scientific reports

Check for updates

OPEN

Identification of genetic variants associated with clinical features of sickle cell disease

Katharine Tsukahara[1,6], Xiao Chang[2,6], Frank Mentch[2], Kim Smith-Whitley[3,4], Anita Bhandari[3,4], Cindy Norris[3,4], Joseph T. Glessner[2,5] & Hakon Hakonarson[2,5]✉

Sickle cell disease (SCD) is an inherited blood disorder marked by homozygosity of hemoglobin S, which is a defective hemoglobin caused by a missense mutation in the β-globin gene. However, clinical phenotypes of SCD vary among patients. To investigate genetic variants associated with various clinical phenotypes of SCD, we genotyped DNA samples from 520 SCD subjects and used a genome-wide association study (GWAS) approach to identify genetic variants associated with phenotypic features of SCD. For HbF levels, the previously reported 2p16.1 locus (*BCL11A*) reached genome significance (rs1427407, $P = 8.58 \times 10^{-10}$) in our GWAS as expected. In addition, we found a new genome-wide significance locus at 15q14 (rs8182015, $P = 2.07 \times 10^{-8}$) near gene *EMC7*. GWAS of acute chest syndrome (ACS) detected a locus (rs79915189, $P = 3.70 \times 10^{-8}$) near gene *IDH2* at 15q26.1. The SNP, rs79915189, is also an expression quantitative trait locus (eQTL) of *IDH2* in multiple tissues. For vasoocclusive episode (VOE), GWAS detected multiple significant signals at 2p25.1 (rs62118798, $P = 4.27 \times 10^{-8}$), 15q26.1 (rs62020555, $P = 2.04 \times 10^{-9}$) and 15q26.3 (rs117797325, $P = 4.63 \times 10^{-8}$). Our findings provide novel insights into the genetic mechanisms of SCD suggesting that common genetic variants play an important role in the presentation of the clinical phenotypes of patients with SCD.

Keywords Genome-wide association study, Sickle cell anemia, Hemoglobinopathies, Fetal hemoglobin

Abbreviations

ACS	Acute chest syndrome
EHR	Electronic health record
eQTL	Expression quantitative trait locus
GWAS	Genome-wide association study
HbF	Hemoglobin F, fetal hemoglobin
HbSS	Hemoglobin SS
SCA	Sickle cell anemia
SCD	Sickle cell disease
SNP	Single nucleotide polymorphism
VOE	Vaso-occlusive episodes

Sickle cell disease (SCD) is a significant public health problem in the United States as well as globally. An estimated 100,000 patients with sickle cell disease live in the United States. These patients face reduced life expectancy by about 30 years and decreased quality of life as a result of the multisystem and long-term effects of this devastating disease[1]. Sickle cell anemia (SCA) is caused by homozygosity of hemoglobin S, a result of a missense mutation in the β-globin gene that substitutes valine for glutamine at the sixth amino acid in the β-globin hemoglobin chain[1]. Among children with same disease-causing mutation, clinical phenotypes vary[1]. Common presentations include acute pain events such as vaso-occlusive episodes (VOE) and dactylitis, which confer significant morbidity and are associated with lower quality of life across several domains[2]. Another acute

[1]Division of Pulmonary and Sleep Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [2]The Center for Applied Genomics, Children's Hospital of Philadelphia, Research Institute, Leonard Madlyn Abramson Research Center, Suite 1216E, 3615 Civic Center Blvd, Philadelphia, PA 19104, USA. [3]Division of Hematology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [4]Division of Pulmonary and Sleep Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [5]Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA. [6]These authors contributed equally: Katharine Tsukahara and Xiao Chang. ✉email: hakonarson@chop.edu

presentation of SCD is acute chest syndrome (ACS), defined as fever and respiratory symptoms with new infiltrate on chest radiograph, which is associated with high risk of morbidity and mortality[3]. Among patients, some will never have an episode of ACS, while others will have frequent episodes[4]. This variability is likely due to genetic and non-genetic factors, of which current knowledge is limited.

A few protective factors have been identified, including certain subtypes of SCD and higher levels of fetal hemoglobin (HbF)[5]. Higher HbF levels in erythrocytes decrease the hemoglobin polymerization that leads to erythrocyte "sickling" and complications of SCD[6]. Higher HbF levels have been associated with decreased rates of ACS and VOE, although rates of priapism, stroke and other complications are minimally affected[7]. To investigate the genetic mechanism of HbF levels, multiple genome-wide association studies (GWAS) have been conducted, which are instrumental in discovering genetic associations with a wide variety of clinical conditions, including metabolic, autoimmune, and psychiatric diseases[8]. Uda et al. discovered that *BCL11A* is associated with persistent HbF levels and amelioration of the phenotype of β-thalassemia[9]. Galarneau et al. further reported that polymorphisms at three loci, *BCL11A*, *HBS1L-MYB*, and β-globin, account for approximately 50% of the variation in HbF levels[10]. However, studies to identify underlying genetic drivers of other clinical features such as ACS and vaso-occlusive crises have had limited results[11,12].

In this study, GWAS was performed on a cohort of patients with hemoglobin SS (HbSS) followed by the Children's Hospital of Philadelphia Comprehensive Sickle Cell Center. To better understand the genetic basis of this disease, we planned to study the entire genome in search for new candidate genes conferring increased risk of clinical features of SCA. Our aims were to 1) determine whether variations in the human genome associate with clinical features including ACS, VOE and HbF in children with HbSS, and 2) explore the potential biological mechanisms underlying ACS by identification of genomic variants.

Methods

Study participants

Our study population consists exclusively of patients with HbSS. We conducted a genome-wide association study (GWAS) to investigate the associations between genotype and various clinical phenotypes, including acute chest syndrome (ACS), vaso-occlusive episodes (VOE), hemoglobin levels, pain, and hospital admissions.

Subjects with HbSS were identified from the biorepository at the Center for Applied Genomics (CAG) at the Children's Hospital of Philadelphia. DNA samples of participants were collected at recruitment. The research was performed in accordance with the Declaration of Helsinki. The Children's Hospital of Philadelphia Institutional Review Board approved the study (IRB 16-013,278). Written consent was obtained from the legal guardian(s) of each participant and assent from any child 7 years and older.

Inclusion criteria were patients with at least one encounter at the Children's Hospital of Philadelphia, with documentation of HbSS. Phenotyping was performed by manual chart review and electronic health record (EHR) data analysis for EHR data between January 2007 and December 2020, since the current EHR system was not routinely used prior to 2007. Analysis of clinical data was performed using R. Clinical factors are listed in Table 1. ACS was defined as fever and/or respiratory symptoms with a new infiltrate on chest radiograph. Vaso-occlusive

Characteristic	N = 520[1]
Demographics	
Age at first encounter	1.5 (0.6, 7.2)
Age at last encounter	17 (11, 22)
Patient gender	
Female	238 (48%)
Male	258 (52%)
Patient ancestry	
Black or African American	369 (97%)
Other	8 (2.1%)
White	2 (0.5%)
Declined to answer	1 (0.3%)
Medical history	
Asthma	197 (39%)
Past or current inhaled corticosteroid use	135 (26%)
Past or current hydroxyurea use	410 (80%)
Past or current chronic transfusion	145 (28%)
Acute chest syndrome rate, per year	0.09 (0.00, 0.24)
Vaso-occlusive episode rate, per year	0.50 (0.15, 1.33)
Pain episode rate, per year	0.52 (0.16, 1.36)
Admissions, per year	1.14 (0.58, 2.11)
[1]Median (IQR); n (%)	

Table 1. Demographic and clinical characteristics of the study cohort.

episodes were defined as acute pain without alternative identified cause, e.g., trauma or constipation. Manual chart review identified episodes of ACS, VOE, and dactylitis requiring an emergency room (ER) visit or hospital admission. The number of ACS episodes per patient were compared to EHR codes, and outliers were manually reviewed. International Classification of Disease 9th (ICD9) and 10th (ICD10) codes for ACS and pneumonia. Hemoglobin F (HbF) levels were excluded within 120 days following blood transfusion or following the initiation of hydroxyurea, which is known to increase HbF expression[5]. HbF levels stabilize between the ages 3–5 years[13], thus the most recent HbF level was used for analysis to best represent baseline HbF.

Genotyping, imputation, and association analysis

The discovery and replication cohorts were genotyped using the Illumina Infinium Global Screening (GSA) and HumanHap550/610 single nucleotide polymorphism (SNP) arrays respectively. EIGENSTRAT was used to detect potential substructures and outliers[14]. Participants with self-reported African ancestry were further confirmed by comparing principal component analysis results of participants and reference populations from Hapmap3 (Figure S1). Samples with chip-wide genotyping failure rate greater than 5% were excluded. SNP markers with minor allele frequencies less than 1%, genotyping failure rates greater that 2%, and Hardy–Weinberg *P* values less than 1×10^{-6} were removed before genotype imputation. Pairwise identity-by-descent values were calculated by PLINK to remove cryptic relatedness and duplicated samples[15]. Genotype imputation was performed with the TOPMed Imputation Server using minimac4 imputation algorithm[16]. The whole genome sequencing data from the Trans-Omics for Precision Medicine (TOPMed) program were used as the imputation reference panel, which achieved a significant improvement in imputation qualities and accuracies of variants from African populations[17]. Common variants (minor allele frequencies > 1%) with high imputation confidence (Rsq > 0.3) were retained for association analysis. Association analyses were performed using logistic regression with an additive model on the imputed dosage of the effect allele while adjusting for sex and the first five PCs. Meta-analysis was performed by PLINK. Fixed-effects P values were reported. Although no significant genomic inflation was detected (Figure S2), we calculated the genomic inflation factor (λ) to ensure the accuracy of our findings. Genomic control correction was applied by dividing each observed chi-squared value by λ, and recalculating the *P*-values from these corrected chi-squared values. This adjustment helps to reduce false-positive associations by correcting for any systematic bias in the data. The corrected *P*-values are included in Table 1 and were used in all subsequent analyses. To enhance the reproducibility and transparency of our research, we have deposited the summary statistics of our results in a public database. The data can be accessed at the following link: https://zenodo.org/records/11948788.

Gene-based association and enrichment analysis

We performed gene-based association analysis using MAGMA (Multi-marker Analysis of GenoMic Annotation), a tool designed for integrating GWAS data to identify genes associated with disease traits[18]. This approach accounts for gene size, linkage disequilibrium patterns, and SNP localization within genes, providing a comprehensive assessment of gene-level associations from our GWAS summary statistics. Subsequently, genes with significant associations ($P < 0.01$) were selected for further analysis. To elucidate the biological implications of these significant genes, we conducted enrichment analysis using DAVID (Database for Annotation, Visualization and Integrated Discovery)[19]. This step allowed us to identify enriched biological processes, pathways, and functional categories, offering deeper insights into the biological mechanisms potentially driving the disease pathogenesis and highlighting possible therapeutic targets. This combined analysis ensures a robust exploration of genetic findings and their functional relevance to the disease.

Results

Cohort characteristics are shown in Table 1. Complete data are available for 520 subjects. The median age of first EHR encounter is 1 year. EHR data are available for each patient over a median period of 11 years (min 1 year, max 14 years). Among all subjects, 38.6% have a diagnosis of asthma at any point, and 26.5% have been prescribed an inhaled corticosteroid. Hydroxyurea was prescribed at least once for 80.4% of the cohort. 91.5% have received a blood transfusion. Six patients have undergone stem cell transplant, of which one underwent in utero haploidentical stem cell transplantation with no evidence of engraftment.

Subjects have a mean of 0.16 episodes of ACS per year. In this regard, 306 patients have at least 1 episode of ACS. For pain episodes, subjects have a mean of 0.98 episodes per year of VOE requiring admission or ER visit. When including dactylitis with pain episodes, subjects have a mean of 1.01 episodes per year. Overall, subjects have a mean of 1.58 total admissions per year for any cause. HbF levels meeting the above criteria are available for 467 subjects. The mean HbF level is 15.25%. As previously described, there is a downtrend in HbF levels over the first 2 years of life, followed by stability.

The discovery cohort and the replication cohort included 391 and 129 individuals, respectively. GWAS analysis was performed for ACS, VOE, and pain (defined as VOE or dactylitis) episodes per year, total number of admissions per year, and HbF levels. Since no genome-wide significant signal was detected in either cohort alone, we then performed a meta-analysis combining the two studies. Table 2 lists the clinical outcomes of interest reporting statistically significant loci. Corresponding regional plots and Manhattan plots are shown in Fig. 1 and S3 and Figure S4.

To understand the possible functional role of significant SNPs, we queried the Broad Institute's HaploReg v4.1 database (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php)[20]. HaploReg annotations indicated the tested SNPs were highly associated in regulatory regions of the genome including enhancer activities, promoter histone markers, DNAse I hypersensitive regions and regulatory motifs (Table S1).

HbF levels showed a significant association with SNPs centered at two loci. As expected, the previously reported 2p16.1 locus[21–23] demonstrated the highest significance. There are two SNPs reaching genome-wide

Trait	Locus	Gene	SNP	EA/NEA	$BETA_{dis}$	P_{dis}	$BETA_{rep}$	P_{rep}	$BETA_{all}$	P_{all}	$P_{correct}$
HbF	2p16.1	*BCL11A*	rs1427407	T/G	4.28	6.06E−04	4.08	9.55E−07	4.14	8.58E−10	8.84E−10
			rs766432	C/A	3.32	3.88E−03	4.11	9.06E−07	3.83	6.72E−09	6.92E−09
	15q14	*EMC7*	rs8182015	C/G	7.23	1.46E−04	5.11	6.02E−05	5.80	2.07E−08	2.13E−08
Acute chest syndrome	15q26.1	*IDH2*	rs79915189	C/G	0.43	2.36E−04	0.25	3.02E−05	0.28	3.70E−08	3.85E−08
Pain	2p25.1	*ASAP2*	rs62118798	A/G	0.81	4.90E−03	0.58	3.35E−06	0.62	3.99E−08	4.19E−08
	5q35.1	*SLIT3*	rs71605708	C/T	0.31	0.12	0.56	1.62E−07	0.51	4.28E−08	4.49E−08
	15q26.1	*ZNF710*	rs62020555	C/T	2.45	9.31E−05	2.31	3.76E−06	2.37	4.64E−10	4.87E−10
	15q26.3	*LRRK1*	rs117797325	T/C	2.66	1.83E−02	1.95	4.95E−07	2.03	1.78E−08	1.87E−08
Vaso-occlusive episodes	2p25.1	*ASAP2*	rs62118798	A/G	0.81	4.83E−03	0.56	3.47E−06	0.60	4.27E−08	4.48E−08
	15q26.1	*ZNF710*	rs62020555	C/T	2.46	7.91E−05	2.09	1.67E−05	2.23	2.04E−09	2.14E−09
	15q26.3	*LRRK1*	rs117797325	T/C	2.66	1.74E−02	1.83	1.18E−06	1.91	4.63E−08	4.86E−08
Admission	17q21.31	*ETV4*	rs76714145	A/G	2.24	4.59E−04	1.43	9.72E−06	1.60	1.59E−08	1.64E−08

Table 2. Association results of detected risk variants. *SNP* single nucleotide polymorphism, *EA* effect allele, *NEA* non-effect allele.

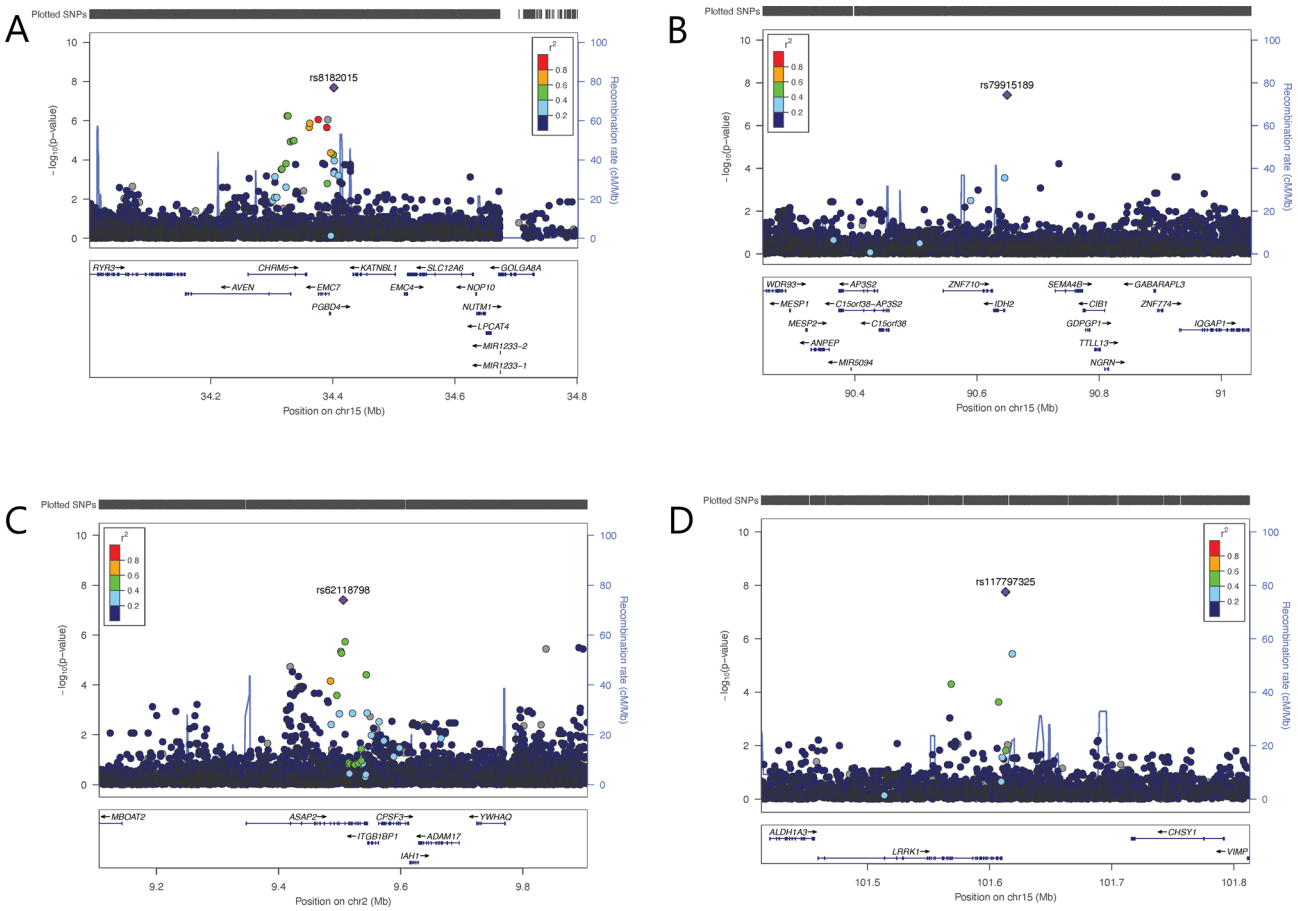

Fig. 1. Regional plots of the four genome-wide significant loci associated with clinical phenotypes of SCD using LocusZoom[42] listed in the GTEx portal. Purple diamond indicates the most significantly associated SNP, and circles represent the other SNPs in the region, with coloring from blue to red corresponding to r^2 values from 0 to 1 with the index SNP. (A) Hemoglobin F, 15q14; (B) acute chest syndrome: 15q26.1; (C) pain, 2p25.1; (D) pain, 15q26.3. All images in the figures were generated using a combination of R and PowerPoint. The data visualizations and analyses were conducted using R (version 4.3.1), which can be accessed at https://www.r-project.org/. The final assembly and labeling of the figures were completed using Microsoft PowerPoint (Microsoft 365 version).

significance at 2p16.1 (rs1427407, $p = 8.58 \times 10^{-10}$, and rs766432, $p = 6.72 \times 10^{-9}$, Table 2). Both are located within the *BCL11A* gene, and the linkage disequilibrium LD between these two SNPs is $R^2 = 0.79$. This strong LD suggests that these SNPs are not independent and likely represent a single signal. In addition, we found a new genome-wide significance locus at 15q14 (rs8182015, $p = 2.07 \times 10^{-8}$, Table 2), which is close to gene, *EMC7*. Interestingly, the GTEx Portal indicated that rs8182015 is an expression quantitative trait locus (eQTL) of other nearby genes including *CHRM5* and *GOLGA8A* (Table 3).

For ACS, one significant locus was identified at 15q26.1. The lead SNP, rs79915189 ($p = 3.70 \times 10^{-8}$, Table 2), is located 3649 bp upstream of the *IDH2* gene, which encodes an isocitrate dehydrogenase expressed in mitochondria[24]. According to GTEx Portal, rs79915189, is an eQTL of nearby genes including *IDH2*, *CIB1*, *ZNF774* (Table 3).

Episodes of VOE were most strongly associated with the 15q26.1 locus (rs62020555, $p = 2.04 \times 10^{-9}$, Table 2). The lead SNP is located approximately 3 kb upstream of *ZNF710*, a zinc finger protein. Other significant loci associated with VOE rate were rs117797325 ($p = 4.63 \times 10^{-8}$, Table 2), approximately 3 kb downstream of *LRRK1*, and rs62118798 ($p = 4.27 \times 10^{-8}$, Table 2), within the *ASAP2* gene. The inclusion of dactylitis episodes, a pain presentation more common in younger children, did not significantly change results. When including dactylitis in addition to VOE as a pain outcome, we found one additional locus met the threshold for significance (rs71605708, $p = 4.28 \times 10^{-8}$, Table 2). The rate of hospital admissions, a strong surrogate marker for overall morbidity, was associated only with locus17q21.31. The lead SNP is 23 kb upstream of a transcription factor coded gene *ETV4*.

Our gene-based association and enrichment analysis further illustrates a significant role for immune response-related pathways, particularly those involving MHC class II protein complexes, across different clinical traits in HbSS patients (Table S2). We found marked enrichment for processes associated with antigen processing and presentation, immune system regulation, and cellular transport mechanisms in traits such as HbF, ACS, pain, and

Locus	SNP	Gene	*P*-Value	NES	Tissue
15q14	rs8182015	CHRM5	3.00E−11	0.76	Nerve—Tibial
		CHRM5	0.0000032	0.53	Breast—Mammary Tissue
		CHRM5	0.0000063	0.71	Stomach
		CHRM5	0.000016	0.83	Brain—Cortex
		CHRM5	0.00004	0.61	Artery—Aorta
		CHRM5	0.000043	0.44	Thyroid
		GOLGA8A	0.00002	−0.37	Brain—Frontal Cortex (BA9)
		GOLGA8A	0.000022	−0.36	Artery—Aorta
15q26.1	rs79915189	CIB1	7.80E−07	−0.24	Whole Blood
		CIB1	0.000018	−0.25	Lung
		CIB1	0.000036	−0.37	Brain—Cerebellar Hemisphere
		CIB1	0.000099	−0.19	Muscle—Skeletal
		IDH2	0.000095	0.25	Skin—Not Sun Exposed (Suprapubic)
		IDH2	0.00012	0.41	Esophagus—Mucosa
		ZNF774	1.80E−07	−0.3	Artery—Tibial
2p25.1	rs62118798	ADAM17	0.0000014	0.37	Brain—Cortex
		ASAP2	0.000018	−0.13	Nerve—Tibial
		CPSF3	4.50E−11	−0.22	Nerve—Tibial
		CPSF3	8.50E−09	−0.2	Skin—Sun Exposed (Lower leg)
		CPSF3	3.60E−08	−0.21	Adipose—Subcutaneous
		CPSF3	3.10E−07	−0.22	Skin—Not Sun Exposed (Suprapubic)
		CPSF3	0.0000021	−0.14	Thyroid
		IAH1	2.20E−12	0.24	Skin—Sun Exposed (Lower leg)
		IAH1	1.80E−09	0.24	Skin—Not Sun Exposed (Suprapubic)
		IAH1	5.50E−09	0.4	Pancreas
		IAH1	1.40E−08	0.21	Thyroid
		IAH1	0.000031	0.18	Colon—Transverse
		ITGB1BP1	6.10E−09	0.31	Muscle—Skeletal
		ITGB1BP1	1.30E−07	0.39	Heart—Left Ventricle
		ITGB1BP1	0.000021	0.33	Heart—Atrial Appendage
		ITGB1BP1	0.000035	0.17	Thyroid
		ITGB1BP1	0.000053	0.22	Esophagus—Mucosa
		RP11-400L8.2	0.0000047	0.27	Muscle—Skeletal
15q26.3	rs117797325	RP11-505E24.2	2.10E−09	0.99	Testis

Table 3. Variants associated with gene expression levels according to GTEx portal ***.

VOE. For instance, traits like HbF and ACS showed a pronounced enrichment in MHC class II-related activities, which are critical for immune function. Additionally, pathways involved in the regulation of T cell activation and immune response were notably enriched, highlighting their potential contribution to the pathophysiology of these conditions.

Discussion

There is significant variation in the clinical presentation among patients with sickle cell disease that remains to be elucidated. Here we present a GWAS of a cohort of 520 pediatric patients with HbSS. Significant loci were identified for all outcomes of interest. HbF findings are consistent with prior studies demonstrating *BCL11A* as a driver of HbF levels[21–23]. *BCL11A* is known to decrease HbF production through direct repression of the γ-globin genes and subsequently has a role in the transition from fetal to adult hemoglobin production[25]. More recently, Esrick et al. reported on six patients treated with short hairpin RNA (shRNA) knockdown of *BCL11A*, of which none had an episode of vaso-occlusive pain or ACS following treatment[26]. In addition, Frangoul, et al. reported on CRISPR-Cas9 editing of the *BCL11A* enhancer[27]. This highlights the importance of identifying genetic drivers of severe phenotypes, which may have potential as therapeutic targets. In addition, we detected a new significant locus (15q14) associated with HbF levels. While it is not clear how this locus may influence HbF production, Gtex data indicate that genotypes of the lead SNP rs8182015 are correlated with the expression level of nearby genes suggesting a regulatory role of this SNP. Consistent with this result, HaploReg show that rs8182015 is located in the region of regulatory motifs suggesting that this locus may warrant further investigation.

This study also identified a locus (15q26.1) of interest in ACS. The lead SNP rs79915189 is close to gene *IDH2*. It is also an eQTL of *IDH2* in epithelial tissues (Table 3). Moreover, multiple lines of evidence suggest a potential role of *IDH2* in respiratory function and lung injury[28–31].

Pain episodes were most strongly correlated with a SNP on chromosome 15, located upstream of *ZNF710*. *ZNF710* codes for a zinc finger protein that has a role in transcription regulation and has been implicated in clear cell renal cell carcinoma[32], but the functional role of *ZNF710* in pain episodes is not clear.

LRRK1 on chromosome 15 codes for leucine rich repeat kinase 1. Several studies have reported an association with bone disease. Homozygous mutations in the *LRRK1* gene can lead to osteosclerotic metaphyseal dysplasia, a disease characterized by skeletal dysplasia and abnormal bone density distribution in the long bones, ribs, vertebrae and iliac crests[33–35]. Interestingly, vaso-occlusive pain most commonly localizes to the back, legs, or hips[36–38]. With the inclusion of dactylitis as a pain outcome, a locus associated with *SLIT3* met the significance threshold. *SLIT3* may also play a role in skeletal homeostasis. Mouse studies have found SLIT3 knockout mice had decreased long bone ossification and length[39] and decreased fracture healing due to reduced skeletal angiogenesis[40].

Complications and complication rates vary significantly among patients with SCD, even among those with the same hemoglobin genotype. The identified genetic variants have significant potential to impact clinical management, risk stratification, and personalized treatment for SCD. As demonstrated in prior *BCL11A* studies, understanding the genetic regulation of HbF levels by genes such as *BCL11A* can inform targeted therapies to increase HbF production, thereby reducing the severity of SCD symptoms. Therapies that boost HbF levels could be particularly beneficial for patients with genetic variants associated with lower HbF production, offering a tailored approach to reducing disease complications. The association of ACS with loci near *IDH2* suggests potential pathways for therapeutic intervention in respiratory complications of SCD. Targeting these pathways could mitigate the frequency and severity of ACS episodes, providing a focused strategy for managing this life-threatening complication. Furthermore, identifying loci linked to pain episodes offers valuable insights into the underlying mechanisms of pain in SCD, potentially leading to the development of new pain management and prevention strategies tailored to the genetic profiles of individual patients, enhancing their quality of life. Incorporating these genetic findings into clinical practice could significantly enhance risk assessment and prognosis for patients with SCD. Patients with genetic variants associated with higher HbF levels might have a more favorable prognosis. Conversely, those with variants linked to severe ACS or frequent pain episodes could be identified as high-risk and receive closer monitoring and more aggressive interventions. This personalized approach to treatment based on genetic profiling represents a significant advancement in the management of SCD, potentially improving outcomes and reducing morbidity.

This study has limitations. Being a single-center study conducted at an urban academic center, the generalizability of our findings is limited. To enhance generalizability and identify additional clinically relevant loci, future research should involve multicenter studies with larger and more diverse populations. Additionally, we lacked data on environmental factors, which may contribute to the risk of acute complications in sickle cell disease. Including such data in future studies will help provide a more comprehensive understanding of disease variability. Acute chest syndrome (ACS) and vaso-occlusive episodes (VOE) present further challenges due to the lack of objective diagnostic criteria and the involvement of multiple pathophysiologic processes, which may limit the ability to detect significant loci. Developing and validating more objective diagnostic criteria for ACS and VOE will standardize outcomes and improve the accuracy of future studies. In addition, approximately 9 months of the study period overlapped with the COVID-19 pandemic, during which respiratory viruses, a known trigger of ACS[3], dropped precipitously among children[41]. However, this represented a minority of the study period, thus SCD complication rates were unlikely to be significantly affected. Another limitation of our study is the absence of colocalisation analysis between GWAS and eQTL data. This decision was due to the population differences: our GWAS data is from African populations, while the GTEx eQTL data is primarily European. Conducting colocalisation under these conditions could introduce biases and affect result validity. Future research should aim to use population-matched eQTL datasets to enable accurate colocalisation analyses.

In conclusion, our GWAS on over 500 pediatric patients with HbSS has provided valuable insights into HbF levels and identified several loci associated with various clinical phenotypes in SCD. These genetic variants have important clinical implications, offering opportunities to enhance clinical management, risk assessment, and therapeutic interventions. Understanding the genetic regulation of HbF levels can guide therapies to increase HbF production, reducing SCD severity. Loci associated with acute chest syndrome and pain episodes offer pathways for therapeutic interventions and personalized pain management strategies. Incorporating these findings into clinical practice can improve risk stratification and prognosis for SCD patients, advancing personalized medicine and enhancing patient outcomes. Our study also suggests that future GWAS with larger sample sizes will likely uncover more clinically relevant loci.

Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Received: 30 January 2024; Accepted: 22 August 2024
Published online: 29 August 2024

References

1. Piel, F. B., Steinberg, M. H. & Rees, D. C. Sickle cell disease. *N. Engl. J. Med.* **376**, 1561–1573. https://doi.org/10.1056/NEJMra1510865 (2017).
2. Dampier, C. *et al.* Health-related quality of life in children with sickle cell disease: A report from the Comprehensive Sickle Cell Centers Clinical Trial Consortium. *Pediatr. Blood Cancer* **55**, 485–494. https://doi.org/10.1002/pbc.22497 (2010).
3. Jain, S., Bakshi, N. & Krishnamurti, L. Acute chest syndrome in children with sickle cell disease. *Pediatr. Allergy Immunol. Pulmonol.* **30**, 191–201. https://doi.org/10.1089/ped.2017.0814 (2017).
4. Castro, O. *et al.* The acute chest syndrome in sickle cell disease: Incidence and risk factors: The cooperative study of sickle cell disease. *Blood* **84**, 643–649 (1994).
5. Kato, G. J. *et al.* Sickle cell disease. *Nat. Rev. Dis. Primers* **4**, 18010. https://doi.org/10.1038/nrdp.2018.10 (2018).
6. Poillon, W. N., Kim, B. C. & Castro, O. Intracellular hemoglobin S polymerization and the clinical severity of sickle cell anemia. *Blood* **91**, 1777–1783 (1998).
7. Steinberg, M. H. *et al.* Fetal hemoglobin in sickle cell anemia: A glass half full?. *Blood* **123**, 481–485. https://doi.org/10.1182/blood-2013-09-528067 (2014).
8. Visscher, P. M. *et al.* 10 Years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005 (2017).
9. Uda, M. *et al.* Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc. Natl. Acad. Sci. USA* **105**, 1620–1625. https://doi.org/10.1073/pnas.0711566105 (2008).
10. Galarneau, G. *et al.* Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat. Genet.* **42**, 1049–1051. https://doi.org/10.1038/ng.707 (2010).
11. Klings, E. S. & Steinberg, M. H. Acute chest syndrome of sickle cell disease: Genetics, risk factors, prognosis, and management. *Expert Rev. Hematol.* **15**, 117–125. https://doi.org/10.1080/17474086.2022.2041410 (2022).
12. Chaturvedi, S. *et al.* Genome-wide association study to identify variants associated with acute severe vaso-occlusive pain in sickle cell anemia. *Blood* **130**, 686–688. https://doi.org/10.1182/blood-2017-02-769661 (2017).
13. Ngo, D. A. *et al.* Fetal haemoglobin levels and haematological characteristics of compound heterozygotes for haemoglobin S and deletional hereditary persistence of fetal haemoglobin. *Br. J. Haematol.* **156**, 259–264. https://doi.org/10.1111/j.1365-2141.2011.08916.x (2012).
14. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909. https://doi.org/10.1038/ng1847 (2006).
15. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum Genet.* **81**, 559–575. https://doi.org/10.1086/519795 (2007).
16. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet.* **48**, 1284–1287. https://doi.org/10.1038/ng.3656 (2016).
17. Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.* **15**, e1008500. https://doi.org/10.1371/journal.pgen.1008500 (2019).
18. de Leeuw, C. A. *et al.* MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* **11**, e1004219. https://doi.org/10.1371/journal.pcbi.1004219 (2015).
19. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221. https://doi.org/10.1093/nar/gkac194 (2022).
20. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877-881. https://doi.org/10.1093/nar/gkv1340 (2016).
21. Bhatnagar, P. *et al.* Genome-wide association study identifies genetic variants influencing F-cell levels in sickle-cell patients. *J Hum Genet.* **56**, 316–323. https://doi.org/10.1038/jhg.2011.12 (2011).
22. Mtatiro, S. N. *et al.* Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. *PLOS ONE.* **9**, e111464. https://doi.org/10.1371/journal.pone.0111464 (2014).
23. Solovieff, N. *et al.* Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. *Blood.* **115**, 1815–1822. https://doi.org/10.1182/blood-2009-08-239517 (2010).
24. PubChem. IDH2 - isocitrate dehydrogenase (NADP(+)) 2 (human). National Center for Biotechnology Information, National Library of Medicine. 2022. https://pubchem.ncbi.nlm.nih.gov/gene/IDH2/human (accessed 7 June 2022).
25. Sankaran, V. G. & Orkin, S. H. The switch from fetal to adult hemoglobin. *Cold Spring Harb. Perspect. Med.* **3**, a011643. https://doi.org/10.1101/cshperspect.a011643 (2013).
26. Esrick, E. B. *et al.* Post-transcriptional genetic silencing of BCL11A to treat sickle cell disease. *N. Engl. J. Med.* **384**, 205–215. https://doi.org/10.1056/NEJMoa2029392 (2021).
27. Frangoul, H. *et al.* CRISPR-Cas9 Gene editing for sickle cell disease and β-thalassemia. *N. Engl. J. Med.* **384**, 252–260. https://doi.org/10.1056/NEJMoa2031054 (2021).
28. Han, Y. K. *et al.* Oxidative stress following acute kidney injury causes disruption of lung cell cilia and their release into the bronchoaveolar lavage fluid and lung injury, which are exacerbated by Idh2 deletion. *Redox Biol.* **46**, 102077. https://doi.org/10.1016/j.redox.2021.102077 (2021).

29. Yeung, B. H. Y. *et al.* Role of isocitrate dehydrogenase 2 on dna hydroxymethylation in human airway smooth muscle cells. *Am. J. Respir. Cell Mol. Biol.* **63**, 36–45. https://doi.org/10.1165/rcmb.2019-0323OC (2020).
30. Willis-Owen, S. A. G. *et al.* COPD is accompanied by co-ordinated transcriptional perturbation in the quadriceps affecting the mitochondria and extracellular matrix. *Sci. Rep.* **8**, 12165. https://doi.org/10.1038/s41598-018-29789-6 (2018).
31. Park, J. H. *et al.* Disruption of IDH2 attenuates lipopolysaccharide-induced inflammation and lung injury in an α-ketoglutarate-dependent manner. *Biochem. Biophys. Res. Commun.* **503**, 798–802. https://doi.org/10.1016/j.bbrc.2018.06.078 (2018).
32. Li, G. *et al.* Overexpression of antisense long non-coding RNA ZNF710-AS1-202 promotes cell proliferation and inhibits apoptosis of clear cell renal cell carcinoma via regulation of ZNF710 expression. *Mol. Med. Rep.* **21**, 2502–2512. https://doi.org/10.3892/mmr.2020.11032 (2020).
33. Iida, A. *et al.* Identification of biallelic LRRK1 mutations in osteosclerotic metaphyseal dysplasia and evidence for locus heterogeneity. *J. Med. Genet.* **53**, 568–574. https://doi.org/10.1136/jmedgenet-2016-103756 (2016).
34. Guo, L. *et al.* Identification of a novel LRRK1 mutation in a family with osteosclerotic metaphyseal dysplasia. *J. Hum. Genet.* **62**, 437–441. https://doi.org/10.1038/jhg.2016.136 (2017).
35. Chorin, O. *et al.* Broadening the phenotype of LRRK1 mutations: Features of malignant osteopetrosis and optic nerve atrophy with intrafamilial variable expressivity. *Eur. J. Med. Genet.* **65**, 104383. https://doi.org/10.1016/j.ejmg.2021.104383 (2022).
36. Dampier, C. *et al.* Characteristics of pain managed at home in children and adolescents with sickle cell disease by using diary self-reports. *J. Pain.* **3**, 461–470. https://doi.org/10.1054/jpai.2002.128064 (2002).
37. McClish, D. K. *et al.* Pain site frequency and location in sickle cell disease: The PiSCES project. *Pain.* **145**, 246–251. https://doi.org/10.1016/j.pain.2009.06.029 (2009).
38. Vijenthira, A. *et al.* Benchmarking pain outcomes for children with sickle cell disease hospitalized in a tertiary referral pediatric hospital. *Pain Res. Manag.* **17**, 291–296. https://doi.org/10.1155/2012/614819 (2012).
39. Kim, H. *et al.* SLIT3 regulates endochondral ossification by β-catenin suppression in chondrocytes. *Biochem. Biophys. Res. Commun.* **506**, 847–853. https://doi.org/10.1016/j.bbrc.2018.10.167 (2018).
40. Xu, R. *et al.* Targeting skeletal endothelium to ameliorate bone loss. *Nat. Med.* **24**, 823–833. https://doi.org/10.1038/s41591-018-0020-z (2018).
41. Sherman, A. C. *et al.* The effect of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) mitigation strategies on seasonal respiratory viruses: A tale of 2 large metropolitan centers in the United States. *Clin. Infect. Dis.* **72**, e154–e157. https://doi.org/10.1093/cid/ciaa1704 (2021).
42. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337. https://doi.org/10.1093/bioinformatics/btq419 (2010).

Acknowledgements

Cassandre-Michel Sebastien of Children's Hospital of Philadelphia Research Institute for assistance with electronic health record data collection.

Author contributions

H.H. and K.T. contributed to the study conception and design. Material preparation and data collection were performed by K.S., A.B. and C.N. Data analysis was performed by K.T., X.C. and F.M. The first draft of the manuscript was written by K.T. and X.C. All authors commented on previous versions of the manuscript. H.H. and J.G. supervised the study. All authors read and approved the final manuscript.

Funding

This study was funded by Institutional Development Fund to the CAG (HH), by The Children's Hospital of Philadelphia Endowed Chair in Genomic Research (HH) and by a donation from the Neff Family Foundation (HH). KT is supported by the National Institutes of Health T32 HG009495.

Competing interests

KSW is the Chief Medical Officer of the Sickle Cell Reproductive Health Education Directive, and a board member of the Sickle Cell Disease Association of America, both nonprofit organizations. She is a current employee of Pfizer/Global Blood Therapeutics and has received stock and/or stock options in this role. The remaining authors have no conflicts of interest to disclose.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-70922-5.

Correspondence and requests for materials should be addressed to H.H.