NIH Public Access
Author Manuscript
Blood Cells Mol Dis. Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:
Blood Cells Mol Dis. 2015 March ; 54(3): 224–230. doi:10.1016/j.bcmd.2015.01.001.

BCL11A Enhancer Haplotypes and Fetal Hemoglobin in Sickle Cell Anemia:

BCL11A Enhancers and HbF

P. Sebastiani[1], **J.J. Farrell**[2], **A. Alsultan**[3], **S. Wang**[1], **H. L. Edward**[2], **H. Shappell**[1], **H. Bae**[4], **J. N. Milton**[1], **C.T. Baldwin**[2], **A.M. Al-Rubaish**[5], **Z. Naserullah**[6], **F. Al-Muhanna**[5], **A. Alsuliman**[7], **P. K. Patra**[8], **L.A. Farrer**[2], **D. Ngo**[2], **V. Vathipadiekal**[2], **D.H.K. Chui**[2], **A.K. Al-Ali**[9], and **M.H. Steinberg**[2]

[1] Department of Biostatistics, Boston University School of Public Health, Boston, MA

[2] Department of Medicine, Boston University School of Medicine, Boston, MA

[3] Sickle Cell Disease Research Center and Department of Pediatrics, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[4] College of Public Health and Human Sciences, Oregon State University, Corvallis, OR

[5] Department of Internal Medicine, College of Medicine, University of Dammam, Dammam, Kingdom of Saudi Arabia

[6] Department of Pediatrics, Maternity & Child Hospital, Dammam, Kingdom of Saudi Arabia

[7] Department of Hematology, King Fahd Hospital, Hafof, Al-Ahsa, Kingdom of Saudi Arabia

[8] Deptartment of Biochemistry, Pt. J.N.M. Medical College, Raipur, Chattisgarh, India

[9] Prince Mohammed Center for Research & Consultation Studies, University of Dammam, Dammam, Kingdom of Saudi Arabia

Abstract

Background—Fetal hemoglobin (HbF) levels in sickle cell anemia patients vary. We genotyped polymorphisms in the erythroid-specific enhancer of *BCL11A* to see if they might account for the very high HbF associated with the Arab-Indian (AI) haplotype and Benin haplotype of sickle cell anemia.

Methods and Results—Six *BCL112A* enhancer SNPs and their haplotypes were studied in Saudi Arabs from the Eastern Province and Indian patients with AI haplotype (HbF ~20%), African Americans (HbF ~7%), and Saudi Arabs from the Southwestern Province (HbF ~12%). Four SNPs (rs1427407, rs6706648, rs6738440, and rs7606173) and their haplotypes were

© 2015 Published by Elsevier Inc.

Corresponding Author: Paola Sebastiani, Department of Biostatistics, Boston University School of Public Health 801 Massachusetts Avenue, Boston MA 02118, sebas@bu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

consistently associated with HbF levels. The distributions of haplotypes differ in the 3 cohorts but not their genetic effects: the haplotype TCAG was associated with the lowest HbF level and the haplotype GTAC was associated with the highest HbF level and differences in HbF levels between carriers of these haplotypes in all cohorts was approximately 6%.

Conclusions—Common HbF *BCL11A* enhancer haplotypes in patients with African origin and AI sickle cell anemia have similar effects on HbF but they do not explain their differences in HbF.

Introduction

Fetal hemoglobin (HbF) is the predominant modulator of the phenotype of sickle cell anemia. By inhibiting sickle hemoglobin (HbS) polymerization it reduces the tissue injury and hemolytic anemia exemplifying this disease.[1-4] HbF levels and the distribution of HbF concentrations in sickle erythrocytes are highly variable. The genetic regulation of HbF was first associated with the haplotype of the β-globin gene (*HBB*) cluster suggesting the importance of cis-acting regulation.[5, 6] In the Middle East and in India the HbS gene is often on an autochthonous Arab-Indian (AI) *HBB* haplotype that is associated with HbF levels more than twice as high as those found with African origin haplotypes, see **Figure 1.**[7, 8] This is clinically important as the youngest individuals with the AI haplotype have the mildest phenotype of all sickle cell anemia patients, although when their HbF level falls from about 30% in children to 15-20% in adults the disease becomes more severe.[9-14] Saudi Arabs with the African origin Benin haplotype have HbF levels nearly twice that of non-Arab patients with this haplotype.[15] Within each haplotype group there is considerable heterogeneity of HbF,[5, 6, 16] suggesting that trans-acting elements also effect γ-globin gene (*HBG*) expression.

One trans-acting element is *BCL11A*, a repressor of γ-globin gene expression.[10, 17-21] Functional studies have shown that *BCL11A* expression is regulated by erythroid-specific enhancers in its 2nd intron. The enhancer elements contain 3 DNase hypersensitive sites (DHS) lcated +62, +58 and +55 kb from the transcription initiation site.[10] Two SNP haplotypes of the enhancer elements were associated with HbF levels in African American patients with sickle cell anemia. The strongest association with HbF levels in African Americans with sickle cell anemia was with rs1427407 in DHS +62.

The cause of high HbF in the AI haplotype and the Saudi Benin haplotype is unexplained and could be due to increased *HBG* expression mediated by cis- or trans-acting regulators. We focused on *BCL11A* enhancer polymorphisms and examined the association of their haplotypes with HbF levels in Saudi and Indian patients with the AI haplotype and Saudi HbS homozygotes with the Benin haplotype, and compared these results to those of African American patients with sickle cell anemia.

Materials and Methods

Study Populations and HbF Measurement

SNPs in *BCL11A* enhancers were genotyped directly or imputed from genome-wide SNP analysis in the following cohorts (**Table 1**):

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

1. 894 African American HbS homozygotes, diverse haplotypes, from the Cooperative Study of Sickle Cell Disease (CSSCD), aged > 5 years.[22]
2. 96 Saudi HbS homozygotes mostly with the Benin haplotype from the Southwestern Province of Saudi Arabia (Saudi W), aged 4-55 years, not taking hydroxyurea. (Saudi W)
3. 110 Saudi HbS homozygotes all with AI haplotype from the Eastern Province of Saudi Arabia, aged 11-59 years, not taking hydroxyurea (Saudi E).[16]
4. 44 Indian HbS homozygotes all with the AI haplotype, age 10-32 years, not taking hydroxyurea.

HbF: HbF was measured in all Saudi samples using high performance liquid chromatography (HPLC) or capillary electrophoresis. HbF in the CSSCD was measured by alkali denaturation.[23] HbF in Indian patients was measured by HPLC both in India and at Boston University.

Genotyping

HbS mutation and HBB haplotype: Homozygosity for the HbS gene was confirmed using amplification refractory mutation system analysis.[24] HbS homozygosity in the CSSCD cohort was based on clinical and hematologic studies. The AI haplotype was ascertained by analysis of rs7482144 (Xmn1 C-T restriction site 158 bp 5' to *HBG2*), rs3834466 (Hinc2 restriction site 5' to *HBE1* and the C-T SNP 68 bp 5' to *HBD*.[25] *SNPs:* Targeted genotyping of *BCL11A* enhancer SNPs was done with tetra-primer ARMS-PCR, TaqMan assays and Sanger sequencing.

Imputation of genotypes: To derive haplotypes of these SNPs in cohorts where direct genotyping was not done we imputed to 1000 Genomes level data from genome-wide SNP data obtained using Illumina technology. (**Supplementary material**)

Data analysis

Data are described by mean and standard deviation. Single SNP associations were estimated using sex and age adjusted linear regression with additive genetic effects, and the B allele in the forward strand was the coded allele (**Table 2**). A mixed effect model with kinship coefficients implemented in the coxme package of the R statistical software was used to analyze the association between HbF levels and SNPs of Saudi W samples, since some subjects were related. HbF levels were approximately normally distributed in the Saudi E and Indian samples (**See Supplement Figure S1**), and a cubic root transformation was used for normalizing the HbF levels of CSSCD samples as in.[22] HbF levels of Saudi W samples were also normalized using a cubic root transformation, Linkage disequilibrium was evaluated using the program HaploView. Haplotype analysis was conducted using the haplo.stats package in the R software.[26] The program uses a two-step EM algorithm to iteratively update the probability of subjects haplotypes, based on coefficient of the age and sex adjusted regression model, and to update the coefficients of the regression model based on the posterior probability of a subject haplotype. The analysis was conducted in the combined Saudi E and Indian samples and, separately, in the Saudi W and CSSCD samples

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

using linear regression adjusted for age and sex. Haplotype pairs for each subject were inferred based on the most likely haplotypes and the distribution of HbF was displayed using boxplots. To test if the effect of haplotype pairs changed substantially in the different cohorts, a multivariable regression model of the HbF levels versus haplotype pairs was fitted, using age, sex, indicators for the haplotype pairs, an indicator of the cohort type (CSSCD, Saudi W, and Saudi E + Indians), and an interaction between indicator variables of the haplotype pairs and study cohorts. Lack of a statistically significant interaction indicated no change of haplotype pairs effects in the 3 cohorts. All analyses were conducted using the statistical software R.

These studies were approved by the Institutional Review Boards of the participation institutions.

Results

Patients

Table 1 summarizes patients' characteristics. All patients were homozygous for the HbS mutation. Saudi E patients and Indian patients were homozygous for the AI haplotype; Saudi W patients were mostly homozygous for the Benin haplotype. Forty-seven percent of African American patients were Benin haplotype homozygotes, 27% Benin/Bantu compound heterozygotes, 10% Benin/Senegal compound heterozygote and 16% had other haplotypes. Eight CSSCD cases were Senegal haplotype homozygotes with HbF levels of 9.2, 2.3, 4.6, 16.0, 16.3, 11, 5.9 and 3.4% and ages of 24, 30, 16, 6, 24, 27, 44 and 23 years, respectively. Saudi E and Indian AI haplotype patients had the highest HbF of all cohorts. HbF in Saudi W patients was intermediate between the AI haplotype and CSSCD African haplotype patients.

SNP genotypes

The region of *BCL11A* DHS and the SNPs genotypes are shown in **Figure 2** and **Table S2** in the supplementary material. The association between SNPs in *BCL11A* and HbF levels are summarized in **Table 2**. Four SNPs (rs1427407, rs6706648, rs6738440 and rs7606173) were significantly associated with varying levels of HbF. **Figure 3** displays the distribution of HbF by genotypes for the 4 significant SNPs and highlights the similar trend of HbF in the four study cohorts. The minor alleles of rs1427407 in DHS +62 and rs7606173 in DHS +55 are individually associated with increased HbF while the minor alleles of rs6706648 and rs6738440 in DHS +58 are individually associated with lower HbF. **Figure 3** also shows the systematically higher levels of HbF in the Saudi E and Indian AI samples relative to the African American cohort.

Bioinformatic analysis of possible changes to transcription factor binding as a result of DHS polymorphisms (http://compbio.cs.queensu.ca/F-SNP/) suggests that rs7606173 in DHS +55 changes a binding site for *MZF1* (myeloid zinc finger 1), a gene active in hematopoietic cells, including K562 cells. Rs6706648 in DHS +58 is also near a MZFI binding site; both SNPs are near binding sites for c-MYB. (**Table S2** in supplement material)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Haplotypes of BCL11A hypersensitive sites

The 4 significant SNPs are located in the DHS +62, +58 and +55 (**Figure 2**). **Figure 4** shows that there is moderate linkage disequilibrium (LD) between rs6706648, rs6738440 and rs7606173, while rs1427407 appears to be independent of the other 3 SNPs. To assess the joint effects of combinations of these SNPs on levels of HbF, we conducted a haplotype analysis of the 4 SNPs. Data from the Saudi E + Indian cohorts were combined, given the similarity of HbF phenotype and *HBB* haplotype and study design, while the data from the other cohorts were analyzed separately. Fourteen distinct haplotypes were identified (**Table 3**). They include 4 common haplotypes (haplotypes 2, 5, 7 and 10 in **Table 3** with haplotype frequency (HF) >0.05). Compared to the African American sample, the combined Saudi E + Indian samples also carry an additional haplotype, (haplotype 1: GCAC), with a frequency of 0.06. This haplotype was carried by 11.33% of Saudi W participants. Haplotype 2 (GCAG) was more frequent in the Saudi E + Indian sample (HF = 53%) than the African American samples (HF = 32%). Haplotypes 5, 7 and 10 were less common in the combined Saudi E + Indian samples (HF = 38%), and Saudi W samples (HF = 45%) compared with African American samples (HF = 64%). The difference in frequency was statistically significant ($p\text{-}^2$ test < 10-5). Frequencies of the haplotypes 2, 5 and 7 in the Saudi W samples were between those of the African American samples and the Saudi E + Indian samples, while haplotype 10 was more common in the Saudi W samples then the other two groups.

Haplotype association analysis summarized in **Table 4** and Supplement Table S3 show that only haplotype 10 is associated with an increase HbF relative to the referent haplotype GCAG in the combined Saudi E and Indian samples, and the Saudi W samples. This result is consistent with the work of Bauer et al,[10] who reported 2 SNP haplotypes of rs1427407-rs7606173 in African Americans where GC was associated with HbF of 4.05±3.10%. The haplotype pairs TG/GT was associated with HbF of 7.08±4.50% and TG was associated with HbF 11.21±4.73%. The average HbF value in carriers of the referent haplotype (GCAG) in the Saudi E+ Indian samples was estimated to be 20%, at the mean age of 27 years, and carriers of haplotype 10 (TCAG) had an average increase of 1.74% compared with carriers of the referent haplotype (**Table 4**). In the CSSCD, the average HbF levels of carriers of the haplotype 10 was about 8%, while carriers of the common haplotype GCAG had an average HbF level of 5.5%. In the Saudi W, the frequency of haplotype 10 was about 30%, and carriers of the haplotype 10 had an average increase of HbF of 2.4%. The other haplotypes were associated with a decrease of HbF in the range of 3.4% and 3.6% in the Saudi E + Indian samples and a similarly negative effect in both African American samples.

The significantly different effects of the haplotypes 1, 5, and 7 suggest that varying combinations of the 4 SNPs haplotype are associated with varying levels of HbF. To better clarify this point, we examined the variation of HbF levels based on pairs of haplotypes. The most likely haplotype pair in each subject was determined based on the highest posterior probability, and the assignment of haplotype pairs was essentially unambiguous (mean posterior probability in CSSCD = 0.996; in Saudi E + Indian = 0.98). The plot in **Figure 5** shows a difference in distribution of HbF levels per haplotype pairs that ranges between a median value of 3.5% HbF in CSSCD carriers of the haplotypes 5, 7, and 10.5% HbF in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

carriers of the 10, 10 haplotype pair. The difference in mean HbF was statistically significant (p-value from t-test < 10-15). In the Saudi E + Indian samples, carriers of the haplotypes 5, 7 had an average HbF of 10% while carriers of the 10, 10 haplotype pair had average HbF of 23%. The difference was statistically significant (p-value from t-test < 0.02). The distribution of HbF in Saudi W samples was between the CSSCD and Saudi E + Indian samples.

The plot also suggests that the effect of haplotype 10 associated with the highest HbF is mediated by the second haplotype and carriers of the 5,10 haplotype pairs have an average 2.5% less HbF in the Saudi E + Indian samples compared to carriers of the 10,10 haplotype pairs, although the difference did not reach statistical significance. The same difference of 2.5% HbF between CSSCD patients carrying the 5, 10 or 10, 10 haplotype pairs reached statistical significance (p-value from t-test < 0.01). In addition, the plot highlights the similarity of the genetic effects of the haplotype pairs in all study populations, and suggests that variants in *BCL11A* explain variations in HbF levels (linear trend in **Figure 5**) but cannot explain the baseline higher HbF levels in the Saudi W samples and the even higher HbF levels in the Saudi E+ Indian sample (shift to the right). To confirm this result, we fitted a multivariable regression model with a haplotype-pair x cohort interaction and the interaction effects did not reach statistical significance (minimum p-value for interaction 0.17)

The differential distributions of haplotypes in the Saudi E + Indian cohorts compared with African Americans can explain the distribution of HbF levels in the two groups (**Figure S1**). The skewed distribution of HbF in the African American samples is due to the higher frequency of haplotypes associated with lower HbF.

Discussion

Aboriginal populations of the Indus Valley Harappa culture might have introduced the HbS gene on the AI *HBB* cluster haplotype to the Eastern Province of Saudi Arabia (**Figure 1**).[27, 28] Carriers of this haplotype typically have HbF levels 3 to 4 times as high as patients with African-origin haplotypes. In contrast, the HbS gene on the Benin *HBB* haplotype was introduced to the Southwestern Province from Africa (**Figure 1**). The genetic population structure of Southwestern Province patients is similar to other Arab populations.[15, 22, 29] Their HbF levels are twice that of carriers of the Benin haplotype of African-origin but half that of patients with the AI haplotype (**Table 1**). These observations suggest that (1) high HbF in the AI haplotype is a result of cis-acting regulators and unique “Arab” trans-acting regulation, (2) Saudi patients from the Southwestern Province with the Benin haplotype lack the cis-acting elements of the AI haplotype but have similar “Arab” trans-acting factors, (3) African patients with the Senegal haplotype have high HbF and cis-acting elements similar to those of the AI haplotype but lack the “Arab” trans-acting elements, and (4) most other African patients with Benin and Bantu *HBB* haplotypes lack both the cis- and trans-acting elements needed for high HbF expression.

As *BCL11A* is associated with HbF levels in many different populations, albeit with different effect sizes, we examined polymorphisms in the DHS of its erythroid-specific

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

enhancers to see if they might explain different HbF levels in the AI and other haplotypes of sickle cell anemia.

Significant differences were present in the distribution of *BCL11A* enhancer haplotypes in AI and African-origin haplotype patients. In all cohorts, haplotypes 5 and 7 were associated with lower HbF than the common referent haplotype 2. Haplotypes 5 and 7 were less frequent in the Saudi E + Indian cohorts than in African American cohorts. The different frequencies of *BCL11A* haplotypes are likely to account for the different distribution of HbF levels in patients with African-origin compared with AI haplotypes (**Figure S1**). In the AI haplotype, HbF is normally distributed with a mean of about 20%; individuals with African haplotypes have a "skewed" HbF distribution with a mean of about 6% and fewer patients with "high" HbF perhaps because of the lower frequency of haplotypes associated with higher HbF.

When haplotype pairs are analyzed, haplotype 10 was associated with the highest HbF. While haplotype pairs are associated with HbF they do not appear to explain the differences in HbF level between AI and African-origin haplotype populations (**Figure 3**). This is consistent with the observation that although *BCL11A* polymorphisms are associated with HbF in AI haplotype patients as in other racial and ethnic groups, in the AI haplotype these polymorphisms explain less of the variation of HbF than studies of the other cohorts.[19, 20, 30-33] It is therefore likely that other loci have a predominant role in the differential expression of HbF in the AI compared with other *HBB* haplotypes. The nature of putative "Arab" trans- and cisacting regulator(s) remain unknown.

Haplotypes capture the combinatorial importance [10, 30] of the *BCL11A* enhancer elements more comprehensively than single SNP analysis. In previous studies of *BCL11A* haplotypes and HbF in African Americans, 3 SNPs in *BCL11A*, rs10189857, rs4671393 and rs7599488 formed 4 haplotypes that were more strongly associated with HbF than single SNP analysis.[30] These SNPs were not in the DHS of the *BCL11A* enhancers. Rs7599488 in DHS +62 was not associated with HbF in AI haplotype patients. Using imputed data from GWAS, Bhatnagar et al,[34] reported similar associations of DHS SNP haplotypes with F-cell levels in African American children. In the CSSCD, Bauer et al genotyped 2 SNP haplotypes of rs1427407-rs7606173 in DHS +62 and +55. Haplotypes TG (24.5%), TC (0.85%), GC (42.3%) and GG (33.1%) were associated with 4.05, 7.08 and 11.2% HbF.[10]

Stimulating HbF gene expression is an attractive therapeutic approach for sickle cell anemia and β thalassemia. A better understanding of the genetic basis of HbF regulation has uncovered novel therapeutic targets [10] for drug development. Finding elements responsible for high HbF in the AI haplotype and Saudi Benin haplotypes might further inform this effort.[10, 35, 36]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Acknowledgments

This work was funded by the University of Dammam, SP 11/2011, Office of Collaboration and Knowledge Exchange, University of Dammam, NIH/NHLBI HL101212, HL068970, HL87681, and National Institute of General Medical (NIGMS) Interdisciplinary Training Grant for Biostatisticians (T32 GM74905) and Training Grant for Hematology (5T32 HL007501)

References

1. Nagel RL, Bookchin RM, Johnson J, Labie D, Wajcman H, Isaac-Sodeye WA, Honig GR, Schiliro G, Crookston JH, Matsutomo K. Structural bases of the inhibitory effects of hemoglobin F and hemoglobin A2 on the polymerization of hemoglobin S. Proc Natl Acad Sci U S A. 1979; 76:670–2. [PubMed: 284392]
2. Poillon WN, Kim BC, Rodgers GP, Noguchi CT, Schechter AN. Sparing effect of hemoglobin F and hemoglobin A2 on the polymerization of hemoglobin S at physiologic ligand saturations. Proc Natl Acad Sci U S A. 1993; 90:5039–43. [PubMed: 7685112]
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; 87:217–9. [PubMed: 22139998]
4. Steinberg MH, Sebastiani P. Genetic modifiers of sickle cell disease. Am J Hematol. 2012; 87:795–803. [PubMed: 22641398]
5. 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; 69:1026–30. [PubMed: 2435339]
6. Nagel RL, Fabry ME, Pagnier J, Zohoun I, Wajcman H, Baudin V, Labie D. Hematologically and genetically distinct forms of sickle cell anemia in Africa. The Senegal type and the Benin type. N Engl J Med. 1985; 312:880–4. [PubMed: 2579336]
7. Kulozik AE, Wainscoat JS, Serjeant GR, Kar BC, Al-Awamy B, Essan GJ, Falusi AG, Haque SK, Hilali AM, Kate S, et al. Geographical survey of beta S-globin gene haplotypes: evidence for an independent Asian origin of the sickle-cell mutation. Am J Hum Genet. 1986; 39:239–44. [PubMed: 3752087]
8. Labie D, Srinivas R, Dunda O, Dode C, Lapoumeroulie C, Devi V, Devi S, Ramasami K, Elion J, Ducrocq R, et al. Haplotypes in tribal Indians bearing the sickle gene: evidence for the unicentric origin of the beta S mutation and the unicentric origin of the tribal populations of India. Hum Biol. 1989; 61:479–91. [PubMed: 2480325]
9. Perrine RP, Brown MJ, Clegg JB, Weatherall DJ, May A. Benign sickle-cell anaemia. Lancet. 1972; 2:1163–7. [PubMed: 4117591]
10. 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; 342:253–7. [PubMed: 24115442]
11. Perrine RP, Pembrey ME, John P, Perrine S, Shoup F. Natural history of sickle cell anemia in Saudi Arabs. A study of 270 subjects. Ann Intern Med. 1978; 88:1–6. [PubMed: 619731]
12. Pembrey ME, Wood WG, Weatherall DJ, Perrine RP. Fetal haemoglobin production and the sickle gene in the oases of Eastern Saudi Arabia. Br J Haematol. 1978; 40:415–29. [PubMed: 749927]
13. Alsultan A, Alabdulaali MK, Griffin PJ, Alsuliman AM, Ghabbour HA, Sebastiani P, Albuali WH, Al-Ali AK, Chui DH, Steinberg MH. Sickle cell disease in Saudi Arabia: the phenotype in adults with the Arab-Indian haplotype is not benign. Br J Haematol. 2014; 164:597–604. [PubMed: 24224700]
14. Italia K, Kangne H, Shanmukaiah C, Nadkarni AH, Ghosh K, Colah RB. Variable phenotypes of sickle cell disease in India with the Arab-Indian haplotype. Br J Haematol. 2014
15. Alsultan A, Solovieff N, Aleem A, AlGahtani FH, Al-Shehri A, Osman ME, Kurban K, Bahakim H, Al-Momen AK, Baldwin CT, Chui DH, Steinberg MH. Fetal hemoglobin in sickle cell anemia: Saudi patients from the Southwestern province have similar HBB haplotypes but higher HbF levels than African Americans. Am J Hematol. 2011; 86:612–4. [PubMed: 21630302]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

16. Ngo D, Bae H, Steinberg MH, Sebastiani P, Solovieff N, Baldwin CT, Melista E, Safaya S, Farrer LA, Al-Suliman AM, Albuali WH, Al Bagshi MH, Naserullah Z, Akinsheye I, Gallagher P, Luo HY, Chui DH, Farrell JJ, Al-Ali AK, Alsultan A. Fetal hemoglobin in sickle cell anemia: genetic studies of the Arab-Indian haplotype. Blood Cells Mol Dis. 2013; 51:22–6. [PubMed: 23465615]
17. 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; 39:1197–9. [PubMed: 17767159]
18. 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; 322:1839–42. [PubMed: 19056937]
19. 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; 41:255–8. [PubMed: 18691915]
20. 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; 105:1620–5. [PubMed: 18245381]
21. 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; 334:993–6. [PubMed: 21998251]
22. 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; 115:1815–22. [PubMed: 20018918]
23. Gaston M, Rosse WF. The cooperative study of sickle cell disease: review of study design and objectives. Am J Pediatr Hematol Oncol. 1982; 4:197–201. [PubMed: 7114401]
24. Little S. Amplification-refractory mutation system (ARMS) analysis of point mutations. Curr Protoc Hum Genet. 2001 Chapter 9:Unit 9 8.
25. Alsultan A, Ngo DA, Farrell JJ, Akinsheye I, Solovieff N, Ghabbour HA, Al-Ali A, Alsuliman A, Al-Baghshi M, Albu-Ali W, Alabdulaali M, Baldwin CT, Farrer LA, Luo H, Melista E, Safaya S, Nwaru M Jr. Chui DH, Steinberg MH. A functional promoter polymorphism of the delta-globin gene is a specific marker of the Arab-Indian haplotype. Am J Hematol. 2012; 87:824–6. [PubMed: 22641479]
26. Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, Schaid DJ. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. Hum Hered. 2003; 55:56–65. [PubMed: 12890927]
27. Coon, CS. The Races of Europe. Macmillan; 1939.
28. Facey, W. The Story of the Eastern Province of Saudi Arabia. Stacey International; 1994.
29. Solovieff N, Hartley SW, Baldwin CT, Klings ES, Gladwin MT, Taylor JGt, Kato GJ, Farrer LA, Steinberg MH, Sebastiani P. Ancestry of African Americans with sickle cell disease. Blood cells, molecules & diseases. 2011; 47:41–5.
30. Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet. 2010; 42:1049–1051. [PubMed: 21057501]
31. 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; 120:1961–2. [PubMed: 22936743]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

32. 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; 108:1077–1083. [PubMed: 16861354]

33. 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; 105:11869–74. [PubMed: 18667698]

34. Bhatnagar P, Purvis S, Barron-Casella E, DeBaun MR, Casella JF, Arking DE, Keefer JR. Genome-wide association study identifies genetic variants influencing F-cell levels in sickle-cell patients. J Hum Genet. 2011; 56:316–323. [PubMed: 21326311]

35. 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; 365:807–14. [PubMed: 21879898]

36. Campbell AD, Cui S, Shi L, Urbonya R, Mathias A, Bradley K, Bonsu KO, Douglas RR, Halford B, Schmidt L, Harro D, Giacherio D, Tanimoto K, Tanabe O, Engel JD. Forced TR2/TR4 expression in sickle cell disease mice confers enhanced fetal hemoglobin synthesis and alleviated disease phenotypes. Proc Natl Acad Sci U S A. 2011; 108:18808–13. [PubMed: 22042865]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

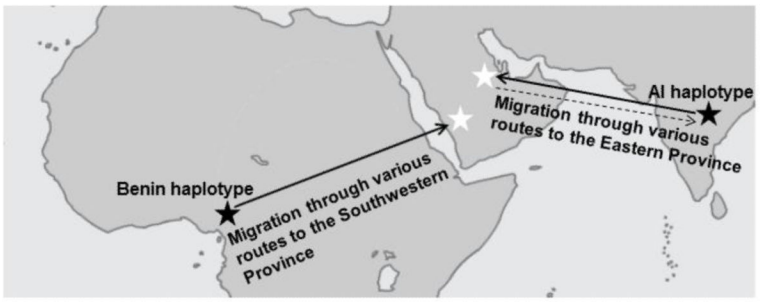

Figure 1.

Origins of the HbS mutation in Africa and India and migration to the Arabian Peninsula. The primary HbS gene-associated haplotype in the Southwestern Province of Saudi Arabia is Benin, which was introduced from Africa. Eastern Province patients have the Arab-Indian (AI) haplotype that might have originated in India (solid arrow) or alternatively, originated in the Middle East and migrated to India (dashed arrow).

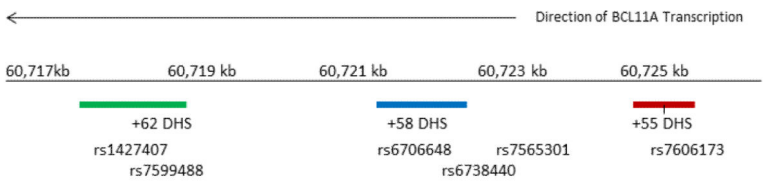

Figure 2.

Location of the 6 SNPs relative to the DNA hypersensitive sites (DHS) labeled as identified in Bauer et al. 2013.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

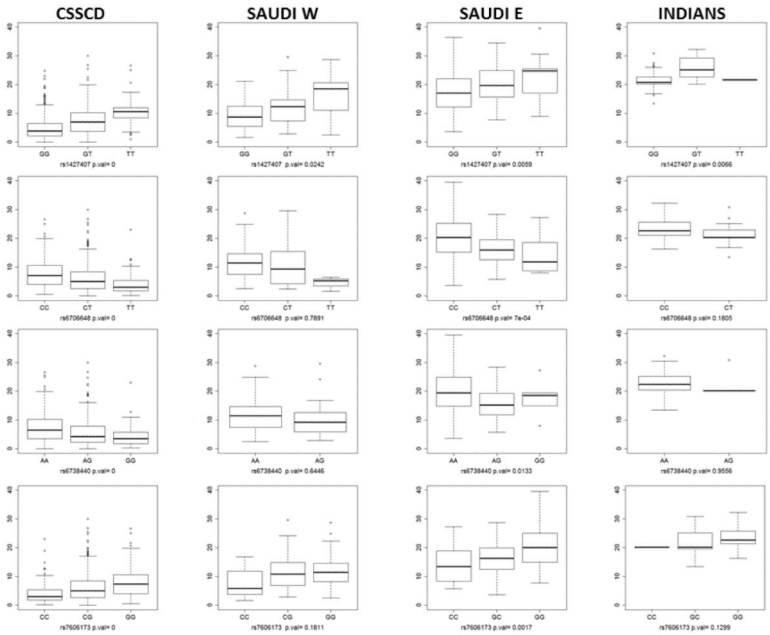

Figure 3.

Distribution of HbF levels in the 4 study populations for different genotypes of the significant SNPs.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

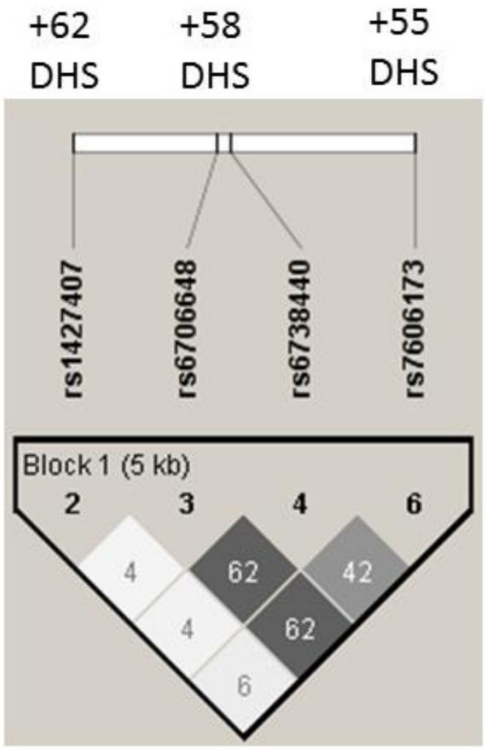

Figure 4.
LD heatmap generated in HaploView to display the linkage disequilibrium (r^2) between the 4 SNPs.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

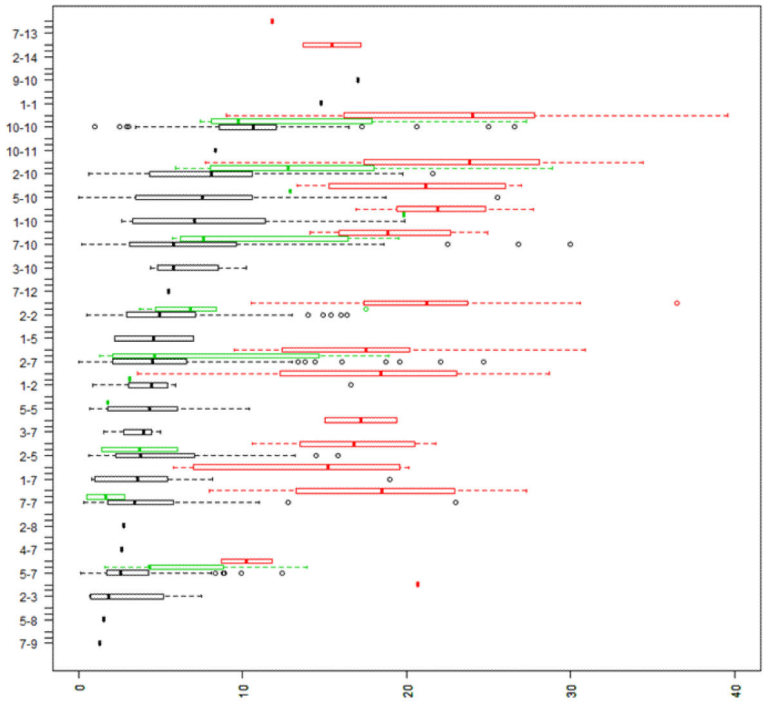

Figure 5.

Distribution of HbF per carriers of haplotype pairs in African Americans sickle cell patients from the CSSCD (black), Saudi W patients (green), and from the combined Saudi E + Indian groups (red). Haplotype pairs are ordered by median HbF in the CCSCD study. Boxplots display 25%, 50% and 75% HbF, and whiskers extends to 1.5 interquartile range.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Cohorts studied.

	N	Age ± SD (yrs.)	HbF ± SD (%)	Genotyping
CSSCD	894	13.6±11.3	5.2±5.6	Illumina/Imputation
Saudi W	99	17.7±9.84	11.4±6.0	Illumina/Imputation
Saudi E: **Al Haplotype**	110	26.7±10.1	18.0±7.0	PCR/Illumina/Imputation
Indian: **Al Haplotype**	44	14.6±4.6	23.0±4.8	PCR

Saudi E and Indian patients are homozygous for the AI haplotype and were not taking hydroxyurea. CSSCD patients have different African haplotypes. Cooperative Study of Sickle Cell Disease (CSSCD). Illumina/Imputation signifies imputation of SNPs not included in the Illumina SNP arrays.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Results of single SNP analysis in the 4 study populations.

		African American (CSSCD)				
SNP	A/B	MAF	AA/AB/BB	b	SE	Pvalue
rs1427407	G/T	**0.26**	**485/355/59**	**0.28**	**0.02**	**1.34E-29**
rs7599488	T/C	0.30	435/383/81	−0.01	0.02	0.5528
rs6706648	C/T	**0.39**	**342/419/138**	−0.2	**0.02**	**3.77E-20**
rs6738440	A/G	**0.27**	**480/351/68**	**−0.18**	**0.02**	**7.96E-14**
rs7565301	A/G	0.26	58/359/482	−0.01	0.03	0.8362
rs7606173	C/G	**0.41**	**153/440/306**	**0.20**	**0.02**	**2.30E-20**
		Saudi W				
SNP	A/B	MAF	AA/AB/BB	b	SE	Pvalue
rs1427407	G/T	**0.32**	**44/42/9**	**0.20**	**0.06**	**0.0007**
rs7599488	T/C	0.42	32/44/16	012	0.06	0.0337
rs6706648	C/T	**0.15**	69/22/3	−0.12	0.07	01142
rs6738440	A/G	**0.10**	74/18/0	−0.05	0.10	0.6446
rs7565301	A/G	0.33	4/36/55	0.05	0.07	0.3928
rs7606173	C/G	**0.27**	6/37/49	−0.06	0.07	0.4042
		Saudi E				
SNP	A/B	MAF	AA/AB/BB	b	SE	Pvalue
rs1427407	G/T	**0.28**	**72/51/13**	**2.5**	**0.89**	**0.0059**
rs7599488	T/C	0.43	38/50/22	0.56	0.95	0.5564
rs6706648	C/T	0.22	**68/36/6**	−3.83	**1.1**	**0.0007**
rs6738440	A/G	0.18	**79/31/5**	**−2.94**	**1.17**	**0.0133**
rs7565301	A/G	0.21	7/27/1964	−1.14	1.18	0.3336
rs7606173	C/G	0.27	**12/28/1957**	**3.27**	**1.01**	**0.0017**
		Indian				
SNP	A/B	MAF	AA/AB/BB	b	SE	Pvalue
rs1427407	G/T	**0.17**	**30/13/1**	**3.21**	**1.12**	**0.00656**
rs7599488	T/C	0.68	5/18/2021	−0.88	0.97	0.37165
rs6706648	C/T	**0.13**	33/11/0	−2	1.47	0.18049
rs6738440	A/G	**0.06**	38/5/0	−0.11	1.96	0.9556
rs7565301	A/G	0.28	4/17/2023	0.33	0.97	0.73827
rs7606173	C/G	**0.14**	1/10/1931	2.03	1.31	0.12992

SNPs significantly associated with HbF are highlighted in bold face. The alleles are ordered by forward strand, and the estimated genetic effect (column beta (b) represents the effect of the B allele in the forward strand. The regression coefficient b in the CSSCD and Saudi W data represents changes in cubic-root transformation of HbF, resulting in smaller values but consistent sign.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Haplotypes of SNPs rs1427407, rs6706648, rs6738440, rs7606173 in *BCL11A* and their frequencies in the study subjects.

Number	Haplotype	HF CSSCD	HF Saudi E+ Indian AI	HF Saudi W
1	GCAC	0.0189	0.0552	0.1099
2*	**GCAG**	**0.3250**	**0.5354**	**0.4279**
3	GCGC	0.0076	0.0110	0.000
4	GCGG	0.0006	-	-
5*	**GTAC**	**0.1257**	**0.0461**	**0.0453**
6	GTAG	0.0001	–	–
7*	**GTGC**	**0.2599**	**0.1303**	**0.1004**
8	GTGG	0.0011	0.0000	-
9	TCAC	0.0020	0.0000	0.0090
10*	**TCAG**	**0.2571**	**0.2069**	**0.3021**
11	TCGC	0.0012	-	0.0053
12	TTGC	0.0009	0.0000	-
13	TTAC	-	0.0071	-
14	TTGG	-	0.0081	-

HF = haplotype frequency.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Frequency of 4 SNPs haplotypes, and mean HbF value per haplotypes in the 3 cohorts.

		Saudi E + Indian			Saudi W			CSSCD		
Haplotype		**Freq**	**HbF**	**Pvalue**	**Freq**	**HbF**	**Pvalue**	**Freq**	**HbF**	**Pvale**
2	**GCAG**	54%	20.4	-	43%	10.0	-	32%	5.2	-
5	**GTAC**	5%	16.8	0.014	4%	6.5	0.384	13%	3.7	0.006
7	**GTGC**	13%	17.0	0.013	10%	8.6	0.990	26%	4.1	0.000
10	**TCAG**	21%	22.8	0.059	30%	12.4	0.002	26%	8.0	0.000

Pvalue= P value for comparing changes in HbF relative to GCAG

Pval is the level of significance of the haplotypes in the multivariable regression model. The numbers in the first column match the numbers in Table 3.