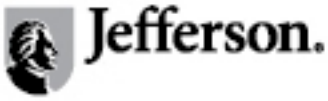

3-27-2016

A case-control genome-wide association study identifies genetic modifiers of fetal hemoglobin in sickle cell disease.

Li Liu
University of Texas at Dallas

Alexander Pertsemlidis
University of Texas Health Science Center at San Antonio

Liang-Hao Ding
University of Texas Southwestern Medical Center

Michael D Story
University of Texas Southwestern Medical Center

Martin H Steinberg
Boston University

See next page for additional authors

Authors
Li Liu, Alexander Pertsemlidis, Liang-Hao Ding, Michael D Story, Martin H Steinberg, Paola Sebastiani, Carolyn Hoppe, Samir K. Ballas, and Betty S Pace

A Case-Control Genome-wide Association Study to Identify Genetic Variations Associated with Fetal Hemoglobin Levels in African American Sickle Cell Disease

Li Liu[1], Alexander Pertsemlidis[2], Ding L[3], Michael Story[3], Carolyn Hoppe[4], Samir K. Ballas[5], Betty S Pace[6,7]*

[1]Department of Biological Sciences, University of Texas at Dallas, Dallas, TX 75083, USA; [2]Greehey Children's Cancer Research Institute; Department of Cellular and Structural Biology; UT Health Science Center at San Antonio; San Antonio, TX USA; Greehey Children's Cancer Research Institute; Department of Pediatrics; UT Health Science Center at San Antonio; San Antonio, TX USA; [3]Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA; [4]Department of Hematology/Oncology, Children's Hospital and Research Center at Oakland, Oakland, CA 94609, USA; [5]Thomas Jefferson University, Department of Internal Medicine, Philadelphia, PA 19107, USA; [6]Department of Pediatric, Augusta University, Augusta, GA 30912, USA; [7]Department of Biochemistry and Molecular Biology, Augusta University, Augusta, GA 30912, USA

Li.Liu2@utdallas.edu
PERTSEMLIDIS@UTHSCSA.edu
michael.story@utsouthwestern.edu
choppe@mail.cho.org
Samir.Ballas@jefferson.edu
bpace@gru.edu

*Betty S. Pace, MD
Professor, Pediatrics and Biochemistry and Molecular Biology
Francis J. Tedesco Distinguished Chair
Augusta University
1120 15th Street, CN4112
Augusta, GA 30912

Phone: 706-721-6893
Fax: 706-721-1670
Email: bpace@gru.edu

Abstract

Sickle cell disease (SCD) is a group of inherited hemoglobinopathies produced by several mutations in the 6th codon of the β-globin (HBB) gene. However, patients with the same genetic mutation display a wide range of clinical symptoms and complications. Fetal hemoglobin (HbF) expression has been demonstrated as an important genetic modifier of SCD leading to improved long-term survival. Therefore to discover genetic factors associated with HbF expression, we performed a genome-wide association study (GWAS) using a case-control experimental design for 244 African American patients with SCD. The case group consisted of subjects with HbF $\geq$8.6% (133 samples) and control group subjects with HbF $\leq$3.2% (111 samples). Our GWAS replicated several SNPs previously identified in an erythroid-specific enhancer region located in the second intron of the *BCL11A* gene, highly associated with HbF expression. Moreover, we identified SNPs in SPARC, GJC1, EFTUD2 and JAZF1 as potential candidate genes associated with HbF level in sickle cell patients. To gain insights into mechanisms of globin gene regulation in the HBB locus, linkage disequilibrium (LD) and haplotype analysis was conducted. We observed strong LD in the low HbF group in contrast to a loss of LD and greater number of haplotypes in the high HbF group surrounding the γ-globin gene region. A search of known HBB locus regulatory elements demonstrated SNPs 5' of δ-globin located in an HbF-silencing region. In particular SNP rs4910736 created a potential binding site for a known transcription repressor GFi1 as a potential gene target for further investigation. Another HbF-associated SNP rs7482144 in the cAMP response element upstream of Gγ-globin was analyzed for functional relevance. Studies performed with siRNA-mediated knockdown of CREB binding protein (CBP) in human

primary erythroid cells demonstrated γ-globin gene activation and HbF induction supporting a repressor role for CBP. The implications of our findings are discussed.

Key words: GWAS, sickle cell disease, fetal hemoglobin, HBB locus, haplotypes, single nucleotide polymorphisms

Introduction

Sickle cell disease (SCD) is a group of inherited hemoglobinopathies produced by different mutations in the 6^{th} codon of the β-globin (HBB) gene on chromosome 11 with the hemoglobin S (HbS) variant being most common. Those individuals with homozygous mutations (HbSS) or heterozygous for the sickle mutation and β^0-Thalassemia have the most severe disease phenotype and are at higher risk for developing clinical complications. SCD is characterized by acute vaso-occlusive events including pain episodes, stroke, splenic dysfunction, and acute chest syndrome among others which contribute to increased morbidity and early mortality. Despite the genetic simplicity of SCD as a Mendelian single gene disorder, sickle cell patients display extreme clinical heterogeneity (1,2). Two major modifiers of SCD clinical phenotypes, fetal hemoglobin (HbF; $\alpha_2\gamma_2$) and α-thalassemia have been demonstrated by basic research and clinical studies and they were subsequently confirmed by genetic studies. HbF protects against many of the hematologic and clinical complications of SCD due to its ability to inhibit HbS polymerization (3,4). The discovery of genetic markers of HbF expression to identify novel targets for therapeutic development for SCD has been a major ongoing research effort.

The Cooperative Study of Sickle Cell Disease (CSSCD), was a multi-institutional natural history study (5,6) that identified HbF as a modifier of mortality and risk factor for early death in sickle cell patients (7). This observations have been corroborated by numerous clinical studies of Europe, India, The Middle East and Africa SCD patients (8) highlighting the importance of understanding the genetic regulation of γ-globin gene expression. The 81-kb HBB locus on chromosome 11 consists of five functional genes including *HBE1* (ε), *HBG2* (Gγ), *HBG1* (Aγ), *HBD* (δ) and *HBB* (β), expressed sequentially from 5' to 3' in a tissue- and development-specific manner during ontogeny (8) a process known as Hb switching. Expression of the β-like globin

genes is controlled in part by the locus control region (LCR) located 6-20 kb upstream of ε-globin (9,10). To achieve normal hemoglobin switching, stage-specific transcription factors bind the LCR and globin promoters during erythropoiesis to activate globin gene expression (11) through a DNA looping mechanism (12). At birth HbF comprises 80-90% of the total hemoglobin synthesized and gradually decreases to <1% by 12 months. The discovery of heritable markers associated the major silencers of γ-globin expression provides a rational approach for precision medicine in SCD.

To develop strategies to induce HbF, it is necessary to define *cis*-regulatory elements and DNA-protein interactions in the HBB locus that modify globin gene expression. In previous studies up to 40% of HbF producing capacity in sickle cell patients has been attributed genetic determinants in Xp22.2 (13), chromosome 8q (14) and genes in the 6q22.3-23.4 quantitative trait loci (15, 16). However later genome-wide association studies (GWAS) have not identified genes on the X chromosome associated with HbF. Thein and colleagues were the first to demonstrate single nucleotide polymorphisms (SNPs) in the *HBS1L-MYB* inter-genic region (HMIP) on 6q23 accounting for 17.6% of HbF variance in Northern Europeans (17) and later they identified the *BCL11A* quantitative trait locus (18) accounting for 15% of HbF variance. Subsequently, Uda et al. (19) confirmed SNPs in the *BCL11A* gene associated with high HbF in Sardinian thalassemia patients, establishing the first potential major repressor of the γ-globin gene. A third locus that contributes to HbF expression is the -158*Xmn1* HBG2 SNP (15,20).

GWAS studies to identify inherited HbF determinants in African American sickle cell patients have been conducted using CSSCD patients (21-24). The GWAS conducted by (23) confirmed *BCL11A* (rs766432) and identified the *ORB1B5/OR51B6* (rs4910755) genes associated with HbF levels. A subsequent meta-analysis was conducted using GWAS genotype

data generated in 7 African American SCD cohorts totaling 2040 patients (25). 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, respectively. Recently, the first GWAS was conducted in a Tanzanian population of 1,213 HbSS and HbSβ^0-Thalassemia patients (26); similar to African Americans, SNPs in BCL11A and the HMIP region were replicated. Interestingly the -158*Xmn*1-*HBG2* (rs782144) SNP was not replicated in Tanzanian patients similar to previous results for African American sickle cell patients (23,25). In addition, eight novel SNPs were identified in the Tanzanian population that may be associated with HbF expression.

The second major modifier of clinical phenotype is expression of the α-globin genes (*HBA1/HBA*); approximately one-third of sickle cell patients of African descent have coexisting α-thalassemia due to the common 3.7-kb deletion (-$\alpha^{3.7}$) (1). Coexisting α-thalassemia reduces intracellular hemoglobin concentrations, thereby reducing HbS polymerization, red cell sickling, and decreased hemolysis. A SNP in *NPRL3* (rs7203560) on chromosome 16 was identified by GWAS and validated by targeted genotyping in an independent cohort (27). The *HBA1/HBA2* genes regulatory element, hypersensitive sites (HS)-33, HS-40 and HS-48 is located in introns of *NPRL3*. When adjusting for HbF and α-thalassemia, variants of the *HBA1/HBA2* gene regulatory loci in *NPRL3*, were associated with reduced hemolysis in SCD.

In our current study we conducted a case (high HbF) control (low HbF) GWAS to identify novel genetic modifiers of γ-globin expression. As previously confirm in several populations, SNPs in BCL11A were associated with HbF levels in our population along with four additional genes suggestive of novel candidate loci. High-density SNP mapping of the 81-kb HBB locus revealed haplotype tagging-SNPs that define unique haplotype patterns associated

with high and low HbF groups. Knockdown of the transcription factor CREB1 binding protein which interacts upstream of the Gγ-globin gene enhances HbF expression. The implications of our findings are discussed.

Methods and Materials

DNA samples sources – A total of 254 genomic samples were analyzed in this study consisting of 155 samples collected during the CSSCD. These samples were obtained after approval from the Biologic Specimen and Data Repository Information Coordinating Center at the National Heart Lung and Blood Institute. Thirty samples were obtained from the Comprehensive Sickle Cell Centers Collaborative Data (CDATA) study and the remaining samples were contributed by Dr. Samir K. Ballas, Thomas Jefferson University.

Genotype determination - The Illumina Omni1-Quad System (San Diego, CA) was used to genotype the DNA samples since this chip contains 1,140,419 SNPs with 76% coverage of SNPs for people of African ancestry. Genotype data quality control and SNP calls were generated with Illumina's Genome Studio software. For the HBB locus haplotype analysis, genotype data for SNPs rs2855121 and rs2855122 were determined using TaqMan® assay probes (ThermoFisher Scientific, Grand Island, NY) for real-time PCR detection.

GWAS analysis - Allele associations were calculated using the case-control association test as implemented in PLINK version 1.07 (28). PLINK performs all necessary data management tasks for GWAS, including generation of summary statistics for quality control, detection of population stratification, testing for statistical association at the single SNP, multiple SNPs and

haplotype levels. The *P* values were plotted against genomic location using Haploview version 4.2 (29). Allele association with case or control status was assessed by Chi-squared test and the expected and observed distributions of SNPs were determined by statistical analysis. The expected numbers of SNPs under the null hypothesis at the nominal significance threshold were generated.

Haploview analysis - Genotype data were inspected and used to construct HBB locus haplotypes using Haploview 4.2 (http://www.broad.mit.edu/haploview), which utilizes the EM algorithm to calculate linkage disequilibrium (LD) coefficients (D') and to infer haplotypes. Examination of the genotypes was conducted to determine the conformity with Hardy-Weinberg equilibrium before Haploview analysis. The SNPs with statistically significant departures ($P<0.001$) and minor allele frequencies < 5% were excluded from haplotype analysis. SNPs with strong LD ($D' \geq 0.8$) were arranged into haplotype block following Wang's algorithm (30). Haplotype-tagging SNPs (tagSNPs) were selected on a block-by-block basis to identify SNPs that represent non-redundant information about genomic structure.

Two phase primary erythroid cultures - Primary erythroid progenitors were obtained from human peripheral blood mononuclear cells purchased from Carter BloodCare (Fort Worth, TX) in accordance with guidelines of the Institutional Review Board at the University of Texas at Dallas. Expansion and induced erythroid differentiation was conducted according to the procedure previously published by our lab (31). For gene silencing experiments in primary erythroid cells, SMARTpool siRNA for CBP (M-series, GE Dharmacon) and non-targeting

scramble RNA were transfected on day 28 of primary cell culture using a $CD34^+$ Nucleofector kit and Nucleofector device (Amaxa, Allendale, NJ) as described previously by our lab (31).

Reverse transcription-real time polymerase chain reaction (RT-qPCR) - To quantify gene transcription levels, total RNA was extracted using RNA Stat-60™ (TEL-TEST "B" Inc., Friendswood, TX) and used for RT-qPCR for γ-globin, β-globin and GAPDH as reported previously (32); primers to quantify actin and CREB binding protein (CBP) gene expression were purchased from Qiagen (Valencia, CA).

Western blot - To determine CBP knockdown in primary erythroid progenitors, western blot analysis were performed using total protein extracts and anti-CBP antibody (SC-583, Santa Cruz Biotechnology, Dallas, TX) and anti-Actin antibody (MAB1501, Millpore, Billerica, MA) as previously published (33).

Fluorescence immunocytostaining - Cells were centrifuged on a glass slide to obtain a monolayer by Cytospin preparation. Immunocytostaining procedure with anti-γ-globin fluorescein isothiocyanate antibody (Bethyl Laboratories, Montgomery, TX) was conducted as previously published (31).

Enzyme-linked immunosorbent assay (ELISA) - After the different treatments were conducted, total protein lysates were prepared and ELISA performed as previously published (31).

Statistical analysis - Each condition was repeated independently three times with triplicate samples for each experiment; data are shown as the mean ± standard error of the mean (SEM). The student's *t*-test was applied where two experimental conditions were compared to determine the statistical significance at $P< 0.05$.

Results

GWAS confirm association of BCL11A with HbF levels

The ability of HbF to ameliorate the clinical severity of SCD and β-Thalassemia has been well established (34). Furthermore the CSSCD demonstrated milder symptoms and improved survival in SCD patients with HbF >8.6% (7). In other diseases when a phenotype associated with morbidity and mortality has been identified, using a case-control approach has decreased the sample size required to identify SNPs associated with genetic modifiers of the clinical phenotype (35, 36). Therefore, given the findings related to mortality and HbF in the CSSCD, we designed a case-control GWAS in HbSS subjects where patients with HbF $\leq$3.2% were selected as the control group (low HbF) those with HbF $\geq$8.6% were designated in the case group (high HbF). Our goal was to identify genetic variants associated with either high or low HbF expression in SCD.

All DNA samples analyzed in the GWAS met quality control and were used to collect genotype data on the IlluminaOmni1-Quad chip contains 1,140,419 SNPs with 76% SNP coverage for people of African ancestry. We genotyped 254 African American diagnosed with HbSS and HbSβ^0-Thalassemia and HbF $\geq$8.6% (case, n=133), HbF $\leq$3.2% (controls, n=111), and

10 intermediate HbF subjects (Table 1). We excluded 45,142 SNPs as monomorphic, 546 as singletons and 1,346 based on departures from Hardy-Weinberg Equilibrium ($P<0.001$). We further excluded 181,454 SNPs that showed minor allele frequencies < 5%. After frequency and genotyping pruning, 833,760 SNPs were used in the GWAS analysis.

The first genomic analysis was conducted to determine the allele association of SNPs using the case-control association test implemented in PLINK version 1.07. Fig. 1 shows the distribution of P-values (Manhattan plot) and distribution of SNPs on chromosome 2 and 5 significantly associated with HbF levels. Using $P<1x10^{-5}$, we identified 74 SNPs of potential significance (data not shown). The top 4 genes and 2 intergenic regions with genome-wide significance ($P<1X10^{-6}$) associated with high HbF are shown in Table 2. The intron region of the *BCL11A* gene contained 14 SNPs of which rs1896295, rs45606437, rs7584113, rs10172646, and rs1896294 have not been previously reported. This could reflect the differences in the SNP allele frequency in African Americans compared to European or African populations. On the other hand, these SNPs could have high LD with other SNPs validated in the region.

Our GWAS data also identified SNPs in the SPARC, GJC1, EFTUD2 and JAZF1 genes associated with high HbF (Table 2). SPARC (secreted protein acidic and rich in cysteine) is a widely expressed profibrotic protein linked to human obesity, insulin resistance, and diabetic retinopathy (37). The GJC1 gene (Gap Junction Protein Gamma 1; connexin45) belongs to a group of proteins that form intercellular channels to allow exchange of small molecules between cells (38). The elongation factor Tu GTP binding domain containing 2 (EFTUD2) encoded a GTPase which is a component of the spliceosome complex (39). The fourth high HbF-associated SNP was identified in the zinc finger nuclear protein JAZF1 that functions as a transcription repressor (40).

To gain insight into possible role of novel SNPs identified in our study, we performed an alignment between HbF-associated SNPs across the BCL11A gene and functional genomics data generated by the Encyclopedia of DNA Elements (ENCODE) project. We focused our *in silico* analysis to a 15-kb region (hg19; 60,713,000 to 60,728,000) enriched with HbF-associated SNPs (Fig. 2) with the BCL11A gene shown at the top. The ENCODE results for K562 cells showed low level enrichment of enhancer activity represented by histone H3K4me1, but no significant H3K4me3 mark detected. DNase hypersensitivity was detected in this region as well as *in vivo* binding of the transcription factors GATA1 and TAL1. Moreover, three loci were detected with strong GATA1-associated enrichment in erythroblasts corresponding to the erythroid enhancer region reported by Bauer et al. (41) located at +55, +58 and +62 relative to the BCL11A transcription start site. Two of the SNPs identified in this GWAS including rs1427407 (+62 site) and rs6738440 (+58 site) could potentially alter recruitment of transcription factors to the region. Using a case-control approach for GWAS analysis of SCD patients with low and high HbF levels, we captured previously reported and novel SNPs located in the erythroid enhancer region in BCL11A. In addition, we identified novel candidate genes that may be involved in γ-globin gene regulation.

High-density SNP mapping in the HBB locus

Transcriptional regulation of the β-like globin genes depends on trans-acting factors (activators and repressors), *cis*-acting elements (DNA binding regions for binding of trans-acting factors) and epigenetic changes in chromatin structure. The LCR provides essential enhancer activity for high-level sequential expression of the downstream β-like globin genes during developmentally regulated hemoglobin switching (12). Numerous studies support the notion that

DNA sequences within 3-4 kb upstream or downstream of the Aγ-globin and Gγ-globin genes influence their transcription rate during erythroid maturation. Thus we performed a case-control analysis for SNPs in the HBB locus to identify genomic structural divergences between low HbF and high HbF groups using Haploview 4.2. Based on the functional role of SNPs in the Gγ-globin promoter (32,42), we also determined genotypes for rs2855121 and rs2855122 by TaqMan® assay. The combined TaqMan and HBB locus genotype data generated using the Illumina Omni-Quad 1 chip (101 SNPs) were used for the case-control analysis. We identified seven SNPs in the HBB locus (Table S1) with $p<0.04$; 3 SNPs were located within 30-kb of the intergenic region between ε-globin and Gγ-globin (rs4910740, rs10128653 and rs2855122) and 4 SNPs in the region between the ψβ-globin and δ-globin genes (rs16912210, rs4910736, rs4402323 and rs4320977). Historically rs10128653 and rs2855122 have been implicated in forming a pre-G genetic framework that influences HbF expression in the context of HBB locus haplotypes defined in different African populations (43). These SNPs reflect changes at the genomic DNA level which may relate to the persistent expression of HbF in sickle cell patients. Further functional studies are needed to ascertain the biological imprint of these SNPs and the region in which they are located.

Characterization of HBB locus haplotypes and tagSNPs

Next we performed LD and haplotype analysis of the HBB locus with genotype data generated from low HbF (101 subjects) and high HbF (103 subjects) groups using Haploview 4.2. There were 103 SNPs spanning the HBB locus (Fig. 3A) which produced 4 haplotype blocks based upon the 4th Gamete Rule for the low HbF cohorts (Fig. 3B). Overall there was strong LD

across the HBB locus in this group illustrated by the LD plot and the allelic D' values shown between the haplotype blocks (D'>0.8). On the contrary, we observed decreased LD and more diverse haplotypes in the high HbF cohort as shown in Fig. 3C with 8 haplotype blocks inferred. Interestingly, the region between rs10128653 and rs7482144 (-158*XmnI*-HBG2) displayed a lower allelic mean D'<0.8 in the high HbF group.

Haploview analysis also identifies a subset SNPs which carry non-redundant information (tag-SNP) which can be used to define the diversity and total number of haplotypes in a given region. We observed 14 tagSNPs in the low HbF group with 3 that were unique to this group (Table S2) and 18 tagSNPs in the high HbF group with 7 unique tagSNPs. These results support a diversified genomic structure in sickle cell patients with the high HbF providing insight into mechanisms of persistent HbF expression in this group.

HbF-associated SNPs located in a known HbF silencing region

We next analyzed the 7 HbF-associated SNPs listed on Table S1 in the HBB locus to establish haplotypes in the low and high HbF groups. Fig. 4A shows that these 7 SNPs formed one haplotype block with strong LD in the low HbF cohorts with five unique haplotypes at decreasing frequency inferred. Analysis of the same 7 SNPs in the high HbF group demonstrated one haplotype block with seven unique haplotypes inferred (Fig. 4B). Interestingly, we identified SNPs rs4910736, rs4402323, and rs4320977 located within the 3.5-kb HbF silencing region upstream of the δ-globin gene (Fig. 4C) identified by Sankaran et al. studying $\delta\beta^0$-Thalassemia patient samples with deletional hereditary persistence of HbF (44). In that report, BCL11A was shown to silence γ-globin expression by interacting with a transcription co-repressor complex in adult erythroid cells.

To gain evidence for an impact of these SNPs, we analyzed functional genomics data produced by the ENCODE project for the 5-kb region 5' of the δ-globin gene in K562 cells (Fig. 4D). There was an absence of RNA polymerase II signal and increased binding of repressor complexes, such as HDAC1, LSD1, CHD2 and COREST supporting involvement of this region in transcription repression. Furthermore, we performed binding site prediction using Tfsitescan. There was no difference on binding site prediction for rs4402323 with either C or T allele at the SNP site, whereas rs4320977-G allele predicted a binding site for CCAAT and enhancer binding protein (data not shown). Most interestingly, as shown in Fig. 4E, the A allele of rs4910736 produced no significant binding site at the SNP position, whereas the C allele predicted a binding site for Growth Factor Independence 1 (GFi1). GFi1 and its paralogue GFi1B are critical transcription regulators for proliferation and maturation of hematopoietic stem cells (45-47). Although it is unclear whether GFi1/GFi1B plays a role in repressing γ-globin transcription, our analysis provides evidence for a potential GFi1 regulatory sequence in the HbF silencing region.

CBP represses γ-globin expression through the HBB locus tagSNP rs2855122

To gain evidence that tagSNPs play a regulatory role in γ-globin transcription we performed functional analysis of SNP rs2855122 located in the cAMP response element (TGACGTCA) at -1225 in the Gγ-globin promoter (G-CRE). We have demonstrated previously the trans-acting factor CREB1 binds to the G-CRE to induce γ-globin transcription through a p38 mitogen activated protein kinase-dependent mechanism (48). Sequential-chromatin immunoprecipitation in K562 cells has shown previously that CREB1 and CBP bind simultaneously to the G-CRE region *in vivo* (48). In addition, we observed decreased CREB1 binding to the G-CRE site *in vivo* during late erythropoietin-induced erythroid differentiation

(33). Subsequent studies from our lab showed a repressor role of CBP through competition with KLF4 binding in the proximal γ-globin promoter CACCC element in a luciferase reporter system (49). Thus, we investigated the effect of CBP knockdown on globin expression in human primary erythroid progenitors at Day 28 during late stage erythroid maturation when the γ-globin gene is silenced. We observed enhanced γ-globin transcription and HbF production with dose-dependent siRNA-mediated CBP gene knockdown (Fig. 5A-C); the number of HbF positive cells measured by flow cytometry was increased as well (Fig. 5D) supporting a repressor role of CBP in γ-globin regulation in this condition.

Discussion

Mutations in the HBB locus, including those located in the γ-globin gene promoters (50,51), and quantitative trait loci at distant sites on chromosomes 2, 6, 8, and the X chromosome have been associated with high HbF expression in humans (13-19,23,25). GWAS in several populations has identified SNPs in BCL11A, the HMIP region and -158*Xmn1*-HGB2 associated with HbF expression. However, SNPs in *BCL11A* show the strongest effect on γ-globin expression accounting for approximately 11% of HbF variance in African American sickle cell patients (23,25) suggesting additional major genetic modifiers of HbF remain to be discovered. In our study, we performed GWAS analysis using African American sickle cell patients and a case (high HbF)-control (low HbF) experimental design. We identified multiple SNPs in BCL11A to replicate previous studies (18,19,23,25). BCL11A is demonstrated to be a stage-specific repressor of γ-globin expression in human primary erythroid progenitors (52) and is involved in γ-globin gene silencing in adult sickle cell transgenic mice (53). The silencing effect of BCL11A involves reconfiguration of the *HBB* locus through interactions of BCL11A with

GATA1, FOG1, and SOX6 as well as the NuRD deacetylase and remodeling complex (52,54). Genetic studies in a Maltese family showed that *KLF1* p.K288X carriers had reduced *BCL11A* expression (55) and elevated HbF levels. Subsequent studies demonstrate that KLF1 is a direct activator of *BCL11A* in human erythroid progenitors and transgenic mice (55,56). Recently an erythroid specific enhancer was discovered in the second intron of BCL11A by GWAS (41); four SNPs (rs1427407, rs6706648, rs6738440, and rs7606173) were associated with HbF levels in SCD patients of African and Arab-Indian descent (57,58). The erythroid enhancer is an attractive DNA element for lineage-specific BCL11A gene silencing as a strategy for gene therapy in SCD and β-Thalassemia.

Genome-wide studies by Thein and colleagues identified the HMIP region on chromosome 6q23 as a modifier of HbF production in the European population (17,59). A meta-analysis of GWAS data from over 2000 sickle cell patients replicates the association of HbF levels with the HMIP region (25). More recently, GWAS study demonstrated SNPs in the HMIP region to associate with HbF level in SCD patients in Tanzania (26), Northern Brazil, and Cameroon (60,61). A previous study demonstrated the ability of MYB to regulate HbF production in erythroid cells (62). Subsequently, Stadhouders et al. described a functional erythroid-specific enhancer located in the HMIP region which regulates MYB expression (63). In our study we did not observe SNPs in the HMIP region associated with HbF level in the patients tested which most likely is due to small sample size and/or differences in ancestry of our population. A recent publication demonstrated the differences in haplotype structures in the core haplotype block 2 of HMIP locus among European, African, and African American SCD subjects (64).

Studies in healthy individuals with elevated HbF identified SNPs in the promoters of the *HBG1* and *HBG2* genes that produce non-deletional hereditary persistence of HbF; however, molecular mechanisms or trans-activating factors that explain the role of these mutations have not been identified. The most extensively studied is the *XmnI* mutation (C to T) at -158 Gγ-globin (rs7482144), associated with elevated HbF levels in selected populations by GWAS studies (17,25,26). Individuals carrying this mutation have a delay in the Gγ to Aγ switch (65) and an increase in F-cells was demonstrated in a European twin study (66). Similar to other GWAS findings in sickle cell patients of African descent (23,25,26), we did not observed an association of the -158*Xmn1*-HBG2 mutation with HbF levels in our cohort.

We also identified SNPs in the SPARC, GJC1, EFTUD2, and JAZF1 genes associated with HbF expression. Multiple SNPs were found in the SPARC gene which is expressed in bone marrow stromal and hematopoietic cells. Gene expression analysis of stem cells from patients with -5q syndrome showed decreased SPARC expression among other proteins associated with a lower platelet count and impaired ability to form erythroid burst-forming units (67). Recent studies in SPARC-null mice demonstrated SPARC plays a crucial role in the regulation of early B lymphopoiesis (68); however SPARC has not been demonstrated to regulate globin gene expression. We also identified the transcription repressor JAZF1 associated with high HbF expression although at a lower genome-wide significance. SNPs in JAZF1 have been associated with type 2 diabetes through a non-coding signal in the first intron, which regulate islets cell function (69). Studies mapping transcriptional activity showed increased binding of a repressor protein complex to a *cis*-regulatory element in JAZF1 intron 1. Additional studies are needed to confirm a functional role of these proteins identified by GWAS in γ-globin regulation in sickle

cell patients. Additional studies are needed to confirm a functional role of these proteins in γ-globin regulation in sickle cell patients.

In contrast to other published GWAS studies we performed a detailed analysis of the HBB locus to define haplotypes associated with high and low HbF expression. Historically, five major β-haplotypes, including Senegal, Benin, Central African Republic (Bantu), Cameroon, and Asian (Indian/Saudi-Arabian) (34,51,70-73), have been defined in African SCD populations; unfortunately, these haplotypes are not predictive of HbF levels. We recently published data demonstrating that β-haplotypes generated using traditional methods are insufficient to define the genomic structure in HBB locus of SCD patients (74). These findings were germane to our experimental design using a case-control approach to analyze genotype data for SNPs identified in the HBB locus which did reach significance at the genome-wide level. However, the difference in haplotype composition in low- and high-HbF groups may provide insights into genomic structural change in the HBB locus between these groups.

When comparing haplotypes in the case-control groups, we observed more inferred haplotypes in the high-HbF group suggesting less LD among SNPs in the HBB locus; in particular SNPs between the Gγ-globin and δ-globin genes. Recent studies demonstrated inter-genetic regions in the HBB locus involved in γ-globin regulation through long-range chromatin looping mediated by various transcription factors and non-coding RNA transcripts (54,75,76). Xu et al. demonstrates a repressor complex consisting of BCL11A, HDAC1, LSD1 (Lysine-specific demethylase 1A), and CoREST (Corepressor for Element-1-silencing transcription factor) binds the inter-genetic region between the Aγ- and δ-globin genes (76). Subsequently, Kiefer et al. reported that genomic sequence between Aγ- and ψβ-globin encodes BGLT3, a non-coding RNA transcript primarily expressed in γ-globin producing cells (75). By contrast, BGLT3

is not transcribed in late erythropoiesis when β-globin expression is high due to binding of a repressor complex consisting of ETO2 and BCL11A resulting in diminished long-range chromatin looping between γ-globin and the LCR (75). Sankaran and colleagues identified a putative HbF silencing region 3.5 kb 5' of δ-globin where the BCL11A repressor complex binds (44). Additional studies localized the HbF silencing region to a 2.4-kb region outside the BCL11A binding region supporting other mechanisms of γ-globin regulation (77).

Our HBB locus analysis identifies SNPs across potential HbF activating (ε- and Gγ-globin) and silencing regions (5'of δ-globin) and novel SNP rs4910736 that creates a GFi1 binding site. The nuclear proteins GFi1 and GFi1B can recruit LSD1 and CoREST to the targeted chromatin region as well as interact with HDAC1 and G9a (H3K9 methyltransferase). The formation of protein complexes between GFi1/GFi1B and their various interacting partners and subsequent binding to targeted DNA sequences mediates histone modifications associated with transcription silencing and regulation of hematopoiesis (45-47,78). Studies by Vassen and colleagues demonstrated increased transcription of the embryonic globin genes and delayed erythroid maturation in Gfi1B-null mice without changes in BCL11A gene expression (79).

The HBB locus SNP rs2855122 located in G-CRE was also associated with HbF levels. We previously confirmed a functional role of this element in γ-globin regulation and characterized a multi-protein complex composed of trans-activators including CREB1 and ATF2 and repressors such as HDAC1 among others (33). Therefore we conducted gene silencing studies for CBP, a known binding partner of CREB1; the data support CBP as a repressor of γ-globin expression. However, additional studies are warranted to fully define molecular

mechanisms of γ-globin gene regulation mediated by novel HbF-associated SNPs identified in this study.

Acknowledgements

This work was supported by grant R56DK083921-01 from the National Institute of Diabetes Digestive and Kidney Diseases to Dr. Betty Pace. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Statement of author contributions

All authors participated in the design of the genome-wide association study, writing and review of the manuscript; LL conducted the majority of the experiments including the haplotype analysis and functional studies in primary erythroid cultures. AP conducted genotype data pruning and genotype-phenotype analysis. MS and LD processed the DNA samples to generate the genotyping raw data pruning and contributed to writing and editing the paper. CH and SB contributed to study design, DNA samples screening and review of manuscript; BSP designed and supervised all experimental studies and contributed to editing and writing the manuscript.

References

1. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. 2nd Ed ed. Steinberg MH, Forget BG, Higgs DR, Weatherall DJ, editors. Cambridge, UK: Cambridge University Press; 2009.

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

3. Akinsheye I, Solovieff N, Ngo D, Malek A, Sebastiani P, Steinberg MH, Chui DH. Fetal hemoglobin in sickle cell anemia: molecular characterization of the unusually high fetal hemoglobin phenotype in African Americans. Am J Hematol. 2012;87(2):217-9.

4. Steinberg MH, Sebastiani P. Genetic modifiers of sickle cell disease. Am J Hematol. 2012;87(8):795-803.

5. Gaston M, Smith J, Gallagher D, Flournoy-Gill Z, West S, Bellevue R, Farber M, Grover R, Koshy M, Ritchey AK. Recruitment in the Cooperative Study of Sickle Cell Disease (CSSCD). Control Clin Trials. 1987;8(4 Suppl):131S-40S.

6. Platt OS, Thorington BD, Brambilla DJ, Milner PF, Rosse WF, Vichinsky E, Kinney TR. Pain in sickle cell disease. Rates and risk factors. N Engl J Med. 1991;325(1):11-6.

7. Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, Klug PP. Mortality in sickle cell disease. Life expectancy and risk factors for early death. N Engl J Med. 1994;330(23):1639-44.

8. Stamatoyannopoulos G, Grosveld, F. . Hemoglobin switching, in The Molecular Basis of Blood Diseas. 3rd ed. Philadelphia: Saunders; 2001.

9. Forrester WC, Thompson C, Elder JT, Groudine M. A developmentally stable chromatin structure in the human beta-globin gene cluster. Proc Natl Acad Sci U S A. 1986;83(5):1359-63.

10. Tuan D, Abeliovich A, Lee-Oldham M, Lee D. Identification of regulatory elements of human beta-like globin genes. Prog Clin Biol Res. 1987;251:211-20.

11. Engel JD, Tanimoto K. Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. Cell. 2000;100(5):499-502.

12. Stamatoyannopoulos G. Control of globin gene expression during development and erythroid differentiation. Exp Hematol. 2005;33(3):259-71.

13. Chang YP, Maier-Redelsperger M, Smith KD, Contu L, Ducroco R, de Montalembert M, Belloy M, Elion J, Dover GJ, Girot R. The relative importance of the X-linked FCP locus and beta-globin haplotypes in determining haemoglobin F levels: a study of SS patients homozygous for beta S haplotypes. Br J Haematol. 1997;96(4):806-14.

14. Garner C, Silver N, Best S, Menzel S, Martin C, Spector TD, Thein SL. Quantitative trait locus on chromosome 8q influences the switch from fetal to adult hemoglobin. Blood. 2004;104(7):2184-6.

15. Craig JE, Rochette J, Fisher CA, Weatherall DJ, Marc S, Lathrop GM, Demenais F, Thein S. Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. Nat Genet. 1996;12(1):58-64.

16. Wyszynski DF, Baldwin CT, Cleves MA, Amirault Y, Nolan VG, Farrell JJ, Bisbee A, Kutlar A, Farrer LA, Steinberg MH. Polymorphisms near a chromosome 6q QTL area are associated with modulation of fetal hemoglobin levels in sickle cell anemia. Cell Mol Biol (Noisy-le-grand). 2004;50(1):23-33.

17. Thein SL, Menzel S, Peng X, Best S, Jiang J, Close J, Silver N, Gerovasilli A, Ping C, Yamaguchi M, Wahlberg K, Ulug P, Spector TD, Garner C, Matsuda F, Farrall M, Lathrop M. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc Natl Acad Sci U S A. 2007;104(27):11346-51.

18. Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, Heath S, Foglio M, Zelenika D, Boland A, Rooks H, Best S, Spector TD, Farrall M, Lathrop M, Thein SL. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat Genet. 2007;39(10):1197-9.

19. Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, Piras MG, Sestu N, Lai S, Dei M, Mulas A, Crisponi L, Naitza S, Asunis I, Deiana M, Nagaraja R, Perseu L, Satta S, Cipollina MD, Sollaino C, Moi P, Hirschhorn JN, Orkin SH, Abecasis GR,

Schlessinger D, Cao A. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc Natl Acad Sci U S A. 2008;105(5):1620-5.

20. Ballas SK, Talacki CA, Adachi K, Schwartz E, Surrey S, Rappaport E. The Xmn I site (-158, C----T) 5' to the G gamma gene: correlation with the Senegalese haplotype and G gamma globin expression. Hemoglobin. 1991;15(5):393-405.

21. Sebastiani P, Wang L, Nolan VG, Melista E, Ma Q, Baldwin CT, Steinberg MH. Fetal hemoglobin in sickle cell anemia: Bayesian modeling of genetic associations. Am J Hematol. 2008;83(3):189-95.

22.Sedgewick AE, Timofeev N, Sebastiani P, So JC, Ma ES, Chan LC, Fucharoen G, Fucharoen S, Barbosa CG, Vardarajan BN, Farrer LA, Baldwin CT, Steinberg MH, Chui DH. BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. Blood Cells Mol Dis. 2008;41(3):255-8.

23. Solovieff N, Milton JN, Hartley SW, Sherva R, Sebastiani P, Dworkis DA, Klings ES, Farrer LA, Garrett ME, Ashley-Koch A, Telen MJ, Fucharoen S, Ha SY, Li CK, Chui DH, Baldwin CT, Steinberg MH. Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. Blood. 2010;115(9):1815-22.

24. Lettre G, Sankaran VG, Bezerra MA, Araujo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, Orkin SH. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc Natl Acad Sci U S A. 2008;105(33):11869-74.

25. Bae HT, Baldwin CT, Sebastiani P, Telen MJ, Ashley-Koch A, Garrett M, Hooper WC, Bean CJ, Debaun MR, Arking DE, Bhatnagar P, Casella JF, Keefer JR, Barron-Casella E, Gordeuk V, Kato GJ, Minniti C, Taylor J, Campbell A, Luchtman-Jones L, Hoppe C, Gladwin MT, Zhang Y, Steinberg MH. Meta-analysis of 2040 sickle cell anemia patients: BCL11A and HBS1L-MYB are the major modifiers of HbF in African Americans. Blood. 2012;120(9):1961-2.

26. Mtatiro SN, Singh T, Rooks H, Mgaya J, Mariki H, Soka D, Mmbando B, Msaki E, Kolder I, Thein SL, Menzel S, Cox SE, Makani J, Barrett JC. Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. PLoS One. 2014;9(11):e111464.

27. Milton JN, Rooks H, Drasar E, McCabe EL, Baldwin CT, Melista E, Gordeuk VR, Nouraie M, Kato GR, Minniti C, Taylor J, Campbell A, Luchtman-Jones L, Rana S, Castro O, Zhang Y, Thein SL, Sebastiani P, Gladwin MT, Steinberg MH. Genetic determinants of haemolysis in sickle cell anaemia. Br J Haematol. 2013;161(2):270-8.

28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

29. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-5.

30. Wang N, Akey JM, Zhang K, Chakraborty R, Jin L. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. Am J Hum Genet. 2002;71(5):1227-34.

31. Muralidhar SA, Ramakrishnan V, Kalra IS, Li W, Pace BS. Histone deacetylase 9 activates gamma-globin gene expression in primary erythroid cells. J Biol Chem. 2011;286(3):2343-53.

32. Sangerman J, Lee MS, Yao X, Oteng E, Hsiao CH, Li W, Zein S, Ofori-Acquah SF, Pace BS. Mechanism for fetal hemoglobin induction by histone deacetylase inhibitors involves gamma-globin activation by CREB1 and ATF-2. Blood. 2006;108(10):3590-9.

33. Liu L, Karmakar S, Dhar R, Mahajan M, Choudhury A, Weissman S, Pace BS. Regulation of Ggamma-globin gene by ATF2 and its associated proteins through the cAMP-response element. PLoS One. 2013;8(11):e78253.

34. Stamatoyannopoulos G. The molecular basis of blood diseases. 3rd ed. Philadelphia: W.B. Saunders; 2001.

35. Daly AK. Candidate gene case-control studies. Pharmacogenomics. 2003;4(2):127-39.

36. Pereira C, Relvas L, Bento C, Abade A, Ribeiro ML, Manco L. Polymorphic variations influencing fetal hemoglobin levels: association study in beta-thalassemia carriers and in normal individuals of Portuguese origin. Blood Cells Mol Dis. 2015;54(4):315-20.

37. Kos K, Wilding JP. SPARC: a key player in the pathologies associated with obesity and diabetes. Nat Rev Endocrinol. 2010;6(4):225-35.

38. Kanter HL, Saffitz JE, Beyer EC. Molecular cloning of two human cardiac gap junction proteins, connexin40 and connexin45. J Mol Cell Cardiol. 1994;26(7):861-8.

39. Lines MA, Huang L, Schwartzentruber J, Douglas SL, Lynch DC, Beaulieu C, Guion-Almeida ML, Zechi-Ceide RM, Gener B, Gillessen-Kaesbach G, Nava C, Baujat G, Horn D, Kini U, Caliebe A, Alanay Y, Utine GE, Lev D, Kohlhase J, Grix AW, Lohmann DR, Hehr U, Bohm D, Majewski J, Bulman DE, Wieczorek D, Boycott KM. Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. Am J Hum Genet. 2012;90(2):369-77.

40. Fogarty MP, Panhuis TM, Vadlamudi S, Buchkovich ML, Mohlke KL. Allele-specific transcriptional activity at type 2 diabetes-associated single nucleotide polymorphisms in regions of pancreatic islet open chromatin at the JAZF1 locus. Diabetes. 2013;62(5):1756-62.

41. Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, Sabo PJ, Vierstra J, Voit RA, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettre G, Orkin SH. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science. 2013;342(6155):253-7.

42. Kodeboyina S, Balamurugan P, Liu L, Pace BS. cJun modulates Ggamma-globin gene expression via an upstream cAMP response element. Blood Cells Mol Dis. 2010;44(1):7-15.

43. Pissard S, Beuzard Y. A potential regulatory region for the expression of fetal hemoglobin in sickle cell disease. Blood. 1994;84(1):331-8.

44. Sankaran VG, Xu J, Byron R, Greisman HA, Fisher C, Weatherall DJ, Sabath DE, Groudine M, Orkin SH, Premawardhena A, Bender MA. A functional element necessary for fetal hemoglobin silencing. N Engl J Med. 2011;365(9):807-14.

45. Moroy T, Vassen L, Wilkes B, Khandanpour C. From cytopenia to leukemia: the role of Gfi1 and Gfi1b in blood formation. Blood. 2015;126(24):2561-9.

46. van der Meer LT, Jansen JH, van der Reijden BA. Gfi1 and Gfi1b: key regulators of hematopoiesis. Leukemia. 2010;24(11):1834-43.

47. Zeng H, Yucel R, Kosan C, Klein-Hitpass L, Moroy T. Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. EMBO J. 2004;23(20):4116-25.

48. Ramakrishnan V PB. Regulation of γ-globin gene expression involves signaling through the p38 MAPK/CREB1 pathway. Blood Cells Mol Dis. 2011;47(1):12-22.

49. Kalra IS, Alam MM, Choudhary PK, Pace BS. Kruppel-like Factor 4 activates HBG gene expression in primary erythroid cells. Br J Haematol. 2011;154(2):248-59.

50. Miller BA, Salameh M, Ahmed M, Olivieri N, Antognetti G, Orkin SH, Huisman TH, Nathan DG. Analysis of hemoglobin F production in Saudi Arabian families with sickle cell anemia. Blood. 1987;70(3):716-20.

51. Miller BA, Olivieri N, Salameh M, Ahmed M, Antognetti G, Huisman TH, Nathan DG, Orkin SH. Molecular analysis of the high-hemoglobin-F phenotype in Saudi Arabian sickle cell anemia. N Engl J Med. 1987;316(5):244-50.

52. Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science. 2008;322(5909):1839-42.

53. Xu J, Peng C, Sankaran VG, Shao Z, Esrick EB, Chong BG, Ippolito GC, Fujiwara Y, Ebert BL, Tucker PW, Orkin SH. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science. 2011;334(6058):993-6.

54. Xu J, Sankaran VG, Ni M, Menne TF, Puram RV, Kim W, Orkin SH. Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. 2010;24(8):783-98.

55. Borg J, Papadopoulos P, Georgitsi M, Gutierrez L, Grech G, Fanis P, Phylactides M, Verkerk AJ, van der Spek PJ, Scerri CA, Cassar W, Galdies R, van Ijcken W, Ozgur Z, Gillemans N, Hou J, Bugeja M, Grosveld FG, von Lindern M, Felice AE, Patrinos GP, Philipsen S. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nat Genet. 2010;42(9):801-5.

56. Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM. KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. Nat Genet. 2010;42(9):742-4.

57. Pule GD, Ngo Bitoungui VJ, Chetcha Chemegni B, Kengne AP, Antonarakis S, Wonkam A. Association between Variants at BCL11A Erythroid-Specific Enhancer and Fetal Hemoglobin Levels among Sickle Cell Disease Patients in Cameroon: Implications for Future Therapeutic Interventions. OMICS. 2015;19(10):627-31.

58. Sebastiani P, Farrell JJ, Alsultan A, Wang S, Edward HL, Shappell H, Bae H, Milton JN, Baldwin CT, Al-Rubaish AM, Naserullah Z, Al-Muhanna F, Alsuliman A, Patra PK, Farrer LA, Ngo D, Vathipadiekal V, Chui DH, Al-Ali AK, Steinberg MH. BCL11A enhancer haplotypes and fetal hemoglobin in sickle cell anemia. Blood Cells Mol Dis. 2015;54(3):224-30.

59. Creary LE, Ulug P, Menzel S, McKenzie CA, Hanchard NA, Taylor V, Farrall M, Forrester TE, Thein SL. Genetic variation on chromosome 6 influences F cell levels in healthy individuals of African descent and HbF levels in sickle cell patients. PLoS One. 2009;4(1):e4218.

60. Wonkam A, Ngo Bitoungui VJ, Vorster AA, Ramesar R, Cooper RS, Tayo B, Lettre G, Ngogang J. Association of variants at BCL11A and HBS1L-MYB with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. PLoS One. 2014;9(3):e92506.

61. Cardoso GL, Diniz IG, Silva AN, Cunha DA, Silva Junior JS, Uchoa CT, Santos SE, Trindade SM, Cardoso Mdo S, Guerreiro JF. DNA polymorphisms at BCL11A, HBS1L-MYB and Xmn1-HBG2 site loci associated with fetal hemoglobin levels in sickle cell anemia patients from Northern Brazil. Blood Cells Mol Dis. 2014;53(4):176-9.

62. Jiang J, Best S, Menzel S, Silver N, Lai MI, Surdulescu GL, Spector TD, Thein SL. cMYB is involved in the regulation of fetal hemoglobin production in adults. Blood. 2006;108(3):1077-83.

63. Stadhouders R, Aktuna S, Thongjuea S, Aghajanirefah A, Pourfarzad F, van Ijcken W, Lenhard B, Rooks H, Best S, Menzel S, Grosveld F, Thein SL, Soler E. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J Clin Invest. 2014;124(4):1699-710.

64. Menzel S, Rooks H, Zelenika D, Mtatiro SN, Gnanakulasekaran A, Drasar E, Cox S, Liu L, Masood M, Silver N, Garner C, Vasavda N, Howard J, Makani J, Adekile A, Pace B, Spector T, Farrall M, Lathrop M, Thein SL. Global genetic architecture of an erythroid quantitative trait locus, HMIP-2. Ann Hum Genet. 2014;78(6):434-51.

65. Peri KG, Gagnon J, Gagnon C, Bard H. Association of -158 (C-->T) (XmnI) DNA polymorphism in G gamma-globin promoter with delayed switchover from fetal to adult hemoglobin synthesis. Pediatr Res. 1997;41(2):214-7.

66. Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL. Genetic influences on F cells and other hematologic variables: a twin heritability study. Blood. 2000;95(1):342-6.

67. Lehmann S, O'Kelly J, Raynaud S, Funk SE, Sage EH, Koeffler HP. Common deleted genes in the 5q- syndrome: thrombocytopenia and reduced erythroid colony formation in SPARC null mice. Leukemia. 2007;21(9):1931-6.

68. Luo Z, Zhou Y, Luo P, Zhao Q, Xiao N, Yu Y, Yan Q, Lu G, Cheng L. SPARC deficiency affects bone marrow stromal function, resulting in impaired B lymphopoiesis. J Leukoc Biol. 2014;96(1):73-82.

69. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008;40(5):638-45.

70. Kutlar A, Hattori Y, Bakioglu I, Kutlar F, Kamel K, Huisman TH. Hematological observations on Arabian SS patients with a homozygosity or heterozygosity for a beta S chromosome with haplotype #31. Hemoglobin. 1985;9(6):545-57.

71. Nagel RL, Erlingsson S, Fabry ME, Croizat H, Susuka SM, Lachman H, Sutton M, Driscoll C, Bouhassira E, Billett HH. The Senegal DNA haplotype is associated with the amelioration of anemia in African-American sickle cell anemia patients. Blood. 1991;77(6):1371-5.

72. Nagel RL, Rao SK, Dunda-Belkhodja O, Connolly MM, Fabry ME, Georges A, Krishnamoorthy R, Labie D. The hematologic characteristics of sickle cell anemia bearing the Bantu haplotype: the relationship between G gamma and HbF level. Blood. 1987;69(4):1026-30.

73. Powars DR, Meiselman HJ, Fisher TC, Hiti A, Johnson C. Beta-S gene cluster haplotypes modulate hematologic and hemorheologic expression in sickle cell anemia. Use in predicting clinical severity. Am J Pediatr Hematol Oncol. 1994;16(1):55-61.

74. Liu L, Muralidhar S, Singh M, Sylvan C, Kalra IS, Quinn CT, Onyekwere OC, Pace BS. High-density SNP genotyping to define beta-globin locus haplotypes. Blood Cells Mol Dis. 2009;42(1):16-24.

75. Kiefer CM, Lee J, Hou C, Dale RK, Lee YT, Meier ER, Miller JL, Dean A. Distinct Ldb1/NLI complexes orchestrate gamma-globin repression and reactivation through ETO2 in human adult erythroid cells. Blood. 2011;118(23):6200-8.

76. Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu YJ, Yao H, Trowbridge JJ, Mandel G, Orkin SH. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc Natl Acad Sci U S A. 2013;110(16):6518-23.

77. Hein MS, Swanson KC, Lundquist PA, Yungerberg JA, Coon LM, Dawson BD, Jevremovic D, Oliveira AM, Hoyer JD, Oliveira JL. Deletional HPFH Vs. Delta Beta Thalassemia: Closing in on a Possible Hb F Silencer Location. Blood. 2015;126(23):3372-.

78. Thambyrajah R, Mazan M, Patel R, Moignard V, Stefanska M, Marinopoulou E, Li Y, Lancrin C, Clapes T, Moroy T, Robin C, Miller C, Cowley S, Gottgens B, Kouskoff V, Lacaud G. GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. Nat Cell Biol. 2015.

79. Vassen L, Beauchemin H, Lemsaddek W, Krongold J, Trudel M, Moroy T. Growth factor independence 1b (gfi1b) is important for the maturation of erythroid cells and the regulation of embryonic globin expression. PLoS One. 2014;9(5):e9663

Figure Legends

Fig. 1 Analysis of GWAS data by PLINK. The genotype data generated by the entire cohort were subjected to case-control association analysis using PLINK software. The top 200,000 SNPs were graphed using Haploview with logarithm transformed P values (Y axis) and the chromosomes in which the various SNPs reside (X axis). The vertical red box shows significant SNPs in the BCL11A (Chr2) and SPARC (Chr5) genes.

Fig. 2 Functional data for HbF-associated BCL11A SNPs locate in an erythroid-specific enhancer region. Results generated by the ENCODE project for the BCL11A gene were analyzed using the UCSC genome browser (genome.ucsc.edu) with hg19 chromosomal coordinates. The BCL11A gene on Chromosome 2 is shown with its major isoforms on the top and AL833181 is a predicted non-coding transcript. A 15-kb region from 60,713,000 to 60,728,000 was enlarged to depict the functional ChIP-seq results in K562 cells (middle) including DNase hypersensitivity (DNase HS), H3K4me1, H3K4me3, RNA Polymerase II binding (Pol II), GATA1 and TAL1 to the region. ChIP-seq results for GATA1 in peripheral blood derived erythroblasts (PBDE) demonstrate the +55, +58 and +62 erythroid-specific enhancer regions. The box on the bottom, illustrates the top HbF-associated BCL11A SNPs indicated by the bars and respective dbSNP ID, identified in our GWAS.

Fig. 3 High-density SNP-mapping in the HBB locus. A) Shown is a schematic diagram of the distribution of the HBB locus SNPs (101) on the IlluminaOmni1-Quad chip. Globin genes are indicated by boxes. The schematic is not drawn to scale. Abbreviations: LCR, locus control region; HS, hypersensitive site. Symbols: Black dot, IlluminaOmni1-Quad SNPs; Red dot, Taqman SNPs -1280GATA1 (rs2588121) and -1225CRE (rs2588122); Green dot, β^S-globin

SNP, rs334. Haploview was used to infer haplotypes with combined Illumina and Taqman genotype data. The degree of LD is defined by value of D' and LOD (logarithm of the likelihood odds ratio), which is a measure of confidence in the D' value. The boxes are interpreted as follows: red, strong LD (LOD >2, D' = 1), white no LD (LOD < 2, D' < 1), and pink (LOD = 2, D' <1), and blue (LOD < 2, D' =1) indicate intermediate LD. The D'=1 unless indicated in each box; by convention D' is multiplied by 100. SNPs with strong LD define haplotype blocks (black triangle) and the size of the region in LD is shown in parentheses. **B)** Haploview analysis for 101subjects with low HbF (mean HbF=1.84%). The schematic diagram of HBB locus is shown on top. **C)** Haploview analysis was completed for 133 subjects with high HbF (mean HbF=13.83%). The schematic diagram of the HBB locus is shown on top.

Fig. 4 Analysis of HbF-associated HBB locus SNPs potentially involved in γ-globin silencing. A) The LD profile for the low HbF group is shown at the left image. The color scheme for LD pattern is the same as defined in Fig. 2. The inferred haplotypes for these SNPs are depicted on the right. **B)** The LD pattern and inferred haplotypes for the high HbF group are shown. **C)** The HbF-associated SNPs identified in the HBB locus between 3' ε-globin and δ-globin, and HbF silencing region are shown. **D)** K562 ChIP-seq data produced by the ENCODE project were analyzed using UCSC genome browser with hg19 chromosomal coordinates. Shown are the tracks for the proteins indicated between positions 5,254,000 to 5,262,140. Three HbF-associated SNPs are indicated at the bottom. **E)** SNP rs4910736 with 320-bp flanking sequence was analyzed by Tfsitescan software to identify predicted transcription factor binding sites. The results for the A allele and C allele are shown on the top and bottom respectively.

Fig. 5 CBP knockdown induces γ-globin expression in primary erythroid progenitors. Using a two-phase liquid cell culture system erythroid progenitors were generated and transiently transfected at day 28 with siRNA against CBP; changes in γ-globin and HbF expression were quantified (See Materials and Methods). **A)** Shown in the graph is the RT-qPCR data for CBP mRNA normalized by the internal control Actin. A representative image of western blot analysis for CBP and Actin are shown on the bottom; *$p<0.05$. **B)** RT-qPCR analysis of γ-globin, β-globin and GAPD was completed; the fold change of γ-globin expression under the different conditions is shown. **C)** ELISA was performed with anti-HbF and total Hb antibodies. The raw data were normalized by total hemoglobin (T-Hb) and total protein (T-Protein) for each sample. **D)** Immunocytostaining with anti-γ-globin fluorescein isothiocyanate antibody demonstrated an increase in HbF producing cells (photomicrograph on left). The quantitative data showing the % HbF-positive cells under the different conditions are shown in the graph; **$p<0.001$.