To the editor:

Meta-analysis of 2040 sickle cell anemia patients: \textit{BCL11A} and \textit{HBS1L-MYB} are the major modifiers of HbF in African Americans

Fetal hemoglobin (HbF) protects against many but not all of the hematologic and clinical complications of sickle cell anemia.\textsuperscript{1,2} This protection is dependent on the ability of HbF to hinder deoxyHbS polymerization. HbF level is variable and highly heritable. Previous genetic association studies found single nucleotide polymorphisms (SNPs) in regions of \textit{BCL11A} (chromosome 2p), in the \textit{HBS1L-MYB} intergenic polymorphism (HMIP; chromosome 6q), and linked to \textit{HBB} (chromosome 11p) that were associated with HbF (reviewed in Akinsheye et al\textsuperscript{3}). Our aim was to perform a meta-analysis of genome-wide association studies (GWAS) to find genetic loci with modest effect sizes that were associated with HbF when a larger sample size was examined.

Common SNPs (585, 563 total) from 7 cohorts totaling 2040 patients were meta-analyzed using the software Meta Analysis Helper (METAL)\textsuperscript{3} with inverse variance method, where effect estimates are weighted in proportion to their precisions.\textsuperscript{4} The 7 cohorts included in the meta-analysis are: Cooperative Study of Sickle Cell Disease (CSSCD: n = 841), Multicenter Study of Hydroxyurea (MUSH: n = 178), Pulmonary Hypertension and the Hypoxic Response in Sickle Cell Disease (PUSH) study (n = 73), Comprehensive Sickle Cell Centers Collaborative Data (C-data) project (n = 127), Treatment of Pulmonary Hypertension and Sickle Cell Disease with Sildenafil Treatment (Walk-PHaSST) trial (n = 181), Duke University Outcome Modifying Genes study (n = 152), and Silent Infarct Transfusion (SIT) trial (n = 488).

In each of these studies, patients with the HbS only phenotype, aged 5 years or more when HbF was measured were included. None of the patients were treated with hydroxyurea when HbF was measured. The association between HbF and the genotype for each SNP was tested in a multiple linear regression analysis, adjusting for sex and the top 10 principal components, where appropriate. The additive genetic model, which codes the SNP genotype as the number of minor alleles (0,1,2), was assumed. To obtain a normal distribution of the phenotype for analysis the cubic root of HbF was used as the response variable in the regression analysis, as previously established.\textsuperscript{5}

Table 1 summarizes the results of the meta-analysis for the SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$). Consistent with previous findings, the most significant SNPs were in \textit{BCL11A} (lowest $P$ value of $5.36 \times 10^{-58}$ for rs766432). In addition, 3 SNPs in HMIP reached genome-wide significance. This gene is situated in the interval between the gene \textit{HBS1L} (a G-protein/elongation factor) and the \textit{MYB} an erythroid transcription factor, on chromosome 6q23.3, and was previously shown to be associated with HbF in a microarray study.\textsuperscript{6} SNPs in \textit{OR51B5}, which were associated with HbF in a previous GWAS,\textsuperscript{7} did not reach genome-wide significance. These SNPs almost reached genome-wide significance in the CSSCD, but were not significant in other cohorts at $\alpha = 0.05$ (supplemental Table 1, available on the Blood Web site; see the Supplemental Materials link at the top of the online letter).

In the CSSCD cohort, the most significant SNPs in \textit{BCL11A} (rs766432) and the HMIP region (rs9494145) explained 11.1% and 3.2% of the phenotypic variability in HbF, respectively, and together explain only 14.7% of the variability. Twelve additional SNPs reached statistical significance of $10^{-5}$ or less, although none reached genome-wide significance (supplemental Table 1). HbF is regulated as a complex trait and the “missing heritability” not detected by GWAS has various explanations. These include unaccounted gene interactions, large undetected insertions and deletions, epigenetic factors, regulation by small RNAs, and multiple rare variants with small effects.\textsuperscript{7} For example, SNP rs2033467 in chromosome 5 (supplemental Table 1) is in a novel region but it did not reach genome-wide significance, although it was statistically significant.

\begin{table}[h]
\centering
\caption{Significant SNPs in the meta-analysis}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|}
\hline
Chromosome & SNP & Base pair positions & Minor allele & MAF & N & Beta & SE & $P$ value & Direction$^*$ & Gene \\
\hline
2 & rs766432 & 60719970 & C & 0.276 & 2038 & 0.236 & 0.0147 & $5.36 \times 10^{-58}$ & ++++++++ & BCL11A \\
2 & rs10195871 & 60720589 & A & 0.313 & 2027 & 0.213 & 0.0144 & $2.142 \times 10^{-50}$ & ++++++++ & BCL11A \\
2 & rs6706648 & 60722040 & A & 0.395 & 2004 & $-0.177$ & 0.0135 & $3.97 \times 10^{-20}$ & ---------- & BCL11A \\
2 & rs6738440 & 60722241 & G & 0.289 & 1976 & $-0.158$ & 0.0159 & $2.945 \times 10^{-23}$ & ---------- & BCL11A \\
2 & rs6709302 & 60726729 & A & 0.320 & 1982 & $-0.142$ & 0.0147 & $3.416 \times 10^{-22}$ & ---------- & BCL11A \\
2 & rs6732518 & 60708997 & G & 0.323 & 2035 & $0.133$ & 0.0156 & $2.197 \times 10^{-17}$ & ++++++++ & BCL11A \\
6 & rs4941145 & 135432552 & G & 0.070 & 2039 & 0.217 & 0.0258 & $4.321 \times 10^{-10}$ & ++++++++ & HMIP \\
6 & rs93919137 & 135419018 & G & 0.060 & 2027 & 0.236 & 0.0294 & $1.172 \times 10^{-15}$ & ++++++++ & HMIP \\
6 & rs4895441 & 135426573 & G & 0.089 & 2025 & 0.187 & 0.0236 & $2.228 \times 10^{-15}$ & ++++++++ & HMIP \\
2 & rs10184550 & 60729294 & G & 0.319 & 2040 & 0.092 & 0.0147 & $3.831 \times 10^{-10}$ & ++++++++ & BCL11A \\
2 & rs12477097 & 60683937 & C & 0.364 & 1976 & $-0.079$ & 0.0144 & $4.451 \times 10^{-08}$ & ---------- & BCL11A \\
\hline
\end{tabular}
\end{table}

SNPs with $P$ values $< 5 \times 10^{-8}$ are shown.

$^*$Indicates the direction of association (positive or negative beta estimate) between HbF and single SNPs in each cohort. The order of cohorts compiled in the meta-analysis is as follows: (1) Cooperative Study of Sickle Cell Disease; (2) Pulmonary Hypertension and the Hypoxic Response in Sickle Cell Disease; (3) Multicenter Study of Hydroxyurea; (4) Comprehensive Sickle Cell Centers Collaborative Data project; (5) Treatment of Pulmonary Hypertension and Sickle Cell Disease with Sildenafil Treatment; (6) Duke University Pulmonary Hypertension study; and (7) Silent Infarct Transfusion (SIT) trial.
The online version of this letter contains a data supplement.

Acknowledgments: This work was supported by National Institutes of Health grants R01 HL87681 (M.H.S.), RC2 HL101212 (M.H.S.), HL079915 (M.T.), HL68959 (M.T.), 2R2S HL003679-8 (V.G.), R01 HL079912 (V.G.), 2M01 RR10284-10 (V.G.), 5-U01-NS042804-07 (M.R.D.), CIDR HHSN268200782096C, U54 HL090515 (J.F.C.).

The SickleGen Consortium included: Boston University (H.B., C.T.B., M.H.S., and P.S.); Duke University (M.J.T., A.A.-K., and M.G.); Vanderbilt University and CDC (W.C.H., C.J.B., and M.R.D. for the SIT Study, NCT00072761); Johns Hopkins University (D.E.A., P.B., J.F.C., J.R.K., and E.B.-C. for the SIT Study); PUSH Study (V.G., G.J.K., C.M., J.T., A.C., and L.L.-J. NCT00495638); Children’s Hospital Oakland (C.H. representing the Sickle Cell Centers C-Data Project); and The University of Pittsburgh (M.T.G. and Y.Z. representing the Walk-PHAAST investigators NCT00492531; supplemental Table 2).

Conflict-of-interest disclosure: J.F.C. is a consultant to Adventrx Corporation and has received an honorarium and travel expenses from Adventrx Corporation for assisting them with a possible clinical trial of an agent to treat vasooclusive crisis in sickle cell disease. The remaining authors declare no competing financial interests.

Correspondence: Martin H. Steinberg, Department of Medicine, Boston University School of Medicine, 72 E Concord St, Boston, MA 02118; e-mail: msteinb@bu.edu.

References

Meta-analysis of 2040 sickle cell anemia patients: *BCL11A* and *HBS1L-MYB* are the major modifiers of HbF in African Americans