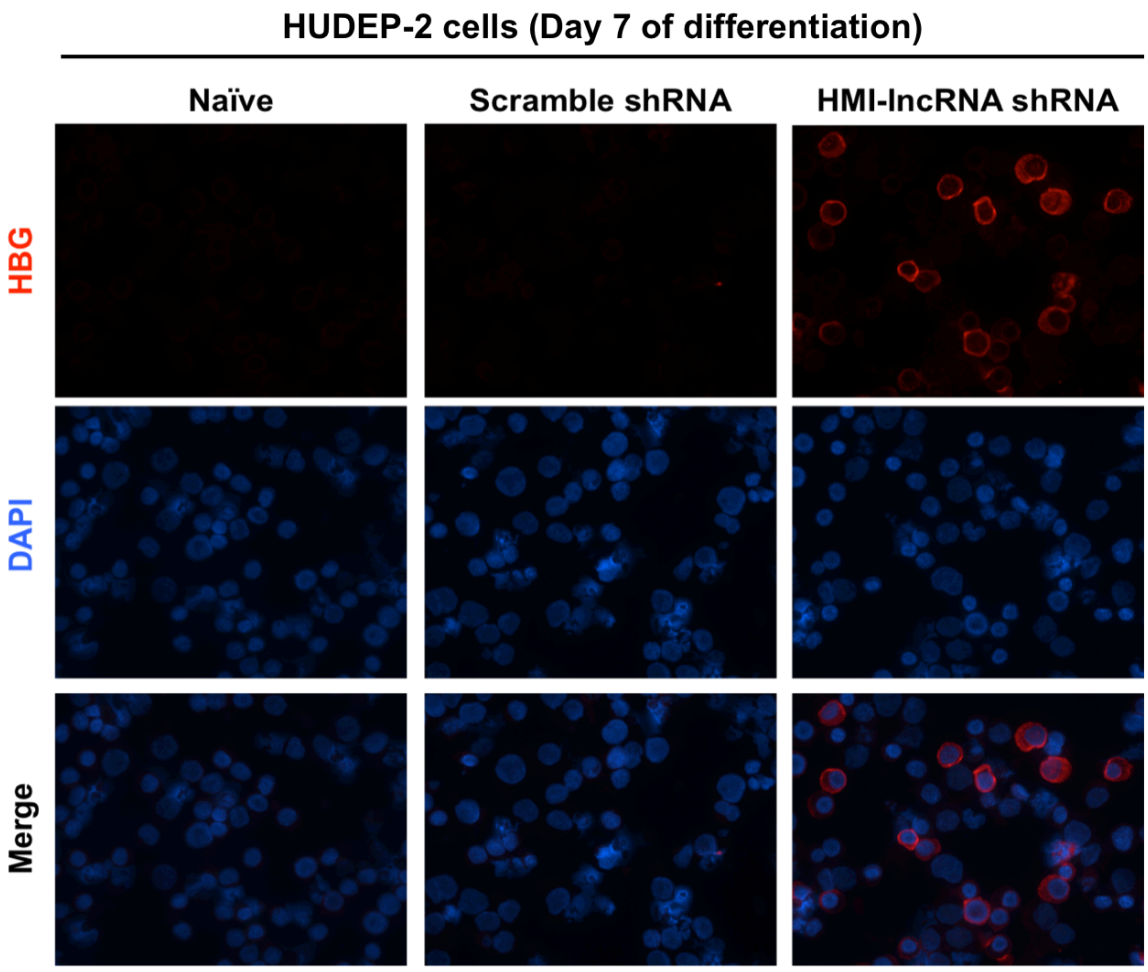
Supplemental Figure 3.5. HBB and HBA expression.

Naïve HUDEP-2 cells, and cells transduced with scramble and HMI-lncRNA shRNAs were stained with (**A**) anti-HBB antibody and (**B**) anti-HBA antibody, followed by secondary antibody labeled with Alexa Fluor-594 (in red), and DAPI to stain nuclei (in blue), and imaged with a fluorescent microscope at 40X magnification.



Supplemental Figure 3.6. Protein expression of c-MYB and hemoglobin with knockdown of *HMI-LNCRNA* during erythroid differentiation of HUDEP-2 cells.

Naïve HUDEP-2 cells, and cells transduced with scramble shRNA and HMI-lncRNA shRNA were placed in differentiation medium for up to 7 days. Protein expression of c-MYB, HBG and HBB were analyzed in cells from (**A**) Day 0, (**B**) Day 5 and (**C**) Day 7 (HUDEP-1 cells were used as control). GAPDH was used as loading control. Densitometry was used to quantify western blot results (measured by ImageJ). p-values: *< 0.05; **< 0.005; ***<0.0005; NS (not significant). p-values were obtained by Student T-test.



Supplemental Figure 3.7. Protein expression of HBG with knockdown of *HMI-LNCRNA* at Day 7 of erythroid differentiation of HUDEP-2 cells.

Naïve HUDEP-2 cells, and cells transduced with scrambled and HMI-lncRNA shRNAs were placed in differentiation medium for 7 days. Cells were stained with anti-HBG antibody, followed by secondary antibody labeled with Alexa Fluor-594 (in red) and DAPI to stain nuclei (in blue), and imaged with a fluorescent microscope at 40X magnification.

LIST OF ABBREVIATED JOURNAL TITLES

Acta Biochim Pol.....Acta Biochimica Polonica

Am J Clin Pathol.....American Journal of Clinical Pathology

Am J Pathol..... American Journal of Pathology

Ann Hematol..... Annals of Hematology

Ann Hum Genet..... Annals of Human Genetics

Ann N Y Acad Sci .....Annals of the New York Academy of Sciences

Annu Rev Genet..... Annual Review of Genetics

Arter Thromb Vasc Biol ..... Arteriosclerosis, Thrombosis, and Vascular Biology

Biochim Biophys Acta.....Biochimica et Biophysica Acta

Biophys Chem.....Biophysical Chemistry

Blood Cells, Mol Dis ..... Blood Cells, Molecules and Diseases

BMC Blood Disord..... BMC Blood Disorders

Br J Haematol .....British Journal of Haematology

Bull N Y Acad Med .....Bulletin of the New York Academy of Medicine

Cell Death Differ.....Cell Death and Differentiation

Cell Rep ..... Cell Reports

Cold Spring Harb Perspect Biol.....Cold Spring Harbor Perspectives in Biology

Cold Spring Harb Perspect Med .....Cold Spring Harbor Perspectives in Medicine

Dev Growth Differ ..... Development, Growth and Differentiation

Eur J Biochem .....European Journal of Biochemistry

Front Med.....Frontiers in Medicine

Front Physiol .....Frontiers in Physiology

Genes Dev ..... Genes and Development

Gene Ther.....Gene Therapy

Genome Biol ..... Genome Biology

Genome Res ..... Genome Research

Hum Genet .....Human Genetics

Hum Mol Genet ..... Human Molecular Genetics

Int J Hematol..... International Journal of Hematology

Int J Mol Sci.....International Journal of Molecular Sciences

JAMA.....Journal of the American Medical Association

J Blood Med.....Journal of Blood Medicine

J Clin Invest .....Journal of Clinical Investigation

J Clin Lab Anal .....Journal of Clinical Laboratory Analysis

Leuk Res .....Leukemia Research

Med Res Rev.....Medicinal Research Reviews

Methods Mol Biol.....Methods of Molecular Biology

Mol Cell ..... Molecular Cell

Mol Cell Biol .....Molecular and Cellular Biology

Mutat Res - Rev Mutat Res.....Mutation Research/Reviews in Mutation Research

Nat Commun..... Nature Communications

Nat Genet .....Nature Genetics

Nat Med .....Nature Medicine

Nat Methods.....Nature Methods

Nat Rev Genet.....Nature Reviews Genetics

N Engl J Med ..... The New England Journal of Medicine

PLoS Genet .....PLOS Genetics

PLOS Med .....PLOS Medicine

PNAS ..... Proceedings of the National Academy of Sciences

Proc Nat Acad Sci USA.....

.....Proceedings of the National Academy of Sciences of the United States of America

Proc Natl Acad Sci ..... Proceedings of the National Academy of Sciences

RNA Biol ..... RNA Biology

Sci Signal ..... Science Signaling

Semin Immunol ..... Seminars in Immunology

Trends Biochem Sci ..... Trends in Biochemical Sciences

Trends Genet ..... Trends in Genetics

Trends Pharmacol Sci ..... Trends in Pharmacological Sciences

YALE J Biol Med ..... Yale Journal of Biology and Medicine

BIBLIOGRAPHY

1. Chiabrando D, Mercurio S, Tolosano E. Heme and erythropoieis: More than a structural role. *Haematologica*. 2014;99(6):973-983. doi:10.3324/haematol.2013.091991.

2. Palis J. Primitive and definitive erythropoiesis in mammals. *Front Physiol*. 2014;5(3):1-9. doi:10.3389/fphys.2014.00003.

3. Baron MH, Isern J, Fraser ST. The embryonic origins of erythropoiesis in mammals. *Blood*. 2012;119(21):4828-4837. doi:10.1182/blood-2012-01-153486.

4. Kingsley PD, Greenfest-Allen E, Frame JM, et al. Ontogeny of erythroid gene expression. *Blood*. 2013;121(6):e5-e13. doi:10.1182/blood-2012-04-422394.

5. Palis J, Mcgrath KE, Kingsley PD. Initiation of Hematopoiesis and Vasculogenesis in Murine Yolk Sac Explants. *Blood*. 1995;86(1):156-163. http://www.bloodjournal.org/content/bloodjournal/86/1/156.full.pdf. Accessed July 16, 2017.

6. Palis J. Of mice and men. *Blood*. 2014;123(22):3367-3368. doi:10.1038/ng1490.

7. Kingsley PD, Malik J, Fantauzzo KA, Palis J. Yolk sac–derived primitive erythroblasts enucleate during mammalian embryogenesis. *Blood*. 2004;104(1):19-25. doi:10.1182/blood-2003.

8. Fraser ST, Isern J, Baron MH. Maturation and enucleation of primitive erythroblasts during mouse embryogenesis is accompanied by changes in cell-surface antigen expression. *Blood*. 2007;109(1):343-352. doi:10.1182/blood-2006-03-006569.

9. Stephenson JR, Axelrad AA, Mcleod DL, Shreeve MM. Induction of Colonies of Hemoglobin-Synthesizing Cells by Erythropoietin In Vitro. *Proc Nat Acad Sci USA*. 1971;68(7):1542-1546. https://www-ncbi-nlm-nih-gov.ezproxy.bu.edu/pmc/articles/PMC389236/pdf/pnas00082-0163.pdf. Accessed July 16, 2017.

10. Hu J, Liu J, Xue F, et al. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood*. 2013;121(16):3246-3253. doi:10.1182/blood-2013-01.

11. Chasis JA, Mohandas N. Erythroblastic islands: niches for erythropoiesis. *Blood*. 2008;112(3):470-478. doi:10.1182/blood-2008.

12. Koury MJ, Bondurant MC. The molecular mechanism of erythropoietin action. *Eur J Biochem*. 1992;210(3):649-663. doi:10.1111/j.1432-1033.1992.tb17466.x.

13. Marsee DK, Pinkus GS, Yu H. CD71 (Transferrin Receptor) An Effective Marker for Erythroid Precursors in Bone Marrow Biopsy Specimens. *Am J Clin Pathol*. 2010;134:429-435. doi:10.1309/AJCPCRK3MOAOJ6AT.

14. Peschle C, Mavilio F, Care A, et al. Haemoglobin switching in human embryos: asynchrony of zeta----alpha and epsilon----gamma-globin switches in primitive and definite erythropoietic lineage. *Nature*. 1985;313(17):235-238.

15. Forget BG. Progress in Understanding the Hemoglobin Switch. *N Engl J Med*. 2011;365(9):852-853. http://www.nejm.org.ezproxy.bu.edu/doi/pdf/10.1056/NEJMe1106969. Accessed July 9, 2017.

16. 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;111(4):2400-2408. doi:10.1182/blood-2007.

17. Kingsley PD, Malik J, Emerson RL, et al. "Maturational" globin switching in primary primitive erythroid cells. *Blood*. 2006;104(4):1665-1672. doi:10.1182/blood-2005-08-3097.

18. Higgs DR. The molecular basis of Alpha-thalassemia. *Cold Spring Harb Perspect Med*. 2013;3(5):1-16. doi:10.1101/cshperspect.a011700.

19. Thein SL. Molecular basis of β thalassemia and potential therapeutic targets. *Blood Cells, Mol Dis*. 2017:1-12. doi:10.1016/j.bcmd.2017.06.001.

20. Li B, Ding L, Yang C, et al. Characterization of transcription factor networks involved in umbilical cord blood CD34+ stem cells-derived erythropoiesis. *PLoS One*. 2014;9(9):e107133. doi:10.1371/journal.pone.0107133.

21. Li B, Ding L, Li W, Story MD, Pace BS. Characterization of the transcriptome profiles related to globin gene switching during in vitro erythroid maturation. *BMC Genomics*. 2012;13(153):1-18. doi:10.1186/1471-2164-13-153.

22. 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. *Proc Natl Acad Sci USA*. 1996;93:12355-12358. http://www.pnas.org.ezproxy.bu.edu/content/93/22/12355.long. Accessed July 16, 2017.

23. Miller IJ, Bieker JJ. A Novel, Erythroid Cell-Specific Murine Transcription Factor

That Binds to the CACCC Element and Is Related to the Kriippel Family of Nuclear Proteinst. *Mol Cell Biol*. 1993;13:2776-2786. https://www-ncbi-nlm-nih-gov.ezproxy.bu.edu/pmc/articles/PMC359658/pdf/molcellb00017-0150.pdf. Accessed July 16, 2017.

24. Tolhuis B, Palstra R-J, Splinter E, Grosveld F, De Laat W. Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus. *Mol Cell*. 2002;10:1453-1465. http://ac.els-cdn.com.ezproxy.bu.edu/S1097276502007815/1-s2.0-S1097276502007815-main.pdf?_tid=8c4e79c6-6a82-11e7-a12f-00000aab0f6c&acdnat=1500249639_adb3c300b804bba98d1ec514ca2c930a. Accessed July 16, 2017.

25. Jiang J, Best S, Menzel S, et al. cMYB is involved in the regulation of fetal hemoglobin production in adults. *Blood*. 2006;108(3):1077-1083. doi:10.1182/blood-2006-01-008912.

26. Mucenski ML, McLain K, Kier AB, et al. A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell*. 1991;65(4):677-689. doi:10.1016/0092-8674(91)90099-K.

27. Greig KT, Carotta S, Nutt SL. Critical roles for c-Myb in hematopoietic progenitor cells. *Semin Immunol*. 2008;20:247-256. doi:10.1016/j.smim.2008.05.003.

28. Bengtsen M, Klepper K, Gundersen S, et al. C-Myb binding sites in haematopoietic chromatin landscapes. *PLoS One*. 2015;10(7):1-25. doi:10.1371/journal.pone.0133280.

29. Bianchi E, Bulgarelli J, Ruberti S, et al. MYB controls erythroid versus megakaryocyte lineage fate decision through the miR-486-3p-mediated downregulation of MAF. *Cell Death Differ*. 2015;22(12):1906-1921. doi:10.1038/cdd.2015.30.

30. Yi Z, Cohen-Barak O, Hagiwara N, et al. Sox6 Directly Silences Epsilon Globin Expression in Definitive Erythropoiesis. *PLoS Genet*. 2006;2(2):e14. http://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.0020014&type=printable. Accessed June 23, 2017.

31. Masuda T, Wang X, Maeda M, et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science (80-)*. 2016;351(6270):285-289.

32. Xu J, Bauer DE, Kerenyi MA, et al. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci*. 2013;110(16):6518-6523. doi:10.1073/pnas.1303976110.

33. Masuda T, Wang X, Maeda M, et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science (80-)*. 2016;351(6270):285-289. doi:10.1126/science.aad3312.

34. Chen Z, Luo H yuan, Steinberg MH, Chui DHK. BCL11A represses HBG transcription in K562 cells. *Blood Cells, Mol Dis*. 2009;42(2):144-149. doi:10.1016/j.bcmd.2008.12.003.

35. Xu J, Sankaran VG, Ni M, et al. Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev*. 2010;24(8):783-789. doi:10.1101/gad.1897310.

36. Stamatoyannopoulos G, Majerus PW, Perlmutter RM, Varmus H. *The Molecular Basis of Blood Diseases*. 3rd ed. (Strauss M, Stringer T, Ehlers J, eds.). W.B. Saunders Company; 2001.

37. DeLoughery T. Microcytic anemia. *N Engl J Med*. 2014;371(14):1324-1331. doi:10.1016/B978-0-323-05405-8.00144-3.

38. Herrick JB. Peculiar Elongated and Sickle-shaped Red Blood Corpuscles in a Case of Severe Anemiaa. *YALE J Biol Med*. 2001;74:179-184. https://www-ncbi-nlm-nih-gov.ezproxy.bu.edu/pmc/articles/PMC2588723/pdf/yjbm00012-0035.pdf. Accessed July 16, 2017.

39. Wailoo K. Sickle Cell Disease — A History of Progress and Peril. *N Engl J Med*. 2017;379(9):805-807. doi:10.1056/NEJMp1002530.

40. Frenette PS, Atweh GF. Sickle cell disease: old discoveries, new concepts, and future promise. *J Clin Invest*. 2007;117(4):850-858. doi:10.1172/JCI30920.850.

41. Eaton WA. Linus Pauling and sickle cell disease. *Biophys Chem*. 2003;100:109-116. doi:10.1016/S0301-4622(02)00269-7.

42. Pauling L. Molecular Disease and Evolution. *Bull N Y Acad Med*. 1964;40:334-342. https://www-ncbi-nlm-nih-gov.ezproxy.bu.edu/pmc/articles/PMC1750607/pdf/bullnyacadmed00290-0020.pdf. Accessed July 16, 2017.

43. Ingram VM. Abnormal human haemoglobins. I. The comparison of normal human and sickle-cell haemoglobins by fingerprinting. *Biochim Biophys Acta*. 1958;28:539-545. http://ac.els-cdn.com.ezproxy.bu.edu/000630025890516X/1-s2.0-000630025890516X-main.pdf?_tid=063d8dfc-6a89-11e7-9f64-00000aacb35e&acdnat=1500252420_797d339fe58a9bdd1f5bc0978e75e68c. Accessed July 16, 2017.

44. Piel FB, Steinberg MH, Rees DC. Sickle Cell Disease. *N Engl J Med*. 2017;376(16):1561-1573. doi:10.1056/NEJMra1510865.

45. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. *Lancet*. 2010;376:2018-2031. doi:10.1016/S0140-6736(10)61029-X.

46. Hebbel RP, Yamada O, Moldow CF, Jacob HS, White JG, Eaton JW. Abnormal Adherence of Sickle Erythrocytes to Cultured Vascular Endothelium: possible mechanism for microvascular occlusion in sickle cell disease. *J Clin Invest*. 1980;65(1):154-160. https://www-ncbi-nlm-nih-gov.ezproxy.bu.edu/pmc/articles/PMC371350/pdf/jcinvest00685-0162.pdf. Accessed July 16, 2017.

47. Hoover R, Rubin R, Wise G, Warren R. Adhesion of Normal and Sickle Erythrocytes to Endothelial Monolayer Cultures. *Blood*. 1979;54(4):872-876. http://www.bloodjournal.org/content/bloodjournal/54/4/872.full.pdf. Accessed July 16, 2017.

48. Steinberg MH. Pathophysiologically based drug treatment of sickle cell disease. *Trends Pharmacol Sci*. 2006;27(4):204-210. doi:10.1016/j.tips.2006.02.007.

49. Dé F, Piel RB, Hay SI, Gupta S, Weatherall DJ, Williams TN. Global Burden of Sickle Cell Anaemia in Children under Five, 2010–2050: Modelling Based on Demographics, Excess Mortality, and Interventions. *PLOS Med*. 2013;10(7):e1001484. doi:10.1371/journal.pmed.1001484.

50. Piel FB, Patil AP, Howes RE, et al. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat Commun*. 2010;1(104):1-7. doi:10.1038/ncomms1104.

51. Pleasants S. Epidemiology: A moving target. *Nature*. 2014;515(7526):S2-S3. http://www.nature.com.ezproxy.bu.edu/nature/journal/v515/n7526_supp/pdf/515S2a.pdf. Accessed July 16, 2017.

52. Data and Statistics | Sickle Cell Disease | NCBDDD | CDC. https://www.cdc.gov/ncbddd/sicklecell/data.html. Accessed July 16, 2017.

53. Platt OS, Brambilla DJ, Rosse WF, et al. Mortality in Sickle Cell Disease. Life expectancy and risk factors for early death. *N Engl J Med*. 1994;330(23):1639-1644. http://www.nejm.org.ezproxy.bu.edu/doi/pdf/10.1056/NEJM199406093302303. Accessed July 17, 2017.

54. Castro O, Brambilla DJ, Thorington B, et al. The Acute Chest Syndrome in Sickle Cell Disease: Incidence and Risk Factors. *Blood*. 1994;84(2):643-649.

http://www.bloodjournal.org/content/bloodjournal/84/2/643.full.pdf. Accessed July 17, 2017.

55. Platt OS, Thorington BD, Brambilla DJ, et al. Pain in Sickle Cell Disease. Rates and risk factors. *N Engl J Med*. 1991;325(1):11-16. http://www.nejm.org.ezproxy.bu.edu/doi/pdf/10.1056/NEJM199107043250103. Accessed July 17, 2017.

56. Letvin NL, Linch DC, Beardsley P, McIntyre KW, Nathan DG. Augmentation of fetal-hemoglobin production in anemic monkeys by hydroxyurea. *N Engl J Med*. 1984;310(14):869-873. http://www.nejm.org.ezproxy.bu.edu/doi/pdf/10.1056/NEJM198404053101401. Accessed July 17, 2017.

57. Hankins JS, Ware RE, Rogers ZR, et al. Long-term hydroxyurea therapy for infants with sickle cell anemia: the HUSOFT extension study. *Blood*. 2005;106(7):2269-2275. doi:10.1182/Blood.

58. Steinberg MH, Barton F, Castro O, 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;289(13):1645-1651. http://dx.doi.org/10.1001/jama.289.13.1645.

59. Rodgers GP, Dover GJ, Noguchi CT, Schechter AN, Nienhuis AW. Hematologic responses of patients with sickle cell disease to treatment with hydroxyurea. *N Engl J Med*. 1990;322(15):1037-1045. http://www.nejm.org.ezproxy.bu.edu/doi/pdf/10.1056/NEJM199004123221504. Accessed July 17, 2017.

60. Charache S, Dover GJ, Moore RD, et al. Hydroxyurea: Effects on Hemoglobin F Production in Patients With Sickle Cell Anemia. *Blood*. 1992;79(10):2555-2565. http://www.bloodjournal.org/content/bloodjournal/79/10/2555.full.pdf. Accessed July 17, 2017.

61. Plant OS, Orkin SH, Dover G, Beardsley GP, Miller B, Nathan DG. Hydroxyurea Enhances Fetal Hemoglobin Production in Sickle Cell Anemia. *J Clin Invest*. 1984;74:652-656. https://www-ncbi-nlm-nih-gov.ezproxy.bu.edu/pmc/articles/PMC370519/pdf/jcinvest00710-0348.pdf. Accessed July 17, 2017.

62. Niihara Y, Matsui NM, Shen YM, et al. L-Glutamine therapy reduces endothelial adhesion of sickle red blood cells to human umbilical vein endothelial cells. *BMC Blood Disord*. 2005;5(4):1-7. doi:10.1186/1471-2326-5-4.

63. Estcourt LJ, Fortin PM, Hopewell S, Trivella M, Wang WC. Blood transfusion for preventing primary and secondary stroke in people with sickle cell disease. In:

Estcourt LJ, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd; 2017. doi:10.1002/14651858.CD003146.pub3.

64. DeBaun MR, Gordon M, McKinstry RC, et al. Controlled trial of transfusions for silent cerebral infarcts in sickle cell anemia. *N Engl J Med*. 2014;371(8):699-710. doi:10.1056/NEJMoa1401731.

65. Steinberg MH. More Blood for Sickle Cell Anemia? *N Engl J Med*. 2014;371(8):775-776. doi:10.1056/NEJMe1405776.

66. Ark W Alters MC, Elinda Atience MP, Endy Eisenring WL, et al. Bone Marrow Transplantation for Sickle Cell Disease. *N Engl J Med*. 1996;335(6):369-376. http://www.nejm.org.ezproxy.bu.edu/doi/pdf/10.1056/NEJM199608083350601. Accessed July 17, 2017.

67. Hoban MD, Cost GJ, Mendel MC, et al. Correction of the sickle cell disease mutation in human hematopoietic stem / progenitor cells. *Blood*. 2015;125(17):2597-2604. doi:10.1182/blood-2014-12-615948.The.

68. Ribeil J-A, Hacein-Bey-Abina S, Payen E, et al. Gene Therapy in a Patient with Sickle Cell Disease. *N Engl J Med*. 2017;376(9):848-855. doi:10.1056/NEJMoa1609677.

69. Thein SL, Menzel S. Discovering the genetics underlying foetal haemoglobin production in adults. *Br J Haematol*. 2009;145(4):455-467. doi:10.1111/j.1365-2141.2009.07650.x.

70. Craig JE, Barnetson RA, Prior J, Raven JL, Thein SL. Rapid Detection of Deletions Causing delta-beta Thalassemia and Hereditary Persistence of Fetal Hemoglobin by Enzymatic Amplification. *Blood*. 1994;83(6):1673-1682. http://www.bloodjournal.org/content/bloodjournal/83/6/1673.full.pdf. Accessed July 11, 2017.

71. Forget BG. Molecular Basis of Hereditary Persistence of Fetal Hemoglobin. *Ann N Y Acad Sci*. 1998;850:38-44. doi:10.1111/j.1749-6632.1998.tb10460.x.

72. Sankaran VG, Xu J, Byron R, et al. A Functional Element Necessary for Fetal Hemoglobin Silencing. *N Engl J Med*. 2011;365(9):807-814. doi:10.1056/NEJMoa1103070.

73. Fleenor DE, Kaufman RE. Characterization of the DNase I Hypersensitive Site 3' of the Human Beta Globin Gene Domain. *Blood*. 1993;81(10):2781-2790. http://www.bloodjournal.org/content/bloodjournal/81/10/2781.full.pdf. Accessed July 17, 2017.

74. Smith EC, Orkin SH. Hemoglobin genetics: recent contributions of GWAS and gene editing. *Hum Mol Genet*. 2016;25(R2):R99-R105. doi:10.1093/hmg/ddw170.

75. Gilman JG, Huisman THJ. DNA Sequence Variation Associated With Elevated Fetal G-gamma Globin Production. *Blood*. 1985;66(4):783-787. http://www.bloodjournal.org/content/bloodjournal/66/4/783.full.pdf. Accessed July 17, 2017.

76. Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of -thalassemia. *Proc Natl Acad Sci*. 2008;105(5):1620-1625. doi:10.1073/pnas.0711566105.

77. Lettre G, Sankaran VG, Bezerra MAC, et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *PNAS*. 2008;105(33):11869-11874. doi:10.1073/pnas.0804799105.

78. Mikobi TM, Tshilobo Lukusa P, Aloni MN, et al. Protective BCL11A and HBS1L-MYB polymorphisms in a cohort of 102 Congolese patients suffering from sickle cell anemia. *J Clin Lab Anal*. March 2017:e22207. doi:10.1002/jcla.22207.

79. Upadhye D, Jain D, Trivedi Y, Nadkarni A, Ghosh K, Colah R. Influence of single nucleotide polymorphisms in the BCL11A and HBS1L-MYB gene on the HbF levels and clinical severity of sickle cell anaemia patients. *Ann Hematol*. 2016;95:1201-1203. doi:10.1007/s00277-016-2675-1.

80. Menzel S, Rooks H, Zelenika D, et al. Global Genetic Architecture of an Erythroid Quantitative Trait Locus, HMIP-2. *Ann Hum Genet*. 2014;78:434-451. doi:10.1111/ahg.12077.

81. Thein SL, Menzel S, Peng X, et al. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *PNAS*. 2007;104(27):11346-11351. doi:10.1073/pnas.0611393104.

82. Farrell JJ, Sherva RM, Chen ZY, et al. A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. *Blood*. 2011;117(18):4935-4945. doi:10.1182/blood-2010-11-317081.

83. Stadhouders R, Aktuna S, Thongjuea S, et al. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest*. 2014;124(4):1699-1710. doi:10.1172/JCI71520.

84. Canver M, Lessard S, Pinello L, et al. Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat Genet*. 2017:1-10. doi:10.1038/ng.3793.

85. Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: Past, present, and future. *Genetics*. 2013;193(3):651-669. doi:10.1534/genetics.112.146704.

86. Yang L, Froberg JE, Lee JT. Long noncoding RNAs: Fresh perspectives into the RNA world. *Trends Biochem Sci*. 2014;39(1):35-43. doi:10.1016/j.tibs.2013.10.002.

87. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015;47(3):199-208. doi:10.1038/ng.3192.

88. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol*. 2013;10(6):925-933. doi:10.4161/rna.24604.

89. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22(9):1775-1789. doi:10.1101/gr.132159.111.

90. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell*. 2012;43(6):904-914. doi:10.1016/j.molcel.2011.08.018.Molecular.

91. Kornienko AE, Dotter CP, Guenzl PM, et al. Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. *Genome Biol*. 2016;17(14):1-23. doi:10.1186/s13059-016-0873-8.

92. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915-1927. doi:10.1101/gad.17446611.

93. Alvarez-Dominguez JR, Hu W, Yuan B, et al. Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood*. 2014;123(4):570-581. doi:10.1182/blood-2013-10-530683.The.

94. Natoli G, Andrau J-C. Noncoding Transcription at Enhancers: General Principles and Functional Models. *Annu Rev Genet*. 2012;46:1-19. doi:10.1146/annurev-genet-110711-155459.

95. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*. 2011;12:283-293. doi:nrg2957 [pii]\r10.1038/nrg2957.

96. Vučićević D, Corradin O, Ntini E, Scacheri PC, Ørom UA. Long ncRNA

expression associates with tissue-specific enhancers. *Cell Cycle*. 2015;14(2):253-260. doi:10.4161/15384101.2014.977641.

97. Rothschild G, Basu U, Feng Y, et al. Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends Genet*. 2017;33(2):143-154. doi:10.1016/j.tig.2016.12.002.

98. Kim TK, Hemberg M, Gray JM. Enhancer RNAs: A class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harb Perspect Biol*. 2015;7:a018622. doi:10.1101/cshperspect.a018622.

99. Ren B. Transcription: Enhancers make non-coding RNA. *Nature*. 2010;465:173-174. doi:10.1038/465173a.

100. Alvarez-Dominguez JR, Hu W, Gromatzky AA, Lodish HF. Long noncoding RNAs during normal and malignant hematopoiesis. *Int J Hematol*. 2014;99:531-541. doi:10.1007/s12185-014-1552-8.

101. Ding N, Xi J, Li Y, et al. Global transcriptome analysis for identification of interactions between coding and noncoding RNAs during human erythroid differentiation. *Front Med*. 2016;10(3):297-310. doi:10.1007/s11684-016-0452-0.

102. Doss JF, Corcoran DL, Jima DD, Telen MJ, Dave SS, Chi J-T. A comprehensive joint analysis of the long and short RNA transcriptomes of human erythrocytes. *BMC Genomics*. 2015;16:952. doi:10.1186/s12864-015-2156-2.

103. Gallagher PG. Long noncoding RNAs in erythropoiesis. *Blood*. 2014;123(4):465-466.

104. Han B-W, Chen Y-QY-Q. Potential Pathological and Functional Links Between Long Noncoding RNAs and Hematopoiesis. *Sci Signal*. 2013;6(289):re5. doi:10.1126/scisignal.2004099.

105. Kulczyńska K, Siatecka M. A regulatory function of long non-coding RNAs in red blood cell development. *Acta Biochim Pol*. 2016;63:1-6. doi:10.18388/abp.2016_1351.

106. Morlando M, Ballarino M, Fatica A. Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia. *Front Med*. 2015;2(23):1-5. doi:10.3389/fmed.2015.00023.

107. Nobili L, Lionetti M, Neri A. Long non-coding RNAs in normal and malignant hematopoiesis. *Oncotarget*. 2016;7(31):50666-50681. doi:10.18632/oncotarget.9308.

108. Paralkar VR, Mishra T, Luan J, et al. Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood*. 2014;123(12):1927-1937. doi:10.1182/blood-2013-12-544494.

109. Satpathy AT, Chang HY. Long Noncoding RNA in Hematopoiesis and Immunity. *Immunity*. 2015;42:792-804. doi:10.1016/j.immuni.2015.05.004.

110. Wang C, Wu X, Shen F, Li Y, Zhang Y, Yu D. Shlnc-EC6 regulates murine erythroid enucleation by Rac1-PIP5K pathway. *Dev Growth Differ*. 2015;57:466-473. doi:10.1111/dgd.12225.

111. Zhang X, Lian Z, Padden C, et al. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood*. 2009;113(11):2526-2534. doi:10.1182/blood-2008-06-162164.

112. Hu W, Yuan B, Flygare J, Lodish HF. Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes Dev*. 2011;25:2573-2578. doi:10.1101/gad.178780.111.

113. Bertani S, Sauer S, Bolotin E, Sauer F. The non-coding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting Mll1 to chromatin. *Mol Cell*. 2011;43(6):1040-1046. doi:10.1038/jid.2014.371.

114. Feng Y, Hu X, Zhang Y, Zhang D, Li C, Zhang L. Methods for the Study of Long Noncoding RNA in Cancer Cell Signaling. *Methods Mol Biol*. 2014;1165:115-143. doi:10.1007/978-1-4939-0856-1.

115. Haemmerle M, Gutschner T. Long non-coding RNAs in cancer and development: Where do we go from here? *Int J Mol Sci*. 2015;16:1395-1405. doi:10.3390/ijms16011395.

116. Li H, Yu B, Li J, et al. Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget*. 2014;5(8):2318-2329. doi:10.18632/oncotarget.1913.

117. Li X, Wu Z, Fu X, Han W. LncRNAs: Insights into their function and mechanics in underlying disorders. *Mutat Res - Rev Mutat Res*. 2014;762:1-21. doi:10.1016/j.mrrev.2014.04.002.

118. Pastori C, Peschansky VJ, Barbouth D, Mehta A, Silva JP, Wahlestedt C. Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome. *Hum Genet*. 2014;133:59-67. doi:10.1007/s00439-013-1356-6.

119. Zhou Q, Chung ACK, Huang XR, Dong Y, Yu X, Lan HY. Identification of novel long noncoding rnas associated with TGF-β/Smad3-mediated renal inflammation and fibrosis by RNA sequencing. *Am J Pathol*. 2014;184(2):409-417. doi:10.1016/j.ajpath.2013.10.007.

120. Faghihi MA, Modarresi F, Khalil AM, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of b-secretase. *Nat Med*. 2008;14(7):723-730. http://www.nature.com.ezproxy.bu.edu/nm/journal/v14/n7/pdf/nm1784.pdf. Accessed July 17, 2017.

121. Perez DS, Hoage TR, Pritchett JR, et al. Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Mol Genet*. 2008;17(5):642-655. doi:10.1093/hmg/ddm336.

122. Ji P, Diederichs S, Wang W, et al. MALAT-1, a novel noncoding RNA, and thymosin b4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*. 2003;22:8031-8041. doi:10.1038/sj.onc.1206928.

123. Lin R, Maeda S, Liu C, Karin M, Edgington T. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene*. 2007;26:851-858. doi:10.1038/sj.onc.1209846.

124. Helgadottir A, Thorleifsson G, Manolescu A, et al. A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. *Science (80-)*. 2008;316:1491-1493.

125. Holdt LM, Beutner F, Scholz M, et al. ANRIL Expression Is Associated With Atherosclerosis Risk at Chromosome 9p21. *Arter Thromb Vasc Biol*. 2010;30:620-627. http://atvb.ahajournals.org/content/30/3/620.abstract.

126. Thein SL. Genetic association studies in beta hemoglobinopathies. *Hematology*. 2013:354-361. doi:10.1136/thx.2005.040790.

127. Ngo D, Bae H, Steinberg MH, et al. Fetal hemoglobin in sickle cell anemia: Genetic studies of the arab-indian haplotype. *Blood Cells, Mol Dis*. 2013;51:22-26. doi:10.1016/j.bcmd.2012.12.005.

128. Bentley DR. The Human Genome Project-an overview. *Med Res Rev*. 2000;20(3):189-196. blob:http://www.readcube.com/52f3ed55-94b3-4aa7-a9ec-298a52c5ff2a. Accessed July 17, 2017.

129. Watson JD. The Human Genome Project: Past, Present, and Future. *Science (80-)*. 1990;248:44-49.

130. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57-63. http://www.nature.com.ezproxy.bu.edu/nrg/journal/v10/n1/pdf/nrg2484.pdf. Accessed July 17, 2017.

131. Rapid amplification of 5' complementary DNA ends (5' RACE). *Nat Methods*. 2005;2(8):629-630. doi:10.1038/nmeth0805-629.

132. Koeffler HP, Golde DW. Human myeloid leukemia cell lines: a review. *Blood*. 1980;56(3):344-350.

133. Kurita R, Suda N, Sudo K, et al. Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. *PLoS One*. 2013;8(3):e59890. doi:10.1371/journal.pone.0059890.

134. Zhang Y, Qu Z, Kim S, et al. Down-modulation of cancer targets using locked nucleic acid (LNA)-based antisense oligonucleotides without transfection. *Gene Ther*. 2011;18:326-333. doi:10.1038/gt.2010.133.

135. Hu X, Moscinski LC, Hill BJ, et al. Characterization of a unique factor-independent variant derived from human factor-dependent TF-1 cells: A transformed event. *Leuk Res*. 1998;22:817-826. doi:10.1016/S0145-2126(98)00073-3.

136. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol Cell*. 2011;44:667-678. doi:10.1016/j.molcel.2011.08.027.

137. Canver MC, Smith EC, Sher F, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015;527:192-197. doi:10.1038/nature15521.

138. Wienert B, Martyn GE, Kurita R, Nakamura Y, Quinlan KGR, Crossley M. KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood*. 2017. doi:10.1182/blood-2017-02-767400.

139. Park S, Gianotti-Sommer A, Molina-Estevez FJ, et al. Stem Cell Reports Resource A Comprehensive, Ethnically Diverse Library of Sickle Cell Disease-Specific Induced Pluripotent Stem Cells. *Stem Cell Reports*. 2017;8:1076-1085. doi:10.1016/j.stemcr.2016.12.017.

140. Paralkar VR, Taborda CC, Huang P, et al. Unlinking an lncRNA from Its Associated cis Element. *Mol Cell*. 2016;62(1):104-110. doi:10.1016/j.molcel.2016.02.029.

141. Zhang B, Arun G, Mao YS, et al. The lncRNA malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep*. 2012;2:111-123. doi:10.1016/j.celrep.2012.06.003.

142. Guo G, Kang Q, Zhu X, et al. A long noncoding RNA critically regulates Bcr-Abl-mediated cellular transformation by acting as a competitive endogenous RNA. *Oncogene*. 2015;34:1768-1779.

143. McColl B, Vadolas J. Animal models in beta hemoglobinopathies: utility and limitations. *J Blood Med*. 2016;7:263-274.

144. Steinberg MH, Chui DHK, Dover GJ, Sebastiani P, Alsultan A. Fetal hemoglobin in sickle cell anemia: a glass half full? *Blood*. 2014;123(4):481-485. doi:10.1182/blood.

CURRICULUM VITAE