DOMAIN-FOCUSED CRISPR-SCREEN IDENTIFIES HRI AS A REGULATOR OF FETAL HEMOGLOBIN IN ADULT ERYTHROID CELLS

Jeremy D. Grevet

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

_____

Dr. Gerd A. Blobel, Professor of Pediatrics,
Perelman School of Medicine, University of Pennsylvania

Graduate Group Chairperson

_____

Dr. Daniel S. Kessler, Associate Professor of Cell and Molecular Biology,
Perelman School of Medicine, University of Pennsylvania

Dissertation Committee

Dr. Stephen A. Liebhaber, Professor, Department of Genetics
Perelman School of Medicine, University of Pennsylvania

Dr. Peter S. Klein, Professor of Medicine
Perelman School of Medicine, University of Pennsylvania

Dr. Junwei Shi, Assistant Professor of Cancer Biology
Perelman School of Medicine, University of Pennsylvania

Dr. Wei Tong, Associate Professor of Pediatrics
Children's Hospital of Philadelphia

Dedicated to my parents, Alice Little Grevet and Jean Grevet; my sister Catherine Grevet Delcourt, my brother-in-law Emile Delcourt, my nephew Mylo Delcourt; and my grandparents, Mamie-Nane et Papi-Nane, and Grandma and Grandpa.

ACKNOWLEDGMENTS

I would like to thank my advisor Gerd Blobel. Gerd provided so much support to me during my thesis, I am extremely grateful to him. I have had many struggles and made many mistakes throughout my PhD, and he helped me through all of them and kept believing in me when I no longer believed in myself. Gerd has set up such a wonderful group, I feel extremely fortunate to have done my thesis in his lab.

I would also like to thank Junwei Shi who has essentially been a co-advisor to me on this project. Working with Junwei has been one of the absolute highlights of my PhD. His energy and optimism are second to none, and he lifted me at a time when I was particularly struggling. I am extremely grateful for all the help he has given me during my thesis.

My family has been a constant source of support for me during my PhD. My sister Catherine, and my brother-in-law Emile have lifted me up time and time again. My nephew Mylo is a constant source of joy. I would like to thank all my extended family, Deb and Bill, Abbey and Nathaniel, Danny and Linda, Daniel, Jacob and Leah, Ben and Elaine, et aussi la famille française, Marie-Georges et Philippe, Margaux et Christophe, Marie-Pierre et Jean, Matthieu et Hélène, Jean-Louis et Béatrice, Nicolas, Julien et Maxence. I am so grateful for all the support you have given me, and only wish I could see all of you more often. I especially want to thank my parents, Alice and Jean, who have been there with me through every struggle and very difficult times, I could never have done this without you, you have given me more opportunities and support than I deserve, I love you so much. Je remercie aussi Mami-Nane à qui je pense souvent, et Papi-Nane qui me manque tant. Finally, this is also in remembrance of Cynthia, Grandma and Grandpa, who gave me the best family I could ask for and I miss you every day.

I want to thank my labmates who are all listed below, but in particular Nicole Hamagami for fighting and sticking with me in the trenches, Vivek Behera for being such a great lab-bay mate, Sarah Hsu for setting such a great example for us, and Xianjiang Lan and Scott Peslak for being such wonderful sources of help and support. I also want to thank Ian Slack, Zandra Walton, Elizabeth Loy, and Dana Bellissimo, and Katie France, Eric Hitimana and Sahil Mehra, for being friends whose value is truly beyond the price of rubies.

In addition I would like to thank the following people:
Blobel Lab, past and present: Vivek Behera, Mike Werner, Nicole Hamagami, Xianjiang Lan, Sarah Hsu, Scott Peslak, Laavanya Sankar, Chris Edwards, Carolyne Face, Saurabh Bhardwaj, Caroline Bartman, Perry Evans, Peng Huang, Eugene Khandros, Kiwon Lee, Jing Luan, Marit Vermunt, Aoi Wakabayashi, Hongxin Wang, Jennifer Yano, Haoyue Zhang, Di Zhang, Aaron Stonestrom, Stephan Kadauke, Rena Zheng, Amy Campbell.
Thesis Committee: Steve Liebhaber, Peter Klein, Wei Tong, Junwei Shi, Jennifer Cremins.
Shi Lab: Eri Arai, Zhendong Cao, Emily Duffner, Rodrigo Gier.
Liebhaber Lab: Xinjun Ji.
Hardison Lab (Penn State): Ross Hardison, Cheryl Keller, Belinda Giardine
UPenn MSTP: Skip Brass, Maggie Krall, Maureen Kirsch, David Bittner, Amy Nothelfer
Funding: This work was supported by the NIH grant from NIDDK 1F30DK107055-01 and the Penn Genetics training grant T32 GM008216.

ABSTRACT

DOMAIN-FOCUSED CRISPR SCREEN IDENTIFIES HRI AS A REGULATOR OF FETAL HEMOGLOBIN IN HUMAN ERYTHROID CELLS

Jeremy D. Grevet

Gerd A. Blobel

Increasing fetal hemoglobin (HbF) levels in adult red blood cells provides clinical benefit to patients with sickle cell disease and some forms of β-thalassemia. To identify potentially druggable HbF regulators in adult human erythroid cells, we employed a protein kinase-domain focused CRISPR/Cas9-based genetic screen with a newly optimized sgRNA scaffold. The screen uncovered the heme-regulated inhibitor HRI (also known as EIF2AK1), an erythroid-specific kinase that controls protein translation, as an HbF repressor. HRI depletion markedly increased HbF production in a specific manner and reduced sickling in cultured erythroid cells. Diminished expression of the HbF repressor BCL11A accounted in large part for the effects of HRI depletion. Taken together, these results suggest HRI as a potential therapeutic target for hemoglobinopathies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

In 1949, Linus Pauling published on the discovery of sickle cell anemia as the first molecular disease (*1*, *2*). He first heard about the disease from hematologist William Castle who told him about patients of his whose red blood cells sickled in venous circulation but not in arterial blood (*2*). Pauling postulated that if the disease's manifestation was determined by oxygen levels, it may reflect a molecular defect in hemoglobin. In this paper, Pauling and his group showed that hemoglobin from normal patients had a different electrophoretic mobility than hemoglobin from sickle patients (*1*). The cause for this difference though was unknown. In 1956, Vernon Ingram, working along with Francis Crick at the Cavendish Laboratories, showed by protein fingerprinting (essentially two-dimensional chromatography) that the amino acid composition of sickle hemoglobin differed from its normal counterpart (*3*). This not only pinpointed the molecular etiology of sickle cell disease, but it was also a major contribution to Crick's formulation of the Central Dogma in showing that inherited genetic mutations can lead to altered protein products (*3*).

The pathophysiology of sickle cell disease has been well established for decades now. It is the classic example of a Mendelian disorder. Since Pauling's paper, researchers have made great strides towards developing improved treatment modalities for this disease, some of which are excitingly reaching patients. But sickle cell disease still remains a remarkably challenging disease to manage clinically, as SCD patients continue to face severe pain crises, and complications that affect nearly every organ. There is also much that remains to be learned about more basic aspects of erythroid cells. How red blood cells develop, how their gene expression profiles are regulated, and how these may be

manipulated for therapeutic purposes are all still open questions. Seventy years after Pauling's publication, erythroid biology continues to be a fascinating area of research.

In this introduction, we will overview how erythroid cells are produced during the course of development, how they differentiate from progenitor cells to form fully mature red blood cells that can carry out oxygen transport, and the gene regulation programs that underlie these processes. We will focus on the regulation of the globin genes as a key process of erythropoiesis. This will bring us to the concept of hemoglobin switching, whereby globin expression transitions from fetal to adult type genes shortly after birth. We will discuss the clinical implications of globin switching in the context of hemoglobinopathies. This will then lead us to a discussion of the molecular regulators of hemoglobin switching, and how this knowledge is being used for the development of new therapeutic strategies for diseases such as thalassemias and sickle cell anemia.

1.1 Erythropoiesis and globin switching

1.1.1 Ontogeny of erythropoiesis

During the course of development, the site of production of red blood cells transitions between different organs. It begins in the yolk sac, then takes place in the ventral aortic wall, subsequently the fetal liver, and finally in the bone marrow. This is a complex process that is only briefly summarized here. Erythroid cells are first produced in the yolk sac of the developing embryo in cellular structures termed blood islands (*4*, *5*). These groups of cells contain both erythroid and endothelial cells and arise from a common precursor with bilineage potential termed the hemangioblast (*4*). Red blood cells produced

from the yolk sac are released in the circulation as precursors and undergo differentiation in the periphery (*4*, *5*). Thus, in these early stages of development, circulating red cells are a heterogeneous population of undifferentiated, mitotic, and nucleated erythroid precursors, and mature, enucleated red blood cells (*5*). These first stages of red blood cell production are collectively termed primitive erythropoiesis.

Definitive erythroid cells are in contrast smaller in size, anucleate and mitotically inactive, and are produced from early stages of fetal development throughout adult life (*5*). It is thought that these cells arise from two origins (*5*). The first site of definitive erythropoiesis is the yolk sac where these cells are thought to arise from common erythro-myeloid progenitors or endothelial cells (*5*). Definitive progenitors produced in the yolk sac may also contribute to seeding the fetal liver in its earliest stage of development (*4*).

The second site of origin of definitive erythroid cells is a specific region of the ventral wall of the aorta termed the aorto-gonadal-mesonephros (AGM) (*5*, *6*). The first hematopoietic stem cells (HSCs) are thought to be produced in this region (*5*, *6*). HSCs derived from the AGM have been shown to have the ability to both self-renew and differentiate into all lineages of the hematopoietic system as evidenced by bone marrow transplantation assays (*6*). The close association of HSCs and the ventral aorta suggests that these cells may be of endothelial origin. This has been shown to have important implications for cellular reprogramming. Two recent studies have shown it is possible to directly convert adult endothelial cells to HSCs (*7*, *8*). One of these studies demonstrated the direct conversion of murine adult endothelium to HSCs culture *in vitro* by the transduction of a set of four transcription factors (*7*). The derived HSCs were shown to be able to reconstitute all lineages of the hematopoietic system after transplantation into

lethally irradiated mice. The second study used human iPSCs as a starting material, which were converted to endothelial cells and then to HSCs by the transduction of seven transcription factors (*8*). Similarly, these derived cells were able to produce essentially all hematopoietic lineages in transplanted mice. RNA-seq showed that the transcriptome of the converted cells was remarkably similar to bona-fide HSCs, though the correlation was not perfect (*8*). Further improvements to these reprogramming protocols will certainly ameliorate the efficiency of conversion and the similarity of derived cells to their true counterparts. But these studies show great progress towards the goal of readily producing patient-derived HSCs for therapeutic purposes, and also highlight the close relationship between HSCs and the endothelium.

HSCs derived in the AGM migrate to the fetal liver where they support erythropoiesis and the production of all other hematopoietic lineages until birth (*4*, *5*). In humans, the fetal liver is the site of hematopoiesis for the second two trimesters of gestation (*4*). At birth, fetal liver HSCs migrate to the bone marrow, which is the site of hematopoiesis for the remainder of life (*4*).

1.1.2 Definitive erythropoiesis

In adults, red blood cells are the most common cell type of the blood. They have a life span of 120 days, and must be constantly replenished. Erythroid cells are produced from HSCs in the bone marrow which also give rise to all lineages of the blood system (*4*). Before giving rise to erythroid cells, HSCs differentiate to a common myeloid precursor (CMP) and subsequently to a megakaryocyte-erythroid progenitor (MEP) (Figure 1.1) (*4*). It is important to note that while hematopoiesis has traditionally been conceived of in this

stepwise fashion, recent studies are suggesting that lineage decisions may occur much earlier in this tree than previously appreciated (*9*, *10*). Furthermore, these differentiation processes most likely do not occur in discrete steps but rather over a continuum of cell states as shown by recent single-cell approaches.

Once committed to the erythroid lineage, MEPs undergo a well-defined process of differentiation to terminal red blood cells. This process is characterized by specific stages of maturation that can be described in 3 main levels: erythroid progenitors which comprise burst forming units (BFU-Es) and colony forming units (CFU-Es), erythroid precursors which include pro-erythroblast (ProEs), basophilic (Baso), polychromatic (PolyE), and orthochromatic (OrthoE) erythroblast, and terminal erythroid cells which are reticulocytes and mature red blood cells (Figure 1.1) (*5*). Erythroid progenitors are defined by their ability to grow colonies on methylcellulose (*4*). BFU-Es are characterized by their rapid proliferation. CFU-Es have less proliferative capacity but are 5 to 8-fold more abundant in the bone marrow than BFU-Es and can still undergo up to 5 cellular divisions (*5*). Erythroid progenitors require erythropoietin for their growth as well as several other cytokines including stem cell factor, thrombopoietin, intertleukin 3 (IL-3), IL-11 and FLT3-ligand (*5*).

Erythroid precursors (ProE, Baso, PolyE, OrthoE) undergo multiple stages of differentiation to become terminal erythrocytes and these take place in a specific structure of the bone marrow termed the erythroblastic island (*4*). These consist of a central macrophage with cellular protrusions along which erythroid precursors mature. Erythroid differentiation is characterized by several morphological changes including a reduction in cellular and nuclear size, cellular membrane changes, a decrease in total RNA content, and

an increase in heme and globin protein levels (*4*). The changes in composition of nucleic acid and protein levels lead to staining patterns that have classically defined these different stages. These are also accompanied by changes in expression of cell surface markers. CD71 (Transferrin Receptor 1) is abundant in ProEs and decreases as cells mature. CD235a (glycophorin A) is lowly expressed in ProEs and becomes more abundant during differentiation (*4*).

OrthoEs enucleate in the erythroblastic island to produce reticulocytes through a complex process that involves the movement of nuclei to the cytoplasmic membrane and their release from the cell (*5*). Evicted nuclei are phagocytosed by the central macrophage (*5*). Reticulocytes are released in the peripheral circulation and undergo further changes until become fully mature red blood cells (*4*). Their volume is reduced by approximately 20% as they degrade ribosomes and mitochondria. Red blood cells thus rely on glycolysis as an energy source, and this explains why glucose-6-phosphate dehydrogenase deficiency is a common cause of hemolytic anemia. Maturing erythroid cells also position their cytoskeleton in close association with their cytoplasmic membrane to adopt their characteristic biconcave shape (*11*). This confers an optimal surface area for gas exchange given the small volume of terminal red blood cells (*11*). The membrane of terminally differentiated red blood cells is composed of a meshwork of spectrin filaments and actin junctions closely associated with lipid rafts (*11*). Recent super-resolution imaging studies by STORM microscopy have provided additional evidence for a triangular lattice organization of these spectrin filaments with actin nodes positioned at regular (~80 nm) intervals of each other (*11*). This regular organization of the cytoskeleton, as well as its

close proximity to the cellular membrane may confer optimal biophysical properties to erythroid cells that allow them to properly circulate through microcapillaries.

1.1.3 Transcriptional regulation of erythropoiesis

Erythroid cells undergo profound changes as they progress from immature, rapidly dividing progenitors to terminally differentiated red blood cells mostly composed of hemoglobin. This differentiation program is orchestrated by a core set of transcription factors, with the DNA binding protein GATA1 being considered one of the master regulators of erythropoiesis (*6*). Its homolog GATA2 is expressed in HSCs and progressively decreases in levels as cells differentiate to MEPs. GATA1 is reciprocally induced at this stage in a process termed GATA switching (*12*). This marks a first commitment to the erythroid lineage as GATA1 begins to occupy many sites previously bound by GATA2, and *de novo* sites as well (*12*). As a key determinant of erythropoiesis, GATA1 instructs both the repression of genes that support cellular proliferation and other hematopoietic lineages, as well as the activation of erythroid specific genes (*6*). These include cytoskeletal and membrane proteins, cell cycle regulators, components of the heme biosynthetic pathway, and globins (*12*). As such, GATA1 plays roles both in transcriptional repression and activation.

How GATA1 accomplishes these differing functions is too complex to fully review here, but suffice it to say it performs these diverse roles in part by associating with different co-factors (*13*). In a highly simplified view, GATA1 can be found in two main types of mutually exclusive complexes, GATA1/TAL1/LMO2 and GATA1/FOG1 complexes (*6*). The first of these are found at many genes activated by GATA1, including the globin genes

(*12*), which will be discussed in greater detail below. Most of the GATA1/TAL1/LMO2 complexes are found in intergenic or intronic regions, and associate with the self-dimerizing looping factor LDB1 (*12*, *14*). These complexes may thus primarily function through long-range enhancer-promoter interactions. These sites also tend to be co-occupied by the transcription factor KLF1, which has been shown to also be required for the differentiation of red blood cells. FOG1 is a co-factor that directly associates with GATA1 (*6*). It further recruits transcriptional regulators including the NuRD complex which is composed of histone deacetylases HDAC1/2, MTA1/2, and the ATP-dependent chromatin remodelers CHD3/4 (*15*). These regulators are typically associated with transcriptional repression, and thus GATA1/FOG1 complexes may be thought of as primarily functioning to silence gene expression. These complexes though are much more complex, as the NuRD complex, and thus GATA1/FOG1 complexes, can function both as activators or repressors in a highly context-dependent fashion (*16*). How these different roles are determined still remains to be fully elucidated. Several studies have investigated the global changes in transcript levels during the course of erythroid differentiation (*17*, *18*). These suggest that GATA1 functions quite extensively in transcriptional repression, with approximately 5,000 genes repressed and 200 activated during GATA1 induction of erythroid maturation in murine systems normalized by spike-in controls (*18*).

1.1.4 Proteome of differentiating erythroid cells

A recent study further investigated the proteome of differentiating erythroid cells by label free whole cell mass spectrometry (*19*). Though mass spectrometry is more limited in its dynamic range than RNA-seq, and further complicated by the high levels of

hemoglobin present in erythroid cells, the authors were able to quantify up to 6,000 proteins from individual stages of erythropoiesis (*19*). These data revealed firstly that global protein levels decrease during erythroid maturation (*19*), similarly to the decrease in transcript levels observed previously. Proteins that decreased in levels involve general cellular processes such as RNA splicing, protein translation, and metabolic pathways. A subset of proteins was induced during the course of differentiation and these included globins, AHSP, and BAND3, among others (*19*). In particular, when globins and AHSP were subtracted, the total level of proteins was estimated to decrease by about 75% from BFU-Es to OrthoEs (*19*), further showing the extent to which the proteome of erythroid cells is remodeled during differentiation.

The decrease in total protein levels may reflect the reduced transcriptional activity of differentiating erythroid cells. Transcript and protein levels though were not closely correlated across these different stages in this study (*19*). This may reflect technical differences between RNA-seq and mass spectrometry in their sensitivities and accuracies in measurement. But it does appear that post-transcriptional mechanisms play a role in remodeling the proteome of differentiating red blood cells. Autophagy has long been implicated in the clearance of cellular components and proteins as factors of this pathway are induced during erythropoiesis (*20*). Two recent studies showed that the ubiquitin-conjugating enzyme UBE2O plays important roles in protein clearance during reticulocyte maturation to terminal erythroid cells (*21*, *22*). Deleting this enzyme prevents removal of many proteins including ribosomal subunits. Finally, translation may also be an important level at which genes are regulated during erythroid differentiation. A recent study showed that a number of transcripts in differentiating murine erythroid cells are subject to

translational control, with a subset of these potentially being regulated by upstream open-reading frames in their 5'UTRs (*23*).

1.1.5 Heme and globin production

The production of heme and globin proteins is one of the most important aspects of red blood cell maturation as these cells' primary raison d'être is to transport oxygen in the circulation. Heme is produced in a clearly defined metabolic process involving 7 enzymatic reactions taking place either in the cytoplasm or mitochondria, which converts 8 molecules of glycine and 8 molecules of succinyl CoA to the tetrapyrrole structure of heme (*24*). The first of these steps is mediated by *δ*-aminolevulinate synthase 2 (ALAS2) located in the mitochondrial matrix which condenses glycine and succinyl CoA to form ALA and CO_2 (*24*). This is the rate limiting step of this metabolic pathway and thus a major regulatory point for heme production (*24*). In particular, ALAS2 expression is tightly controlled by available iron levels (*25*). This may be intuitive as iron is a key substrate for oxygen binding by heme, but the mechanism by which this regulation occurs is particularly interesting and a core feature of cellular iron metabolism regulation. The 5' UTR of the ALAS2 transcript contains a hairpin loop motif termed an iron-responsive element (IRE) (*25*). This element is bound by an iron-responsive protein (IRP) in low iron conditions and blocks translation of the downstream ALAS2 open reading frame. In iron replete conditions, IRPs bind iron which prevents their association with the IRE (*25*). It is also thought that high iron levels activate an E3 ligase which ubiquinates IRPs to direct them for proteasomal degradation. Thus, increased levels of iron stimulate ALAS2 synthesis, which in turn begins heme production (*26*).

Iron is transported into erythroid cells relatively early during maturation as the transferrin receptor (CD71) is expressed in ProEs. Thus heme synthesis is activated during this earliest stage of erythropoiesis by the import of iron (*26*). Shortly thereafter, globins begin to be expressed. This is in part assisted by heme which inhibits the transcriptional repressor BACH1 that occupies the globin genes (*27*), and also by inhibition of the heme-regulated inhibitor (HRI) which leads to translation of globin transcripts (described further below).

The balanced production of heme and globins presents a notable challenge to erythroid cells as the presence of either of these in excess can lead to oxidative stress and be toxic to cells (*24*). As stated above, heme production begins at early stages of erythroid maturation, and before globins are synthesized (Figure 1.2) (*26*). This presents a short time window during which heme is in excess to globins (*26*) . Erythroid cells have multiple systems in place for overcoming this. The first is the feline leukemia virus subgroup C receptor (FLVCR), a membrane transport protein, which exports heme and is expressed early in maturation (*24*). CFU-Es and Pro-Es critically rely on the FLVCR as depleted it is notably toxic to these cells and prevents proper differentiation, further supporting the concept that heme is in excess at these early stages (*26*). The second system in place for balancing heme and globin levels is the translation initiation regulator heme-regulated inhibitor (HRI) (*28*). HRI is a kinase that phosphorylates the translation initiation factor eIF2α which blocks its proper function in initiating translation globally (*28*). HRI is inhibited by heme which in turn favors translation (*28*). Thus, as heme levels progressively increase in cells, HRI inhibition stimulates protein translation to favor balanced levels of heme and globin proteins. Of note, several erythroid diseases including Diamond-Blackfan

anemia (DBA) and a specific form of myelodysplastic syndrome (MDS) termed 5q minus syndrome have been linked to genetic defects in the translation machinery, and in particular ribosomal proteins (*26*, *29*). It is thought that the pathology underlying these diseases may in part be due to reduced translation of globin proteins, and excess heme leading to ineffective erythropoiesis (*26*, *29*). These disorders thus further highlight the importance of balanced production of these two key components of red blood cells.

1.1.6 Hemoglobin switching

Definitive erythroid cells of the fetal liver and the adult bone marrow show a great degree of overlap in their transcriptomes and chromatin signatures (*30*). However, there are a number of important differences between these two different stages of erythroid ontogeny. This is most notably exemplified by the expression of the globin genes (*31*). As presented above, the production of globins is a key process of erythroid differentiation. Hemoglobin A is a tetramer of 2 α-chains and 2 β-type globin chains ($\alpha_2\beta_2$). The β-type globin genes are encoded on chromosome 11 and there are multiple forms of these genes: embryonic globin (ε-globin, HBE), two fetal globin genes (γ-globin, HBG1/2), and two adult globin genes (*δ*- and β-globin, HBD and HBB), with β-globin being the major form that is expressed during adulthood (Figure 1.4) (*31*). In primitive erythroid cells, embryonic globins are primarily expressed. Fetal type globins are the major forms expressed in definitive erythroid cells of the fetal liver. It is thought that fetal hemoglobin has a stronger affinity for oxygen that adult hemoglobin which may facilitate oxygen extraction to the fetal circulation in the placenta. Shortly after birth, fetal globin expression decreases as adult globins increase in levels in a process commonly referred to as hemoglobin switching

(*31*). It is important to note that fetal globin genes specifically evolved in non-human primates, and thus are not found in lower mammals including mice, which instead have two embryonic type globin genes.

The locus control region (LCR) is a powerful regulatory element that lies upstream of β-globin locus, and long-range interactions between the LCR and globin promoters have been implicated as playing a key role in globin regulation (*31*). During erythroid differentiation, GATA1/TAL1/LMO2 complexes binds globin promoters to activate their transcription (*12*, *32*). Chromosomal conformation capture (3C) assays showed that GATA1 occupancy at the globin promoters is accompanied by a significant increase in long-range contacts between the LCR and globin promoters as these genes are activated (*32*). The interactions between the LCR and globin promoters were further shown to causally underlie the transcriptional activation of these genes. Tethering the looping factor LDB1, which normally associates with GATA1/TAL1/LMO2 complexes, to the β-globin promoter by the use of artificial zinc-fingers was shown to be sufficient both to induce long-range interactions between the LCR and the β-globin promoter, and to activate transcription (*33*). Tethering LDB1 to fetal globin promoters can reactivate the expression of these genes in adult erythroid cells, and is further accompanied by the formation of long-range contacts with the LCR (*34*). The LCR thus plays a critical role in the regulation of globin genes during development.

In most individuals, fetal hemoglobin is strongly silenced as most adult express approximately 0.5% HbF as a percentage of total β-type globins. The extent to which this switch occurs though can vary considerably, and this has important clinical implications, as high levels of fetal globin are beneficial to patients with genetic disorders of the adult

β-globin genes (*31*). This has provided great motivation for understanding this basic process of gene regulation, and how it may be manipulated, over the past decades (*31*).

1.3 Hemoglobinopathies

Mutations in the adult globin genes underlie a group of disorders collectively termed hemoglobinopathies. This group of disorders encompasses some of the most common genetic disorders worldwide and typically manifest themselves after birth when adult globins are expressed. There are two main types of hemoglobinopathies affecting the β-type globin genes: β-thalassemia and Sickle Cell Disease (SCD).

1.1.1 β-Thalassemia

Thalassemias arise from mutations resulting in quantitative imbalances of α- to β-globin expression. A wide range of mutations lead to these disorders, with up to 120 α-thalassemia mutations (*35*) and over 250 β-thalassemia mutations described (*36*).

β-thalassemia is typically caused by point mutations rather than deletions. These mutations can affect any regulatory point of β-globin expression, leading to decreased production of β-globin (*36*). The extent to which beta-globin is reduced determines the severity of the disorder. Mutations that lead to a complete loss of β-globin expression are termed β0-thalassemia. Such mutations typically affect β-globin splicing, transcript stability, or translation (mutated initiation codon, nonsense or frameshift mutations). These can also involve deletions of the β-globin genes, though these types of mutations are found less commonly (*36*). β+ and β ++ thalassemia refer to mutations that only partially reduce β -globin expression. These can involve the beta-globin polyadenylation signal (β+), β-

globin promoter (β++), or the 5'UTR (β++) (*36*). Though less common, some deletions involve the β-globin gene or the locus control region and result in β0-thalassemia.

While the specific genetic etiologies of thalassemias are diverse, all result in an excess of free-globin chains which are toxic. The downstream pathological effects of these excess α-globin chains are quite complex (*36*). Free α-globin chains auto-oxidize, leading to the formation of reactive oxygen species (ROS) (*36*). This in turn leads to oxidation of membranes proteins which causes hemolysis, and activation of apoptotic pathways resulting in ineffective erythropoiesis.

1.2.2 Sickle Cell Disease

Sickle Cell Disease (SCD) is one of the most common genetic disorders worldwide. It is caused by a point mutation in the β-globin gene (Glu6Val, rs334) and is inherited as an autosomal recessive Mendelian trait (*37*). Heterozygotes are typically unaffected, but homozygous patients suffer from both acute and chronic complications. In deoxygenated conditions, this mutation leads to polymerization of hemoglobin, causing sickling of red blood cells. Vaso-occlusion due to altered red blood cell shape and hemolysis constitute the basis for the pathophysiology of this disease. Endothelial inflammation and aberrant expression of adhesion protein on sickle red blood cells further promote vaso-occlusive events (*37*). As a result, nearly every organ can be affected by SCD and the complications are thus numerous. These include acute chest syndrome, stroke, right heart failure, hyposplenism, increased risk of infection, renal failure, priapism, and retinopathy (*37*). Furthermore, acute red blood cell sickling leads to ischemic events that cause severe pain crises, which are one of the most difficult complications to manage for patients with SCD.

While treatment modalities such as the anti-sickling agent hydroxyurea are available, the median life expectancy for SCD patients is still of less than 40 years of age (*37*).

An estimated 100,000 patients with SCD live in the United States, and approximately 2 to 3 million people live with the disease worldwide (*37*). The prevalence is highest in sub-Saharan Africa, South America, the Middle East, and India. The regions in which SCD is most prevalent are endemic for malaria as sickle cell trait is known to have protective effects against *Plasmodium* parasites, which carry out part of their life cycle in erythroid cells (*37*). The exact mechanism of protection though is still unknown. The sickle mutation has thus been maintained in the population by the selective pressure of malaria and is thus a classic example of a balanced polymorphism (*37*). Several haplotypes of the sickle mutation, defined by specific restriction sites in the globin locus, have been identified in various geographical locations (*37*). These include the Bantu, Benin, Cameroon, Senegal, and Arab-Indian haplotypes, and suggested the sickle cell mutation had originated independently in these disparate geographical locations. A recent study that fine mapped variants surrounding the sickle variant using 1000 Genomes Project whole-genome sequencing data put the multi-center hypothesis into question (*38*). This study found that the sickle mutation in these different haplotypes is in stronger linkage disequilibrium with variants that are more closely located to it than the previously described haplotypes (*38*). This suggests that the sickle mutation may thus have arisen in sub-Saharan Africa in a single population when malaria was already present, and was subsequently disseminated throughout the African continent and the world by migration and slave routes. Future work will further clarify whether this truly reflects the evolutionary history of this mutation.

SCD prevalence is projected to increase over the next decades (*39, 40*). In developed countries, this is largely due to the improvements in care for SCD patients which include newborn screening, antibiotic prophylaxis, stroke prevention, hydroxyurea treatment, and allogeneic bone-marrow transplantation which still remains the only curative option for patients with SCD (*39*, *40*). The median survival for SCD patients in high-income countries is now of more than 60 years of age (*37*). In developing countries, approximately 90% of SCD deaths occur in the first 5 years of life due to a lack of access to care and in particular antibiotic prophylaxis (*37*). Improvements in basic aspects of healthcare in these countries will certainly lead to an increase in the number of people living with SCD in these countries. Genetic testing may offer opportunities to curb the increased prevalence of SCD, but having the proper resources to put these types of programs in place, and especially in developing countries where cultural differences can go against such efforts, it is still difficult to envisage such strategies being truly effective.

While SCD is caused in all patients by the same point mutation in the β-globin gene, its severity varies considerably between individuals (*37*). A number of environmental factors have been implicated as modifiers of the disease. High temperatures, high humidity, low air quality, smoking, and high altitude have all been suggested to favor red cell sickling, though the specific mechanisms underlying these effects remain to be fully elucidated (*37*). Genetic modifiers of SCD have been characterized in far greater detail. One of these is α-thalassemia, in which one or more copies of the α-globin genes are defective or deleted (*35*). α-thalassemia leads to reduced levels of hemoglobin in the cell and thus less detrimental polymerization, and is co-inherited in up to a third of SCD patients of African origin and up to half of those in the Mediterranean basin and India (*37*). The co-

inheritance of α-thalassemia with SCD is associated with reduced hemolytic events, but may favor pain crises and vaso-occlusion. Lower levels of α-globin are thought to lead to reduced tetramerization of sickle hemoglobin and thus reduce its polymerization. A number of other genetic modifiers have been implicated in SCD, including genes that play important roles in the development of cerebrovascular disease and stroke, but many of these associations still remain to be further elucidated (*37*).

1.3 Genetic modifiers of HbF silencing

1.3.1 GWAS studies

The best characterized genetic modifier of SCD and β-thalassemia is fetal hemoglobin (HbF). As explained above, fetal hemoglobin is expressed at low levels in most adults, but there is considerable inherited natural variation in the extent to which this occurs. Inherited mutations leading to significantly increased levels of HbF are referred to as Hereditary Persistence of Fetal Hemoglobin (HPFH) (*31*). In the context of hemoglobinopathies, HPFH has considerable clinical benefits. Observational studies have shown that SCD patients with high levels of HbF have fewer complications and significantly improved survival outcomes (Figure 1.4) (*41*). Uncovering the genetic underpinnings of HbF silencing in adult erythroid cells has thus been a major goal in the field. GWAS studies have been instrumental towards this by identifying genetic variants associated with high HbF levels (*42*). Three loci have emerged from these studies, with variants mapping to the HBS1L-MYB locus, BCL11A, and the β-globin locus itself (Figure 1.5).

HBS1L-MYB

The variants of the HBS1L-MYB locus associated with high HbF levels map to a regulatory element in this region that is also associated with a number of other erythroid traits such as mean corpuscular hemoglobin, mean corpuscular volume, and red blood cell count (*43*). Knockdown of MYB has been shown in primary erythroid cultures to induce HbF, though the mechanism by which it represses the fetal globin genes still remains to be fully elucidated (*31*). Targeting MYB for therapeutic purposes may present challenges though as this factor plays diverse roles in the hematopoietic system, and its depletion may have adverse effects on erythroid viability and differentiation (*43*).

BCL11A

BCL11A is a transcription factor that has now been extremely well characterized as a major repressor of fetal hemoglobin. Initial studies showed that a specific isoform of BCL11A is expressed in adult erythroid cells (*44*) (BCL11A XL, which contains a C-terminal zinc-finger required for HbF silencing (*45*)) and knockdown of BCL11A in these cultures induces fetal hemoglobin. The effects on fetal hemoglobin were remarkably specific and did not impair cell viability or erythroid differentiation (*44*). Further murine in vivo models showed that Bcl11a knockout also induced mouse embryonic type globins (*46*), as well as a correction of red blood cell sickling in a humanized mouse model of SCD (*47*). The mechanism by which BCL11A functions to repress the fetal globin genes had remained elusive until recently. BCL11A had previously been shown to associate with multiple members of the NuRD complex (*48*), which are all known transcriptional repressors. It had remained unclear though whether BCL11A directly recruits these factors

to the fetal globin promoters, or other regions within the globin locus. This was primarily due to technical challenges as BCL11A is a notoriously difficult factor on which to perform traditional chromatin immunoprecipitation. Two recent studies have shown either by chromatin immunoprecipitation using an ER-tagged BCL11A expressing cDNA cassette or Cut-&-Run immunoprecipitation that BCL11A occupies multiple regions in the β-globin locus (*45*, *49*). Notably, these studies showed that BCL11A occupies a specific region (-115 cluster) in the proximal promoter of the γ-globin genes (Figure 1.6). BCL11A may thus function by recruiting transcriptional repressors including the NuRD complex to transcriptionally silence these genes. This was confirmed by mutating the BCL11A binding sites in these promoters, which abrogated BCL11A occupancy (*49*), and induced fetal globin expression (*45*, *49*).

BCL11A is expressed in multiple tissues, including the erythroid lineage, B-cells, and neurons. In erythroid cells, BCL11A expression is controlled by a regulatory element that lies in its second intron (*50*). The GWAS variants originally implicating BCL11A in HbF repression map to this particular element, which is characterized by active enhancer marks in adult erythroid cells including H3K27ac and H3K4me1 (*50*). Chromatin conformation capture further showed the element forms long-range contacts with the BCL11A promoter (*50*). Deleting this element in murine and human erythroid systems leads to a significant decrease in BCL11A expression, and concomitantly induces HbF levels (*50*). Saturating mutagenesis screens of the BCL11A enhancer showed with greater precision that a GATA element in this enhancer plays a key role in regulating BCL11A expression (*51*).

β-globin locus

A number of HPFH variants map directly to the β-globin locus, some of which lie in the promoters of the γ-globin genes. One of these variants creates a *de novo* TAL1 binding site at the -175 site upstream of the γ-globin gene promoters (*52*). Knocking-in this variant into cells that normally express low HbF levels led to transcriptional activation of these genes (*52*). This was accompanied by the recruitment of the looping factor LDB1 and the formation of long-range contacts with the LCR, further showing the importance of this regulatory element in controlling globin expression (*52*). Another recent study showed that a CCAAT box approximately 100bp upstream of the γ-globin TSS plays important roles in repressing the fetal globin genes (*53*). Directed mutagenesis of these regions led to a significant increase in HbF in a human erythroid cell line as well as primary erythroid cultures (*53*). Deep sequencing of these edited regions showed a strong enrichment for a specific 13bp deletion, most likely due to the fact that there is a region of microhomology within these elements which may favor these precise edits (*53*). The fact that such precision can be achieved by bulk editing offers exciting opportunities for the therapeutic targeting of these regions. Mechanistically, this region is a BCL11A binding site as described above (*45*, *49*), suggesting that deleting this element may abrogate the occupancy of this factor and in turn allow for transcriptional activation.

Other HPFH variants map to a region between the γ-globin and the adult *δ*- and β-globin genes. This region contains the pseudogene HBBP1 and the non-coding gene BGLT3 (*54*). Two HPFH variants are located within the second intron of HBBP1 (*54*). A recent study showed that this region plays an important role in controlling higher-order chromatin structure within the β-globin locus during these different developmental

timepoints. Hi-C studies showed that the overall chromatin conformation of the locus is very similar between fetal and adult stage, and that the HBBP1/BGLT3 region lies at a sub-TAD boundary (*54*). Finer resolution conformation capture studies by Capture-C showed that during fetal stages the HBBP1/BGLT3 region interacts with two sites that flank the β-globin locus (HS5 and 3'HS1). This positions the LCR and the γ-globin genes in the same sub-TAD, thus favoring fetal hemoglobin expression. During adult stages, the HBBP1/BGLT3 region interacts with the ε-globin promoter, thus preventing the LCR from interacting with the γ-globin genes. Deletion of the HBBP1/BGLT3 region lead to transcriptional activation of γ-globin, further supporting this region plays key roles in regulating hemoglobin switching (*54*). Two recent studies showed that there is a strong BCL11A binding site within this region (*45*, *49*). The HBBP1/BGLT3 region was shown to interact with HS5 and 3'HS1 upon BCL11A knockout in adult erythroid cells, further suggesting that BCL11A may play a role in regulating the chromatin structure of the globin locus through this element (*54*).

Several large deletions within the β-globin locus have also been shown to be associated with high fetal globin levels in adulthood (*55*). The Corfu deletion is one such example, and is proximal to the adult *δ*-globin gene (*31*). There are also larger deletions removing both the *δ*- and β-globin genes identified in Sri-Lankan families who also show increased levels of fetal hemoglobin (*55*). In contrast, a number of $(\delta\beta)^0$-thalassemia deletions have been described that also involve both adult genes but do not show elevated HbF levels (*55*). This may be due to the fact that these types of deletions do not involve the region overlapping the Corfu deletion. ChIP studies have suggested that this region upstream of

the δ-globin promoter may be a BCL11A binding site (*55*), but how this element controls the preferential expression of adult globins still remains to be fully elucidated.

1.3.2 Additional factors involved in fetal globin repression

While GWAS was instrumental in identifying regulators of fetal hemoglobin, other factors that were not originally identified in these studies also appear to play roles in regulating fetal hemoglobin. Recently, a study showed that the transcription factor ZBTB7A, also known as LRF, is also a major repressor of fetal hemoglobin (*56*). LRF knockout in murine models lead to a significant increase in either murine embryonic type globins (*56*). Similarly, knocking out LRF in human erythroid cells, and knocking it down in primary human erythroid cultures elevated HbF, though depleting this factor affects erythroid maturation (*56*). LRF silences fetal hemoglobin independently of BCL11A as double knockouts for these two factors show additive effects on HbF expression (*56*). Mechanistically, LRF associates with members of the NuRD complex, but some of its co-factors are distinct from BCL11A (*56*). Recent chromatin immunoprecipitation studies have shown that LRF occupies a regulatory element in the gamma-globin promoters (-200 cluster), which is upstream of the BCL11A binding site identified at the -115 cluster (Figure 1.6) (*49*). These distinct occupancy patterns further support the fact that these two factors operate independently of each other to repress fetal globin.

A number of other factors have also been implicated in HbF repression as well. The transcription factor KLF1, which plays key roles in erythroid maturation, has been shown to have HPFH variants mapping to it (*31*). Knockdown experiments in primary human erythroid cultures have shown that depleting this factor leads to induction of fetal

hemoglobin (*57*). This may be in part due to a reduction in BCL11A levels. It is possible that KLF1 regulates BCL11A as it has been shown to occupy its erythroid specific enhancer (*31*). These experiments though are confounded by the fact that KLF1 depletion adversely affects erythroid maturation, but nonetheless implicate KLF1 in as playing a role in globin switching.

The methyl-CpG binding domain MBD2 has also been implicated in repressing fetal hemoglobin as knocking-out this factor in murine models leads to a robust induction of embryonic type globins, and its knockdown in primary erythroid cultures leads to significant HbF elevation (*58*). DNA methylation may play an important role in maintaining the fetal globin genes in a repressed state as DNMT1 has also been shown to play a role in this process (*48*), and BCL11A depletion leads to aberrant methylation within the γ-globin gene promoters (*47*). Several other transcriptional repressors have also been implicated in this process including COUPTFII which occupies direct repeat elements in close proximity to the CCAAT box at the -115 cluster of the γ-globin promoters (*31*). It still remains to be determined to what extent these factors specifically regulate the fetal globin genes, and how they carry out these functions mechanistically.

1.4 Current efforts to raise HbF in adult patients

While treatments for hemoglobinopathies have improved over the past decades with the use of antibiotic prophylaxis and bone marrow transplantation, these diseases still remain challenging to manage clinically. Hydroxyurea is currently the only FDA-approved drug for inducing HbF for these patients, its efficacy is still limited and mechanism of action still unclear (*59*). A number of strategies are being pursued to develop improved

treatment modalities for hemoglobinopathies. These either rely on the permanent modification of HSCs, or pharmacologic approaches (*60*).

1.4.1 Gene therapy

Several gene therapy strategies are currently being developed and tested clinically for the treatment of thalassemias and SCD. The first of these approaches relies on the lentiviral transduction of a T87Q anti-sickling form of adult β-globin cassette in patient-derived HSCs (*61*). The polymerization of sickle hemoglobin requires the incorporation of two β^S subchains in tetramers to interface with other sickle globin subchains in fibers. Tetramers mixed for β^S and a non-sickle β-type globin polymerize much less efficiently. These antisickling effects are most pronounced for γ- and δ-globin as these have fewer residues that can interact with β^S-globin. Specific residues within the δ-globin protein were found to be most potent in conferring anti-sickling properties, which lead to the development of the T87Q variant of β-globin which is very effective in preventing sickling when overexpressed (*61*).

Several vectors have been developed to robustly express the T87Q β-globin variant *in vivo*, and use a mini-LCR in tandem with the β-globin promoter to drive the expression of this transgene to high enough levels to prevent red cell sickling. These vectors are transduced into patient-derived HSCs which are then re-administered to patients by bone marrow transplantation. The pharmaceutical company Bluebird Bio is now conducting clinical trials to test the safety and efficacy of these approaches (drug product LentiGlobin BB305) in small cohorts of SCD patients (*62*) and thalassemia patients (*63*). One treated SCD patient experienced no sickle-related complications 15 months post-transplant and

showed significant reduction in red cell sickling (*62*). In thalassemia patients, treatment with the BB305 drug product lead to a significant or complete reduction in the need for transfusions (*63*). No adverse complications due to vector integration have been reported thus far, and further observation will be required to fully assess the safety profile of this approach. Another gene therapy approach for the treatment of hemoglobinopathies uses a vector expressing γ-globin as a transgene (*64*). This was shown to be effective in a humanized pre-clinical murine model of SCD by bone marrow transplantation. This approach has yet to be tested clinically, but further demonstrates the potential for these types of modalities.

1.4.2 Gene editing

As overviewed above, BCL11A is regulated by an enhancer in its second intron that confers erythroid specific expression (*50*). Saturating mutagenesis screens have narrowed down specific regions within this element required for its activity (*51*). Editing these sites in HSCs would thus confer a permanent abrogation of BCL11A expression once engrafted in patients by bone marrow transplantation. Phase I clinical trials led by Vertex/CRISPR Therapeutics and Bluebird Bio are beginning this year to test the safety of this approach. One important aspect to assess for these approaches will be off-target effects. While CRISPR editing has been shown to have minimal off-target effects by GUIDE-seq and CIRCLE-seq, it still remains to be determined to what extent these occur *in vivo*, and whether any clonal events will occur. Furthermore, recent studies have shown that gene editing may lead to the selection of p53 negative clones. P53 is normally activated in response to double strand breaks to favor repair and block progression through the cell

cycle. Cells that can progress with edits may be able to do so in part due to lower levels or reduced function of this transcription factor (*65*). These will all be important aspects to consider for these approaches, but nonetheless editing of the BCL11A erythroid specific enhancer offers promising opportunities for the permanent correction of these diseases.

1.4.3 Pharmacologic approaches

Gene therapy and gene editing approaches offer exciting opportunities for the treatment of hemoglobinopathies and are rapidly reaching the clinic. Non-myeloablative conditioning regimens using CD117 (cKit) monoclonal antibodies to target recipient HSCs are one example of how bone marrow transplantation approaches can further improve in the short-term (*66*). But these methods still require significant resources which are still currently not available to the global population of patients with hemoglobinopathies. Pharmacologic agents are thus still needed to manage these diseases clinically.

Prior to hydroxyurea's (HU) discovery as an HbF-inducing drug, the primary compound known to induce HbF was the DNA methylase inhibitor 5-azacytidine which reduces methylation of the fetal globin genes (*67*). Its off-target and cytotoxic effects though prevented its clinical use. HU was first shown to elevate HbF levels in non-human primate models approximately 30 years ago with limited cytotoxicity, and further shown to have clinical benefits for patients with SCD (*68*). The mechanism of action of HU is still unknown, though it has been proposed to induce HbF through cell cycle effects. HU is also metabolized to nitric oxide which may ameliorate the vascular pathology seen with SCD (*67*). HU is currently the only available FDA-approved drug for inducing HbF for patients with SCD.

Several other pharmacologic agents have been shown to induce HbF in erythroid culture systems. One of these is the H3K9me1/2 histone methyltransferase EHMT1/EHMT2 (also known as G9a) inhibitor UNC0638 (*69*). Treatment of erythroid cells with this compound leads to an increase in HbF levels with little effect on erythroid maturation. This was accompanied by a global decrease in H3K9me1/2 levels, including at the fetal globin genes, with a reciprocal increase in H3K9ac. RNA-seq assays showed that γ-globin transcripts were amongst the most strongly increased, but also showed broader perturbations to gene expression profiles upon treatment with UNC0638 (*69*). This compound is now being developed for clinical purposed by Epizyme, and further work will determine its safety, efficacy, and off-target profile *in vivo*.

Another compound recently shown to induce HbF levels is the drug pomalidomide (*70–72*). A derivative of thalidomide, it is currently FDA-approved for the treatment of multiple myeloma. It binds cereblon to activate its E3-ligase activity, which leads to ubiniquination and proteasomal degradation of Ikaros and Aiolos (IKZF1 and IKZKF3, respectively), which promotes cyotoxic effects in myeloma cells (*72*). Treatment of erythroid cells with pomalidomide led to a significant increase in HbF levels (*72*). While this led to a delay in early erythroid maturation at the BFU-E to CFU-E transition, later stages of erythroid maturation were relatively unperturbed (*72*). Pomalidomide downregulated a number of regulators of fetal hemoglobin, including a pronounced down-regulation of BCL11A (*72*). The specificity of these effects and mechanisms underlying them still remain to be determined, but may be independent of IKZF1/3.

A number of pharmacologic approaches have been identified to raise HbF levels for the treatment of hemoglobinopathies. Given the rising global health burden that these

diseases represent, there is still a strong need to develop improved small molecule strategies for these types of disorders. In the next chapters, we will overview an approach we have used towards this goal, showing how genetic screening approaches using CRISPR-Cas9 can be used to identify regulators of fetal hemoglobin that might serve as therapeutic targets or combined with other known pharmacologic inducers to increase HbF levels.

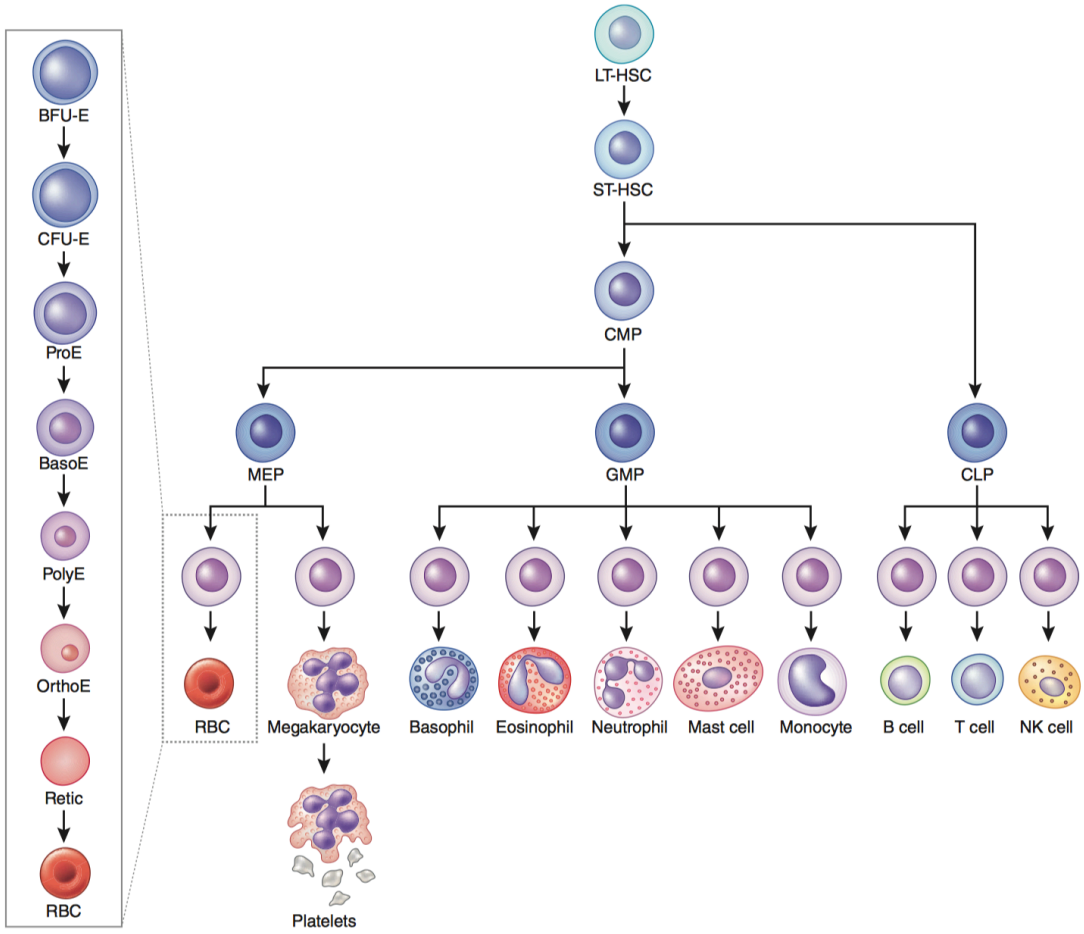

Figure 1.1 – Hematopoeitic tree from (*60*). Hematopoietic lineages arise from a hematopoietic stem cell (HSC) which can self-renew and differentiate to a Common Lymphoid Progenitor (CLP) which gives rise to lineages of the adaptive immune system, and a Common Myeloid Progenitor (CMP). CMPs give rise to lineages of the innate immune system through a Granulocyte-Macrophage Progenitor (GMP), and to the erythroid and megakaryocyte linages through a Megakaryocyte-erythroid progenitor (MEP). Erythroid differentiation is shown on the left.

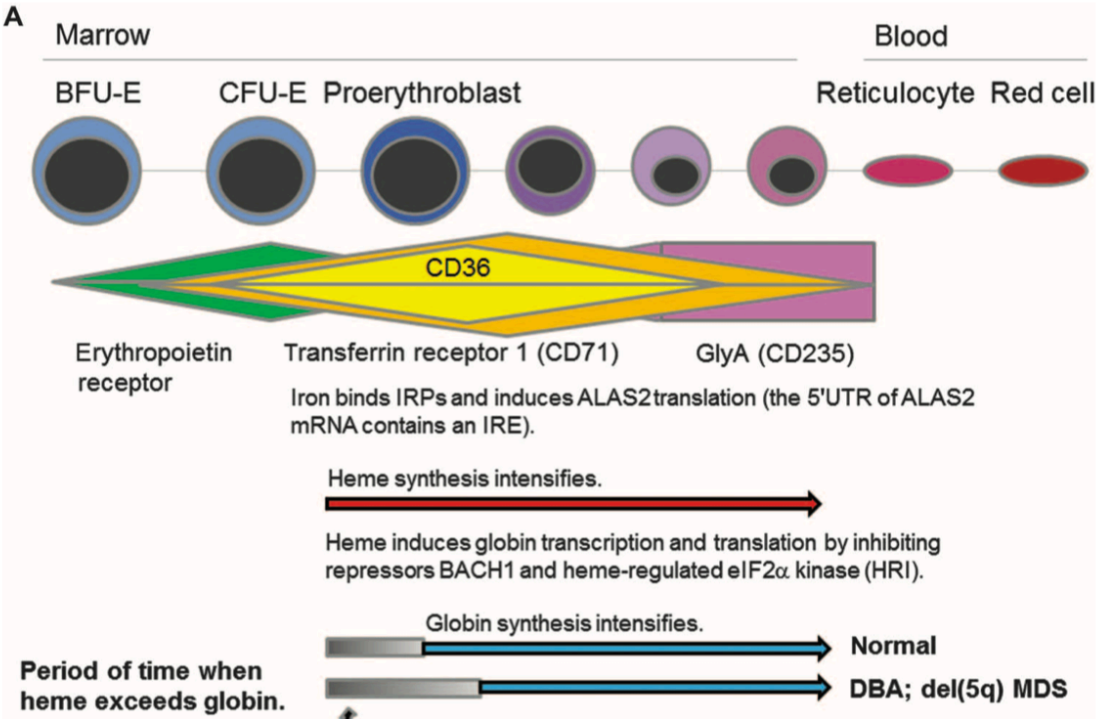

Figure 1.2 – Timing of heme and globin production during erythroid maturation, from (*26*). Stages of erythroid maturation are illustrated at the top. Surface marker expression profiles during erythroid maturation are shown immediately below. Heme synthesis begins early in erythroid maturation, and shortly thereafter is met with increased globin production. There is a short window early in erythroid maturation where heme levels exceed globin levels. In the context of disorders affecting globin protein synthesis by impairment of the translational machinery such as DBA and 5q minus MDS, globin synthesis does not match heme levels and this leads to impaired erythropoiesis.

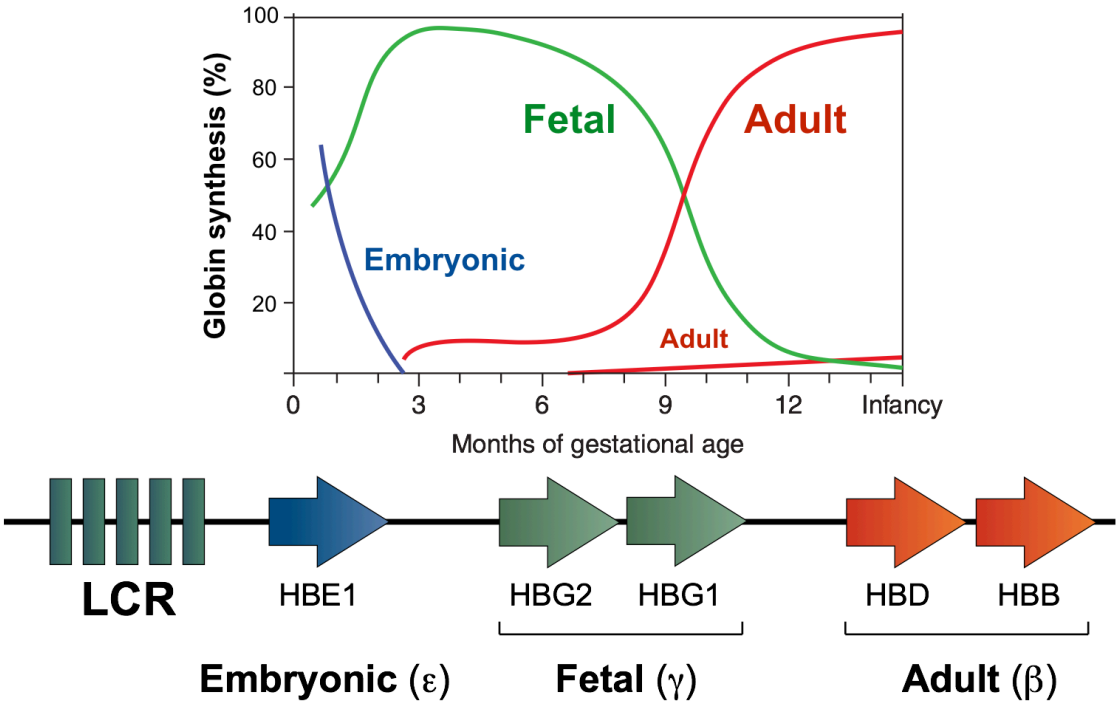

Figure 1.3 – Organization of the β-globin locus and hemoglobin switching (*31*). The β-globin locus encodes multiple β-type globin genes that are expressed during specific developmental time frames. These genes are regulated by a regulatory element termed the Locus Control Region (LCR) that lies upstream of the locus. The fetal type genes are primarily expressed during gestation. Shortly after birth expression shifts to adult type genes. In most adults HbF constitutes only 0.5% of total β-type globin synthesis. There is however natural variation in the extent to which HbF is silenced in adulthood, which has important clinical implications in the context of genetic disorders that affect adult type globin genes, such as β-thalassemia and SCD.

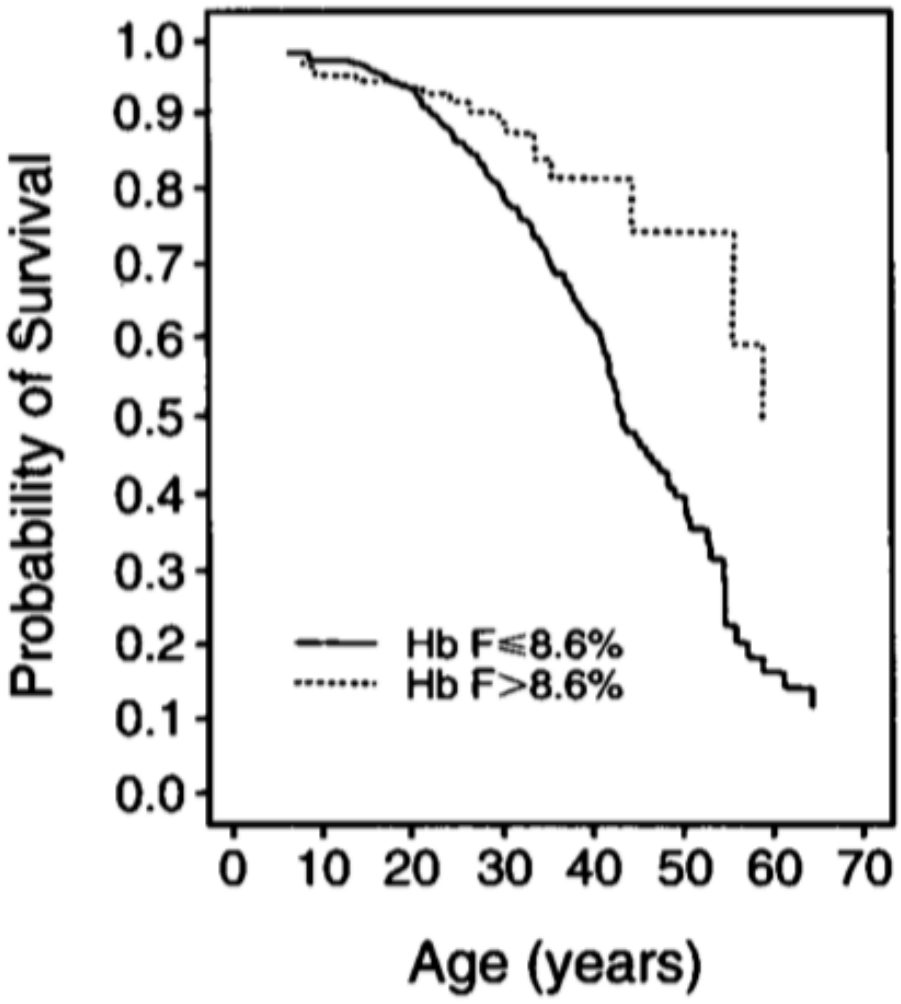

Figure 1.4 – HbF levels and survival outcomes for SCD patients (*41*). These results were obtained from an observational study of SCD patients who were stratified into 2 groups as either having high levels of HbF (above 8.6%) and low levels of HbF (below 8.6%). Survival outcomes are shown for each group. High levels of HbF are associated with significantly improved survival outcomes for these SCD patients groups.

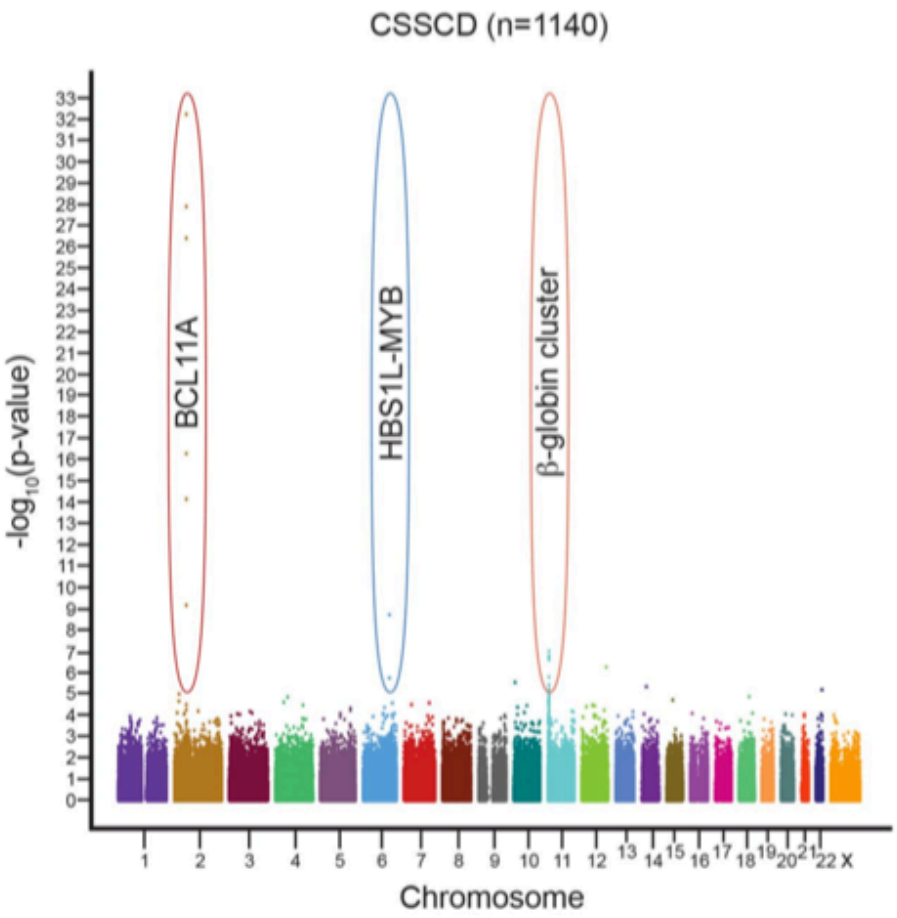

Figure 1.5 – GWAS for HbF as a quantitative trait from (*42*). These studies identified 3 loci with variants associated with high HbF levels: HBS1L-MYB, BCL11A and the β-globin locus.

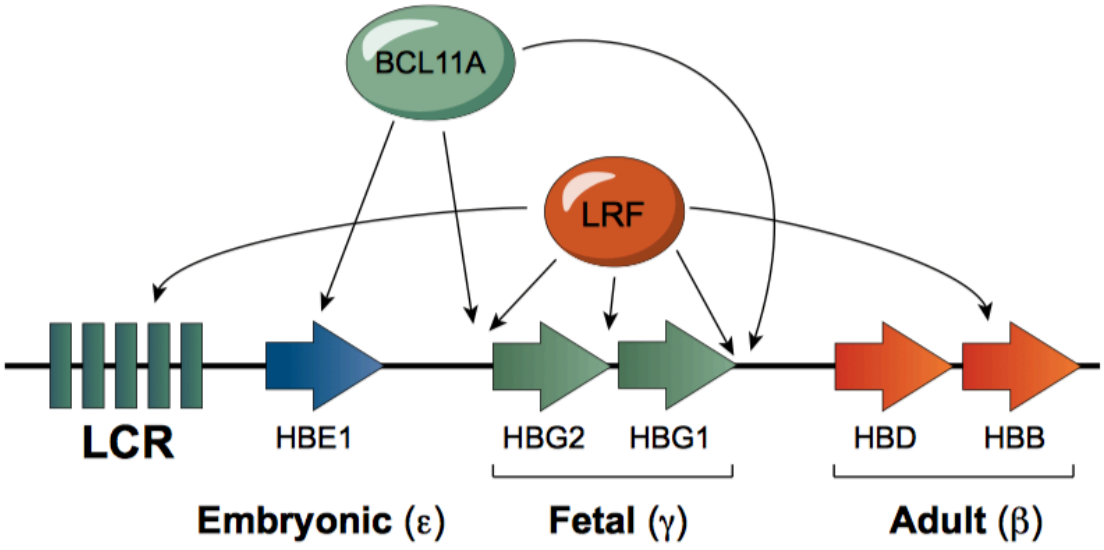

Figure 1.6 – Model for repression of fetal globin genes by BCL11A and LRF, adapted from (47), BCL11A and LRF occupy multiple regions within the globin locus to silence fetal globin genes, including by direct occupancy of their promoters.

CHAPTER 2: MATERIALS AND METHODS

Cell culture

HUDEP2 cells were cultured as described (*73*). Cells were maintained in StemSpan™ Serum Free Medium (SFEM, StemCell Technologies, Cat. #09650) supplemented with 50ng/ml human Stem Cell Factor (hSCF, Peprotech, Cat. #300-07), 10μM dexamethasone (Sigma, Cat. #D4902), 1μg/ml doxycycline (Sigma, Cat. #D9891), 3IU/ml erythropoietin (Amgen, Cat. #55513-144-10), 1% penicillin/streptomycin (ThermoFisher, Cat. #15140122). Cells were kept at a density of less than 1.0 million/ml. Differentiation was achieved by growing cells for 7 days (one medium change at day 3 or 4) in IMDM (Mediatech, Cat. #MT10016CV) supplemented with 50ng/ml hSCF, 3IU/ml erythropoietin, 2.5% fetal bovine serum, 250μg/ml holo-transferrin (Sigma, Cat. #T4132), 1% penicillin/streptomycin, 10ng/ml heparin (Sigma, H3149), 10μg/ml insulin (Sigma, Cat. #I9278), 1μg/ml doxycycline.

CD34+ cultures. Peripheral blood mononuclear cells (PBMCs) were obtained from the University of Pennsylvania Human Immunology Core. CD34+ cells were purified from using MACS MicroBead kit (Miltenyi, Cat #130-100-453) and cultured in three phases in IMDM supplemented with 3IU/ml erythropoietin, 2.5% human male AB serum (Sigma, Cat. #H4522), 10ng/ml heparin, 10μg/ml insulin. Phase-I medium was supplemented with 100ng/ml hSCF, 5ng/ml IL-3 (Peprotech, Cat. #200-03), 250μg/ml holo-transferrin. Phase-II medium was supplemented with 100ng/ml hSCF and 250μg/ml holo-transferrin. Phase-III medium was supplemented with 1.25mg/ml holo-transferrin. Cells were maintained in

Phase-I from the day of collection until day 8. Cells were spin-infected (see virus preparation and infections below for full details) and transitioned to Phase-II media, and finally transitioned to Phase-III media on day 13 of culture. RNA samples for RT-qPCR were harvested on day 13 of culture for optimal RNA yield. Western Blot, HPLC, and HbF FACS samples were harvested on day 15 of culture.

Cell lines for sgRNA-scaffold optimization. A-375 cells were cultured in DMEM, HCT116 and A549 cells were cultured in DMEM/F-12, K562 cells were cultured in RPMI1640. All cell lines used in this study were mycoplasma free.

Plasmids

The Cas9 expression vector was constructed by inserting the 5′-3xFLAG-tagged human-codon optimized Cas9 cDNA from *Streptococcus pyogenes* (Addgene: #49535) into a lentiviral EFS-Cas9-P2A-Puro expression vector using the In-Fusion cloning system (Clontech: #638909). All sgRNA encoding oligonucleotides were inserted into a lentiviral U6-sgRNA-EFS-GFP expression vector (LRG, Addgene: #65656) using a BsmBI restriction site. Oligonucleotides corresponding to the shRNAs were annealed and cloned into the same vector as were sgRNAs with the sgRNA scaffold removed by a BsmBI-EcoRI double digestion resulting in LRG-U6-shRNA-EFS-GFP. See Table S4 for a list of shRNA sequences used for this study.

Virus preparation and infections

Virus was produced in HEK293T cells grown in DMEM supplemented with 10% Fetal Bovine Serum, 2% penicillin/streptomycin, 1% L-glutamine, 100μM sodium pyruvate. HEK293Ts were plated in 10cm plates 18-24h prior to transfection such that cells were 90-100% confluent by the time of transfection. Expression vectors were mixed in a 4:3:2 ratio with packaging plasmids, PAX2, and VSVG envelope plasmid, respectively, in 500μl OPTI-MEM (ThermoFisher Scientific, Cat #31985070), and added to 500ul of 160μg/ml polyethylenimine (PEI, Polysciences Cat. #23966) in OPTI-MEM for precipitation. Plasmid precipitations were added to HEK293T cells and transfections were incubated at 37°C, 5% CO_2 for 6 hours, after which fresh media was placed on the cells. Viral supernatants were harvested 24 hours and 48 hours post-transfection and pooled.

One million HUDEP2 cells were infected with Cas9- or single sgRNA-containing virus by spin-infection with 1ml of viral supernatant supplemented with 8μg/ml polybrene and 10mM HEPES buffer. Spin-infections were carried out at 2,250 rpm for 1.5 hours at room temperature in 12-well plates. HUDEP2-Cas9 cells were selected starting 48-hours post-infection for 5 days in 1μg/ml puromycin. sgRNAs were transduced consistently above 90% GFP+ levels. See below for full details on transductions for the kinase domain sgRNA-library.

For infection of primary CD34+ derived erythroid cells with shRNA containing virus, viral particles required concentration due to lower infectivity of these cells. Viral supernatants were concentrated with LentiX-concentrator (Clontech, Cat. #631232) per manufacturer specifications. Viral pellets were resuspended in Phase-I media. Cells were spin-infected with concentrated virus from one 10cm plate (24-hour and 48-hour viral supernatants pooled) supplemented with 8μg/ml polybrene and 10mM HEPES per 1

million cells. Spin-infections were carried out at 2,250rpm for 1.5 hours at room temperature in 12-well plates, followed by incubation at 37°C, 5% CO_2 overnight. Cultures were then transitioned to Phase-II media.

TIDE analysis

CRISPR-Cas9 targeted regions were amplified by PCR from genomic DNA (gDNA) from HUDEP2 cells transduced with a sgRNA, as well as from an uninfected HUDEP2 control cell line. Primers were designed to flank targeted regions 250-300bp on either side. Amplifications were done with AccuPrime™ Pfx Supermix (ThermoFisher Scientific Cat. #12344040) per manufacturer specifications. Amplicons were purified with the QIAquick PCR purification kit (Qiagen, Cat. #28106), and Sanger sequenced at the Children's Hospital of Philadelphia sequencing core (NAPCore facility). Sequencing chromatograms were analyzed with the TIDE R-package (*74*). See Table S5 for a list of primers used for TIDE.

HbF flow cytometry and FACS

For HbF analysis cells were fixed in 0.05% glutaraldehyde for 10min, washed 3 times with PBS/0.1%BSA (Sigma, Cat. #A7906), and permeabilized with 0.1% Triton X-100 (Life Technologies Cat. #HFH10, prepared in PBS/0.1%BSA) for 5 minutes. Following one wash with PBS/0.1% BSA, cells were stained with HbF-APC conjugate antibody (Invitrogen, Cat. #MHFH05). For HUDEP2 samples, 2 to 5 million cells were incubated with 2μg HbF-APC antibody for 30 minutes in the dark at room temperature. For primary CD34+ erythroid cells, 2 to 5 million cells were incubated with 0.4μg HbF-

APC antibody for 10 minutes in the dark at room temperature. Cells were then washed twice with PBS/0.1%BSA. Flow cytometry was carried out on a BD FACSCanto™ and cell sorting on a BD FACSJazz™ at the Children's Hospital of Philadelphia flow cytometry core.

Kinase domain-focused CRISPR sgRNA library construction

A comprehensive human kinase gene list was taken from a previous study (*14*), and catalytic enzymatic domains retrieved from the NCBI database of conserved domain annotations. sgRNAs were designed against 496 kinase domains of 482 kinase genes, with 6 sgRNAs per domain. 26 sgRNAs targeting essential genes and 50 non-targeting sgRNAs were spiked-in as positive and negative controls, respectively. In total, the library contains 3052 sgRNAs. All the sgRNAs were designed using design principles previously reported and were filtered based on predicted off-target effects (*75*). sgRNAs were synthesized in a pooled format on an array platform (Twist Bioscience) and then PCR cloned into the LRG vector with sgRNA2.1 backbone (LRG 2.1 vector, see Supplementary Text for full description). To ensure proper representation of sgRNAs in the pooled lentiviral plasmids, the library was analyzed by deep-sequencing on a MiSeq instrument (Illumina), which confirmed that 100% of the designed sgRNAs were cloned in the LRG2.1 vector with over 95% of sgRNAs being within 5-fold of the mean read count (data not shown).

CRISPR-Cas9 screen

The kinase domain-focused sgRNA library was packaged in lentivirus as indicated above. Viral titer was determined by serial dilution. HUDEP2-Cas9 cells were transduced with the kinase-domain library at a low multiplicity of infection (MOI 0.3-0.5) such that 30 to 50% of cells were GFP positive (MOI for these studies was of ~0.4, see Fig. S2E). 9 million cells were infected in total to yield 1000x coverage of the sgRNA library in the GFP+ population. GFP+ cells were sorted by FACS on day 2 post-infection.

GFP+ cells were then cultured in HUDEP2 media (as indicated above) for an additional 6 days (total 8 days post-infection). On day 8 post-infection, cells were transitioned to differentiation media (as indicated above), and differentiated for 7 days. On day 15 post-infection, cells were stained for HbF as indicated above, and sorted into HbF high and HbF low populations (see Fig. 1B) as previously described (*51*).

Genomic DNA (gDNA) was harvested from these samples by phenol/chloroform (Fisher Scientific Cat #BP1752I-100) extractions per standard methods. sgRNAs were amplified with Phusion Flash High Fidelity Master Mix Polymerase (ThermoFisher Scientific, Cat. #F-548L) per manufacturer specifications with the LRG F2/R2 primer pair (see Table S6). Reactions were done with 23 cycles of amplification with 100ng of gDNA, and 50 parallel reactions were performed to maintain sgRNA library representation. PCR reactions were then pooled for each sample and column purified with QIAGEN PCR purification kit. PCR products were subjected to Illumina MiSeq library construction and sequencing. First, PCR products were end repaired with T4 DNA polymerase (New England BioLabs, NEB), DNA polymerase I (NEB), and T4 polynucleotide kinase (NEB). Then, an A-overhang was added to the end-repaired amplicons using Klenow DNA Pol Exo- (NEB). The A-overhang DNA fragment was ligated with diversity-increased

barcoded Illumina adaptors followed by seven pre-capture PCR cycles with primer pair PE-5/PE-7 (see Table S6). Samples were then purified with the QIAquick PCR purification kit. sgRNA library concentrations were quantified on a 2100 Bioanalyzer (Agilent). The barcoded libraries were pooled at an equal molar ratio and subjected to massively parallel sequencing through a MiSeq instrument (Illumina) using 75-bp paired-end sequencing (MiSeq Reagent Kit v3; Illumina MS-102-3001).

The sequencing data were de-barcoded and trimmed to contain only the sgRNA sequence, and subsequently mapped to the reference sgRNA library without allowing any mismatches. The read counts were calculated for each individual sgRNA and normalized to total read counts. Normalized read counts of sgRNAs in HbF high and HbF low populations were $\log_2$ transformed in R, and graphical representation was done using the R-package ggplot2.

Western Blot

Western blotting was performed using standard procedures. Primary antibodies: HRI (1:1000, Biosource, Cat. #MBS2526743), eIF2α-Phospho (1:1000, anti-EIF2S1-phospho-S51, Abcam, Cat. #ab32157), total-eIF2α (1:1000, Cell Signaling, Cat. #9722), BCL11A (1:1000, Abcam, Cat. #19489), LRF (1:1000, eBioscience Cat. #), GATA1 (1:1000, Santa Cruz, Cat. #sc-265), γ-globin (1:1000, Santa Cruz, Cat. #sc-21756), B-actin (1:1000, Santa-Cruz, Cat. #sc-47778). Secondary antibodies: anti-rabbit (1:10,000, GE Healthcare, Cat. #NA934V); anti-mouse (1:10,000, GE Healthcare, Cat. #NA931V).

RT-qPCR

RNA samples were harvested in TRIzol™ (ThermoFisher Scientific, Cat #15596018), and purified with the RNeasy Mini kit which included an on-column DNAse treatment to remove genomic DNA (Qiagen, Cat. #74106). cDNAs were prepared by reverse-transcription using iScript Supermix (Bio-Rad, Cat. #1708841). qPCR reactions were prepared with Power SYBR Green (ThermoFisher Scientific) and run on a ViiA7 Real Time qPCR machine (ThermoFisher Scientific). Quantification was performed with the ΔΔCT method. Primers used for RT-qPCR are listed in Table S7.

RNA-seq

Total RNA was harvested in TRIzol™, and purified with the RNeasy Mini kit including a DNase treatment step to remove genomic DNA. Sequencing libraries were then constructed from 100 ng of purified total RNA using the ScriptSeq Complete Kit (Illumina cat# BHMR1224) according to manufacturer's specifications. In brief, the RNA was subjected to rRNA depletion using the Ribo-Zero removal reagents and fragmented. First strand cDNA was synthesized using a 5' tagged random hexamer, and reversely transcribed, followed by annealing of a 5' tagged, 3'-end blocked terminal-tagged oligo for second strand synthesis. The Di-tagged cDNA fragments were purified, barcoded, and PCR-amplified for 15 cycles.

The size and quality of each library were then evaluated by Bioanalyzer 2100 (Agilent Techniologies, Santa Clara, CA), and quantified using qPCR. Libraries were sequenced in paired-end mode on a NextSeq 500 instrument to generate 2 x 76 bp reads using Illumina-supplied kits. The sequence reads were processed using the ENCODE3 long RNA-seq pipeline (https://www.encodeproject.org/pipelines/ENCPL002LPE/). In brief, reads were

mapped to the human genome (hg19 assembly) using STAR, followed by RSEM for gene quantifications. Read counts were further analyzed with the R-package DESeq2.

Proteomics analysis via nLC-MS/MS

Cells were lysed in 100µl of cold lysis buffer (6M urea/2M thiourea, 50 mM ammonium bicarbonate, pH 8.0). Dithiothreitol (DTT) was added to a final concentration of 5 mM for 1 hour at room temperature and alkylated with 20 mM iodoacetamide in the dark for 30 minutes at room temperature. Samples were diluted 5x to reduce urea concentration. Proteins were then digested with trypsin (Promega) at an enzyme-to-substrate ratio of approximately 1:50 overnight at room temperature. Digestion was interrupted by adding 1% trifluoroacetic acid. Samples were desalted using Sep-Pak tC18 Plus Light Cartridge (Waters). The peptide samples were subsequently lyophilized.

Dried samples were resuspended in buffer A (0.1% formic acid) and loaded onto an Easy-nLC system (Thermo Fisher Scientific, San Jose, CA, USA), coupled online with a Q-Exactive mass spectrometer (Thermo Scientific). Peptides were loaded into a picofrit 18 cm long fused silica capillary column (75 µm inner diameter) packed in-house with reversed-phase Repro-Sil Pur C18-AQ 3 µm resin. A gradient of 105 minutes was set for peptide elution from 2-28% buffer B (100% ACN/0.1% formic acid), followed by a gradient from 28-80% buffer B in 5 min and an isocratic 80% B for 10 min. The flow rate was 300 nl/min. The MS method was set up in a data-dependent acquisition (DDA) mode. The full MS scan was performed at 70,000 resolution (FWHM at 200 *m/z*) in the *m/z* range 360-1200 and an AGC target of 10e6. Tandem MS (MS/MS) was performed at a resolution of 17,500 with an HCD collision energy set to 24, an AGC target of 2x10e4, a maximum

injection time to 100 msec, a loop count of 10, an intensity threshold for signal selection at 10e4, including charge states 2-4, and a dynamic exclusion set to 60 sec.

MS raw files were analyzed by MaxQuant software (*76*) version 1.5.5.1. MS/MS spectra were searched by the Andromeda search engine against the Human UniProt FASTA database (version November 2015). All parameters for the search were kept as default. Intensity-based absolute quantification (iBAQ) (*76*) was enabled for label-free quantification. Match between runs was enabled and set to 1 min window.

For data analysis, iBAQ values were $\log_2$ transformed and normalized by subtracting to each value the average value of the respective sample. Statistics was performed by using a two-tails homoscedastic t-test.

Hemoglobin HPLC

Cell were lysed in MilliQ H_2O. Hemolysates were cleared by centrifugation and analyzed for identity and levels of hemoglobin variants (HbF and HbA) by cation-exchange high-performance liquid chromatography (HPLC). We utilized a Hitachi D-7000 Series (Hitachi Instruments, Inc., San Jose, CA), and a weak cation-exchange column (Poly CAT A: 35 mm x 4.6 mm, Poly LC, Inc., Columbia, MD). Hemoglobin isotype peaks were eluted with a linear gradient of phase B from 0% to 80% at A_{410nm} (Mobile Phase A: 20 mM Bis-Tris, 2 mM KCN, pH 6.95; Phase B:20 mM Bis-Tris, 2 mM KCN, 0.2 M sodium chloride, pH 6.55). Cleared lysates from normal human cord blood samples (high HbF content), as well as a commercial standard containing approximately equal amounts of HbF, A, S and C (Helena Laboratories, Beaumont, TX), were utilized as reference isotypes.

Sickling assay

Cells were resuspended in 100μL HEMOX buffer supplemented with glucose (10 mM) and BSA (0.2 %) in a 96-well plate. Suspensions were subsequently incubated under Nitrogen gas at 37°C for 1 hour. 200μL of 2% glutaraldehyde solution was then added to the samples without exposure to air for immediate fixation. Fixed cell suspensions were spread onto glass microslides (Fiber Optic Center) and subjected to microscopic morphological analysis of bright field images (at 40x magnification) of single layer cells on an Olympus BX40 microscope fitted with an Infinity Lite B camera (Olympus), and the coupled Image Capture software. Images were randomized and blinded. The percentage of sickled cells for each condition was obtained by manually counting the number of sickled cells.

Wright-Giemsa stains

Cells were spun onto glass slides with a Cytospin4 (ThermoFisher Scientific, Cat#A78300003) at 1,000rpm for 3 min. Slides were allowed to dry for 5 minutes before staining with May Grünwald (Sigma Aldrich, MG1L-1L) for 2 minutes followed by 1:20 diluted Giemsa stain (Sigma Aldrich, GS-500 500ML) for 10 minutes. The stained slides were rinsed in water and allowed to dry for 10 minutes before a coverslip was sealed on the preparation with Cytoseal 60 (Thermo Scientific Cat #8310-4). The images were captured at 10X resolution on Olympus BX60 microscope using Infinity software (Lumenera corporation).

Polysome profiling

Cells were arrested in Phase II media with 100ug/ml cycloheximide (Sigma CC4859) for 15 minutes at 37C. Cells were then washed twice in ice cold 1x PBS with 100ug/ml cycloheximide, and resuspended in 500ul lysis buffer (10mM Tris-HCl (pH 7.4), 5mM MgCl2, 100mM KCl, 1% Triton X-100, 2mM DTT, 500U/mL RnasinPlus (Promega N2615), protease inhibitor cocktail (Sigma 11873580001)). Lysates were then dounced with a 26G needle 5 times, and cellular debris were cleared by centrifugation at 13,000rpm for 10 minutes. 10% of cleared lysates were set aside for input mRNA quantitation, and samples were snap-frozen in liquid nitrogen.

Cleared polysomal lysates were loaded on sucrose gradients and spun at 40,000rpm for 2 hours at 4C (Beckman SW41 rotor). Sixteen fractions of equal volume (~700ul) were collected and gradient profiles were measured by ultraviolet absorbance at 254nm with a UA-5 detector (ISCO). Fractions were incubated with 1% SDS and proteinase K for 30 minutes at 37C, and stored at -80C until further analysis.

RNA was purified from fractions by phenol-chloroform extraction. ERCC spike-in RNA was added to each fraction (1ul of 1:10 dilution of spike-in control per fraction, ERCC-spike in (ThermoFisher, 4456740)). 400ul of phenol (Phenol/Water 3.75:1 Invitrogen Cat. # 15594-047) was added to each fraction. Samples were vortexed for 30 seconds. 350ul of chloroform (ThermoFisher, Cat. #BP1145-1) was added, and samples were vortexed again for 30 seconds. Samples were spun down at 10,000rpm for 10 minutes, and aqueous layers were transferred to fresh tubes. One additional round of phenol-chloroform extractions were performed. Aqueous phases were split into two tubes with

350ul of sample each. 1.5ul of glycogen carrier, 30ul of 3M sodium acetate pH 5.5 (ThermoFisher, Cat. #AM9740) and 800ul of ice cold 100% ethanol were added to each final aqueous phases, and place at -80C overnight. RNA was precipitate by spinning samples at 13,000rpm for 30 minutes at 4C. Pellets were washed with 1mL of 75% ethanol, and RNA was resuspended in nuclease-free water. cDNAs were prepared using iScript Supermix (Bio-Rad, Cat. #1708841). qPCR reactions were prepared with Power SYBR Green (ThermoFisher Scientific) and run on a ViiA7 Real Time qPCR machine (ThermoFisher Scientific).

BCL11A cDNA rescue experiments

The BCL11A cDNA was a gift from Jian Xu (UT Southwestern) and subcloned into a lentiviral LRG-EFS-BCL11A-IRES-mCherry vector by In-Fusion cloning. Virus was prepared and transduced in respective cell lines as indicated above. To control for expression levels, the low 50% mCherry+ cells were sorted on a BD FACSJazz™ at the Children's Hospital of Philadelphia flow cytometry core. Cells were differentiated and analyzed as indicated above.

Mouse fetal liver cultures

E14.5 mouse fetal livers were obtained from CD1 pregnant females (Charles River strain code 022, req #V00018219:1, IACUC protocol #660). The EasySep™ Mouse Hematopoietic Progenitor Cell Enrichment Kit (Stem Cell Technologies cat. #19756) was used for hematopoietic progenitor isolation. Cells were transduced with shRNAs per spinfection protocol detailed above. Transduced cells were cultured for 48h in StemPro34

with supplement (Thermo Fisher cat #10639011), 1% L-glutamine, 1% P/S, 0.5U/ml erythropoietin, 1μM dexamethasone. Cells were harvested for RT-qPCR and Western Blot as indicated above.

Pomalidomide combination experiments

HUDEP2 cells were incubated with 10μM pomalidomide (Sigma, Cat. # P0018) for 2 days. Cells were then transitioned to differentiation media as indicated above for 7 days. During the course of differentiation cells were maintained in 10μM pomalidomide. Media was changed 3 to 4 days within the differentiation phase with fresh drug supplemented. HbF FACS and RT-qPCR were performed as indicated above.

CHAPTER 3: DOMAIN-FOCUSED CRISPR-SCREEN IDENTIFIES HRI KINASE AS A REGULATOR OF FETAL HEMOGLOBIN IN ADULT ERYTHROID CELLS

3.1 Chapter summary

Research described in this chapter was carried out in collaboration with the laboratory of Dr. Junwei Shi who designed and executed the sgRNA and Cas9 plasmid cloning, sgRNA scaffold optimization, sgRNA library construction and viral packaging. Xianjiang Lan, Nicole Hamagami, Laavanya Sankaranarayanan, Christopher E. Edwards, Saurabh Bhardwaj, and Carolyne Face executed experiments. Dave Posocco, Osheiza Abdulmalik, and Stella Chou carried cultures and assessment of sickle derived primary cultures. Simone Sidoli and Ben Garcia performed mass spectrometry and downstream analysis. Cheryl Keller, Belinda Giardine and Ross Hardison executed RNA-seq and read alignment. Xinjun Ji and Stephen A. Liebhaber helped design and perform the polysome profiling experiments.

The work described here forms the body of a manuscript that was published prior to the writing of this dissertation in Science (2018) with the following reference:

Grevet, J., Lan, X., Hamagami, N., Edwards, C., Sankaranarayanan, L., Ji, X., Bhardwaj, S., Face, C., Posocco, D., Abdulmalik, O., Keller, C., Giardine, B., Sidoli, S., Garcia, B., Chou, S., Liebhaber, S., Hardison, R., Shi, J. & Blobel, G. (2018). Domain-focused CRISPR screen identifies HRI as a fetal hemoglobin regulator in human erythroid cells. *Science* 361(6399), 285–290.

3.2 Introduction

The human β-globin gene cluster comprises an embryonic (ε-globin), two fetal (γ-globin) and two adult type (δ-globin and β-globin) genes (*31*). Diseases affecting the adult type β-globin genes, such as sickle cell disease (SCD) and some types of β-thalassemia manifest themselves after birth following the transition from fetal to adult hemoglobin production. The clinical course of these diseases is ameliorated when fetal hemoglobin (HbF) levels are elevated because of genetic causes or medical intervention (*41*). Genome wide association studies (GWAS) have identified three major loci that are linked to natural variation in HbF levels, including the transcription factor BCL11A (*31*). Patients with heterozygous loss of BCL11A exhibit elevated HbF levels throughout adult life (*77*), and BCL11A depletion in cultured human erythroid precursors raises HbF production (*44*, *46*). Loss of BCL11A in engineered mice that express human sickle hemoglobin stimulates fetal globin expression and ameliorates their disease phenotype (*47*).

Strategies to achieve long-term elevation of HbF levels in patients include gene therapy to force the expression of γ-globin, and gene editing to impede BCL11A expression (*50*, *64*). Current clinical risks, costs, and availability limit such therapies to a fraction of the patient population. The only FDA-approved drug to raise HbF is hydroxyurea which benefits many patients but is limited in its efficacy (*59*). Hence there is a need for pharmacologic approaches to increase HbF levels.

Two transcription factors, BCL11A and LRF (ZBTB7A) and their co-regulators mediate most of γ-globin transcriptional silencing in adult erythroid cells (*56*). However, transcription factors are inherently challenging to inhibit with small molecules. We carried out a CRISPR/Cas9 screen to target protein kinases, because in principle, they are

inherently controllable by small molecules (*78*). It has recently been shown that targeting sgRNAs to functional protein domains can significantly improve screening efficiency, as both in-frame and frameshift mutations contribute to generating hypomorphic alleles (*79*). We designed a library of sgRNAs targeting 482 kinase domains (6 sgRNAs per kinase, and 50 non-targeting sgRNAs as negative controls), which covers almost all annotated kinases in the human genome (*78*).

3.3 Screen optimization

Prior to performing the screen, we optimized the *Streptococcus Pyogenes* Cas9 sgRNA-scaffold for on-target activity (see Supplementary Text for details). This sequence adopts a well-defined secondary structure, with a repeat-antirepeat duplex linked to a tetraloop and 3 stem loops, all of which facilitate binding to Cas9 and stabilize the sgRNA (Fig. S1A,B) (*80*). Previous work showed that an A-U flip in the repeat-anti-repeat duplex removes a putative premature Pol-III stop codon, and that extending the tetraloop significantly improved sgRNA stability and Cas9 activity (*81*). By examining the Cas9 crystal structure, we noticed that stem loop 2 forms few contacts with Cas9 residues, similar to the tetraloop (*80*). Mutations in stem loop 2 are also well tolerated and allow for efficient Cas9 activity (*80*). We thus reasoned that extending stem loop 2 could further improve sgRNA stability similar to what had been done for the tetraloop, as this could further increase the double-stranded portion of the scaffold (Fig. S1B). We compared different lengths of such extensions (sgRNA 2.0 – sgRNA2.3).

To do so, we performed negative selection assays using the bicistronic sgRNA/GFP expression vector detailed above (LRG, Addgene: #65656). Cell lines were transduced

with sgRNAs targeting essential genes such that selection against the sgRNAs would directly reflect editing efficiency for the different sgRNA scaffold designs. Negative selection was estimated by taking the proportion of GFP+ cells remaining over time (flow cytometry analysis was performed on 96-well plates of cells using a Guava Easycyte HT instrument (Millipore), gating was performed on live cells (using forward and side scatter), before measuring of GFP+ cells) (Fig. S1C,D). We found that an extended hairpin structure in both tetraloop and stem loop regions of the sgRNA backbone (termed sgRNA2.1) significantly improved the magnitude of negative selection in various human cell lines induced by sgRNAs targeting DNA replication proteins RPA3 and PCNA (Fig. S1C,D).

We then tested the modified scaffold design in HUDEP2 cells, a human erythroid cell line with low HbF background levels (See Materials and Methods for full details) (*73*) which we used to perform the kinase domain screen (Fig. S2A). Control sgRNAs targeting the known HbF regulators BCL11A and the chromatin remodeler CHD4 (*48*), were introduced via lentivirus into engineered HUDEP2 derivatives stably expressing Cas9 (HUDEP2-Cas9, see Materials and Methods for full details). Indel distributions were assessed by TIDE (see Materials and Methods for full details) (*74*). In the context of the sgRNA 2.1 template we achieved up to 98% editing efficiency as early as 4 days post-infection (Fig. S2B) with most sequence perturbations being deletions of less than 10 base pairs (Fig. S2C). As expected, targeting BCL11A increased the number of HbF positive cells (*51*). Based on these results, we used the sgRNA 2.1 design for all subsequent CRISPR-based experiments.

sgRNA2.1 emerged as the most efficient in mutagenizing target loci in multiple cell lines, achieving up to 98% genome editing efficiency as early as 4 days following introduction into cells (Fig.S1 and Fig.S2). We cloned the kinase-domain sgRNA library into the sgRNA2.1 scaffold and introduced it into HUDEP2 cells stably expressing Cas9 (Figure 1A). HUDEP2 is an immortalized human cell line that can be induced to undergo erythroid differentiation and produce high levels of adult and low levels of fetal type hemoglobin (*13*). We used anti-HbF FACS to isolate the top 10% and bottom 10% HbF expressing cells (Fig. 1A,B) and deep sequenced sgRNAs as described previously (*51*). Non-targeting control sgRNAs were similarly distributed across HbF low and high populations, indicating that the screen did not bias the enrichment in either direction (Fig. 1C).

3.4 Kinase-domain CRISPR screen identifies HRI as a potential regulator of fetal hemoglobin

Heme-regulated inhibitor HRI (also known as EIF2AK1) was the only kinase for which all targeting sgRNAs were overrepresented in the HbF high population (Fig. 1C). HRI is one of four kinases known to phosphorylate the protein translation initiation factor eIF2α (Fig. S3) (*28*). Phosphorylation of eIF2α generally impedes mRNA translation, but enables the translation of transcripts that include one or more upstream open reading frames (uORFs) in their 5' untranslated regions. HRI is enriched in erythroid cells where it can be inhibited by its natural ligand heme (Fig. S4A,B). It is thought that this regulation balances heme levels with production of globin chains, as excess of either of these can be detrimental (*28*). The other eIF2α kinases are expressed at significantly lower levels in erythroid cells

(Fig. S4C). While small molecules have been developed against other members of the eIF2α kinase family, effective HRI-specific inhibitors are lacking.

3.5 Validation of HRI as a fetal hemoglobin regulator in HUDEP2 cells

We validated HRI as a HbF regulator by expressing all six "hit" sgRNAs (Fig. S5) individually in HUDEP2-Cas9 cells, along with a sgRNA against BCL11A as a benchmark, and a non-targeting sgRNA as a negative control. All six HRI targeting sgRNAs increased the fraction of HbF+ cells (Fig. 2A,B). Cells expressing HRI sgRNAs #3 and #5 were studied by Western blot, which revealed that the sgRNAs achieved strong depletion of HRI resulting in decreased eIF2α phosphorylation, and significantly increased γ-globin protein levels (Fig.2C). There were no changes in the protein levels of GATA1, suggesting that HRI loss did not impair cell maturation (Fig. 2C).

EIF2α phosphorylation is thought to broadly control protein translation, suggesting that HRI inhibition might lead to widespread changes in protein levels and secondary changes in gene expression. We measured protein abundances in undifferentiated and differentiated HRI-depleted HUDEP2 pools for sgRNA#3 and #5 by replicate whole cell mass spectrometry studies (see Fig. S6 A-C). Notably, protein amounts were changed relatively little in HRI-depleted differentiated HUDEP2 cells (Fig. 2D, Fig. S7A-D), and γ-globin was one of the most induced proteins (Fig. 2D, Fig. S7A, Table S1). We did observe a modest rise in global protein levels, perhaps due to a non-specific augmentation in protein translation resulting from lower eIF2α phosphorylation (Fig. S7A,B). Whole cell mass spectrometry detected the ~2000 and ~2500 most abundant proteins in differentiated

and undifferentiated samples, respectively. Thus, it is possible that significant changes were missed among low abundance proteins. Nevertheless, overall the effects of HRI depletion were unexpectedly specific for γ-globin.

In parallel, we carried out replicate RNA-seq experiments (see Fig. S8A) in differentiated HRI depleted HUDEP2 pools for sgRNA#3 and #5 and found that global transcript distributions were relatively unperturbed (Fig. 2E). γ-globin mRNA stood out as the most significantly induced in differentiated cells (Fig. 2E, Fig. S8B, Table S1). RT-qPCR experiments confirmed the effects on γ-globin mRNA and primary transcripts, suggesting that HRI depletion increases γ-globin transcription (Fig. S9A,B). We did not observe any noticeable changes in α-globin, ALAS2, BAND3 and GATA1 mRNAs (Fig. S9C), suggesting that erythroid differentiation is essentially normal in HRI deficient cells. These results were further validated in two independent clonal HRI depleted cell lines generated from the HRI sgRNA#5 pool (Fig. S10A-E).

3.6 Validation of HRI as fetal hemoglobin regulator in primary CD34+ derived erythroid cultures

HRI's role as a HbF repressor was further tested in primary erythroblasts derived from a three-phase human CD34+ culture system (see Materials and Methods). We depleted HRI with two independent shRNAs. For comparison, we used the strongest shRNAs against LRF amongst two previously published (*56*). We verified transduction levels by GFP flow cytometry (Fig. S11A). Both HRI shRNAs significantly increased the

proportion of HbF+ cells (Fig. 3A, B). The proportion of γ-globin mRNA as a percentage of total globin transcripts was also significantly increased upon HRI knockdown (Fig. 3C).

RNA-seq in cells with either shRNA demonstrated that the increases in γ-globin mRNA was among the most significant changes (Fig. 3D), similar to the results in HUDEP2 cells, with global transcript distributions for HRI knockdown samples being strongly correlated with those from the scrambled shRNA (Fig. 3D). Similar elevation of HbF was detected at the protein level as measured by HPLC (Fig. 3E, note donor to donor variation in the human culture system). HRI depletion did not appear to delay the maturation of these cells, as evidenced by similar transcript levels for maturation markers (Fig. 3D), cell surface markers CD71 and CD235a (Fig. S11B), and cell morphology (Fig. S11C).

Elevated HbF levels counteract sickling of cells expressing sickle hemoglobin (HbS) (*31*). To assess whether HRI depletion elicits anti-sickling effects, we knocked down HRI in CD34+ cells obtained from patients with sickle cell disease (SCD) (Fig. S12A). This led to comparable increases in γ-globin levels as in CD34+ cultures from healthy donors shown above (Fig. S12B,C), and, as before, did not seem to affect erythroid maturation (Fig. S12D). When grown under low oxygen tension cells expressing HRI shRNAs were less prone to sickling than controls (Fig. 3F), suggesting that HRI depletion might achieve therapeutically relevant levels of HbF.

3.7 Mechanism of HbF elevation upon HRI depletion

BCL11A and LRF are the major direct repressors of γ-globin transcription (*56*), suggesting that the effects of HRI might converge on one or both of these factors as they are expressed during overlapping time intervals of erythroid maturation (Fig. S13A,B). We examined BCL11A and LRF protein levels by Western blot using extracts from CD34+ derived erythroblasts expressing HRI shRNAs. Notably, BCL11A but not LRF was strongly depleted in both HRI knock-down samples (Fig. 4A) Western blots confirmed the loss of HRI, the resulting reduction in phospho-eIF2α, and the increase in γ-globin (Fig. 4A). GATA1 was unaffected, in agreement with ostensibly normal cell maturation under conditions of HRI depletion.

We tested whether HRI controls BCL11A levels directly by governing its translation, or indirectly, for example by controlling the translation of a nuclear factor that promotes BCL11A transcription. Both BCL11A mRNA and primary transcript levels were markedly reduced in HRI depleted cells, indicating that most if not all of the BCL11A loss is a result of transcription inhibition in HRI-depleted primary erythroid cells, HUDEP2 pools and clonal lines (Fig. 4A and S14 A-D). To measure a possible contribution of direct translational control of BCL11A mRNA by HRI we performed polysome profiling. HRI depletion modestly increased overall translation efficiency as polysomal fractions were slightly more abundant relative to monosomes (Fig. S15A). Moreover, in HRI depleted cells BCL11A mRNA levels were significantly reduced across all polysomal fractions, consistent with an overall loss in mRNA levels (Fig. S15B). To better visualize distribution across fractions, results were plotted as percentage of total BCL11A mRNA, which revealed little or no shift in the distribution pattern (Fig. S15C), suggesting that BCL11A

mRNA translation plays only a minor if any role in response to HRI loss. α-globin mRNAs tended to shift slightly to the larger polysome fractions (to the right), consistent with a subtle increase in their translation (Fig. S15C).

To assess to what extent BCL11A loss accounts for the effects of HRI depletion, BCL11A expression was restored in HRI-depleted HUDEP2 clonal cell lines. This reinstated γ-globin repression by 65% to 80% (Fig. 4B, Fig. S16A-D), suggesting that BCL11A is a major HRI effector. Of note, depleting HRI in primary definitive (adult-type) erythroid cells from murine fetal livers did not reduce BCL11A levels (Fig. S17 A-D), suggesting that certain pathways controlled by HRI might be species specific. In sum, BCL11A emerges as a HRI target in the control of HbF production in human adult erythroid cells.

As a proof of principle, we assessed whether HRI's effects on HbF can be amplified with pharmacologic HbF inducers. Pomalidomide has been recently shown to induce HbF in part by reducing BCL11A transcription (*72*). We tested the effects of pomalidomide treatment on two HRI depleted HUDEP2 populations and control HUDEP2 cells. For comparison we also included HUDEP2 cells in which BCL11A transcription was reduced using a sgRNA targeting the erythroid BCL11A enhancer. These combinations did not impair cell viability or maturation (Fig. S18A,D). All conditions alone elevated the fraction of HbF+ cells and increased γ-globin transcription as expected (Fig. 4C, Fig. S18B,C,E). The strongest cooperativity was observed when HRI depletion was combined with pomalidomide.

3.8 Discussion

Using an improved CRISPR-Cas9 domain-based screening approach we identified the erythroid specific kinase HRI as a potentially druggable target that is involved in HbF silencing. Our findings contrast with previous reports showing that stabilization of eIF2α-phosphorylation with the drug salubrinal can stimulate HbF synthesis in cultured human erythroid cells (*82*, *83*). Salubrinal acts downstream of HRI, and the mechanism underlying the induction of HbF remains to be fully elucidated. Hematopoietic stress can be associated with increased fetal hemoglobin levels. However, we believe that the effects of HRI depletion are not due to stress, since there are seemingly little adverse effects on cell viability and cell maturation as judged by cell morphological, surface phenotyping, transcriptome, as well as proteome analysis. Moreover, silencing of HbF in HRI depleted cells is largely restored upon BCL11A re-expression, consistent with a specific function of HRI.

Mice in which the HRI gene has been knocked out appear largely normal, including their hematological indices (*84*). This agrees with our results in human erythroid cell cultures, and suggests that HRI loss is generally well tolerated. However, HRI mice displayed impaired adaptation to erythropoietic stress induced by iron deficiency, and exacerbated protoporphyria and thalassemia phenotypes (*85*, *86*). It remains to be seen whether HRI inhibition in SCD patients would elevate HbF levels sufficiently to improve outcomes. HRI inhibition elevated HbF levels to a point at which it reduced cell sickling in culture, suggesting that pharmacologic HRI inhibitors might provide clinical benefit in SCD patients. Moreover, based on our results, combining HRI inhibition with an additional pharmacologic HbF inducer might improve the therapeutic index.

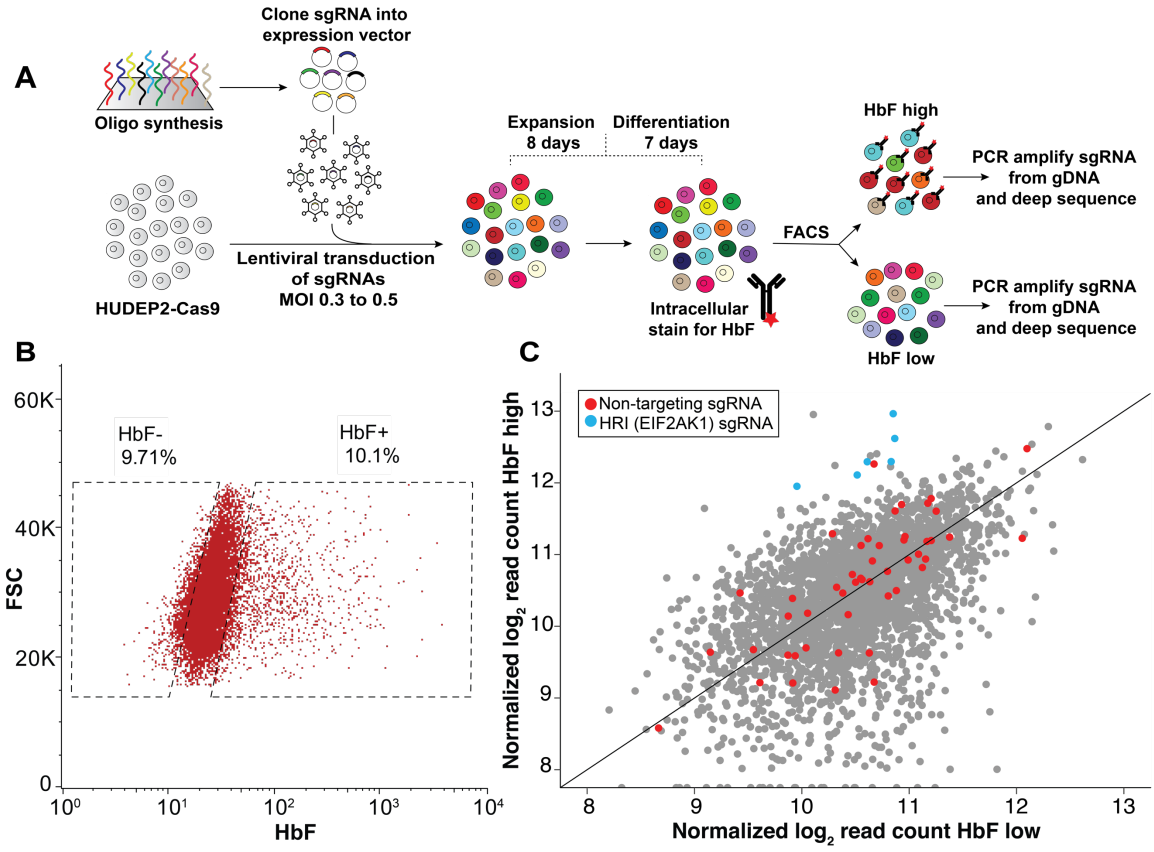

Fig. 1. A kinase-domain selective CRISPR-Cas9 screen identifies HRI as a novel fetal globin repressor.

(**A**) Experimental outline. See Materials and Methods for full description. (**B**) Representative HbF FACS gating strategy for sorting HbF high and HbF low populations for screens. (**C**) sgRNA representation in HbF low vs. HbF high populations as $\log_2$ transformed normalized read counts.

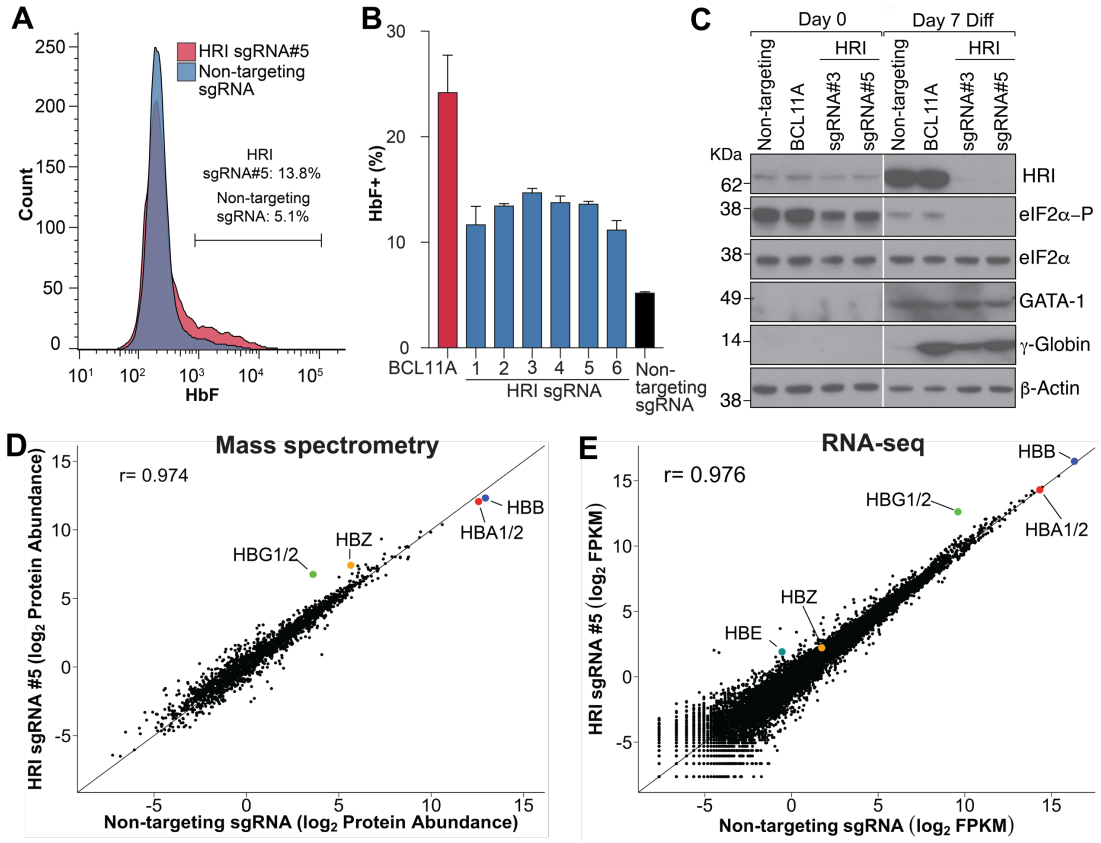

Fig. 2. HRI depletion elevates γ-globin in HUDEP2 cells. (**A**) Representative HbF FACS for HUDEP2-Cas9 sgRNA pools. (**B**) Summary of HbF flow cytometry of cells expressing indicated sgRNAs. Mean is shown ± standard deviation (SD) from (n=2) biological replicates. (**C**) Western blots with antibodies against indicated proteins in cell pools expressing sgRNAs listed at the top. (**D**) Mass spectrometry for HRI sgRNA #5 pool from (n=3) biological replicates. $\log_2$ mean-normalized protein abundances in cells with HRI sgRNA #5 and non-targeting sgRNA, r-value denotes Pearson correlation coefficient. (**E**) RNA-seq of cell pools with HRI sgRNA #5, $\log_2$ FPKM values averaged from (n=2) biological replicates.

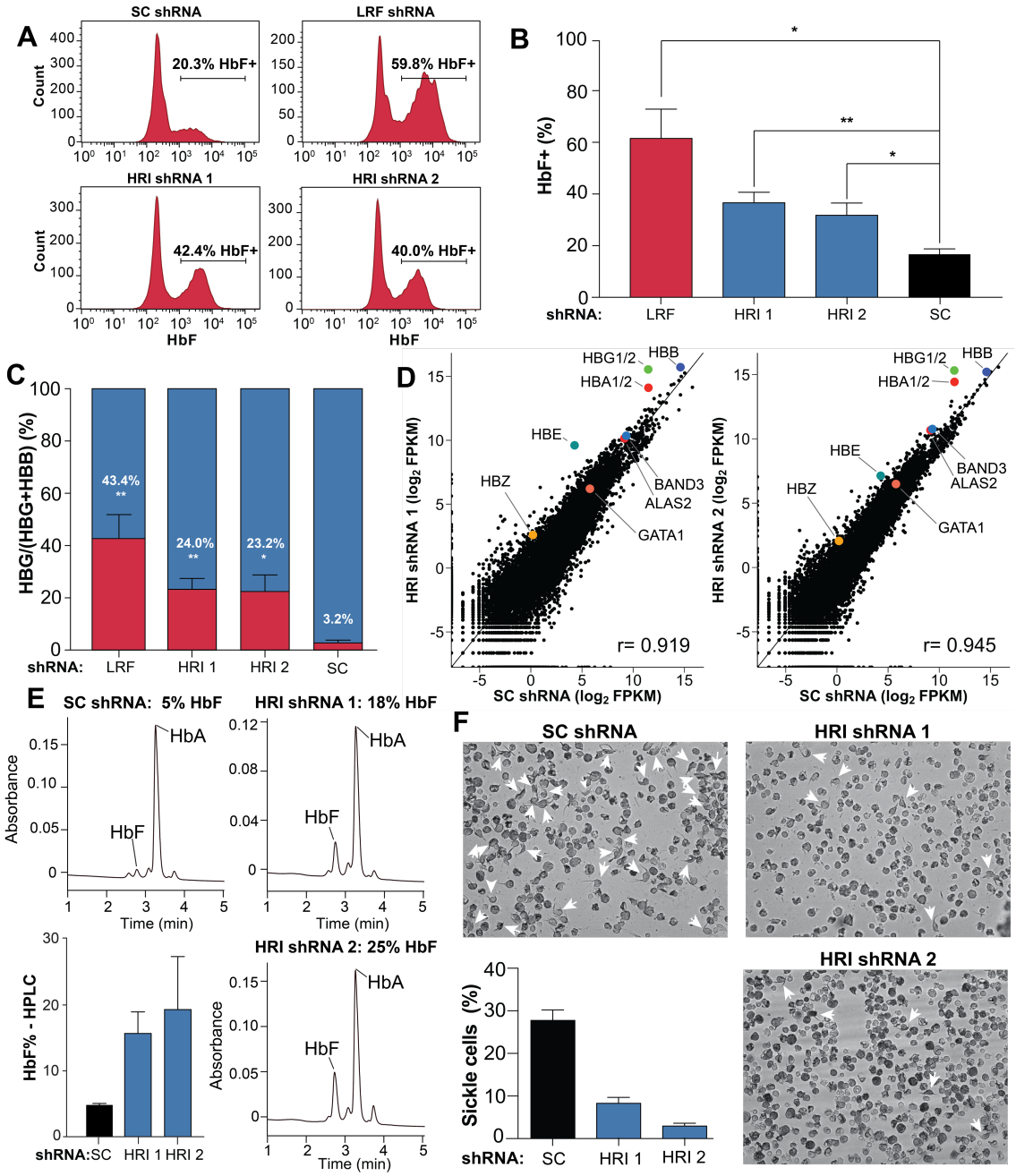

Fig. 3. HRI depletion elevates γ-globin in primary erythroid CD34+ cells. (**A**) Representative HbF flow cytometry on day 15 of CD34+ erythroid differentiation. (**B**) Summary of HbF flow cytometry experiments. Error bars: SEM from 3 independent donors (* denotes p-value<0.05, ** denotes p-value<0.01 from unpaired student T-tests). (**C**) γ-globin mRNA as fraction of γ-globin + β-globin by RT-qPCR on day 13. Error bars: SEM from 3 independent donors (* denotes p-value<0.05, ** denotes p-value<0.01 from unpaired student T-tests). (**D**) RNA-seq for CD34+ cells on day 13. R-value denotes Pearson correlation coefficient. Data was obtained from one patient donor. (**E**) HPLC analysis of samples fomr cell expressing annotated shRNAs, at day 15 of differentiation HbA: hemoglobin A (adult form); HbF: fetal hemoglobin. Error bars: standard deviation from 2 independent donors. (**F**) Images of sickle cell patient-derived erythroid cultures. Arrow heads mark cells with sickle-like

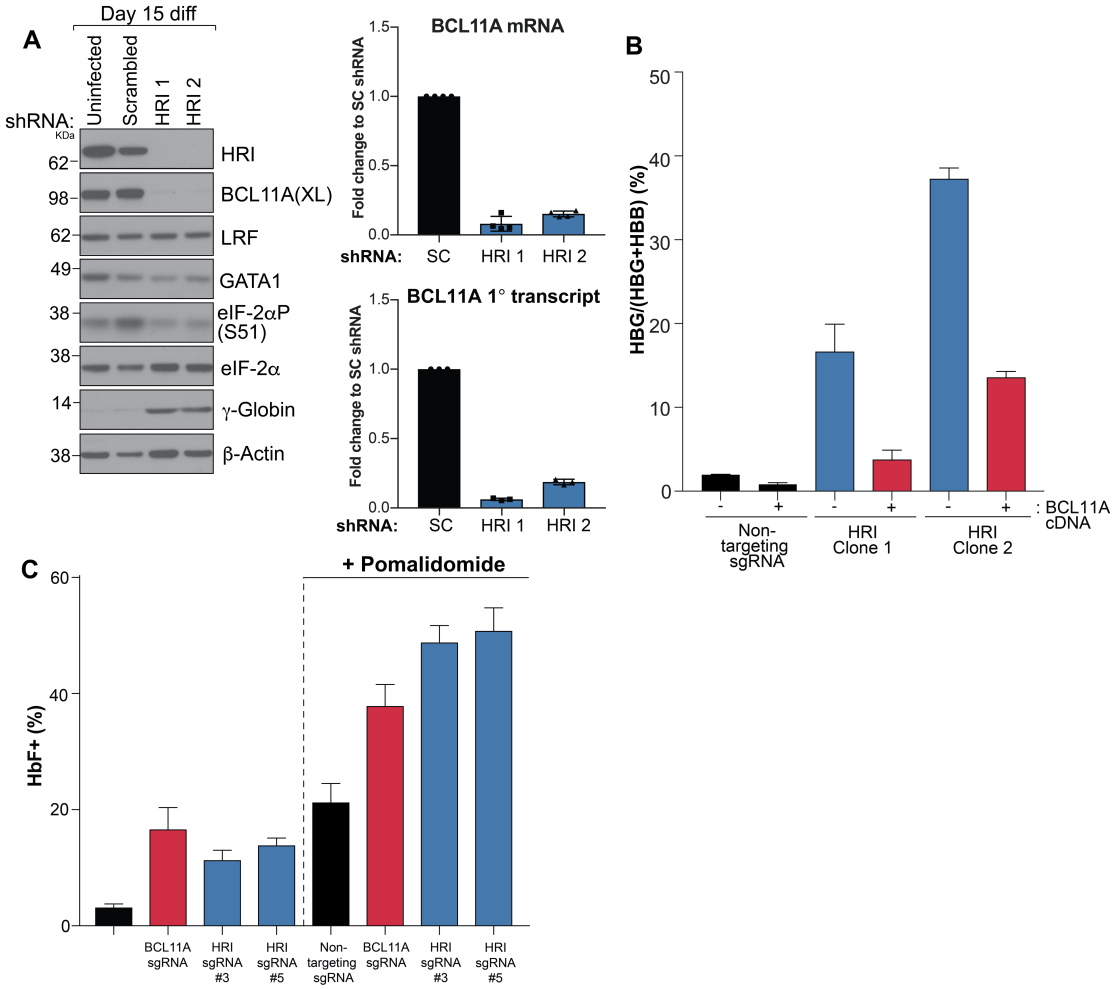

Fig. 4. HRI regulates BCL11A protein levels.
(**A**) Left: Western blot with indicated antibodies in cells uninfected, or infected with virus expressing scrambled or HRI shRNAs on day 15 of differentiation. Right: BCL11A mRNA (n=4 independent donors) and primary transcript (n=3 independent donors) levels in HRI knock down cells as fraction of SC shRNA expressing cells. Error bars are standard deviation from biological replicates. (**B**) BCL11A cDNA rescue in HRI depleted HUDEP2 clonal cell lines. Data is RT-qPCR for γ-globin as a fraction of γ-globin + β-globin. Error bars: standard deviation from two biological replicates (**C**) HbF flow cytometry of indicated HUDEP2 sgRNA pools grown in the presence or absence of pomalidomide. Error bar: SEM of 3 biological replicates.

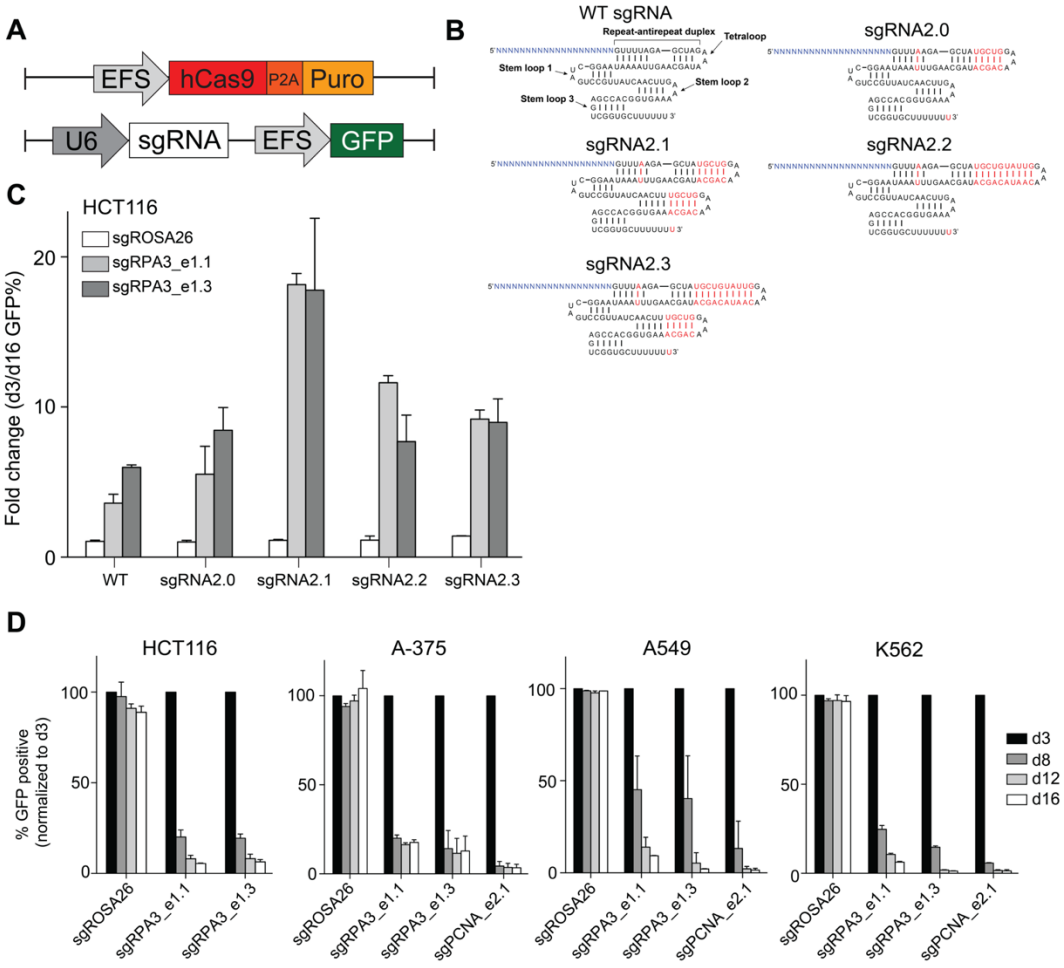

Fig. S1. sgRNA scaffold optimization. (See Supplementary Text for full description). (**A**) Cas9 and sgRNA plasmid designs. (**B**) sgRNA scaffolds tested, modifications to WT sgRNA-scaffold are highlighted in red. sgRNA 2.0 was previously described (*26*). (**C**) sgRNA editing efficiency assessed in HCT116 cells. GFP+ populations were measured at day 3 and day 16 post-sgRNA infection with sgRNAs targeting the essential DNA replication gene RPA3, and the ROSA26 locus as a negative control. Fold depletion was used to quantify editing efficiency for each of the sgRNA- scaffold designs. (**D**) sgRNA editing efficiency was similarly assessed with sgRNA-scaffold design 2.1 in HCT116, A-375, A549 and K562 cells lines. sgRNAs against the essential DNA replication genes RPA3 and PCNA were used, as well as an sgRNA targeting the ROSA26 locus as a negative control.

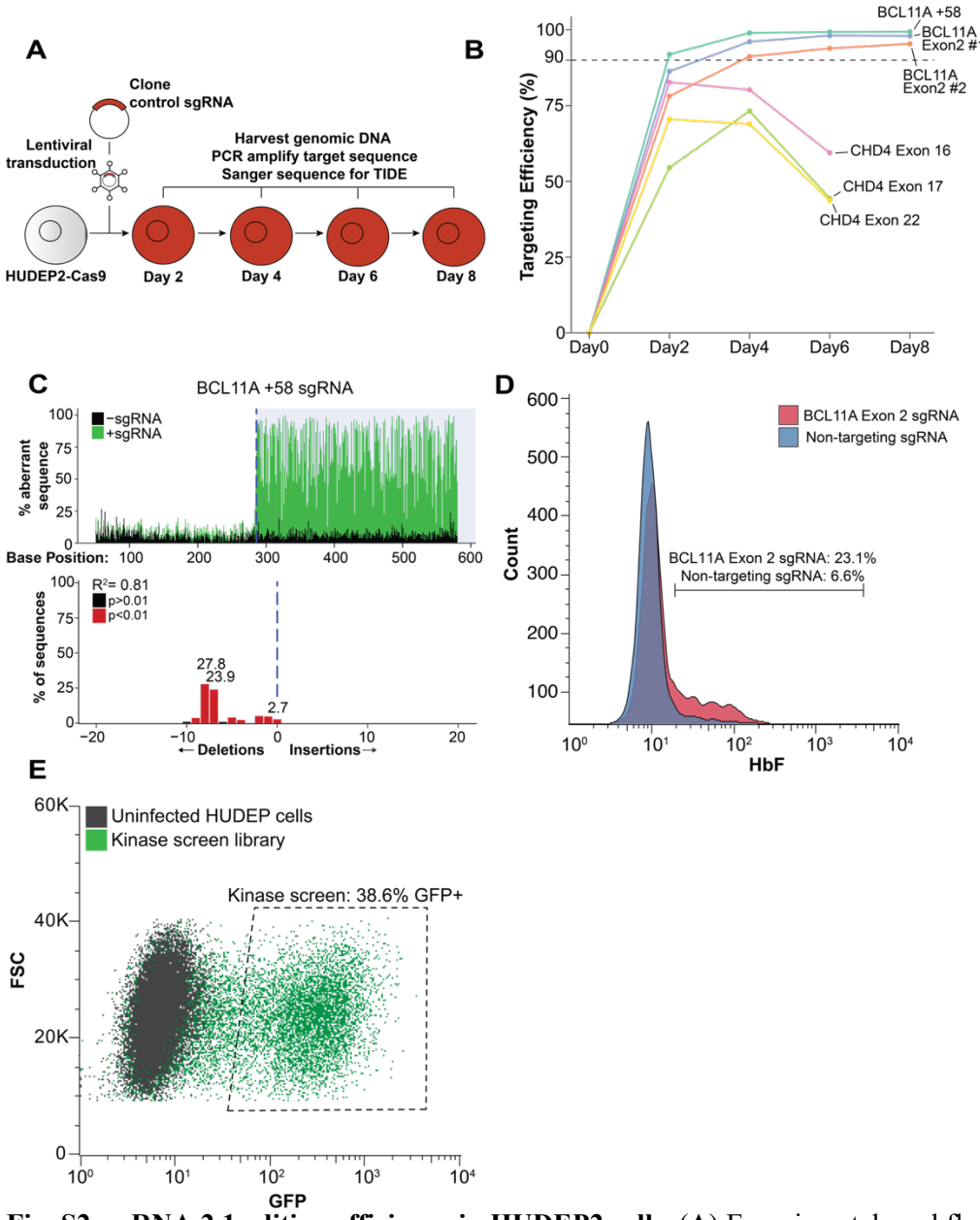

Fig. S2. sgRNA 2.1 editing efficiency in HUDEP2 cells. (**A**) Experimental workflow to verify editing efficiency in HUDEP2-Cas9 cells with the sgRNA 2.1 scaffold design. After transducing sgRNAs, we harvested genomic DNA every 2 days over the course of a total of 8 days. We amplified targeted regions by PCR, Sanger sequenced the amplicons, and used TIDE to quantify editing efficiency. (**B**) Summary of targeting efficiencies for time course. The proportion of edited CHD4 (*48*) alleles declined over time, most likely because CHD4 is essential for cell growth or viability. (**C**) Representative TIDE plot. Proportion of WT alleles remaining in population (0 bases deleted or inserted) was taken to determine proportion of alleles that were edited. (**D**) Representative HbF flow cytometry for BCL11A sgRNA control. (**E**) Representative GFP flow cytometry at day 2 of screening protocol.

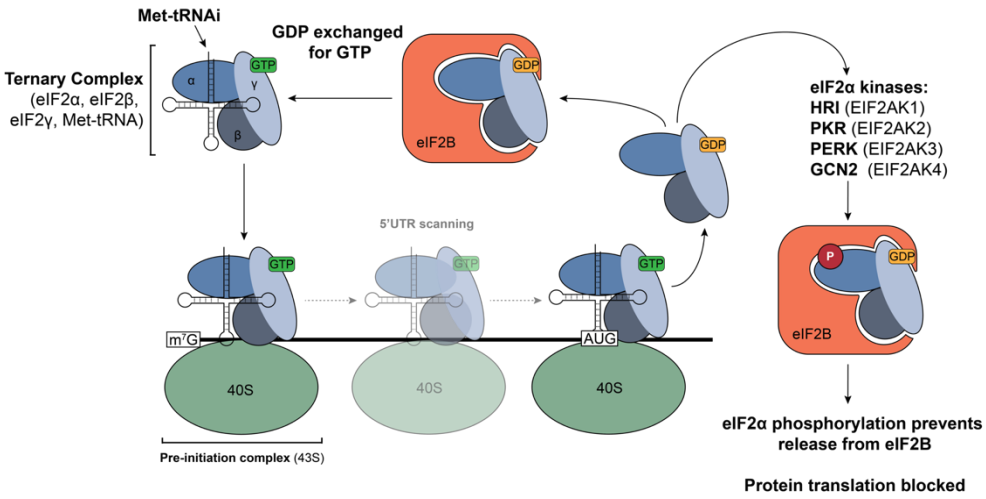

Fig. S3. HRI's role in regulating protein translation initiation. HRI is one of four kinases known to phosphorylate the protein translation initiation factor eIF2α (*28*). In its unphosphorylated form, eIF2α bound to GTP as part of a ternary complex with eIF2β and eIF2γ recruits the first methionine tRNA to the 40S ribosomal subunit to form a pre-initiation complex (PIC). The PIC then scans along the 5'UTR until it recognizes a start codon, at which point the GTP bound to eIF2α is hydrolyzed to GDP and the eIF2 factors are released from the PIC, allowing for formation of the 80S ribosomal unit. GDP is released from the eIF2 complex by the guanine exchange factor eIF2B, which allows for these factors to participate in further rounds of translational initiation. The phosphorylation of eIF2α prevents its release from eIF2B, thus blocking the recycling of this factor and further rounds of translation initiation. As such, the family of eIF2α kinases function as general repressors of protein translation.

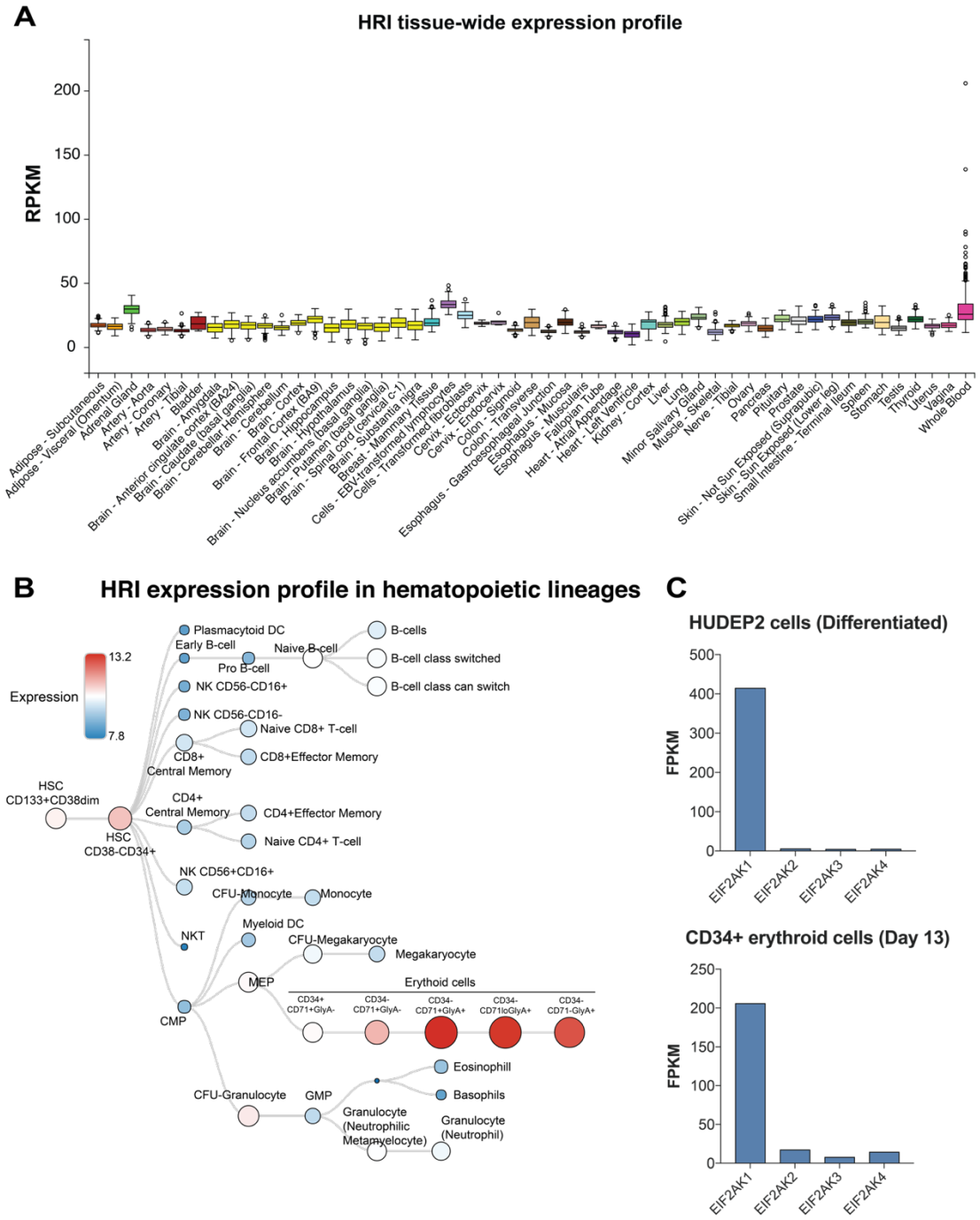

Fig. S4. Tissue expression for HRI. (**A**) Tissue-wide expression profile for HRI from GTEx. Expression levels for HRI are highest in whole blood. (**B**) Hematopoietic expression profile for HRI from BloodSpot (http://servers.binf.ku.dk/bloodspot/) (*97*). Expression levels for HRI are highest in the erythroid lineage. (**C**) Expression levels of eIF2α kinases in HUDEP2 cells and CD34+ erythroid cells. HRI is the highest expressed eIF2α kinase in these systems. It is worth noting that while the other eIF2α kinases were included in the CRISPR-screen shown in Fig. 1, they were not enriched in the HbF population, suggesting they are not expressed at high enough levels to impact on HbF induction.

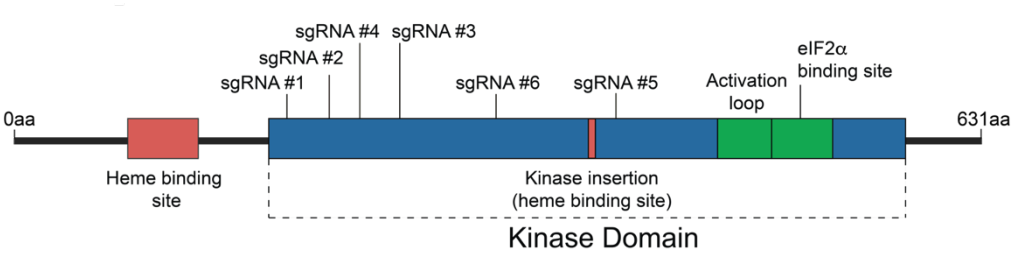

Fig. S5. HRI domain architecture. Position of the six "hit" HRI sgRNAs within the kinase domain.

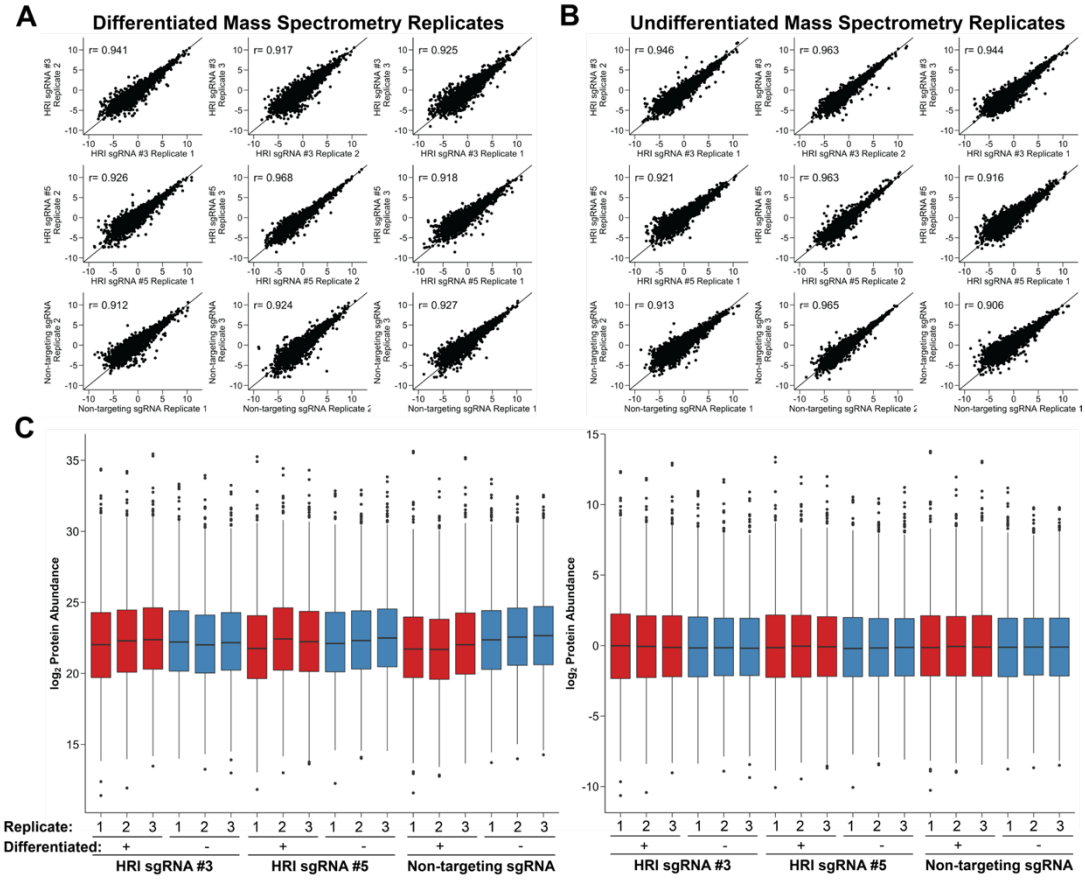

Fig. S6. Replicate concordance for mass spectrometry samples and normalization. (**A**) Replicate experiments in differentiated HUDEP2 cells. (n=3) biological replicates were generated each for HRI sgRNA #3, HRI sgRNA #5, and non-targeting sgRNA HUDEP2-Cas9 cell pools. All 3 pairwise replicate comparisons are shown for each condition. R-values denote Pearson correlation coefficients. Plots show $\log_2$ protein abundance for each replicate displayed. (**B**) Replicate experiments in undifferentiated HUDEP2 cells. (n=3) biological replicates were generated each for HRI sgRNA #3, HRI sgRNA #5, and non-targeting sgRNA HUDEP2-Cas9 pools. All 3 pairwise replicate comparisons are shown for each sample. R-values denote Pearson correlation coefficients. Plots show $\log_2$ protein abundance for each replicate displayed. (**C**) Normalization for mass spectrometry data. Protein abundances were $\log_2$ transformed (left panel), and normalized to their mean (right panel).

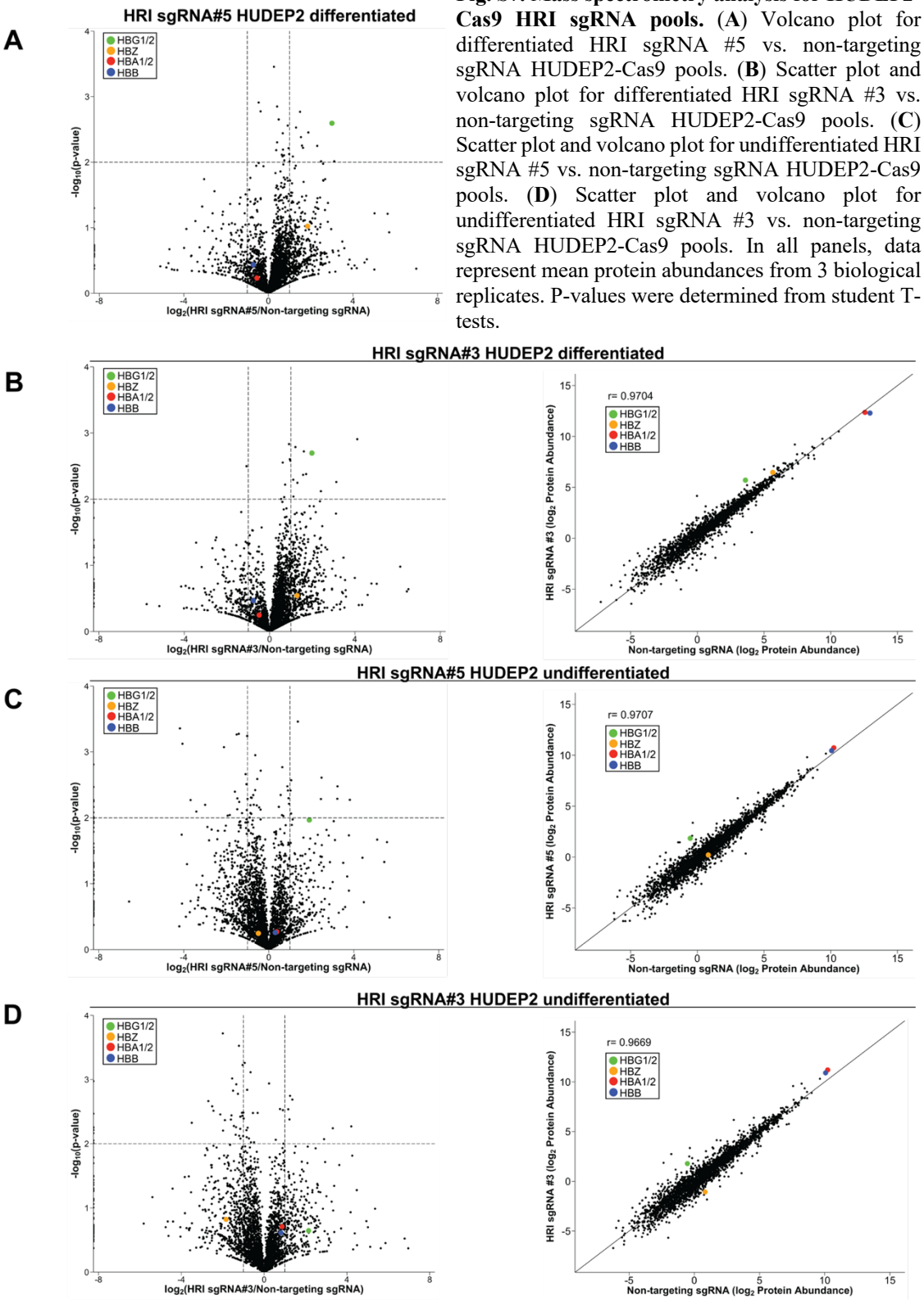

Fig. S7. Mass spectrometry analysis for HUDEP2-Cas9 HRI sgRNA pools. (**A**) Volcano plot for differentiated HRI sgRNA #5 vs. non-targeting sgRNA HUDEP2-Cas9 pools. (**B**) Scatter plot and volcano plot for differentiated HRI sgRNA #3 vs. non-targeting sgRNA HUDEP2-Cas9 pools. (**C**) Scatter plot and volcano plot for undifferentiated HRI sgRNA #5 vs. non-targeting sgRNA HUDEP2-Cas9 pools. (**D**) Scatter plot and volcano plot for undifferentiated HRI sgRNA #3 vs. non-targeting sgRNA HUDEP2-Cas9 pools. In all panels, data represent mean protein abundances from 3 biological replicates. P-values were determined from student T-tests.

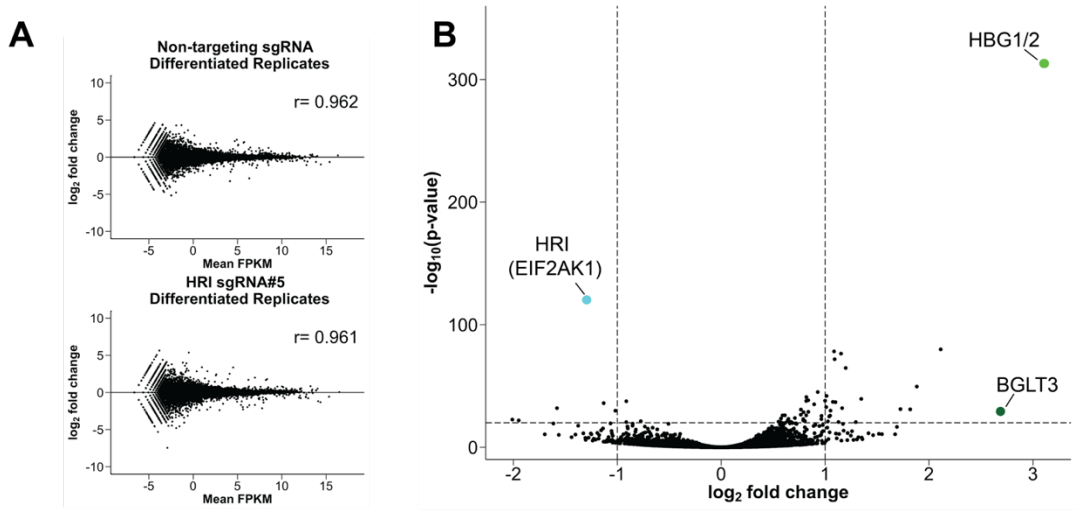

Fig. S8. RNA-seq for HUDEP2 cells. (**A**) MA plots for HRI sgRNA #5 and non-targeting sgRNA HUDEP2-Cas9 pools. 2 RNA seq biological replicates were generated each for differentiated cells. R-values denote Pearson correlation coefficients across the two indicated replicates. (**B**) Volcano plot for differentiated HRI sgRNA #5 and non-targeting sgRNA pools. Statistical significance was determined using the R-package DESeq2, with a negative binomial as a null distribution to determine statistical significance. P-values were FDR corrected for multiple hypothesis testing. Vertical lines denote 2 fold changes, and the horizontal line denotes a significance level of 10^{-20}.

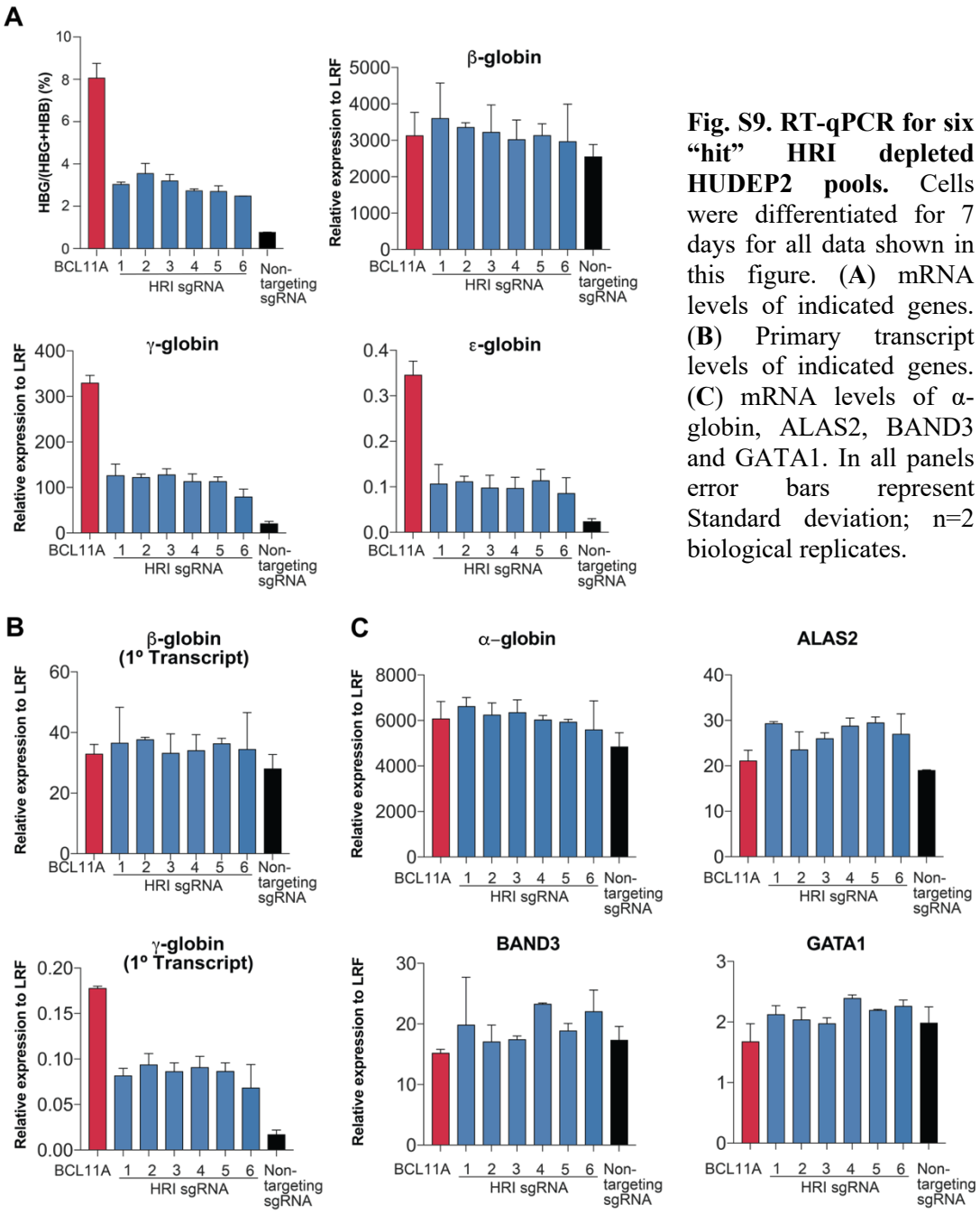

Fig. S9. RT-qPCR for six "hit" HRI depleted HUDEP2 pools. Cells were differentiated for 7 days for all data shown in this figure. (**A**) mRNA levels of indicated genes. (**B**) Primary transcript levels of indicated genes. (**C**) mRNA levels of α-globin, ALAS2, BAND3 and GATA1. In all panels error bars represent Standard deviation; n=2 biological replicates.

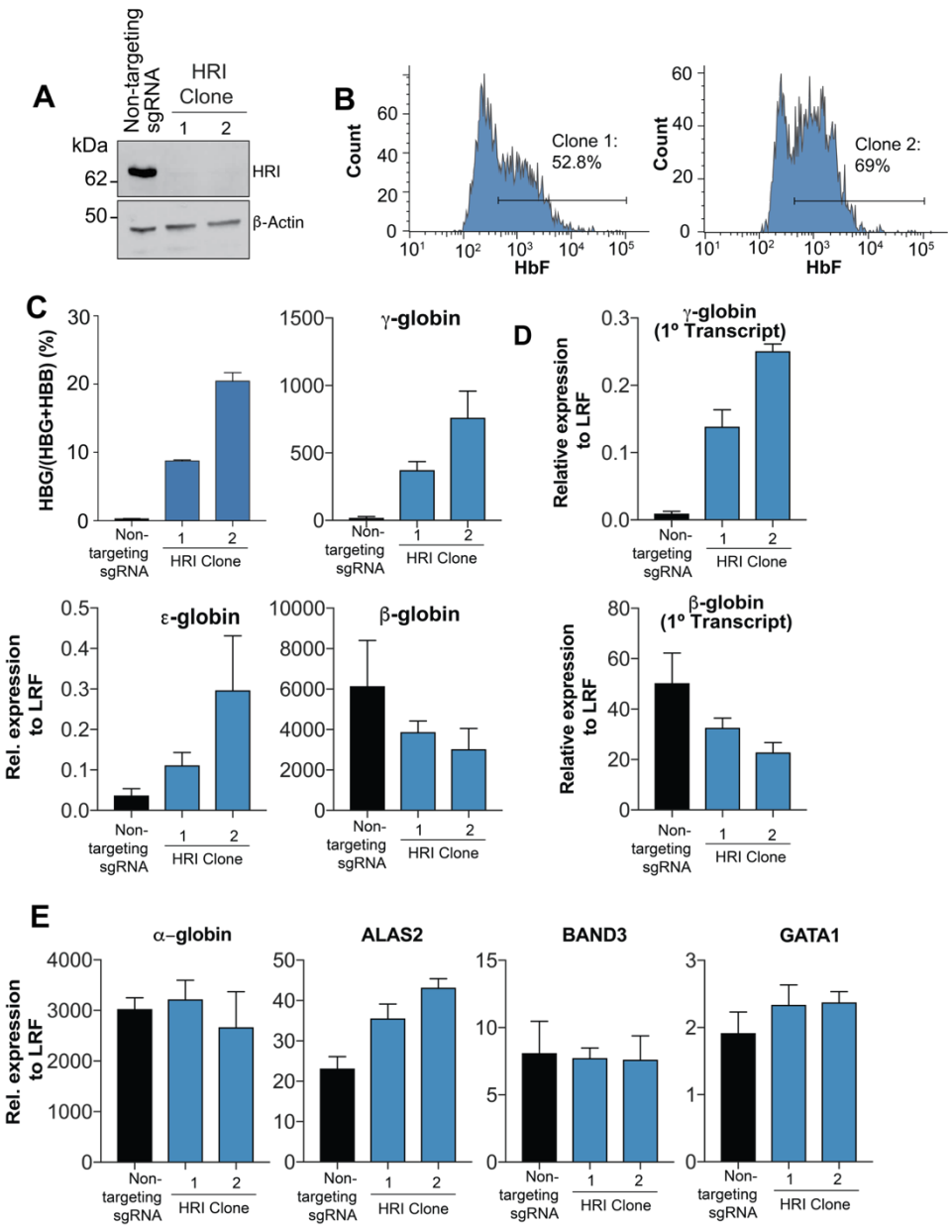

Fig. S10. HRI depleted HUDEP2 clones. Given the heterogeneity of mutations induced by sgRNAs in pools (see Fig. S2), we generated clonal HUDEP2 HRI depleted cell lines from the HRI sgRNA#5 pool. Cells were differentiated for 7 days for all data shown in this figure. In all panels error bars represent SEM; n=3 biological replicates. (**A**) Anti-HRI Western blot from lysates of two clonal HRI depleted cell lines and control cells. (**B**) HbF flow cytometry for HRI depleted clonal cell lines (gates were set on negative control as in Fig 2A). (**C**) RT-qPCR for globins for HRI depleted clonal cell lines, mean is shown ± standard error of the mean (SEM) from (n=3) biological replicates. (**D**) Primary transcripts of β-type globins. (**E**) mRNA levels of indicated genes.

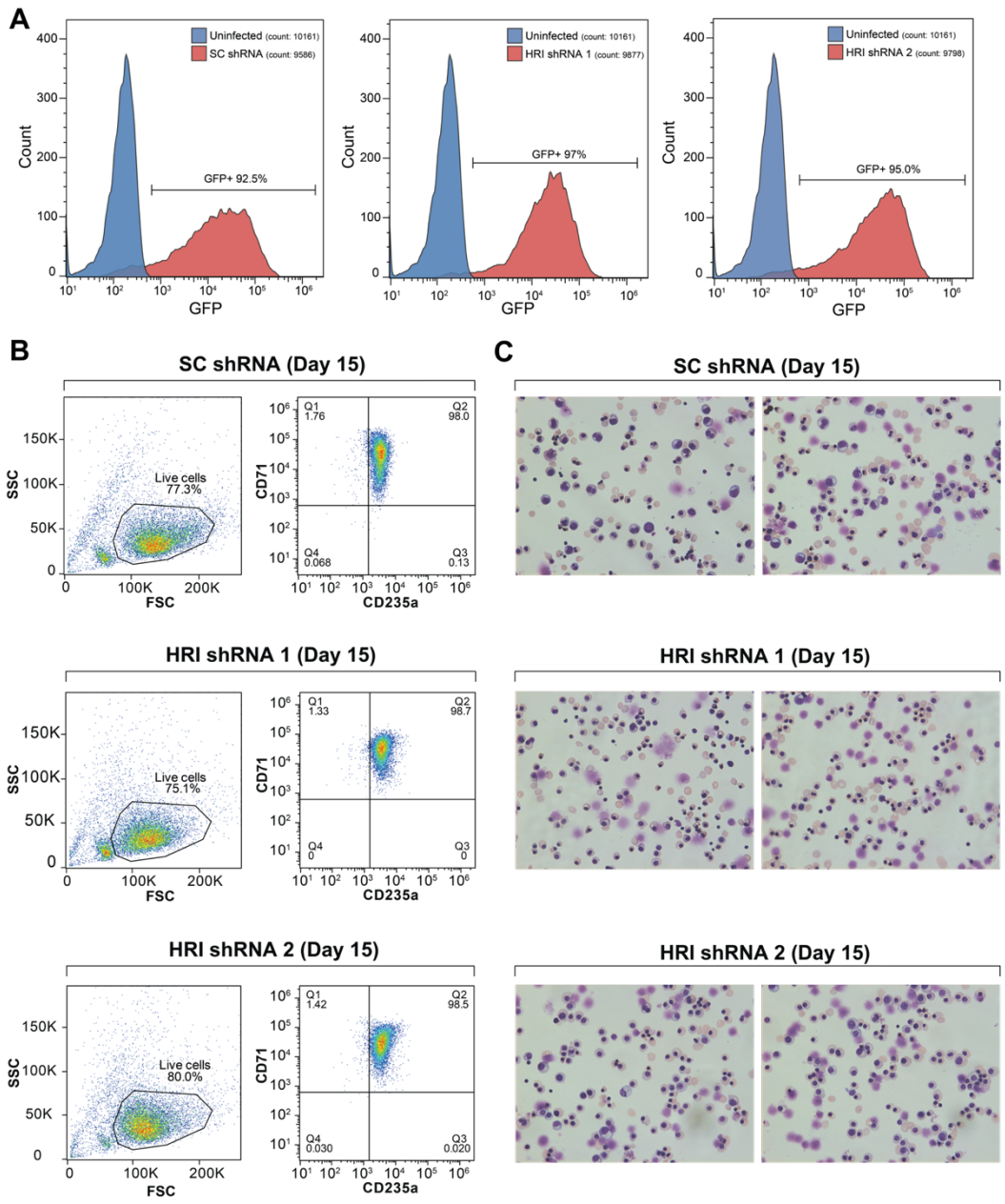

Fig. S11. Cell surface markers and cell morphology for HRI depleted CD34+ cells. (**A**) Representative transduction levels for shRNAs by GFP flow cytometry (Day 11). Uninfected were used as a gating control (blue) and compared to shRNA transduced populations (red). The uninfected control is the same for all panels shown. (**B**) Representative CD71/CD235a staining for SC shRNA and HRI shRNAs infected CD34+ cells at day 15 of culture. (**C**) Giemsa staining for SC shRNA control and HRI shRNAs infected CD34+ cells at day 15 of culture. 2 fields of view are shown for each sample.

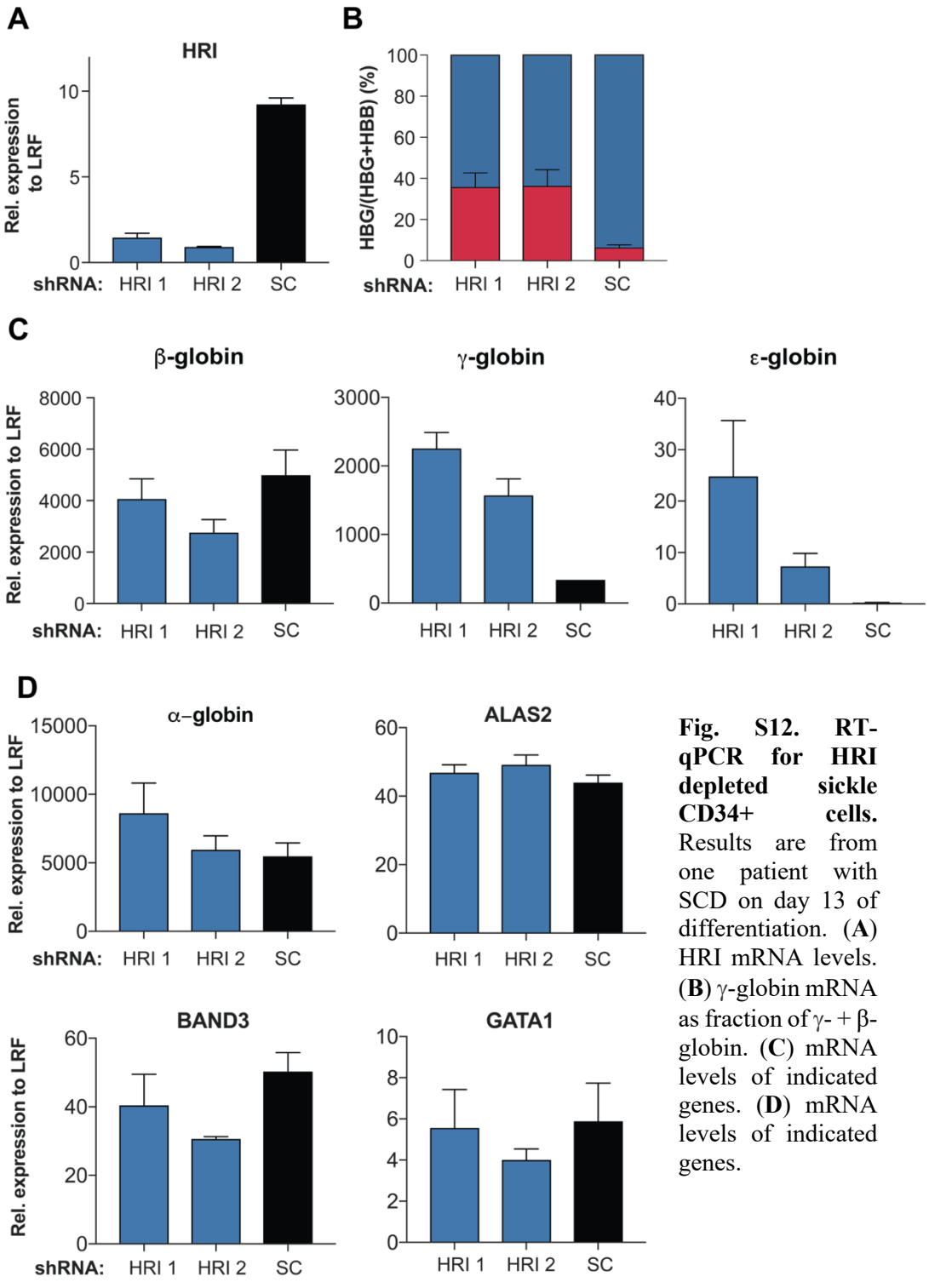

Fig. S12. RT-qPCR for HRI depleted sickle CD34+ cells. Results are from one patient with SCD on day 13 of differentiation. (**A**) HRI mRNA levels. (**B**) γ-globin mRNA as fraction of γ- + β-globin. (**C**) mRNA levels of indicated genes. (**D**) mRNA levels of indicated genes.

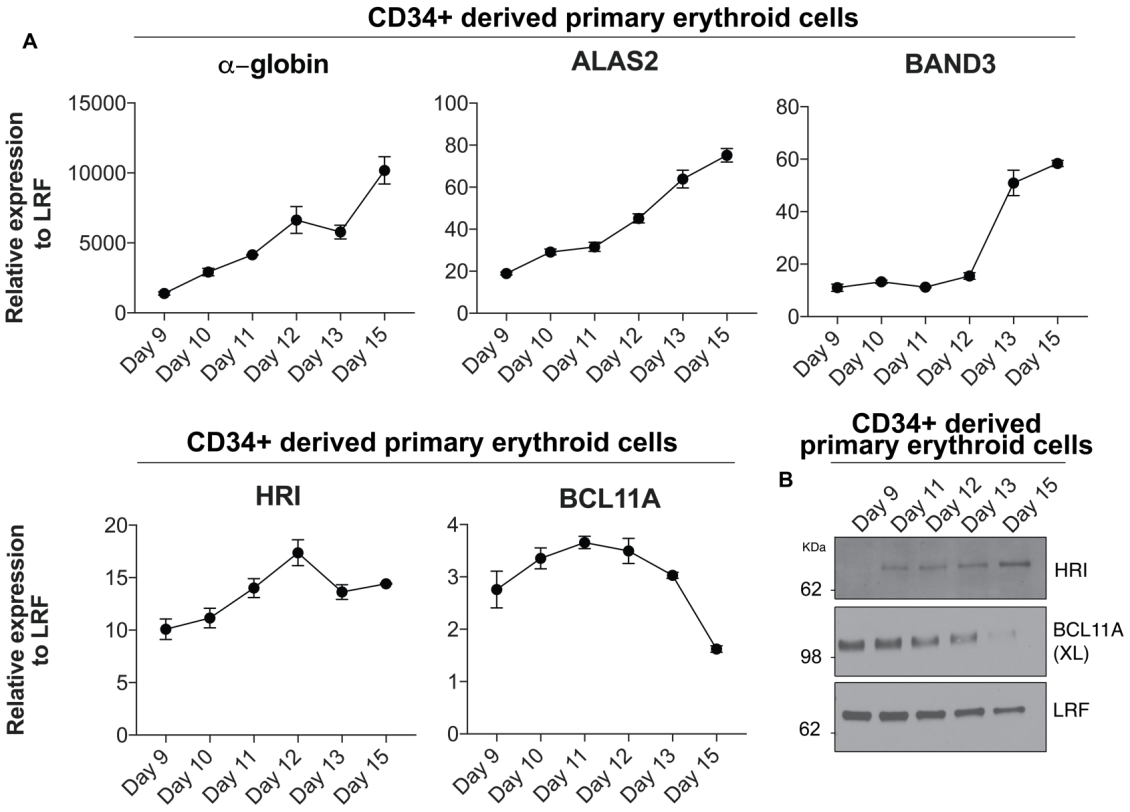

Fig. S13. HRI and BCL11A expression during erythroid maturation**.** (**A**) RT-qPCR for indicated primer sets for (n=2) technical replicates from one patient donor. Maturation markers α-globin, ALAS2 and BAND3 increase during erythroid maturation. HRI progressively increases during maturations. BCL11A increases early in maturation and then declines. (**B**) Western Blot for HRI, BCL11A and LRF during erythroid maturation. HRI increases in protein levels during maturation. BCL11A protein levels increase and then progressively decrease as shown previously (*98*). LRF remains relatively unchanged during maturation, as shown previously (*56*). These results suggest that BCL11A expression partly precedes HRI expression. The anti-correlated expression levels of HRI and BCL11A might appear in conflict with a requirement of HRI for BCL11A expression. However, please note that while HRI levels increase during erythroid maturation, its activity seems to decline as shown by western blots measuring eIF2α-P levels (Fig.2C), most likely due to the production of heme as a result of erythroid differentiation (*26*).

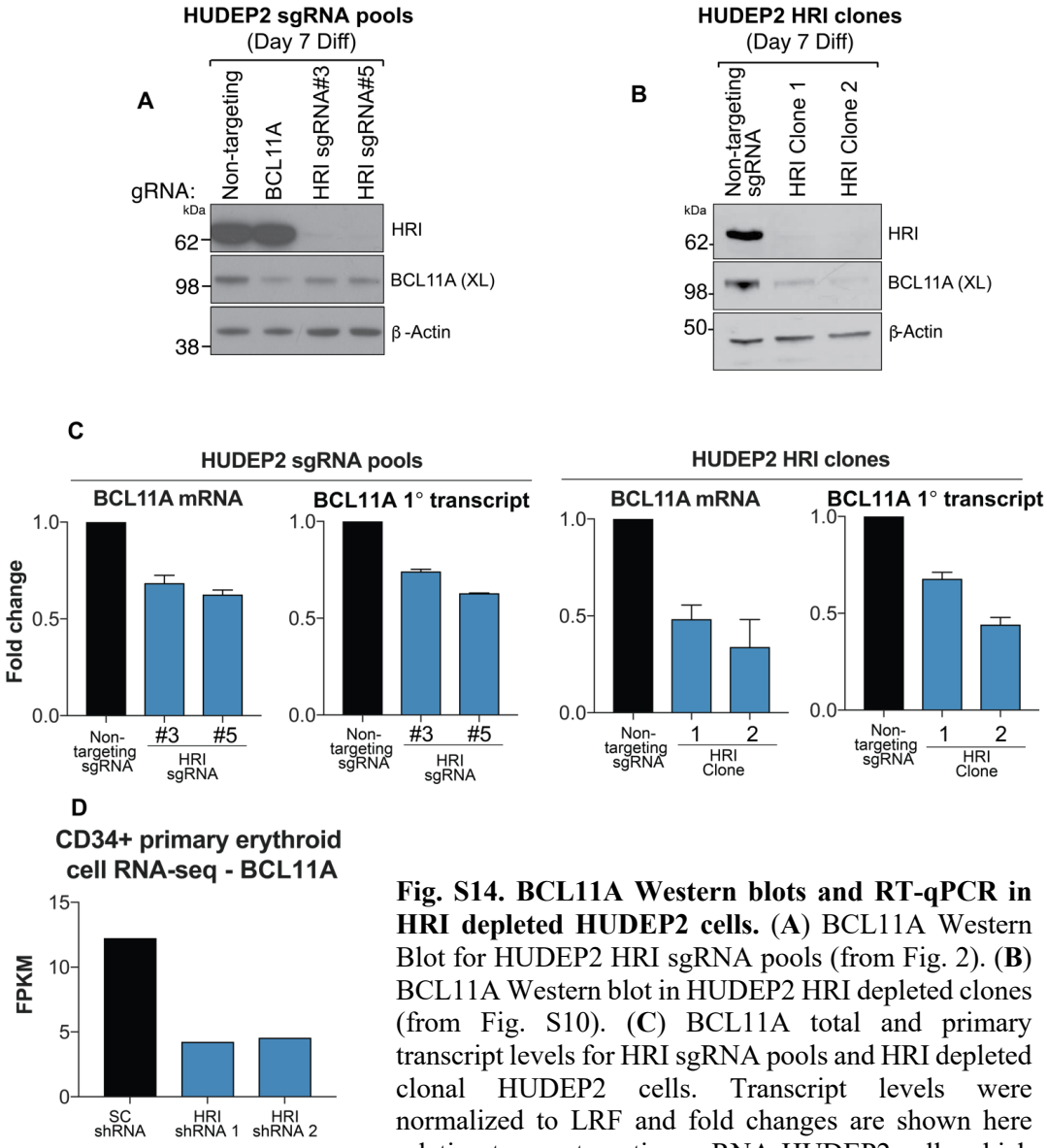

Fig. S14. BCL11A Western blots and RT-qPCR in HRI depleted HUDEP2 cells. (**A**) BCL11A Western Blot for HUDEP2 HRI sgRNA pools (from Fig. 2). (**B**) BCL11A Western blot in HUDEP2 HRI depleted clones (from Fig. S10). (**C**) BCL11A total and primary transcript levels for HRI sgRNA pools and HRI depleted clonal HUDEP2 cells. Transcript levels were normalized to LRF and fold changes are shown here relative to non-targeting sgRNA HUDEP2 cells which served as a negative control. (**D**) BCL11A transcript levels in RNA-seq from Fig. 3F.

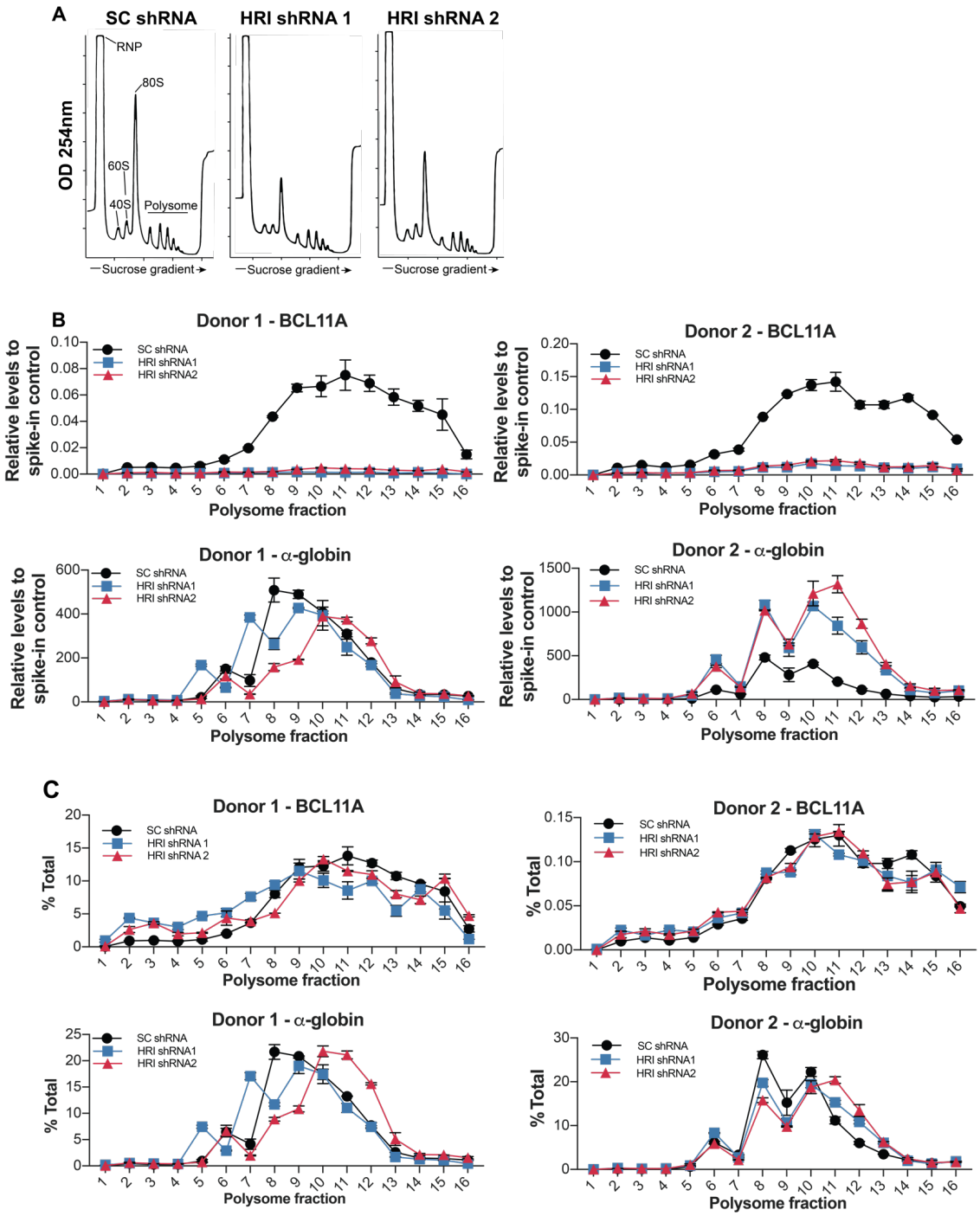

Fig. S15. Polysome profiling in primary erythroid cells. (**A**) Polysome tracings for shRNA transduced CD34+ erythroid cultures (Day 13). (**B**) Transcript abundances for each polysomal fraction normalized to spike-in control. (**C**) Transcript distribution across polysomal fractions. The proportion of transcripts for each fraction was determined by dividing the transcript abundance by the total amount of transcript across all the fractions (shown here as % of total). Polysome profiling was performed from cultures derived from (n=2) independent patient donors. Error bars are standard deviations for (n=2) technical replicates.

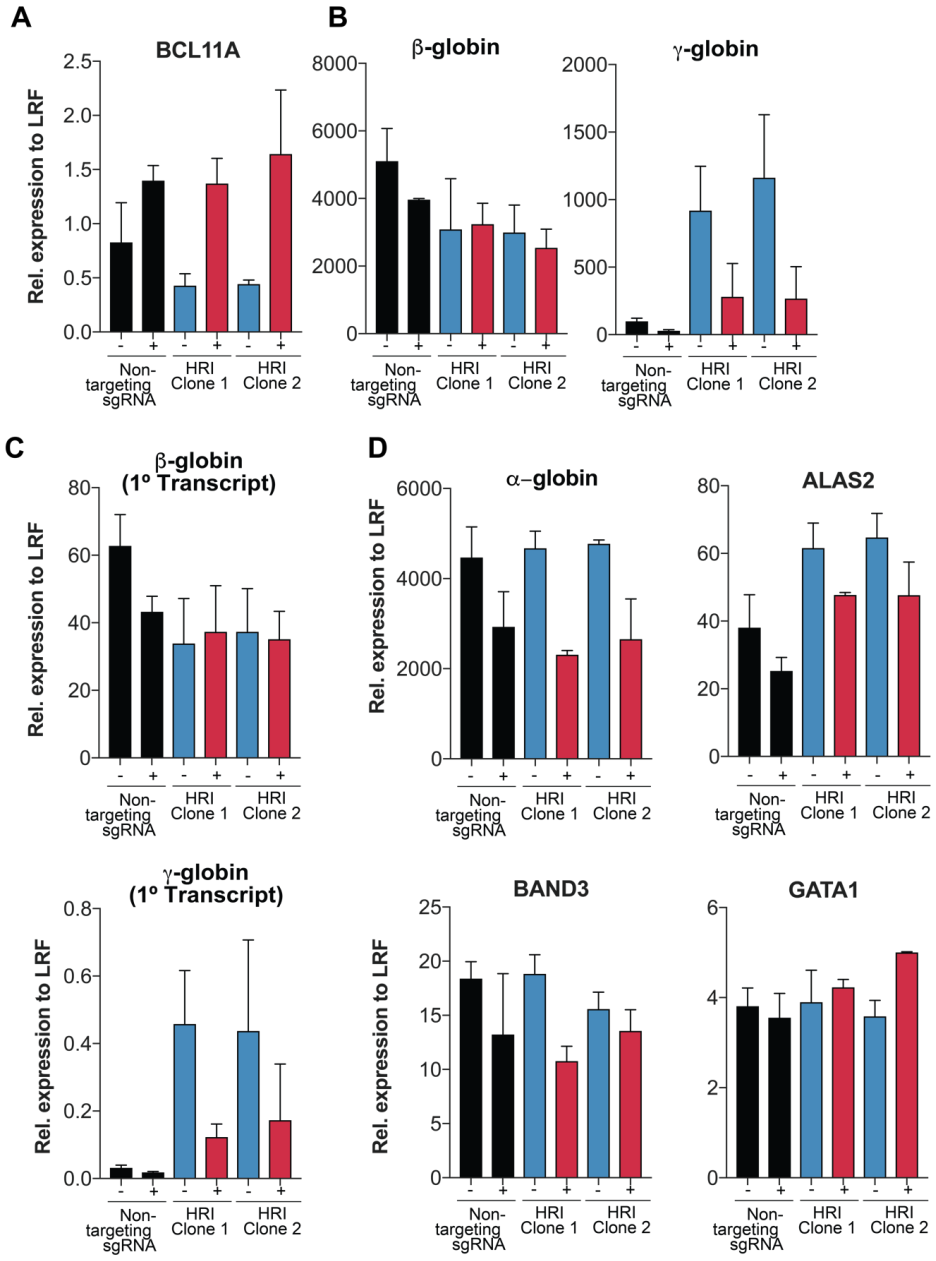

Fig. S16. Gene expression measurements in the BCL11A rescue experiments. Presence or absence of BCL11A cDNA is indicated by + or -, respectively. Cells were differentiated for 7 days for all data shown. Error bars: standard deviation from 2 biological replicates. (**A)** BCL11A transcript levels. While BCL11A transcripts were slightly higher in cells lines expressing exogenous BCL11A cDNA, they remained in a similar range to endogenous levels. (**B)** mRNA levels of indicated genes. BCL11A expression strongly reduced γ- and ε-globin transcript levels. (**C)** Primary-transcript levels of indicated genes. BCL11A expression strongly reduced γ-globin primary transcript levels. (**D)** Transcript levels for maturation markers. We note that expression BCL11A cDNA slightly reduced alpha-globin and BAND3 levels, but were not significantly different to levels from parental HUDEP2 cells.

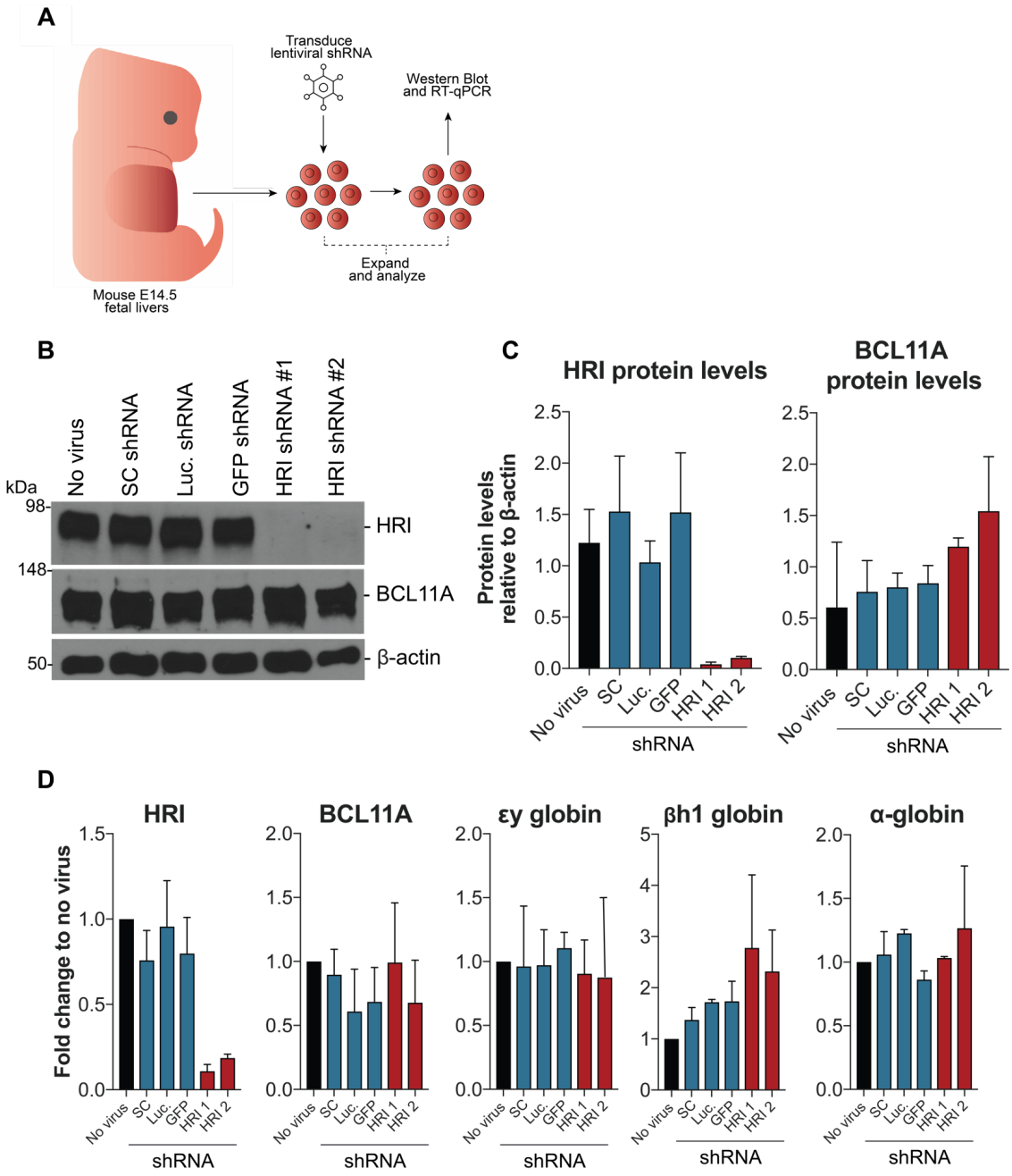

Fig. S17. HRI depletion in fetal liver derived adult-type erythroblasts. (A) Experimental approach, for details see Methods and Materials. **(B)** Western Blot for HRI, BCL11A and β-actin for all denoted conditions. HRI depletion did not result in reduced BCL11A protein levels. **(C)** Image J quantification of protein levels from (n=2) independent biological replicates. Error bars denote standard deviations. **(D)** RT-qPCR for denoted transcripts. Data were collected from (n=2) biological replicates, error bars represent standard deviations. HRI depletion did not affect BCL11A or embryonic-type globin transcript levels, or maturation as evidenced by α-globin levels.

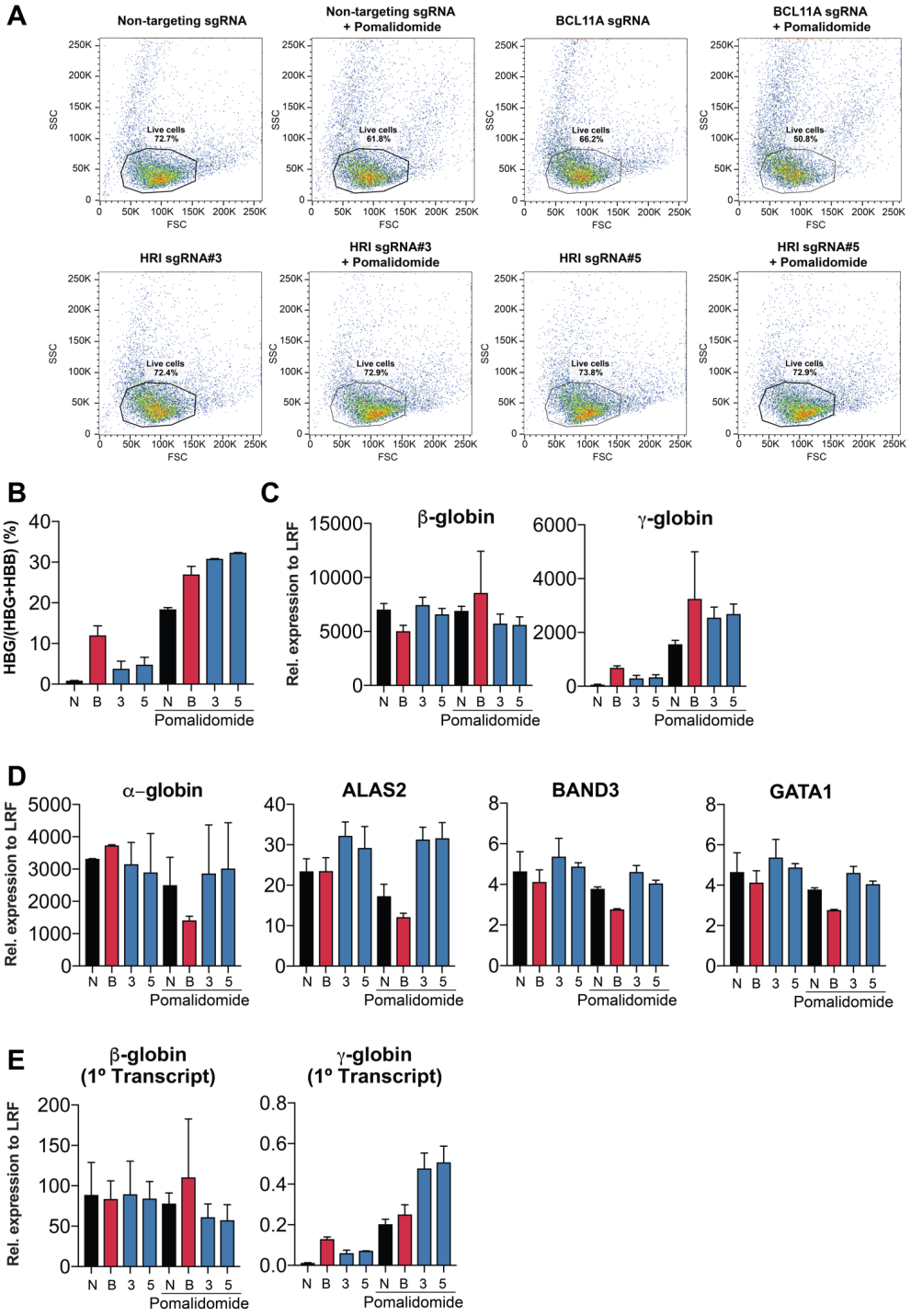

Fig. S18. Gene expression measurements in the pomalidomide combination experiments. Cells were differentiated for 7 days for all data shown with sgRNA HUDEP2-Cas9 pools. N denotes Non-targeting sgRNA, B denotes BCL11A sgRNA, 3 denotes HRI sgRNA #3, and 5 denotes HRI sgRNA#5. (**A**) Cell viability assessed by flow cytometry. (**B**) γ-globin mRNA as fraction of γ- + β-globin. (**C**) Globin profiles (mature transcripts) (**D**) mRNA levels of indicated genes. Error bars: standard deviation from 2 biological replicates. (**E**) Primary-transcript levels of indicated genes.

Protein Names	Gene Names	Protein abundance HRI sgRNA#5	Protein abundance non-targeting sgRNA	log2 (Fold Change)	p-value
Lariat debranching enzyme	DBR1	0.206	0.024	3.118	0.010
Hemoglobin subunit gamma-2	HBG2	100.998	12.559	3.008	0.003
Probable ATP-dependent RNA helicase YTHDC2	YTHDC2	0.077	0.014	2.493	0.010
Rab GTPase-activating protein 1-like, isoform 10	RABGAP1L	0.236	0.047	2.312	0.006
Alpha-hemoglobin-stabilizing protein	AHSP	646.580	174.552	1.889	0.004
Carbonic anhydrase 3	CA3	5.768	1.852	1.639	0.006
Angio-associated migratory cell protein	AAMP	1.361	0.452	1.591	0.006
Synaptosomal-associated protein 23	SNAP23	2.033	0.699	1.541	0.004
Heat shock 70 kDa protein 1B	HSPA1B	94.190	33.318	1.499	0.002
Selenium-binding protein 1	SELENBP1	17.385	7.014	1.310	0.003
Calcyclin-binding protein	CACYBP	20.810	9.323	1.158	0.008
Carbonic anhydrase 2	CA2	176.977	88.204	1.005	0.010

Table S1. Top differentially expressed proteins from HUDEP2 mass spectrometry.

Gene Name	Full Name	HRI sgRNA#5 mean count	Non-targeting sgRNA mean count	log2(Fold Change)	p-value (adjusted)
HBG2	hemoglobin subunit gamma 2	32957.50	4128.50	3.11	0
HBG1	hemoglobin subunit gamma 1	36346.00	4597.00	3.10	0
BGLT3	beta globin locus transcript 3 (non-protein coding)	285.00	14.50	2.69	6.16E-30
ESPN	espin	2434.00	589.50	2.11	1.43E-80
HBBP1	NA	2735.50	768.50	1.88	3.17E-50
FOSL2	FOS like 2, AP-1 transcription factor subunit	942.50	260.50	1.82	1.19E-31
IFIT1B	interferon induced protein with tetratricopeptide repeats 1B	1072.00	325.50	1.73	7.78E-32
RUNDC3A	RUN domain containing 3A	3079.50	1326.50	1.35	3.78E-40
KLHDC8B	kelch domain containing 8B	1549.50	667.00	1.30	2.98E-21
HIST2H2AA4	histone cluster 2 H2A family member a4	13544.50	6623.00	1.20	2.18E-65
MYL4	myosin light chain 4	2156.00	1166.50	1.03	8.04E-22
HIST1H2BB	histone cluster 1 H2B family member b	7694.00	4280.50	1.01	1.38E-42
HIST1H2BL	histone cluster 1 H2B family member l	6537.00	3626.00	1.01	1.72E-33
HIST1H2BG	histone cluster 1 H2B family member g	9231.00	5180.00	1.00	1.14E-38
REXO2	RNA exonuclease 2	4149.50	9857.50	-1.02	1.54E-30
ARG2	arginase 2	3213.00	8292.00	-1.13	9.90E-37
EIF2AK1	eukaryotic translation initiation factor 2 alpha kinase 1	17616.00	50037.00	-1.29	6.06E-121
TRIB3	tribbles pseudokinase 3	485.00	1824.50	-1.58	1.27E-32
SMN2	survival of motor neuron 2, centromeric	129.00	803.50	-1.95	1.28E-22
TNFRSF10B	TNF receptor superfamily member 10b	86.50	580.00	-2.01	2.82E-23

Table S2. Top differentially expressed transcripts from HUDEP2 RNA-seq.

Name	Sequence
BCL11A_Exon2_#1_F	CACCGTGAACCAGACCACGGCCCGT
BCL11A_Exon2_#1_R	AAACACGGGCCGTGGTCTGGTTCAC
BCL11A_Exon2_#2_F	CACCGCATCCAATCCCGTGGAGGT
BCL11A_Exon2_#2_R	AAACACCTCCACGGGATTGGATGC
BCL11A_+58_F	CACCGTTGCTTTTATCACAGGCTCC
BCL11A_+58_R	AAACGGAGCCTGTGATAAAAGCAAC
CHD4_Exon16_F	CACCGAACTCATTCTCTCGGATGA
CHD4_Exon16_R	AAACTCATCCGAGAGAATGAGTTC
CHD4_Exon17_F	CACCGCTGATTGTTCTTCAGCCGAT
CHD4_Exon17_R	AAACATCGGCTGAAGAACAATCAGC
CHD4_Exon22_F	CACCGAAATACGAACGCATCGATGG
CHD4_Exon22_R	AAACCCATCGATGCGTTCGTATTTC
EIF2AK1_sgRNA#1_F	CACCGTCTGGAAGTGCTCTCCGACC
EIF2AK1_sgRNA#1_R	AAACGGTCGGAGAGCACTTCCAGAC
EIF2AK1_sgRNA#2_F	CACCGTATCCACCTTTTCCTAAGA
EIF2AK1_sgRNA#2_R	AAACTCTTAGGAAAAGGTGGATAC
EIF2AK1_sgRNA#3_F	CACCGTTTTAGTTGCACCCTTAATC
EIF2AK1_sgRNA#3_R	AAACGATTAAGGGTGCAACTAAAAC
EIF2AK1_sgRNA#4_F	CACCGTTGTTGGCTATCACACCGCG
EIF2AK1_sgRNA#4_R	AAACCGCGGTGTGATAGCCAACAAC
EIF2AK1_sgRNA#5_F	CACCGATAGTCGAGAGAAACAAGCG
EIF2AK1_sgRNA#5_R	AAACCGCTTGTTTCTCTCGACTATC
EIF2AK1_sgRNA#6_F	CACCGTGGTGAACTTGAGTCGACCC
EIF2AK1_sgRNA#6_R	AAACGGGTCGACTCAAGTTCACCAC
Non_targeting_sgRNA_F	CACCGACCGGAACGATCTCGCGTA
Non_targeting_sgRNA_R	AAACTACGCGAGATCGTTCCGGTC

Table S3. Oligos for sgRNAs.

Name	Sequence
BCL11A_Exon2_#1_TIDE_F	ATGCAAACAGCTTTTCTCCTTGC
BCL11A_Exon2_#1_TIDE_R	CACATTTGTTGCTGTGGGTGTG
BCL11A_Exon2_#2_TIDE_F	CCAGAGAGAAGGTGTCTGATGTGT
BCL11A_Exon2_#2_TIDE_R	AGATGTGCTTCTCCCCTTTCTGTC
BCL11A_+58_TIDE_F	GTAGCTGGTACCTGATAGGTGCC
BCL11A_+58_TIDE_R	GGAAAACAGCCTGACTGTGCC
CHD4_Exon16_TIDE_F	GAGCCCAAAATAGCCATGTCAATG
CHD4_Exon16_TIDE_R	TGTAGTTATGGTGGGACCTTGGG
CHD4_Exon17_TIDE_F	GTGCTGGAGTGAGTAACCATTCAAT
CHD4_Exon17_TIDE_R	CGCATGAAGGTATCTCAGGGGA
CHD4_Exon22_TIDE_F	GCTATTTGAGCTTGAACAAGTCCTGAC
CHD4_Exon22_TIDE_R	TGTTCTCAGGAAGCTCCTAAGATGC

Table S4. Primers used for TIDE assays shown in Fig. S2.

Name	Sequence
HRI_shRNA1_F	CACCGGCAGAGAGCAATGTGGTGTTAACTCGAGTTAACACCACATTGCTCTCTGTTTTTG
HRI_shRNA1_R	AATTCAAAAACAGAGAGCAATGTGGTGTTAACTCGAGTTAACACCACATTGCTCTCTGCC
HRI_shRNA4_F	CACCGGGCAGAAGTTCTAACAGGTTTACTCGAGTAAACCTGTTAGAACTTCTGCTTTTTG
HRI_shRNA4_R	AATTCAAAAAGCAGAAGTTCTAACAGGTTTACTCGAGTAAACCTGTTAGAACTTCTGCCC
LRF_shRNA_F	CACCGGCCACTGAGACACAAACCTATTCTCGAGAATAGGTTTGTGTCTCAGTGGTTTTTG
LRF_shRNA_R	AATTCAAAAACCACTGAGACACAAACCTATTCTCGAGAATAGGTTTGTGTCTCAGTGGCC
SC_shRNA_F	CACCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTTG
SC_shRNA_R	AATTCAAAAACAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGCC
Luciferase_shRNA_F	CACCGGCGCTGAGTACTTCGAAATGTCCTCGAGGACATTTCGAAGTACTCAGCGTTTTTG
Luciferase_shRNA_R	AATTCAAAAACGCTGAGTACTTCGAAATGTCCTCGAGGACATTTCGAAGTACTCAGCGCC
TurboGFP_shRNA_F	CACCGGCGTGATCTTCACCGACAAGATCTCGAGATCTTGTCGGTGAAGATCACGTTTTTG
TurboGFP_shRNA_R	AATTCAAAAACGTGATCTTCACCGACAAGATCTCGAGATCTTGTCGGTGAAGATCACGCC
mHRI_shRNA_1_F	CACCGGTGGTGTTAAAGATAATGAAAGCTCGAGCTTTCATTATCTTTAACACCATTTTTG
mHRI_shRNA_1_R	AATTCAAAAATGGTGTTAAAGATAATGAAAGCTCGAGCTTTCATTATCTTTAACACCACC
mHRI_shRNA_2_F	CACCGGCAGCCATTCGGGACAGAAATGCTCGAGCATTTCTGTCCCGAATGGCTGTTTTTG
mHRI_shRNA_2_R	AATTCAAAAACAGCCATTCGGGACAGAAATGCTCGAGCATTTCTGTCCCGAATGGCTGCC

Table S5. shRNA oligos.

Name	Sequence
LRG_F2	TCTTGTGGAAAGGACGAAACACCG
LRG_R2	TCTACTATTCTTTCCCCTGCACTGT
PE-5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PE-7	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

Table S6. sgRNA sequencing library primers.

Name	Sequence
Gamma_globin_F	TGGCAAGAAGGTGCTGACTTC
Gamma_globin_R	GCAAAGGTGCCCTTGAGATC
Beta_globin_F	TGGGCAACCCTAAGGTGAAG
Beta_globin_R	GTGAGCCAGGCCATCACTAAA
Alpha_globin_F	AAGACCTACTTCCCGCACTTC
Alpha_globin_R	GTTGGGCATGTCGTCCAC
Gamma_globin_PT_F	TTTGTGGCACCTTCTGACTG
Gamma_globin_PT_R	GCCAAAGCTGTCAAAGAA
Beta_globin_PT_F	GCTAGGCCCTTTTGCTAATC
Beta_globin_PT_R	CTGGTGGGGTGAATTCTT
ALAS2_F	CAGCGCAATGTCAAGCAC
ALAS2_R	TAGATGCCATGCTTGGAGAG
BAND3_F	ACCTCTCTCACCTCACCTTCTG
BAND3_R	AACCTGTCTAGCAGTTGGTTGG
GATA1_F	CTGTCCCCAATAGTGCTTATGG
GATA1_R	GAATAGGCTGCTGAATTGAGGG
HRI_F	CACCCGAACAGTTGGAAGGA
HRI_R	AGAGCTCTAGCAGGACCACA
BCL11A_F	ACAAACGGAAACAATGCAATGG
BCL11A_R	TTTCATCTCGATTGGTGAAGGG
ZBTB7A_F	GCTTGGGCCGGTTGAATGTA
ZBTB7A_R	GGCTGTGAAGTTACCGTCGG
Mouse_HRI_F	CTGCTGCAGAGTGAGCTTTTTC
Mouse_HRI_R	CTGAGAAAGGAGGCTTAGTTGC
Mouse_Bcl11a_F	GCCCCAAACAGGAACACATA
Mouse_Bcl11a_R	GGGGCATATTCTGCACTCAT
Mouse_ey-globin_F	ACAGCTTTGGGAACTTGTCCTC
Mouse_ey-globin_R	TCTCCAAAAGCAGTCAGCACC
Mouse_bh1-globin_F	AGGCAGCTATCACAAGCATCTG
Mouse_bh1-globin_R	AACTTGTCAAAGAATCTCTGAGTCCA
Bactin_F	ACACCCGCCACCAGTTC
Bactin_R	TACAGCCCGGGGAGCAT

Table S7. RT-qPCR primers.

CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Conclusions

In these studies, we have used a domain-focused CRISPR-Cas9 screen with which we have identified the kinase HRI as a regulator of fetal hemoglobin. We have shown that depleting this kinase induces HbF in a specific fashion without affecting cell maturation in both HUDEP2 cells and primary erythroid cultures. The effects appear to be mediated in part through a transcriptional downregulation of BCL11A. Finally, depleting the effects of depleting this kinase are amplified in combination with the HbF inducer pomalidomide. These studies thus suggest that HRI may be used as a drug target for hemoglobinopathies, and show that these types of genetic screens may further uncover regulators of fetal hemoglobin that could serve as therapeutic targets for these disorders.

By their nature, these types of screens do not guarantee that HbF regulators will be identified. But it is perhaps not entirely surprising that they have been able to uncover regulators that had not been previously characterized. In light of Johnathan Pritchard's omnigenic model of gene regulation (*87*), in that almost all expressed genes in a cell can virtually affect any trait given how interconnected regulatory networks are, there may be a number of genes that play a role in regulating fetal hemoglobin, and many have already been characterized (*48*). But perhaps the more surprising finding of our studies is the specificity of the induction of HbF. The specific mechanistic reasons for this are still unclear, but it is certainly possible that the screen selects for such hits. Once the sgRNA libraries are transduced, the cells are expanded for 8 days and differentiated for 7 days. The depletions have to be relatively well tolerated for the duration of the screens for any hits to eventually be identified. The screens thus probably select for targets that can target HbF in

a relatively specific fashion. We have now also seen this with other targets we have identified with other screening libraries.

Our current model is that HRI may play a role in regulating BCL11A transcription, which in turn controls HbF levels (Figure 4.1). Given HRI's role as a regulator of translation, we hypothesize that HRI might control one or several transcription factors that transcriptionally regulate BCL11A expression. eIF2α kinases are known to translationally regulate several transcription factors including ATF4 and DDIT3 (*88*). The transcripts for these transcription factors have one or multiple short upstream open reading frames (uORF) in their 5'UTRs. These elements can be quite common in some contexts as identified by genome-wide based on ribosome footprinting assays. Approximately 25% of transcripts in murine embryonic stem cells have been shown to have uORFs with actively engaged ribosomes (*89*). Some studies have suggested that up to 50% of transcripts might have such uORFs (88, *90*), but it remains to be seen to what extent this may reflect technical differences in the analysis of translation efficiencies (*88*). Non-cognate start codons, particularly CUG, are typically found to be most common found in these elements (*89, 94*). One recent study showed that, in murine erythroid cells, approximately 10% of transcripts had uORFs with associating ribosomes (*23*), suggesting that the extent to which such elements can be found in 5'UTRs may also be context specific. It is also important to note that while a broad range of been identified, only a handful of these have been functionally characterized. Generally, uORFs are thought to reduce translation of the downstream reading frame (*88*). This is either due to ribosome dissociation at the stop-codon of the uORF, or the translated uORF peptide can function as an attenuator of translational

elongation and lead to ribosome stalling within the uORF, thus blocking further pre-initiation complexes from progressing.

The phosphorylation of eIF2α inhibits its recycling after one round of translation as detailed above and reduces the available amount of this factor for further protein synthesis. This generally disfavors translation of uORFs, as is the case for the GADD34 which codes for a subunit of protein phosphatase-1 which dephosphorylates eIF2α (*88*). The GADD34 transcript has one such uORF and when eIF2α is phosphorylated, the translation of its downstream coding messenger ORF is diminished. It has been proposed that, when available eIF2α levels are low due to its phosphorylation, the pre-initiation complexes may begin scanning through the 5'UTR without associating with eIF2 factors, and only bind eIF2α further downstream in the 5'UTR once past the uORF (*88*). Alternatively, it is possible phosphorylated eIF2α can associate with pre-initiation complexes but alter the efficiency of recognition of start codons to reduce initiation at the uORF, though this seems to be less likely (*88*). Regardless of the proposed model, the end result is that phosphorylation of eIF2α favors translation of the downstream GADD34 messenger ORF. Once the GADD34 protein is synthesized, it then de-phosphorylates eIF2α, thus providing an auto-regulatory mechanism for eIF2α-phosphorylation. These mechanisms of translation can become much more complex when multiple uORFs are present within a 5'UTR, as some uORFs do allow for translation re-initiation. This is the case for ATF4, which is normally lowly translated, but when eIF2α-phosphorylation is high, the delayed re-initiation of ribosomes re-initiating downstream of its first uORF diverts them from a second inhibitory uORF, which then allows for translation of its messenger ORF (*88*). It is thus possible that HRI controls other transcription factors

translationally through one or multiple of these mechanisms, which may in turn control BCL11A expression.

4.2 Characterizing mechanism of HRI's regulation of HbF

A number of follow-up studies are now of interest to further characterize mechanistically how HRI controls HbF levels, test the specific domains of HRI that are required for this role and further dissecting its regulatory pathway, and investigate drug combinations with HRI depletion that may indicate further therapeutic strategies.

4.2.1 Transcription factor CRISPR screen

Towards an initial attempt to better understand how HRI controls fetal hemoglobin levels mechanistically, one approach we have thought of is to characterize more broadly which transcription factors are involved in regulating fetal hemoglobin. As overviewed in the introduction, a number of transcriptional regulators have been implicated in regulating HbF, with the major repressors being the DNA binding proteins BCL11A and LRF, but it is still unclear which other factors may play a role in this process. Furthermore, as the BCL11A enhancer is being more extensively characterized, it will be of interest to know which set of factors are required for the transcriptional regulation of this factor. Towards this goal, we've made use of another sgRNA library developed by Junwei Shi which targets essentially all annotated DNA binding domains of the human genome (*91*). This library is well distributed across all these domains, and includes approximately 9,000 sgRNAs with 1,400 different transcription factors targeted. We have now tested this library in the

HUDEP2 cultures as we had done the kinase-domain screen. These screens revealed several factors that may function as fetal hemoglobin repressors (Figures 4.2 and 4.3). One of these is the transcription factor NFIA which is part of a family of 4 transcription factors that play diverse roles in regulating transcription during development and in multiple tissues (*92*). HPFH variants have previously been shown to map to one of these family members, NFIX. Initial validation experiments suggest that both NFIA and NFIX regulate HbF, and the mechanism may involve BCL11A or Lin28B (data not shown, generated by Peng Huang, postdoc in our lab) but whether these are direct or secondary effects still remains to be determined. These factors do not seem to be perturbed upon HRI depletion, so they may not be its direct effectors. Another factor that was identified from this screen is WT1. This transcription factor has been studied in the context of Wilms tumor, which is a pediatric renal malignancy (*93*). We are still validating this factor, and its connection to HRI is to be determined. A number of other factors were identified in the screen as well that appeared to have a more intermediate phenotype. Initial validation experiments for REST showed a robust HbF induction (data not shown and generated by Xianjiang Lan, a postdoc in our lab), suggesting it may be worthwhile validating other factors which scored in this intermediate range as well. Some of these may be related to HRI's effects on HbF, but may also be worth pursuing in their own right as well.

4.2.2 Ribosome profiling

As stated above, we are currently hypothesizing that HRI controls one or several transcription factors translationally. Differential eIF2α-phosphorylation levels are known to influence the translation of a number of transcripts, including those for several

transcription factors. A recent study in murine erythroid cultures showed that a number of factors have translated uORFs in their 5'UTRs (*23*), including BCL11A, and may be translationally regulated, but this has yet to be characterized in human systems. Furthermore, the extent to which differential eIFα-phosphorylation levels regulate translation through uORFs globally is still an open question. Ribosome profiling is perfectly suited to study these questions, as it allows to map ribosome occupancy on transcripts with single nucleotide resolution in a genome-wide fashion (*94*). This method is similar to polysome profiling that was presented above. The first step of the method is to arrest translation by treating cells with cycloheximide. Cells are then lysed and ribosomes isolated using a sucrose cushion. Ribosome footprints are generated by nuclease digestion. Fragments protected by ribosomes (typically 28 to 30 nt in length) are not digested, these can be recovered for deep sequencing. The drug harringtonine blocks protein synthesis and leads to the accumulation of ribosomes at translation initiation (*89, 94*). This can be useful to precisely map sites of translation initiation. This method can be adapted to different cell types and different conditions, and thus may be particularly useful in characterizing HRI's downstream targets that may be involved in regulating fetal hemoglobin.

4.3 Domain analysis of HRI

Another set of follow-up experiments that could provide insight into HRI's role in regulating fetal hemoglobin is to further dissect which of its domains and activities are required for this function. Given that our sgRNAs and shRNAs lead to a near complete loss of the total protein, it is still unclear if the effects on HbF are due to a reduction in the total

level of the protein, or is specifically due to its enzymatic activity. To test this, we have tried doing rescue experiments by adding back HRI cDNA to cells in which it has been depleted. The cDNA can then be mutated to abrogate its kinase activity to directly test if this enzymatic function is required for the effects on HbF. We have tried multiple approaches to do these experiments, but HRI has been challenging to overexpress. We are now trying to do this with virus-free approaches, but it still remains to be seen how these will work.

Another approach to characterize HRI's role in regulating HbF is to perform a tiling screen along the coding sequence for this kinase. We have designed such a library which covers it quite well (Figure 4.4). This approach might help pinpoint specific domains within HRI that may be specifically required for HbF regulation. This is limited though by the fact that many sgRNAs lead to a strong depletion of the kinase, and resolving whether effects are due to a specific loss of function of a given domain versus degradation of the protein would be challenging. It might be possible to knock-in GFP into the endogenous HRI gene and then select for cells that have not lost HRI total expression at the end of the screen.

Lastly, while we are currently hypothesizing that HRI mediates its effects on HbF through its role on regulating protein translation, other eIF2α kinases are known to have other targets than eIF2α. To assess to what extent the effects on HbF are mediated through this factor, we could knock-in a S51A mutant into this gene to abrogate its ability to be phosphorylated. This mutant has been characterized in murine models recently (*95*), and could be useful in this context to further dissect whether this canonical pathway is involved in HbF regulation. If this showed that a target other than eIF2α is implicated, then

phosphor-proteomics may allow to elucidate other targets than may be involved in this process.

4.4 Combination therapy experiments

As shown above, while depleting HRI leads to an increase in HbF levels, its effects are partial, and certainly do not reach the same level of induction as BCL11A or LRF. It is thus unclear whether targeting HRI would lead to a strong enough elevation of HbF to be clinically useful. We showed that combining HRI depletion with the HbF inducing drug pomalidomide amplifies the effects on HbF. This suggests that if an inhibitor could be developed to block HRI, or lead to its depletion using a thalidomide derivative for chemical degradation, this approach could be combined with other HbF inducing drugs to lead to a robust induction while hopefully using lower drug dosages to limit off-target effects. It is worth pointing out that combination therapy may be the most adapted approach for the treatment of hemoglobinopathies. The few factors that were identified from the GWAS studies may be the only targets that be adapted for treatment with a single agent target, as these factors' reduction in patients are relatively well tolerated, and small perturbations such as SNPs are enough to give a significant induction of HbF. BCL11A in particular is a potent repressor, and if it could be targeted with a compound using a chemical degradation system previously described (*96*), it could certainly serve as a very useful pharmacologic target if this can be accomplished in a tissue specific fashion. Combination therapy has its own set of challenges, but does offer the possibility to minimize off-target effects of each drug will maximizing on-target activity.

To this end, we are currently further investigating combination treatments with other known HbF inducing drugs including hydroxyurea (HU) and the recently described G9a inhibitor (UNC0638) presented in the introduction (work done by Scott Peslak, clinical fellow in the lab). So far, HU has little additive effect with HRI depletion in primary erythroid cultures, but both pomalidomide and the G9a inhibitor have additive if not synergistic effects on HbF levels. It still remains to be seen to what extent these effects are specific, but the combination of these treatments seems to be well tolerated by the cells and do not affect erythroid maturation.

Another approach to characterize factors that may have additive or synergistic effects with HRI would be to repeat similar CRISPR screens as outlined above, but either in the context of HRI depletion, or not. If sgRNAs show differential enrichment across these different conditions, this may be indicative of such a combinatorial effect. These types of approaches can be easily adapted to testing combination treatments over the background of a single depletion such as HRI. This only involves transducing these libraries in the depleted cell lines. Perhaps the more challenging aspect of these types of screens is the number of possible libraries to screen for, and which of these to use. It would be possible to do a genome-wide screen on the background of an HRI depletion, which may be better adapted for this particular goal. The quality of the screen would depend on how well such libraries deplete their targeted factors, particularly as many of these use early exon targeting sgRNAs as opposed to being domain-targeted. A few approaches also allow multiplexing sgRNAs in the same expression vector. Such approaches allow for essentially all inner combinations of a given library, which can in theory cover more permutations of sgRNAs, but also lead to very large library sizes.

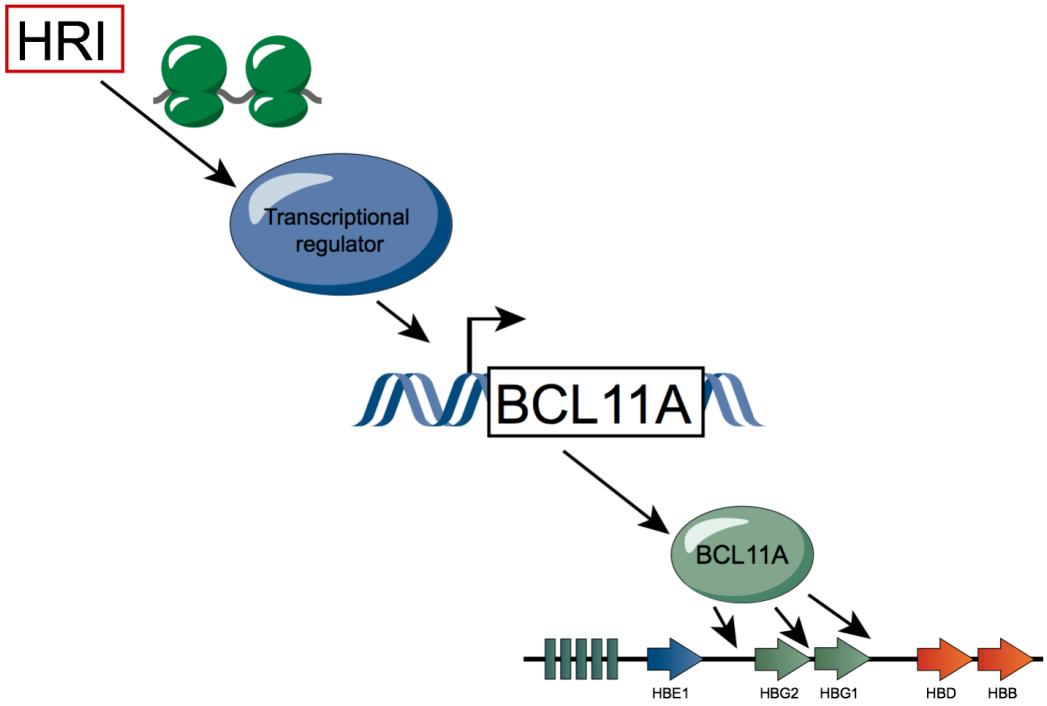

Figure 4.1 – Model for HRI's control of fetal hemoglobin expression. We hypothesize that HRI translationally controls one or a set of transcription factors that in turn regulate BCL11A expression which then influences fetal globin levels.

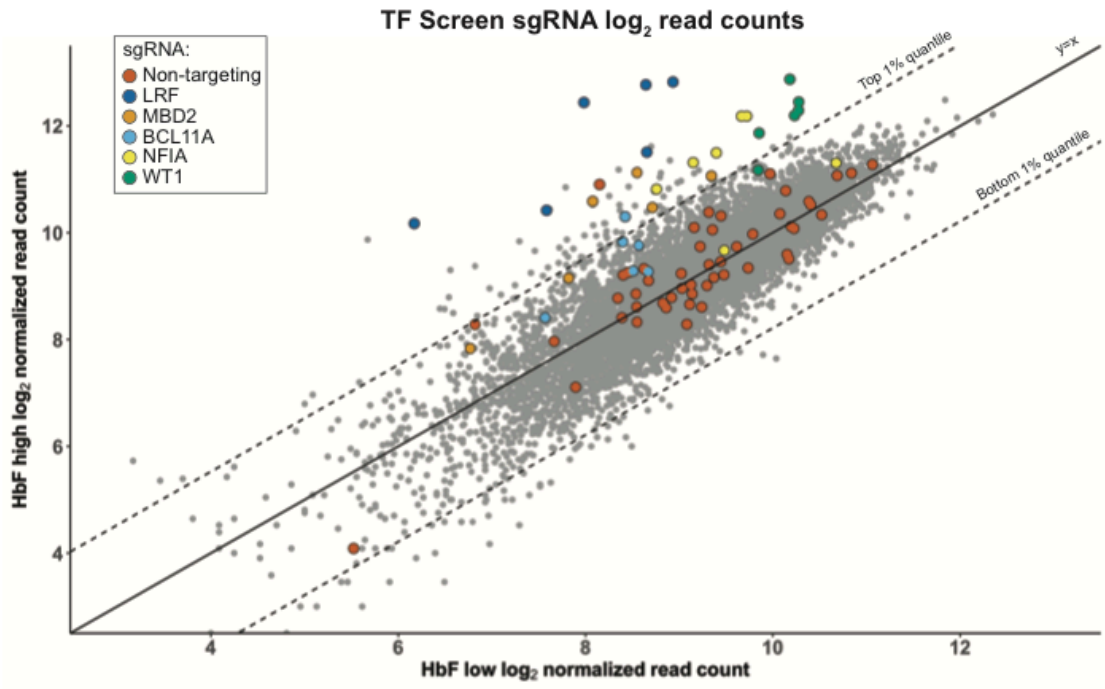

Figure 4.2 – Transcription factor CRISPR screen. Several TFs were identified as being potential fetal hemoglobin repressors including NFIA and WT1. Control sgRNAs are also indicated.

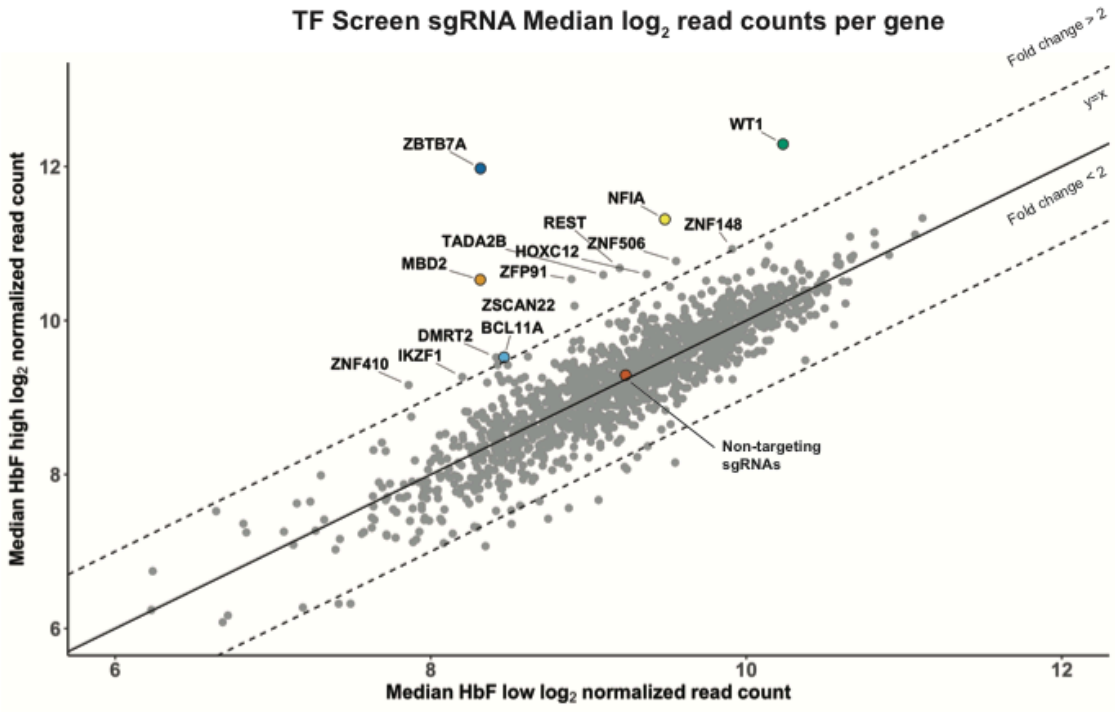

Figure 4.3 – Transcription factor CRISPR screen. This is the same data as in Figure 4.2 but the median read counts for each gene are shown to simplify the display. A number of TFs scored in a more intermediate range in this screen.

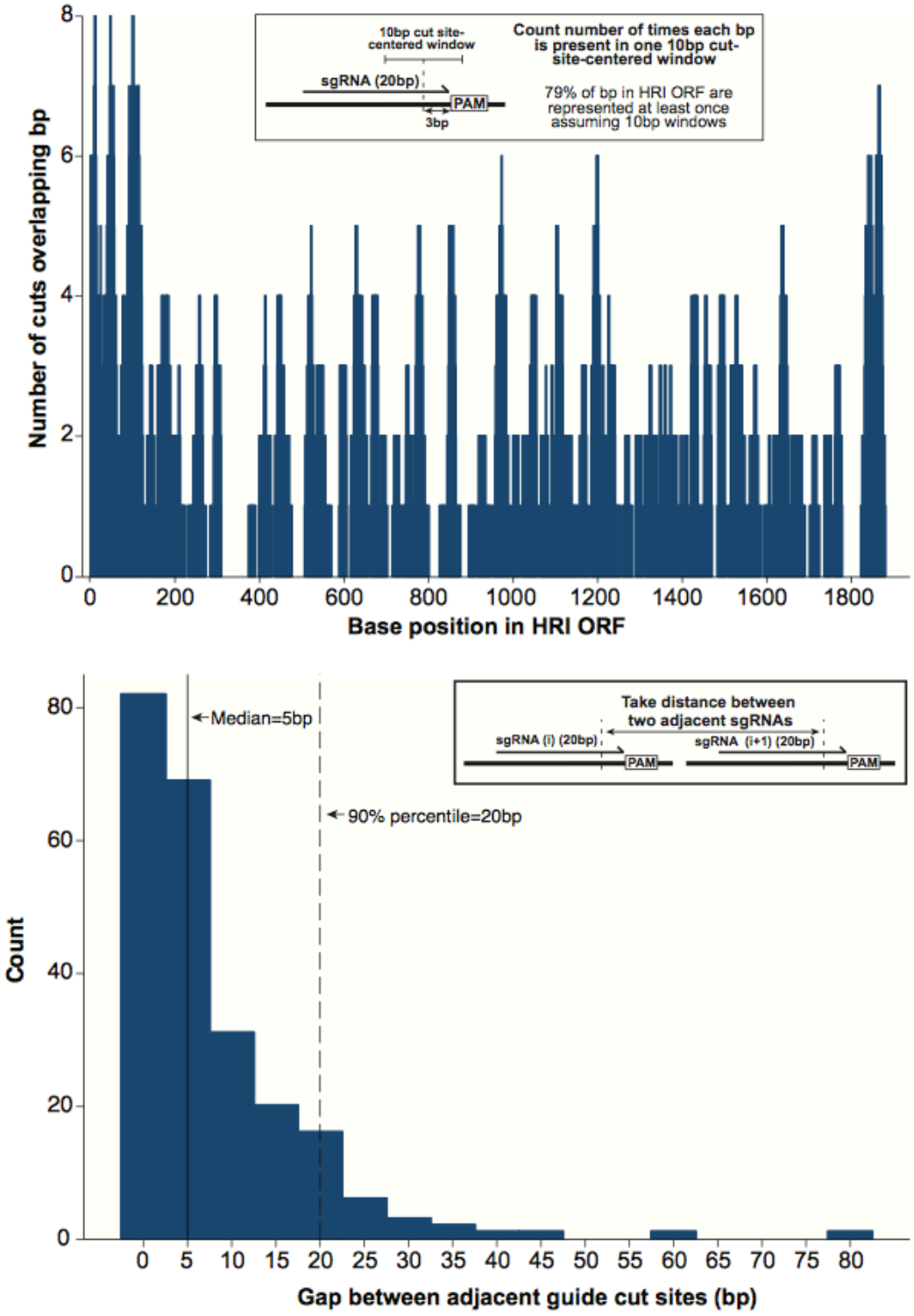

Figure 4.4 – Tiling screen sgRNA distribution along HRI open reading frame. This library contains approximately 300 sgRNAs with over 50% of them being less than 5bp appart, and over 90% being 20bp apart.

References

1. L. Pauling, H. A. Itano, S. Singer, I. C. Wells, Sickle Cell Anemia, a Molecular Disease. *Science*. **110**, 543–548 (1949).

2. W. A. Eaton, Linus Pauling and sickle cell disease. *Biophysical Chemistry*. **100**, 109–116 (2002).

3. H. J. Freedman, The eighth day of creation. *Cold Spring Harbor Laboratory Press* (1979).

4. E. Dzierzak, S. Philipsen, Erythropoiesis: Development and Differentiation. *Cold Spring Harbor Perspectives in Medicine*. **3**, a011601 (2013).

5. J. Palis, Primitive and definitive erythropoiesis in mammals. *Frontiers in Physiology*. **5**, 3 (2014).

6. S. H. Orkin, L. I. Zon, Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*. **132**, 631–644 (2008).

7. R. Lis *et al.*, Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature*. **545**, 439–445 (2017).

8. R. Sugimura *et al.*, Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature*. **545**, 432–438 (2017).

9. J. Sun *et al.*, Clonal dynamics of native haematopoiesis. *Nature*. **514**, 322 (2014).

10. A. E. Rodriguez-Fraticelli *et al.*, Clonal analysis of lineage fate in native haematopoiesis. *Nature*. **553**, 212 (2018).

11. L. Pan, R. Yan, W. Li, K. Xu, Super-Resolution Microscopy Reveals the Native Ultrastructure of the Erythrocyte Cytoskeleton. *Cell Reports*. **22**, 1151–1158 (2018).

12. P. E. Love, C. Warzecha, L. Li, Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends in Genetics*. **30**, 1–9 (2014).

13. K. R. Katsumura, A. W. DeVilbiss, N. J. Pope, K. D. Johnson, E. H.

Bresnick, Transcriptional Mechanisms Underlying Hemoglobin Synthesis. *Cold Spring Harbor Perspectives in Medicine*. **3**, a015412 (2013).

14. L. Li *et al.*, Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood*. **121**, 4575–4585 (2013).

15. G. D. Gregory *et al.*, FOG1 requires NuRD to promote hematopoiesis and maintain lineage fidelity within the megakaryocytic-erythroid compartment. *Blood*. **115**, 2156–2166 (2010).

16. D. L. Letting, Y.-Y. Chen, C. Rakowski, S. Reedy, G. A. Blobel, Context-dependent regulation of GATA-1 by friend of GATA-1. *P Natl Acad Sci Usa*. **101**, 476–481 (2004).

17. J. J. Welch *et al.*, Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*. **104**, 3136–3147 (2004).

18. A. J. Stonestrom *et al.*, Functions of BET proteins in erythroid gene expression. *Blood*. **125**, 2825–2834 (2015).

19. E.-F. Gautier *et al.*, Comprehensive Proteomic Analysis of Human Erythropoiesis. *Cell Reports*. **16**, 1470–1484 (2016).

20. E. Khandros, C. S. Thom, J. D'Souza, M. J. Weiss, Integrated protein quality-control pathways regulate free α-globin in murine β-thalassemia. *Blood*. **119**, 5265–5275 (2012).

21. A. T. Nguyen *et al.*, UBE2O remodels the proteome during terminal erythroid differentiation. *Science*. **357**, eaan0218 (2017).

22. K. Yanagitani, S. Juszkiewicz, R. S. Hegde, UBE2O is a quality control factor for orphans of multiprotein complexes. *Science*. **357**, 472–475 (2017).

23. J. R. Alvarez-Dominguez, X. Zhang, W. Hu, Widespread and dynamic translational control of red blood cell development. *Blood*. **129**, 619–629 (2017).

24. H. A. Dailey, P. N. Meissner, Erythroid Heme Biosynthesis and Its Disorders. *Cold Spring Harb Perspectives Medicine*. **3**, a011676 (2013).

25. M. W. Hentze, M. U. Muckenthaler, B. Galy, C. Camaschella, Two to Tango:

Regulation of Mammalian Iron Metabolism. *Cell*. **142**, 24–38 (2010).

26. Z. Yang *et al.*, Delayed globin synthesis leads to excess heme and the macrocytic anemia of Diamond Blackfan anemia and del(5q) myelodysplastic syndrome. *Science Translational Medicine*. **8**, 338ra67-338ra67 (2016).

27. N. Tanimura *et al.*, Mechanism governing heme synthesis reveals a GATA factor/heme circuit that controls differentiation. *EMBO reports*. **17**, 249–265 (2016).

28. J.-J. J. Chen, Translational control by heme-regulated eIF2α kinase during erythropoiesis. *Current opinion in hematology*. **21**, 172–8 (2014).

29. R. K. Khajuria *et al.*, Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell*. **173**, 90-103.e19.

30. J. Xu *et al.*, Combinatorial Assembly of Developmental Stage-Specific Enhancers Controls Gene Expression Programs during Human Erythropoiesis. *Developmental Cell*. **23**, 796–811 (2012).

31. V. G. Sankaran, S. H. Orkin, The Switch from Fetal to Adult Hemoglobin. *Cold Spring Harbor Perspectives in Medicine*. **3**, a011643 (2013).

32. C. R. Vakoc *et al.*, Proximity among Distant Regulatory Elements at the β-Globin Locus Requires GATA-1 and FOG-1. *Molecular Cell*. **17**, 453–462 (2005).

33. W. Deng *et al.*, Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell*. **149**, 1233–1244 (2012).

34. W. Deng *et al.*, Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. *Cell*. **158**, 849–860 (2014).

35. D. R. Higgs, The Molecular Basis of α-Thalassemia. *Cold Spring Harbor Perspectives in Medicine*. **3**, a011718 (2013).

36. S. Thein, The Molecular Basis of β-Thalassemia. *Cold Spring Harbor Perspectives in Medicine*. **3**, a011700 (2013).

37. F. B. Piel, M. H. Steinberg, D. C. Rees, Sickle Cell Disease. *The New*

England Journal of Medicine. **376**, 1561–1573 (2017).

38. D. Shriner, C. N. Rotimi, Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase. *The American Journal of Human Genetics*, doi:10.1016/j.ajhg.2018.02.003 .

39. F. B. Piel, S. I. Hay, S. Gupta, D. J. Weatherall, T. N. Williams, Global Burden of Sickle Cell Anaemia in Children under Five, 2010–2050: Modelling Based on Demographics, Excess Mortality, and Interventions. *PLoS Medicine*. **10**, e1001484 (2013).

40. D. J. Weatherall, The inherited diseases of hemoglobin are an emerging global health burden. *Blood*. **115**, 4331–4336 (2010).

41. O. Platt *et al.*, Mortality in sickle cell disease. Life expectancy and risk factors for early death. *The New England journal of medicine*. **330**, 1639–44 (1994).

42. D. E. Bauer, S. C. Kamran, S. H. Orkin, Reawakening fetal hemoglobin: prospects for new therapies for the β-globin disorders. *Blood*. **120**, 2945–2953 (2012).

43. M. C. Canver *et al.*, Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nature Genetics*. **49**, 625–634 (2017).

44. V. G. Sankaran *et al.*, Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science (New York, N.Y.)*. **322**, 1839–42 (2008).

45. N. Liu *et al.*, Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell*, doi:10.1016/j.cell.2018.03.016 .

46. V. G. Sankaran *et al.*, Developmental and species-divergent globin switching are driven by BCL11A. *Nature*. **460**, 1093–1097 (2009).

47. J. Xu *et al.*, Correction of Sickle Cell Disease in Adult Mice by Interference with Fetal Hemoglobin Silencing. *Science*. **334**, 993–996 (2011).

48. J. Xu *et al.*, Corepressor-dependent silencing of fetal hemoglobin expression

by BCL11A. *Proceedings of the National Academy of Sciences*. **110**, 6518–6523 (2013).

49. G. E. Martyn *et al.*, Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nature Genetics*, 1–6 (2018).

50. D. E. Bauer *et al.*, An Erythroid Enhancer of BCL11A Subject to Genetic Variation Determines Fetal Hemoglobin Level. *Science*. **342**, 253–257 (2013).

51. M. C. Canver *et al.*, BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. **527**, 192–197 (2015).

52. B. Wienert *et al.*, Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nat Commun*. **6**, 7085 (2015).

53. E. A. Traxler *et al.*, A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nature Medicine*. **22**, 987–990 (2016).

54. P. Huang *et al.*, Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes & Development*. **31**, 1704–1713 (2017).

55. V. G. Sankaran *et al.*, A functional element necessary for fetal hemoglobin silencing. *The New England journal of medicine*. **365**, 807–14 (2011).

56. T. Masuda *et al.*, Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science*. **351**, 285–289 (2016).

57. D. Zhou, K. Liu, C.-W. Sun, K. M. Pawlik, T. M. Townes, KLF1 regulates BCL11A expression and γ- to β-globin gene switching. *Nature Genetics*. **42**, 742–744 (2010).

58. E. P. Kransdorf *et al.*, MBD2 is a critical component of a methyl cytosine-binding protein complex isolated from primary erythroid cells. *Blood*. **108**, 2836–2845 (2006).

59. S. M. Bakanay *et al.*, Mortality in sickle cell patients on hydroxyurea therapy. *Blood*. **105**, 545–547 (2005).

60. V. G. Sankaran, M. J. Weiss, Anemia: progress in molecular mechanisms and therapies. *Nature Medicine*. **21**, 221–230 (2015).

61. K. Takekoshi, Y. Oh, K. Westerman, I. London, P. Leboulch, Retroviral transfer of a human beta-globin/delta-globin hybrid gene linked to beta locus control region hypersensitive site 2 aimed at the gene therapy of sickle cell disease. *Proc National Acad Sci*. **92**, 3014–3018 (1995).

62. J.-A. Ribeil *et al.*, Gene Therapy in a Patient with Sickle Cell Disease. *New Engl J Medicine*. **376**, 848–855 (2017).

63. A. A. Thompson *et al.*, Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia. *The New England Journal of Medicine*. **378**, 1479–1493 (2018).

64. A. Perumbeti *et al.*, A novel human gamma-globin gene vector for genetic correction of sickle cell anemia in a humanized sickle mouse model: critical determinants for successful correction. *Blood*. **114**, 1174–1185 (2009).

65. R. J. Ihry *et al.*, p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat Med*, 1–8 (2018).

66. K. Mallhi *et al.*, Non-myeloablative conditioning for second hematopoietic cell transplantation for graft failure in patients with non-malignant disorders: a prospective study and review of the literature. *Bone Marrow Transpl*. **52**, 726 (2017).

67. D. G. NATHAN, Regulation of Fetal Hemoglobin Synthesis in the Hemoglobinopathies. *Ann Ny Acad Sci*. **445**, 177–187 (1985).

68. O. Platt *et al.*, Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. *J Clin Invest*. **74**, 652–656 (1984).

69. A. Renneville *et al.*, EHMT1 and EHMT2 inhibition induces fetal hemoglobin expression. *Blood*. **126**, 1930–9 (2015).

70. L. A. Parseval *et al.*, Pomalidomide and lenalidomide regulate erythropoiesis and fetal hemoglobin production in human CD34+ cells. *Journal of Clinical Investigation*. **118**, 248–258 (2008).

71. S. E. Meiler *et al.*, Pomalidomide augments fetal hemoglobin production without the myelosuppressive effects of hydroxyurea in transgenic sickle cell mice. *Blood*. **118**, 1109–1112 (2011).

72. B. lmovits *et al.*, Pomalidomide reverses γ-globin silencing through the transcriptional reprogramming of adult hematopoietic progenitors. *Blood*. **127**, 1481–92 (2016).

73. R. Kurita *et al.*, Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. *PLoS ONE*. **8**, e59890 (2013).

74. E. K. Brinkman, T. Chen, M. Amendola, B. van Steensel, Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*. **42**, e168–e168 (2014).

75. P. D. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*. **31**, 827–832 (2013).

76. J. Cox, M. Mann, *Nat Biotechnol*, in press, doi:10.1038/nbt.1511 .

77. A. Basak *et al.*, BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. *Journal of Clinical Investigation*. **125**, 2363–2368 (2015).

78. G. Manning, D. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The Protein Kinase Complement of the Human Genome. *Science*. **298**, 1912–1934 (2002).

79. J. Shi *et al.*, Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nature Biotechnology*. **33**, 661–667 (2015).

80. H. Nishimasu *et al.*, Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*. **156**, 935–949 (2014).

81. B. Chen *et al.*, Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*. **155**, 1479–1491 (2013).

82. C. K. Hahn, C. H. Lowrey, Eukaryotic initiation factor 2α phosphorylation mediates fetal hemoglobin induction through a post-transcriptional mechanism. *Blood*. **122**, 477–485 (2013).

83. C. K. Hahn, C. H. Lowrey, Induction of fetal hemoglobin through enhanced translation efficiency of γ-globin mRNA. *Blood*. **124**, 2730–4 (2014).

84. A. Han *et al.*, Heme-regulated eIF2alpha kinase (HRI) is required for translational regulation and survival of erythroid precursors in iron deficiency. *The EMBO journal*. **20**, 6909–18 (2001).

85. S. Liu *et al.*, Haem-regulated eIF2α kinase is necessary for adaptive gene expression in erythroid precursors under the stress of iron deficiency. *British Journal of Haematology*. **143**, 129–137 (2008).

86. S. Liu *et al.*, Deficiency of heme-regulated eIF2alpha kinase decreases hepcidin expression and splenic iron in HFE-/- mice. *Haematologica*. **93**, 753–6 (2008).

87. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. **169**, 1177–1186 (2017).

88. A. G. Hinnebusch, I. P. Ivanov, N. Sonenberg, Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science*. **352**, 1413–1416 (2016).

89. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*. **147**, 789–802 (2011).

90. S. Lee *et al.*, Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc National Acad Sci*. **109**, E2424–E2432 (2012).

91. S. A. Lambert *et al.*, The Human Transcription Factors. *Cell*. **172**, 650–665 (2018).

92. L. Harris, L. A. Genovesi, R. M. Gronostajski, B. J. Wainwright, M. Piper, Nuclear factor one transcription factors: Divergent functions in developmental versus adult stem cell populations. *Developmental Dynamics*. **244**, 227–238 (2015).

93. L. Yang, Y. Han, S. F. Saiz, Minden, A tumor suppressor and oncogene: the WT1 story. *Leukemia*. **21**, 2404624 (2007).

94. G. A. Brar, J. S. Weissman, Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology*. **16**, 651–664 (2015).

95. S. Zhang *et al.*, HRI coordinates translation by eIF2αP and mTORC1 to mitigate ineffective erythropoiesis in mice during iron deficiency. *Blood*. **131**, 450–461 (2018).

96. G. E. Winter *et al.*, Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science*. **348**, 1376–1381 (2015).

97. N. Novershtern *et al.*, Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. *Cell*. **144**, 296–309 (2011).

98. J. Xu *et al.*, Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes & Development*. **24**, 783–798 (2010).