We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,300 Open access books available
116,000 International authors and editors
130M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter 12

Precision Medicine for Sickle Cell Disease: Discovery of Genetic Targets for Drug Development

Betty S. Pace, Nicole H. Lopez, Xingguo Zhu and Biaoru Li

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/64817

Abstract

Sickle cell disease (SCD) consists of inherited monogenic hemoglobin disorders affecting over three million people worldwide. Efforts to establish precision medicine based on the discovery of genetic polymorphisms associated with disease severity are ongoing to inform strategies for novel drug design. Numerous gene mutations have been associated with the clinical complications of SCD such as frequency of pain episodes, acute chest syndrome, and stroke among others. However, these discoveries have not produced additional treatment options. To date, Hydroxyurea remains the only Food and Drug Administration-approved agent for treating adults with SCD; recently it was demonstrated to be safe and effective in children. The main action of Hydroxyurea is the induction of fetal hemoglobin, a potent modifier of SCD clinical severity. Three inherited gene loci including Xmn1-HBG2, HBS1L-MYB and BCL11A have been linked to HBG expression, however the greatest progress has been made to develop BCL11A as a therapeutic target. With the expanded availability of next generation sequencing, there exist opportunities to discover additional genetic modifiers of SCD. The progress made over the last two decades to define markers of disease severity and the implications for achieving precision medicine to treat the complications of SCD will be discussed.

Keywords: fetal hemoglobin, single nucleotide polymorphism, drug discovery, genome-wide association studies

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
1. Introduction

Sickle cell anemia is caused by an A to T point mutation in the sixth codon of the β-globin (HBB) gene on chromosome 11 leading to the production of hemoglobin S (HbSS) during adult development. When the sickle mutation is combined with one of over 400 additional mutations reported in the HBB locus, different subtypes of sickle cell disease (SCD) are produced. For example, heterozygosity for the sickle HBB gene and hemoglobin C produces HbSC disease [1]. A definitive diagnosis of SCD can be made by hemoglobin electrophoresis, isoelectric focusing, or high-performance liquid chromatography. However, DNA testing is required to detect the presence of β-thalassemia mutations, which when inherited with the sickle HBB causes HbS-β⁰-thalassemia and HbSβ⁺-thalassemia.

About one in 500 African-American and one in 36,000 Hispanic-American children are born with SCD disease [2], which is diagnosed at birth by newborn screening in the United States. The carrier state or sickle cell trait is detected in 1:13 African Americans and 1:100 Hispanic Americans [3] with an estimated 2.5 million Americans with sickle cell trait [4]. Worldwide about 3.2 million people have SCD and 43 million have sickle cell trait [5] with 80% occurring in sub-Saharan Africa mainly as a protective mechanism against malaria. Moreover, the HBB sickle mutation also occurs in Europe, India, the Arabian Peninsula, and Brazil [6].

Hemoglobin is a tetrameric protein, composed of two α-like and two β-like globin polypeptide chains, which transports oxygen to the body tissues. During human development, two switches in the type of hemoglobin synthesized occur, a process known as hemoglobin switching [1]. The first switch at 6–8 weeks of development involves ε-globin gene silencing and activation of the HBG2 and HBG1 genes throughout fetal erythropoiesis, during which Gγ-globin and Aγ-globin fetal hemoglobin (HbF; α₂γ₂) are produced. The second switch occurs shortly after birth when the HBG1/HBG2 genes are silenced and HBB is activated. HbF levels decline to <1% of total hemoglobin by 6–12 months of age [7], and HbF is restricted to a population of erythrocytes called F-cells [8]. During hemoglobin switching, the site of hematopoiesis moves from the yolk sac to the liver/spleen and finally the bone marrow, which becomes the main site of hematopoiesis where adult hemoglobin A (HbA, α₂β₂) is produced in healthy individuals [1]. As the level of HbF decreases around 5–6 months of age, the clinical symptoms of SCD are observed due to high HbS levels and polymerization under deoxygenated conditions producing sickle-shaped red blood cells (RBCs), vascular occlusion, and tissue ischemia. Therefore, precision medicine based on genetic or pharmacologic approaches to maintain high HbF levels is a proven efficacious strategy to treat SCD.

2. Clinical manifestations of sickle cell disease

Over the last 30 years, survival in people living with SCD has improved significantly due to decreased death rates during infancy. However, morbidity remains high due to central nervous system and pulmonary complications during childhood and end-organ damage in adults [9, 10]. The average life expectancy of people with SCD is 50 years in the United States [11].
Individuals with SCD experience a chronic hemolytic anemia caused by HbS polymerization under deoxygenated conditions, which [12] produces RBC membrane damage and a shortened life span of 14–21 days. As a result, HbSS patients have an average hemoglobin level of 6–8 g/dL with an elevated reticulocyte count and plasma lactate dehydrogenase level [13]. Furthermore, the damaged membrane leads to inflexible and dehydrated sickled RBCs and abnormal adhesion to the vascular endothelium producing the vasculopathy observed in persons with SCD [13].

The most common pathophysiology of SCD is vaso-occlusive (VOC) events produced by tissue ischemia leading to pain and acute or chronic injury to the spleen, brain, lungs, kidneys, and bones [13]. Individuals with a severe SCD sub-phenotype have more frequent VOC events, a higher white blood cell count, a lower HbF level, and increased blood vessel flow resistance under deoxygenation conditions [14–16]. The most common clinical manifestation of SCD is acute painful episodes which occur mainly in the extremities, but can involve the abdomen, back, and chest [17, 18].

As HbF falls below protective levels at around 6–12 months of age, dactylitis involving pain and swelling of the hands and feet is an early manifestation of SCD and is a risk factor for disease severity [19]. Splenic sequestration occurs in 30% of children between the ages of 6 months to 3 years, which can cause severe life-threatening anemia and death if not treated promptly. Over time, repeated episodes of VOC in the spleen lead to infarction and a markedly increased risk for infection due to encapsulated bacteria such as Streptococcus pneumonia, Haemophilus influenza, and Staphylococcus aureus among others [20]. To address this significant cause of early mortality, the Prophylactic Penicillin Study I was conducted which demonstrated the ability of prophylactic penicillin to decrease overwhelming sepsis by 90% and improved survival among infants with SCD [21]. This study provided the rationale for establishing newborn screening for SCD in the late 1980s to facilitate the initiation of penicillin prophylaxis in the first few months of life to protect against infection and prevent early mortality. Penicillin prophylaxis has become the standard of care worldwide.

Other types of VOC events include acute chest syndrome [22, 23], silent and acute cerebral infarcts [24, 25], and osteonecrosis of the femoral head. Episodes of acute chest syndrome can be caused by pulmonary VOC, infection, and/or fat emboli from bone marrow infarcts [22]. Long-term damage in the lungs can precede pulmonary hypertension [26] in older children and adults with SCD causing high morbidity and mortality. By adolescents, 50% of individuals with SCD suffer silent cerebral infarcts [27] and 10% of children over the age of 2 experience overt strokes requiring chronic transfusions [28, 29]. The process of VOC can affect any organ system producing a wide variety of complications in SCD involving the heart, liver, gall bladder, kidney, and skin [30].

3. Treatment of vaso-occlusive complications

Blood transfusions are the mainstay of therapy for individuals suffering from acute and chronic complications of SCD. Red blood cell transfusions improve the oxygen-carrying ca-
pacity and prevent sickling by decreasing the HbS level to <30% of total hemoglobin [31–33]. Transfusions are also used for the acute exacerbation of anemia associated with splenic sequestration and aplastic crisis caused by Parvo B19 virus infection [34]. The most common symptom in persons with SCD is acute and chronic pain due to tissue ischemia, which is correlated with long-term survival [35]. Therefore, early aggressive treatment of pain episodes to prevent complications is the standard of care [36]. Recent research has provided insights into mechanisms of pain related to tissue injury (nociceptive), nerve injury (neuropathic), or unknown causes (idiopathic). Effective pain treatment is most often achieved using opioid narcotics combined with nonsteroidal anti-inflammatory drug. To address the long-term effects of repeated pain episodes, extensive research has been conducted to develop drugs that induce HbF, which inhibits HbS polymerization [37] to improve the clinical symptoms of SCD. Based on findings in the Multicenter Study of Hydroxyurea [38], this agent is the only Food and Drug Administration-approved drug for the treatment of adults with SCD [39]. Subsequent studies in children including BABY HUG demonstrated that hydroxyurea (HU) is an effective HbF inducer and can be used safely in the first year of life [40]. Unfortunately, HU has a 30% nonresponse rate in adults, causes bone marrow suppression, and has detrimental effects on fertility [38, 41]. Therefore, the development of novel therapeutic agents based on inherited mutations that alter the expression of the HBG1/HBG2 genes to produce high HbF levels is desired to establish precision medicine for SCD.

4. Genetic modifiers of sickle cell disease severity

While homozygosity for the βS-globin gene mutation (HBB; glu6val) causes sickle cell anemia, the clinical diversity of phenotypes and disease severity are similar to the manifestations of multigenic disorders. Intensive studies have been performed to identify genetic risk factors correlated with SCD complications such as stroke, leg ulcers, pulmonary artery hypertension, priapism, and osteonecrosis. To extend the findings of genome-wide association studies of single nucleotide polymorphisms (SNPs) linked with clinical phenotypes, more advanced genomic techniques including next-generation DNA sequencing provide new opportunities to define mechanisms of SCD complications. A comprehensive review of genetic studies conducted in SCD is beyond the scope of this chapter. Therefore, we focus our discussion on efforts to discover SNPs associated with the clinical sub-phenotypes of SCD including pain severity, acute chest syndrome, pulmonary hypertension, osteonecrosis, priapism, leg ulcers, and nephropathy.

4.1. Vaso-occlusive pain

SCD patients experience a wide variety of clinical pain ranging from acute mild/severe to persistent chronic pain. The underlying mechanisms of differences in pain rates are complex and likely involve a number of genetic polymorphisms in several biological systems. Studies have been conducted that provide insights into SNPs associated with the frequency and
severity of pain in SCD. Jhun et al. [42] identified mutations in the dopamine D3 receptor (Ser9Gly heterozygotes) associated with a lower acute pain rate. The most commonly used opioid medications including codeine and hydrocodone require cytochrome P450 2D6 (CYP2D6) for drug activation, which can impact the efficacy of these agents. The CYP2D6 gene is highly polymorphic, with variant alleles that result in decreased, absent, or ultra-rapid metabolism [43]. Altered CYP2D6 enzymatic activity in CYP2D6*17 (reduced activity), CYP2D6*5 (gene deletion), and CYP2D6*4 (absent function) is correlated with the analgesic response to codeine and hydrocodone. Therefore, genotyping the CYP2D6 gene is a reasonable approach for developing personalized medicine for the treatment of pain in persons with SCD. Moreover, missense or frame-shift mutations in CYP2C9 decrease or abolish enzymatic activity, respectively, which impairs opioid activation [44, 45]. Likewise, an SNP in the promoter of the gene encoding the enzyme uridine 5′-diphospho (UDP)-glucuronosyltransferase 2B7 (−840G/A) responsible for morphine glucuronidation in the liver is associated with lower morphine metabolites in sickle cell patients suggesting that higher doses of morphine may be required to achieve adequate pain control [46].

4.2. Acute chest syndrome/pulmonary hypertension

Acute chest syndrome continues to contribute to significant morbidity and mortality in children and adults with SCD [47]; therefore, the discovery of genetic modifiers of this complication has the potential for high impact and the design of precision medicine. Redha et al. [48] investigated the association of the vascular endothelial growth factor A (VEGFA) 583C/T mutation with acute chest rates in children with SCD. The presence of the 583T/T genotype was associated with increased serum VEGF levels while the VEGFA 583C/T caused reduced VEGF serum levels.

The rate of RBC hemolysis and release of free heme in the circulation are associated with clinical severity of SCD. Heme oxygenase-1 (HMOX1) is the inducible, rate-limiting enzyme in the catabolism of heme which attenuates the severity of VOC and hemolytic events. The (GT)(n) dinucleotide repeat in the promoter of HMOX1 is highly polymorphic, with long repeats linked to decreased gene activation. Bean et al. [49] examined two HMOX1 promoter polymorphisms including −413A/T and the (GT)(n) microsatellite (with allele (GT)(n) length from 13 to 45 repeats). The length of the (GT)(n) allele was associated with acute chest syndrome, but not pain rates in children with SCD.

Over the last decade, numerous studies have been conducted to define risk factors associated with pulmonary artery hypertension [50, 51], which defines a severe sub-phenotype of SCD leading to premature death. SNPs in genes involved in the regulation of endothelial function, which alter the synthesis of the endothelium-derived vasodilators nitric oxide and prostacyclin, have been implicated [52]. An extended screen of 297 SNPs in 49 candidate genes [53] identified mutations in the transforming growth factor (TGF) superfamily including the activin A type II-like 1 receptor (ACVRL1), bone morphogenetic protein (BMP) receptor 2, bone morphogenetic protein 6, and the β-1 adrenergic receptor (ADRB1) associated with pulmonary artery hypertension. A multiple regression model using age and hemoglobin as covariates demonstrated that SNPs in ACVRL1, BMP6, and ADRB1 independently contribute to pulmo-
nary hypertension risk. These findings offer promise for identifying patients at risk for this complication and developing novel therapeutic targets for SCD.

A recent study by Al-Habboubi et al. [54] examined the association between VEGF secretion and VOC rates among 210 individuals with SCD. Mutations in VEGFA including rs2010963 heterozygous and rs833068 and rs3025020 homozygous states were associated with increased pain rates. Moreover, Youssry et al. [55] observed that the homozygous mutant eNOS 786T/T was significantly associated with a high risk of acute chest syndrome. By contrast, the wild-type eNOS 4a/4b genotype was protective against VOC and pulmonary hypertension while the homozygous haplotype (C, 4a) was significantly associated with the risk of VOC pain, acute chest syndrome, and pulmonary hypertension. Thus, eNOS SNPs may be useful as a genetic marker of prognostic value in SCD to predict a severe disease sub-phenotype.

4.3. Cerebral vascular disease

SCD is the most common cause of ischemic stroke occurring in 10% of children under 15 years of age; by contrast, hemorrhagic strokes are observed more commonly in adults over 30 years of age [56]. Genetic polymorphisms in multiple genes have been implicated in childhood stroke risk. For example, a mutation in vascular adhesion molecule-1 (VCAM1) including the G1238C in the coding region was protective and the intronic T1594C SNP predisposed to small-vessel stroke [57–59]. Mutations in the interleukin (IL)4R, tumor necrosis factor (TNF), and ADRB2 genes were found to be independently associated with stroke susceptibility in the large-vessel stroke subgroup, while SNPs in VCAM1 and LDLR NcoI genes were associated with small-vessel stroke risk [59]. Additional genes have been implicated in stroke risk such as the GT-repeat polymorphism in the angiotensinogen gene including alleles A3 and A4, which conferred a fourfold increase in risk [60]. Hoppe et al. [61] identified SNPs in the cystathionine-β-synthase (278thr) and the apoE3 genes that were associated with protection and increased risk for stroke, respectively.

Ischemic stroke is common in children with SCD producing high morbidity and mortality. A meta-analysis by Sarecka-Hujar et al. [62] demonstrated the association of SNP 677C/T in the methylenetetrahydrofolate reductase gene with the risk of stroke. Abnormalities in the coagulation pathway have been implicated in the pathogenesis of cerebral bleeding. For example, protein Z, a vitamin K-dependent glycoprotein structurally related to the vitamin K-dependent coagulation factors, is devoid of catalytic activity and inhibits the generation of thrombin. Mahdi et al. [63] identified three SNPs in the protein Z gene promoter (rs3024718, rs3024719, and rs3024731) and one intronic SNP rs3024735 associated with stroke risk suggesting that reduced protein Z levels produced a procoagulant state and increased risk for thrombotic diseases including ischemic stroke. These studies provide evidence for genetic markers that can be used to assess stroke risk in SCD and targeted for therapeutic intervention.

4.4. Osteonecrosis

Repeated episodes of bone infarction caused by vaso-occlusive events precede osteonecrosis of the head of the femur and humerus, a disabling complication of SCD [64, 65]. The discovery
of SNPs in genes involved in bone morphogenesis, metabolism, and vascular disease will identify individuals at high risk for osteonecrosis. Previously, 233 SNPs in seven genes including BMP6, TGFBR2, TGFBR3, EDN1, ERG, KL, and ECE1 were shown to be associated with this complication. There were 18 SNPs in the KL gene, which encodes the glycosyl hydrolase protein that participates in a negative regulatory network of vitamin D metabolism; moreover, 14 SNPs in BMP6 and six SNPs in ANXA2 were significantly associated with osteonecrosis [66]. A second research group [67] demonstrated the association of rs267196 (BMP6) and rs7170178 (ANXA2) with a higher risk of osteonecrosis. However, additional studies are needed to confirm if these markers are predictive of the clinical risk for this complication.

4.5. Priapism

Thirty percent of males with SCD experience the potentially devastating complication of priapism associated with a clinically severe disease sub-phenotype. Proteins involved in neuro-regulatory and adrenergic pathways, nitric oxide biology, and ion channels have been implicated in the pathophysiology of priapism [68–71]. More recently, clinical studies have identified genetic markers of priapism that produce erectile dysfunction and determine the ability to respond to phosphodiesterase inhibitors. Nolan et al. [72] identified SNPs in the KLOTHO gene including rs2249358, rs211239, rs211234, and rs211239 associated with an increased risk for priapism among 148 males with SCD. To support these findings, Elliott et al. [69] examined polymorphisms in a second group of adult male SCD patients with a 42% history of priapism. Mutations in the nitric oxide biology (NOS2, NOS3, and SLC4A1) and KLOTHO genes were associated with priapism risk providing further evidence for modulating nitric oxide levels as a therapy for this complication.

4.6. Nephropathy

Sickle nephropathy is a serious complication of SCD that can lead to renal failure and is rapidly becoming a major cause of death in adults. In view of the high medical burden and poor health outcome of end-stage renal disease, genetic markers of nephropathy risk are desirable. Youssry et al. [73] identified soluble FMS-like tyrosine kinase-1, a member of the vascular endothelial growth factor receptor family, as a biomarker for sickle nephropathy. In addition, Ashley-Koch et al. [53] demonstrated that the myosin, heavy chain 9, non-muscle (MYH9), and apolipoprotein L1 (APOL1) genes are associated with risk for focal segmental glomerulosclerosis and end-stage renal disease in African Americans. Seven SNPs in MYH9 and one in APOL1 remained significantly associated with proteinuria after multiple testing corrections. The causative role of these proteins in the development of sickle nephropathy needs to be tested further.

4.7. Leg ulcers

Cutaneous leg ulcers occur more often in adult sickle cell patients with low baseline hemoglobin levels and increased hemolysis rates indicated by high lactate dehydrogenase, bilirubin, and reticulocyte levels. The V34L G/T SNP (rs5985) in the factor XIII gene (F13A1) has been associated with leg ulcers [74]. Other studies have implicated factor V Leiden [75], the fibroblast
growth factor receptor [76], and the HLA-B3525 antigen [77] in the pathogenesis of leg ulcers. A larger study involving 243 sickle cell patients [78] examined SNPs in 60 candidate genes that have a putative role in the pathophysiology of SCD. The association of SNPs in KLOTHO, TEK, and the TGF-β/BMP-signaling pathway was implicated in leg ulcer risk. Of these, KLOTHO promotes endothelial nitric oxide production and the TEK receptor tyrosine kinase is involved in angiogenesis. The TGF-β/BMP-signaling pathway modulates wound healing and angiogenesis, among other functions. Hemolysis-driven phenotypes such as leg ulcers could be improved by agents that increase nitric oxide bioavailability.

5. Genetic modifiers of fetal hemoglobin

5.1. HBB locus haplotypes

Inherited genetic mutations that modulate HBG1/HBG2 gene expression enable persons with SCD to maintain high HbF levels, which ameliorates their clinical symptoms and long-term survival [17]. Individual SNPs inherited in set patterns define HBB haplotypes and determine the ancestral origin of the β^S-globin gene mutation in different ethnic and racial groups. Five common haplotypes including Senegal, Benin, Central African Republic (Bantu), Cameroon, and Asian (Indian/Saudi-Arabian) have been identified [1]. HbF levels vary greatly among individuals with different and the same HBB haplotype, which has precluded the establishment of a consistent correlation between the two parameters. However, individuals with the Senegal haplotype generally have higher HbF levels and milder disease [79], whereas individuals with the Benin haplotype tend to have lower HbF levels and more severe disease [80]. To address this limitation, a genomic study by Liu et al. [81] established the complexity of the HBB locus providing insights into the challenges of defining distinct HBB haplotypes for the prediction of disease severity and the development of therapeutic strategies.

5.2. Genome-wide association studies (GWAS)

The normal switch from HbF to HbA synthesis occurs during the first year of life reaching adult levels of HbF <1% by 12 months of age. A group of disorders known as hereditary persistence of HbF expression is caused by inherited deletions in the HBB locus or point mutations in the promoter region of the HBG genes. HbF levels range from 10 to 40% depending on whether heterozygous or homogeneous mutations are inherited. To gain insights into loci outside the HBB locus that control HbF heritability, GWAS to identify quantitative trait loci were conducted [82]. Three major loci were discovered including the Xmn1-HBG2 (Gγ-globin) on chromosome 11, HBS1L-MYB intergenic region (HMIP) on chromosome 6q23, and BCL11A gene on chromosome 2p16 that control up to 40% of HbF variance in different populations [83]. These loci will be discussed subsequently in the context of the development of precision medicine for persons with SCD.
5.3. Xmn1-HBG2

In 1985, the C/T SNP at nucleotide −158 of the HBG2 gene (rs7482144; T/T) was shown to be associated with high HbF levels with an increase in HbF expressing erythrocytes or F-cells (Figure 1A), and a milder disease phenotype in persons with SCD and β-thalassemia [84]. The positive association between the rs7482144 minor alleles (C/T) and HbF levels was replicated in European and Native Indian populations. However, this SNP was not associated with HbF levels in the people of African ancestry [85]. By contrast, the rs7482144 (G/A) allele occurred at a higher frequency in sickle cell patients with the Senegal and Arab-Indian haplotypes suggesting that the A allele is associated with the geographical origin of the study population. The ancestry for African Americans with SCD showed a high degree of European, African, and Native American admixture at 39.6, 29.6, and 30.8%, respectively.

![Diagram showing SNP associations](image)

**Figure 1.** Summary of major single nucleotide polymorphisms (SNPs) associated with inherited genetic modifiers of HbF variance. Genome-wide genetic studies and GWAS identified SNPs associated with inherited levels of HbF in various ethnic and racial groups. Shown are SNPs in the HBB locus (A), the HBS1L-MYB intergenic region (B), and intron 2 of the BCL11A gene (C) associated with HBG regulation.

5.4. HBS1L-MYB (HMIP) region

Early studies conducted in a family of Asian Indian origin using segregation analysis demonstrated a modifier of HBG gene expression independent of the HBB locus [86]. Using a regressive model, a major locus was discovered on chromosome 6q23–q24 in the HMIP region.
Of the three SNPs identified, only rs4895441 was significantly associated with HbF levels, explaining 9.2% of variance. Later studies showed an association of the other two SNPs, rs28384513 and rs9399137, with HbF levels in the Northern European population (Figure 1B). Subsequently, these SNPs were also demonstrated to control HbF expression in African American, Brazilian, African British, and Tanzanian sickle cell patients [87]. The minor allele frequency of rs9399137 (C) is most significantly associated with HbF expression, but is less common in African populations, with a frequency of 1–2% in African sickle cell patients without European admixture. Similarly, a 3-bp (TAC) deletion on chromosome 6q23 is common in non-African populations, whereas the minor allele of rs9399137 occurs at a higher frequency in African Americans with SCD and elevated HbF levels [88].

5.5. BCL11A

After the completion of the Human Genome Project and the development of genome-wide techniques, GWAS became the preferred approach to identify inherited genetic modifiers of disease phenotypes. The first GWAS to identify HbF modifiers utilized a selected genotyping study design, targeting 179 individuals with contrasting extremes of F-cell numbers [89]. The Xmn1-HBG2 and HMIP regions were identified along with a novel locus in the second intron of the oncogene BCL11A located at chromosome 2p16; the A allele of rs4671393 was associated with increased HbF levels. Subsequently, Uda et al. [90] confirmed SNPs in the BCL11A gene associated with high HbF in Sardinian thalassemia patients, establishing the first major repressor of HBG1/HBG2 gene expression (Figure 1C). The majority of GWAS to identify inherited HbF determinants in African Americans with SCD have been conducted using samples collected during the Cooperative Study of Sickle Cell Disease [91–94]. The first GWAS conducted by Solovieff et al. [93] confirmed the BCL11A SNP (rs766432) and identified a polymorphism in the ORB1B5/OR51B6 locus (rs4910755) associated with HbF levels in sickle cell patients (Figure 1A). A subsequent meta-analysis was conducted using GWAS data generated in seven African-American SCD cohorts totaling 2040 patients [95]. The most significant SNPs were identified in BCL11A (rs766432) and the HMIP region (rs9494145), which represented 11.1 and 3.2% of the phenotypic variability in HbF expression, respectively. Recently, the first GWAS was conducted in a Tanzanian population of 1213 individuals with SCD [96]. Similar to African Americans, SNPs in the BCL11A gene and the HMIP region were replicated in Tanzanians. Other studies have shown up to 10% of HbF variance associated with the BCL11A SNP rs4671393 in sickle cell patients from Northern Brazil (Figure 1C).

5.6. Mechanism of regulating HBG expression

Many decades of research have revealed that two types of mechanisms play a major role in modifying HbF levels: (1) direct transactivation of the HBG1/HBG2 genes through the Xmn1-HBG2 site or (2) an indirect effect on HBG1/HBG2 through the repression of silencers such as BCL11A or MYB. The Xmn1-HBG2 variant rs7482144 mediates a direct effect on Gγ-globin gene expression by functioning as a promoter [1]. By contrast, SNPs in the 14-kb second intron of BCL11A produces a strong enhancement of HbF expression. High levels of the short BCL11A isoform are associated with enhanced HbF expression in primitive erythroblasts, whereas full-
length BCL11A isoforms are present in adult-stage erythroblasts when the HBG genes are silenced. BCL11A interacts with several DNA-binding proteins such as the corepressors LSD1/CoREST [97], DNMT1 [98], GATA1/FOG1/NuRD complex [99], and Sox6 [100] to facilitate γ-globin gene silencing through binding in the HbF-silencing region located upstream of the δ-globin gene [101]. Other studies have shown direct binding of BCL11A to a core motif 5′-GGCCGG-3″ in the HBG promoters to form a repressor complex in K562 cells [102]. Recently, an erythroid-specific enhancer was discovered in the second intron of BCL11A [103], which can be targeted to achieve lineage-specific gene silencing to achieve gene therapy for SCD directed at inhibiting BCL11A in erythroid progenitors.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1186868</td>
<td>BCL11A</td>
<td>Baseline HbF</td>
<td>Uda et al. [90]</td>
</tr>
<tr>
<td>rs766432</td>
<td>BCL11A</td>
<td>Baseline HbF</td>
<td>Sedgewick et al. [92]</td>
</tr>
<tr>
<td>rs4671393</td>
<td>BCL11A</td>
<td>Baseline HbF</td>
<td>Lettre et al. [94]</td>
</tr>
<tr>
<td>rs7557939</td>
<td>BCL11A</td>
<td>Baseline HbF</td>
<td>Lettre et al. [94]</td>
</tr>
<tr>
<td>rs7482144</td>
<td>HBB</td>
<td>Baseline HbF</td>
<td>Lettre et al. [94]</td>
</tr>
<tr>
<td>rs10128556</td>
<td>HBB</td>
<td>Baseline HbF</td>
<td>Galameau et al. [110]</td>
</tr>
<tr>
<td>rs3759070</td>
<td>HBE1</td>
<td>Baseline HbF</td>
<td>Sebastiani et al. [91]</td>
</tr>
<tr>
<td>rs5024042</td>
<td>OR51B5/OR51B6</td>
<td>Baseline HbF</td>
<td>Solovieff et al. [93]</td>
</tr>
<tr>
<td>rs4995441</td>
<td>HBS1L-MYB</td>
<td>Baseline HbF</td>
<td>Lettre et al. [94]</td>
</tr>
<tr>
<td>rs9494145</td>
<td>HBS1L-MYB</td>
<td>Baseline HbF</td>
<td>Bae et al. [95]</td>
</tr>
<tr>
<td>rs9399137</td>
<td>HBS1L-MYB</td>
<td>Baseline HbF</td>
<td>Creary et al. [107]</td>
</tr>
<tr>
<td>rs28384513</td>
<td>HBS1L-MYB</td>
<td>Baseline HbF</td>
<td>Galameau et al. [110]</td>
</tr>
<tr>
<td>rs12103880</td>
<td>GLP2R</td>
<td>Baseline F-cells</td>
<td>Bhatnagar et al. [109]</td>
</tr>
<tr>
<td>rs4769558</td>
<td>ALOX5AP</td>
<td>HbF induced by HU</td>
<td>Sebastiani et al. [91]</td>
</tr>
<tr>
<td>rs1867380</td>
<td>AQP9</td>
<td>HbF induced by HU</td>
<td>Sebastiani et al. [91]</td>
</tr>
<tr>
<td>rs17599586</td>
<td>ARGI</td>
<td>HbF induced by HU</td>
<td>Ware et al. [108]</td>
</tr>
<tr>
<td>rs2295644</td>
<td>ARG2</td>
<td>HbF induced by HU</td>
<td>Ware et al. [108]</td>
</tr>
<tr>
<td>rs10453802</td>
<td>AR2G</td>
<td>HbF induced by HU</td>
<td>Ma et al. [105]</td>
</tr>
<tr>
<td>rs2182008</td>
<td>FLT1</td>
<td>HbF induced by HU</td>
<td>Ma et al. [105]</td>
</tr>
<tr>
<td>rs10494223</td>
<td>HAO2</td>
<td>HbF induced by HU</td>
<td>Ma et al. [105]</td>
</tr>
<tr>
<td>rs7130110</td>
<td>HBE1</td>
<td>HbF induced by HU</td>
<td>Sebastiani et al. [91]</td>
</tr>
<tr>
<td>rs7977109</td>
<td>NOS1</td>
<td>HbF induced by HU</td>
<td>Ma et al. [105]</td>
</tr>
<tr>
<td>rs944725</td>
<td>NOS2A</td>
<td>HbF induced by HU</td>
<td>Ma et al. [105]</td>
</tr>
<tr>
<td>rs4282891</td>
<td>SARI1</td>
<td>HbF induced by HU</td>
<td>Kumkhaek et al. [111]</td>
</tr>
<tr>
<td>rs2310991</td>
<td>SARI1</td>
<td>HbF induced by HU</td>
<td>Kumkhaek et al. [111]</td>
</tr>
</tbody>
</table>

HbF, fetal hemoglobin; HU, hydroxyurea.

Table 1. SNPs known to modulate HbF levels and response to hydroxyurea therapy.
The mechanism by which the HMIP region silences $HBG$ expression is less clear. It is known that a 24-kb nonprotein-coding region exists between the $HBS1L$ and $MYB$ oncogenes. A recent study identified a distal regulatory locus HMIP 2, which contains a regulatory element composed of several GATA-1 motifs that coincided with DNaseI-hypersensitive sites associated with intergenic transcripts in erythroid precursor cells [104]. It was suggested that the HMIP 2 element might regulate $MYB$, which is a repressor of the $HGB$ genes.

### 5.7. Genetic modifiers of response to hydroxyurea therapy

Data from the Multicenter Hydroxyurea Study [38] suggest that not all persons with SCD respond to HU treatment with increased HbF expression. Therefore, genetic markers to predict response to HU would support the development of precision medicine by limiting unnecessary exposure to a chemotherapy drug that causes bone marrow suppression and decreased fertility [41]. Although limited, studies have identified genetic modifiers of HbF response to HU. For example, SNPs in the $ARG2$, $FLT1$, $HAO2$, and $NOS1$ genes were associated with increased HbF expression based on HapMap data [105]. Interestingly, 29 genes involved in HU metabolism were located in loci previously reported to be linked to HbF levels including 6q22.3–q23.2, 8q11–q12, and Xp22.2–p22.3 [105, 106]. A novel bioinformatics method Random Forest was used to investigate the association between SNPs and the change in HbF after stable long-term HU therapy. SNPs in the $ARG2$, $FLT1$, $HAO2$, and $NOS1$ genes and 6q22.3–23.2 and 8q11–q12 regions were associated with the HbF response to HU [105]. A summary of the SNP-associated $HBG$ expression at baseline or in response to HU treatment in sickle cell patients is shown in Table 1 [90-92, 94, 95, 107–111].

### 5.8. MicroRNA-mediated control of $HBG$ gene expression

Recent studies have focused on posttranscriptional mechanisms of $HBG$ regulation via microRNA (miRNA) gene expression. For example, Miller and colleagues [112] demonstrated the ability of LIN28 to silence miRNA let-7 to activate HbF in human primary erythroid progenitors. Likewise, miR-15a and miR-16-1 [113] enhance $HBG$ expression through the inhibition of $MYB$ expression. Studies by Walker et al. correlated miR-26b with baseline HbF levels and miR-151-3p expression with the maximal tolerated dose of HU in children with SCD [114].

Other miRNAs have been implicated in $HBG$ regulation including miR-96 [115], miR-486-3p, miR-210 [116], and miR-34a [117]. Recent studies demonstrated the preferential expression of miR-96 in adult erythroid cells and its ability to directly target the open-reading frame of $\gamma$-globin mRNA; the inhibition of miR-96 resulted in a 20% increase in $\gamma$-globin expression in erythroid progenitors [115]. BCL11A is directly targeted by miR-486-3p, and its overexpression reduces BCL11A levels followed by an increase in $\gamma$-globin expression [118]. The role of MYB as a repressor of $\gamma$-globin was demonstrated in children with trisomy 13 where increased miR-15a and miR-16 expression targets MYB expression directly to mediate high HbF levels [113]. By contrast, a subset of miRNAs has been shown to be associated with enhanced $\gamma$-globin expression. For example, miR-210 was elevated in a $\beta$-thalassemia patient with high HbF expression [116]. Similarly, the Pace group recently demonstrated the ability of miR-34a to
exert a positive regulatory effect on the HBG1/HBG2 genes when stably expressed in K562 cells [117] suggesting that these miRNAs target repressor proteins. These studies demonstrate the potential of developing miRNAs as targets for precision medicine and the development of therapeutic options for individuals with SCD.

6. Precision medicine for sickle cell disease

Completion of the Human Genome Project greatly improved efforts to develop gene-based treatment strategies for β-hemoglobinopathies. Early efforts to identify genetic modifiers of clinical severity and sub-phenotypes of disease severity in SCD consisted of candidate gene studies. Insights were gleaned into risk factors for acute VOC pain events such as SNPs in the dopamine D3 receptor [42]. Expanded investigations to understand the wide range of opioid dose required by individual sickle cell patients led to the characterization of mutations in the CYP2D6 gene required for opioid activation and classification of slow, intermediate, and rapid metabolizers [43]. However, additional studies with larger sample sizes and/or direct DNA sequencing are required to develop gene markers of disease severity for the development of precision medicine to inform clinical decision making.

A great urgency exists to identify genetic factors associated with risk for acute chest syndrome, the leading cause of morbidity and mortality in children and adults with SCD. Mutations in VEGF [48] and the HMOX1 [49] genes hold promise since they serve as markers of endothelial damage and hemolysis associated with the release of free heme in the vascular space, respectively. Long-term repeated episodes of acute chest syndrome can lead to pulmonary hypertension and early death. With a paucity of effective therapies for this complication, genetic markers that identify subgroups of sickle cell patients at risk will support efforts to develop precision medicine. For example, SNPs in the TGF superfamily of proteins and the ADRB1 gene can be targeted for drug development to improve clinical outcomes. Likewise, SNPs in the eNOS genes [55] required for maintaining normal nitric oxide levels might serve as excellent targets for pharmacologic modulation. Interestingly, SNPs in the KLOTHO [72] and NOS2/ NOS3 [69] genes have been associated with the occurrence of priapism in SCD. These observations suggest that developing drug therapy-targeting genes involved in nitric oxide regulation might treat multiple complications of SCD. Genome-wide studies involving next-generation DNA sequencing technology will move the field closer to achieving precision medicine in SCD.

Based on the absence of clinical symptoms in infants and the amelioration of symptoms in persons with hereditary persistence of HbF, the most effective strategy to modulate disease severity in persons with SCD is HBG activation. Therefore, understanding molecular mechanisms of HBG1/HBG1 gene silencing during hemoglobin switching is an attractive but challenging strategy adopted by many investigators over the last three decades. Early genome-wide family genetic studies [82] and subsequent GWAS identified the XmnI-HBG2, HBS1-MYB, and BCL11A loci that account for ~40% of inherited Hbf variance [83]. Orkin and colleagues advanced the field significantly by defining mechanisms of BCL11A-mediated γ-
globin gene repression during murine development and correction of the SCD phenotype [119]. Genetic studies in an extended family identified mutations in KLF1 that produce hereditary persistence of HbF [120, 121] suggesting this transcription factor is a viable target for gene therapy. However, the efficacy of targeting transcription factors for therapeutic development remains to be demonstrated.

Additional genetic studies that utilize high-throughput DNA (whole genome and exome) and RNA/miRNA (RNA-seq) sequencing will increase our knowledge of mechanisms involved in HBG regulation. With the expanded availability of genome-wide approaches, novel technologies for gene editing, and preclinical mouse models, the translation of bench research findings to clinical trials will be accelerated to improve treatment options for SCD and β-thalassemia.

Funding source

National Heart Lung and Blood Institute, National Institutes of Health to BSP (R01HL069234).

Author details

Betty S. Pace*, Nicole H. Lopez, Xingguo Zhu and Biaoru Li

*Address all correspondence to: bpace@augusta.edu

Department of Pediatrics, Augusta University, Augusta, GA, USA

References


Lee YT, de Vasconcellos JF, Yuan J, Byrnes C, Noh SJ, Meier ER, Kim KS, Rabel A, Kaushal M, Muljo SA, Miller JL. LIN28B-mediated expression of fetal hemoglobin and...


