

# Variation and impact of polygenic hematologic traits in monogenic sickle cell disease

Thomas Pincez,<sup>1,2</sup> Ken Sin Lo,<sup>1</sup> Anne-Laure Pham Hung d'Alexandry d'Orengiani,<sup>3</sup> Melanie E. Garrett,<sup>4</sup> Carlo Brugnara,<sup>5</sup> Allison E. Ashley-Koch,<sup>4</sup> Marilyn J. Telen,<sup>6</sup> Frédéric Galactéros,<sup>3</sup> Philippe Joly,<sup>7,8</sup> Pablo Bartolucci<sup>3</sup> and Guillaume Lettre<sup>1,9</sup>

<sup>1</sup>Montreal Heart Institute, Montréal, Québec, Canada; <sup>2</sup>Department of Pediatrics, Division of Pediatric Hematology-Oncology, Charles-Bruneau Cancer Center, CHU Sainte-Justine, Université de Montréal, Montréal, Québec, Canada; <sup>3</sup>Red Cell Genetic Disease Unit, Hôpital Henri-Mondor, Assistance Publique-Hôpitaux de Paris (AP-HP), Université Paris Est, IMRB - U955 - Équipe n. 2, Créteil, France; <sup>4</sup>Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA; <sup>5</sup>Department of Laboratory Medicine, Boston Children's Hospital, Boston, MA, USA; <sup>6</sup>Department of Medicine, Division of Hematology, Duke University Medical Center, Durham, NC, USA; <sup>7</sup>Unité Fonctionnelle 34445 'Biochimie des Pathologies Érythrocytaires', Laboratoire de Biochimie et Biologie Moléculaire Grand-Est, Groupement Hospitalier Est, Hospices Civils de Lyon, Bron, France; <sup>8</sup>Laboratoire Inter-Universitaire de Biologie de la Motricité (LIBM) EA7424, Equipe 'Biologie Vasculaire et du Globule Rouge', Université Claude Bernard Lyon 1, Comité d'Universités et d'Établissements (COMUE), Lyon, France and <sup>9</sup>Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, Canada

**Correspondence:** G. Lettre  
guillaume.lettre@umontreal.ca

**Received:** April 5, 2022.

**Accepted:** September 28, 2022.

**Early view:** October 13, 2022.

<https://doi.org/10.3324/haematol.2022.281180>

©2023 Ferrata Storti Foundation

Published under a CC BY-NC license



## Abstract

Several of the complications observed in sickle cell disease (SCD) are influenced by variation in hematologic traits (HT), such as fetal hemoglobin (HbF) level and neutrophil count. Previous large-scale genome-wide association studies carried out in largely healthy individuals have identified thousands of variants associated with HT, which have then been used to develop multi-ancestry polygenic trait scores (PTS). Here, we tested whether these PTS associate with HT in SCD patients and if they can improve statistical models associated with SCD-related complications. In 2,056 SCD patients, we found that the PTS predicted less HT variance than in non-SCD individuals of African ancestry. This was particularly striking at the Duffy/DARC locus, where we observed an epistatic interaction between the SCD genotype and the Duffy null variant (rs2814778) that led to a two-fold weaker effect on neutrophil count. PTS for these HT which are measured as part of routine practice were not associated with complications in SCD. In contrast, we found that a simple PTS for HbF that includes only six variants explained a large fraction of the phenotypic variation (20.5–27.1%), associated with acute chest syndrome and stroke risk, and improved the statistical modeling of the vaso-occlusive crisis rate. Using Mendelian randomization, we found that increasing HbF by 4.8% reduces stroke risk by 39% ( $P=0.0006$ ). Taken together, our results highlight the importance of validating PTS in large diseased populations before proposing their implementation in the context of precision medicine initiatives.

## Introduction

Sickle cell disease (SCD), the most frequent monogenic disease worldwide, is caused by mutations in the  $\beta$ -globin gene.<sup>1</sup> SCD patients present a wide range of complications such as vaso-occlusive crisis (VOC), acute chest syndrome (ACS), stroke, and end-organ dysfunction, and their life expectancy is reduced when compared to the general population.<sup>1</sup> Critically, the causes of this clinical heterogeneity are not fully understood.

Hematologic traits (HT) are among the main factors known to be associated with clinical outcomes in SCD. Fetal hemoglobin (HbF) is a major disease modifier, and

is associated with a reduction in the occurrence of several complications such as VOC, ACS, and death.<sup>2–4</sup> HbF >30% is associated with an almost complete absence of complications in SCD patients.<sup>5</sup> However, most SCD patients have lower HbF levels while not receiving disease-modifying therapy, and for some complications, such as stroke, the risk reduction associated with HbF has not been quantified in large cohorts. Several other HT have been associated with SCD-related complications, notably elevated white blood cell (WBC) count and neutrophil count with survival,<sup>2,6,7</sup> low hemoglobin (Hb) levels with composite severe outcomes and death,<sup>7,8</sup> and platelet (PLT) count with ACS.<sup>9</sup>

Polygenic trait scores (PTS) have been developed in an effort to harness the power of large-scale human genetic studies to make useful clinical predictions, and many studies have already recognized their value in the context of precision medicine initiatives.<sup>10,11</sup> For each participant, PTS are calculated by adding up the number of phenotype-associated alleles across associated variants, each weighted by the effect of the alleles on the phenotype. One widely discussed limitation of PTS is their poor performance when tested in populations that have different ancestral backgrounds than those in which they were optimized.<sup>12</sup> Another equally important aspect that has not been as extensively studied is how well PTS, which are normally calibrated in “healthy” individuals, perform in “diseased” individuals.<sup>13</sup> This is important, because PTS could, in theory, be useful to stratify patients into mild or severe categories. For example, if higher WBC count is associated with an increased risk of death in SCD patients, would a PTS developed to capture WBC count variation in healthy individuals be useful to predict death in SCD patients?

Here, we explored how SCD impacts the performance of HT PTS, and whether these PTS are clinically useful predictors of SCD-related complications. Our analyses had four aims. First, to study the performance of HT PTS in explaining HT variation in SCD patients. Second, to test if these HT PTS are associated with SCD-related complications. Third, to explore whether specific genetic variants included in the HT PTS have reduced impact (i.e., effect size) on HT variation in the context of SCD. Finally, although not one of the main goals of our study, we also

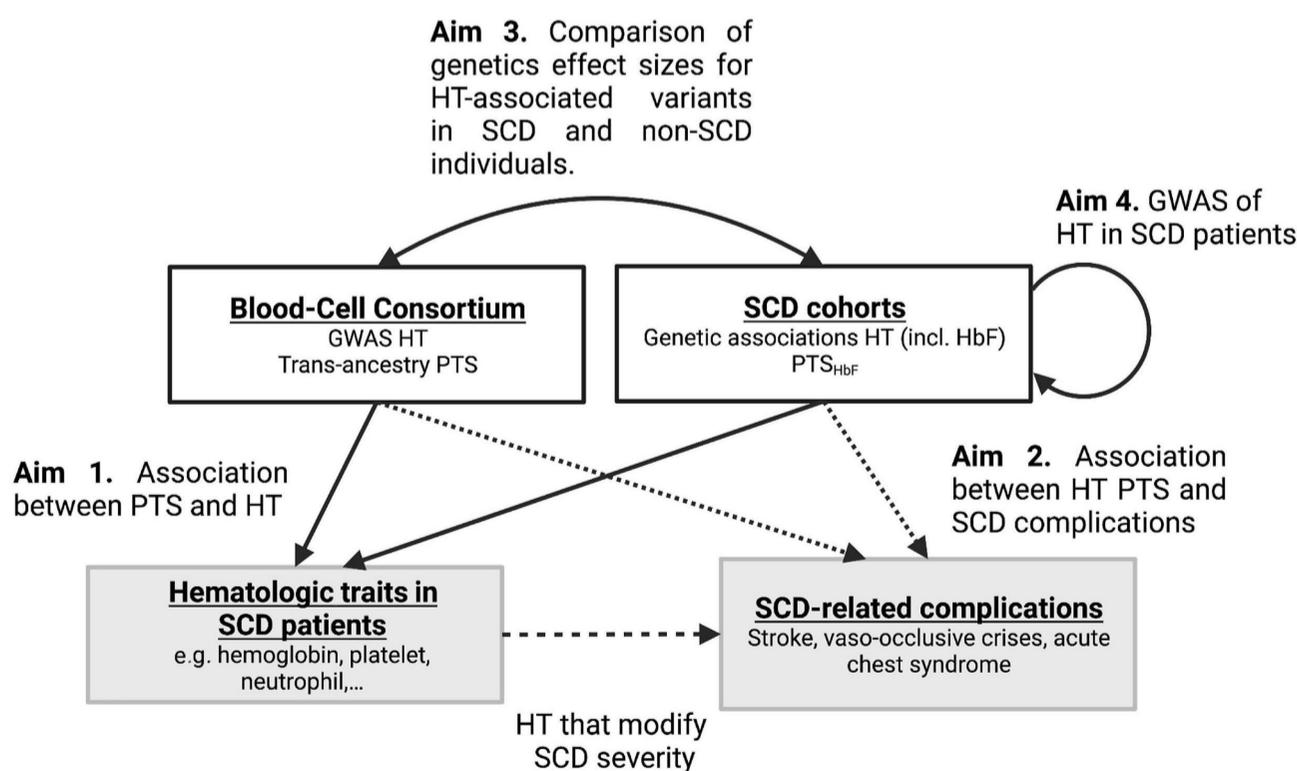
aimed to carry out genome-wide association studies (GWAS) of HT in up to 1,736 SCD patients to identify strong effect variants that could influence blood-cell phenotypes in this patient population. A summary of the study design is shown in Figure 1.

## Methods

Complete details of the methods used are available in the *Online Supplementary Appendix*.

### Study populations

We collected data from three SCD cohorts with SS genotype individuals: the Cooperative Study of Sickle Cell Disease (CSSCD, n=1,278),<sup>14</sup> Genetic Modifier (GEN-MOD, n=406),<sup>15,16</sup> and Mondor/Lyon (n=372)<sup>17</sup> (*Online Supplementary Table S1*). For replication, we tested associations in the Duke University Outcome Modifying Gene (OMG) SCD cohort (n=333), which has been described elsewhere.<sup>18</sup> We collected data according to the principles of the Declaration of Helsinki and the study was approved by the institutional ethics committees. Informed consent was provided by all study participants. For comparison, we also accessed data from non-SCD individuals of African ancestry from the BioMe cohort<sup>19</sup> and the UK Biobank (*Online Supplementary Table S1*).<sup>20</sup> To ensure that the differences found were not due to a difference in the sample size between the SCD cohorts and the UK Biobank, we down-sampled the African-ancestry UK Biobank cohort to



**Figure 1. Study design and aims.** We tested if polygenic trait scores (PTS) developed in non-sickle cell disease (SCD) individuals associate with hematologic traits (HT) (Aim 1) and clinical outcomes (Aim 2) in SCD patients. We also tested if genetic variants associated with HT have different effect sizes when comparing non-SCD and SCD individuals (Aim 3). Finally, we performed genome-wide association studies (GWAS) to identify genetic variants associated with HT in SCD patients (Aim 4). HbF: fetal hemoglobin.

the same number of participants (n=1,278) as in the CSSCD and repeated our analyses.

### Polygenic trait scores for hematologic traits

For all HT except HbF, we used the multi-ancestry PTS obtained by the Blood-Cell Consortium to test for association with HT.<sup>19</sup> For HbF, we derived a PTS by considering the conditional effect sizes of six variants at three loci (*BCL11A*, *HBS1L-MYB*, *HBB*) associated with HbF levels in SCD patients (*Online Supplementary Table S2*).<sup>21-23</sup> We calculated a bootstrapped, empirical *P*-value to compare the variance explained by PTS in SCD and non-SCD individuals. We also tested *g*(HbF), a previously published 4-SNP PTS for HbF,<sup>24</sup> in a subset of the CSSCD cohort (n=816). Finally, we analyzed the respective contribution of  $\alpha$ -thalassemia and PTS on HT variance explained in the CSSCD cohort.

### Association between hematologic traits or polygenic trait scores with SCD-related clinical outcomes

In the CSSCD cohort, we tested the association between PTS and outcomes (VOC rate, ACS rate, stroke), considering only PTS for which the corresponding HT was associated with the outcome (*P*<0.05). We performed an analysis of deviance to determine if the PTS improves the model beyond the measured HT.

### Mendelian randomization

We used a two-sample Mendelian randomization (MR) approach to test if HbF captured by the 6-SNP PTS<sub>HbF</sub>, causally impact SCD-related complications (VOC rate, ACS rate and stroke). We calculated instrument (i.e., PTS for HbF)-exposure (i.e., HbF) and instrument-outcome (i.e., complications) effects from the GEN-MOD and CSSCD cohorts, respectively. We used the multiplicative random-effect inverse variance-weighted (IVW) approach as the main method for each MR analysis.

### Genome-wide association studies of hematologic traits in SCD patients

We performed GWAS for each HT available in the three SCD cohorts separately, then performed a meta-analysis of the results.

### Comparing effect sizes of hematologic trait-associated single nucleotide polymorphism in SCD patients and non-SCD individuals

For each single nucleotide polymorphism (SNP)-HT pair considered in the multi-ancestry PTS models, we used the *t*-statistic to compare the effect in SCD (derived from SCD GWAS meta-analyses) and non-SCD individuals (derived from published multi-ancestry meta-analyses from the Blood-Cell Consortium).<sup>19</sup> We computed *q*-value with a 5% false discovery rate to correct for multiple testing.

## Results

### Performance of hematologic polygenic trait scores in SCD patients

We investigated the phenotypic variance explained by PTS of HT in SCD patients of African ancestry from three cohorts (CSSCD, GEN-MOD, Mondor/Lyon) and in non-SCD African-American individuals from BioMe not included in the discovery effort used to generate the PTS, as well as non-SCD individuals of African ancestry from the UK Biobank (using all African-ancestry participants [n=6,627] or a random subset with a sample size similar to the CSSCD [n=1,278]). In SCD participants, PTS reached significance (*P*<0.05) in at least one cohort for nine of the 12 HT tested (Table 1); this number remained unchanged after adjustment for multiple testing. PTS for hematocrit (Ht), Hb concentration, and lymphocyte count were not significant. The variance explained by significant PTS was 0.5-4.1% for RBC traits, 0.9-3.8% for WBC traits, and 0.7-4.2% for PLT traits. When we compared the performance of these PTS in SCD participants and non-SCD African-ancestry individuals, we found that all PTS with significant association, except the PTS for eosinophils, explained less phenotypic variance in SCD participants (Table 1, Figure 2A). One of the most striking differences was seen for WBC and neutrophil counts: the mean variance explained was 1.9% and 3.3% in SCD participants and 10.3% and 11.8% in non-SCD individuals, respectively (Figure 2A). HbF levels are an important modifier of severity in SCD. Since it is rarely measured in large non-SCD cohorts, the genetics of this trait have not been extensively studied in very large sample sizes. However, smaller GWAS in SCD patients have identified robust associations between HbF levels and genetic variants at three loci: *BCL11A*, *HBS1L-MYB* and the  $\beta$ -globin locus (reviewed by Lettre *et al.*<sup>25</sup>). Given this, we derived PTS<sub>HbF</sub> that includes six conditionally independent variants (*Online Supplementary Table S2*). PTS<sub>HbF</sub> was strongly associated with HbF levels in all three SCD cohorts and explained 20.5-27.1% of the variance (Table 1). We compared this PTS with a previously published 4-SNP model for HbF (*g*(HbF)).<sup>24</sup> We used the subset of the CSSCD cohort (n=816) with genotyping or imputation data available for each SNP. Our 6-SNP PTS<sub>HbF</sub> explained 24.6% of HbF variance (*P*=5.7x10<sup>-51</sup>) whereas *g*(HbF) explained 20.8% of HbF variance (*P*=1.5x10<sup>-42</sup>).

Finally, we analyzed the respective contribution of  $\alpha$ -thalassemia and PTS on HT variance explained in the CSSCD cohort (Figure 2B). Because HbF is such a strong modifier of SCD phenotype, we also assessed whether PTS<sub>HbF</sub> could contribute to the phenotypic variation of other HT.  $\alpha$ -thalassemia was the main contributor of mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) but was also associated to other RBC, WBC, and PLT traits. PTS<sub>HbF</sub> explained a large fraction of several RBC traits, but was also associated with several WBC traits; this is consist-

ent with the known beneficial effect of HbF in normalizing HT in SCD patients. PTS for the corresponding HT was the main contributor for several WBC traits and PLT count.

Taken together, although most of the PTS for HT derived in non-SCD individuals were associated with HT in SCD individuals, they did not perform as well in this patient population. By contrast, a simple PTS for HbF derived from GWAS performed in SCD patients explained a large proportion of the HbF phenotypic variance.

### Associations between hematologic trait polygenic trait scores and SCD-related complications

Variation in HT has been associated with several clinical outcomes observed in SCD patients such as death,<sup>2</sup> stroke,<sup>26-28</sup> VOC,<sup>3</sup> and ACS.<sup>4,9</sup> We were able to reproduce most of these results in the genotyped subset of the CSSCD (*Online Supplementary Table S3*). We asked ourselves whether PTS for HT were also associated with these complications in the largest available SCD cohort (CSSCD). For these analyses, we used significant PTS (Table 1) for which the corresponding

HT was also significantly associated with the complication (*Online Supplementary Table S3*). In total, we tested seven PTS-complication combinations, and found significant associations for PTS<sub>HbF</sub>-stroke and PTS<sub>HbF</sub>-ACS (Table 2).

We extended these analyses to determine if the PTS could improve the association of the models beyond the baseline HT measures. We reasoned that, because PTS capture HT heritable variation, they would more faithfully represent “life-long exposure” and add information that is independent of the imprecisions associated with HT measurement in the lab. The statistical models did not improve for stroke and ACS rate, but we found that PTS<sub>HbF</sub> improved the association with VOC, as described previously (Table 2).<sup>29,30</sup> Interestingly, additional analyses revealed that PTS<sub>HbF</sub> improves the association of the model with VOC in patients with low HbF values (<10%), levels at which HbF is not associated with VOC (*Online Supplementary Table S4*). For high HbF values (>10%), HbF is strongly associated with VOC, and adding PTS<sub>HbF</sub> did not improve the statistical model.

**Table 1.** Hematologic trait variance explained by polygenic trait scores in sickle cell disease patients.

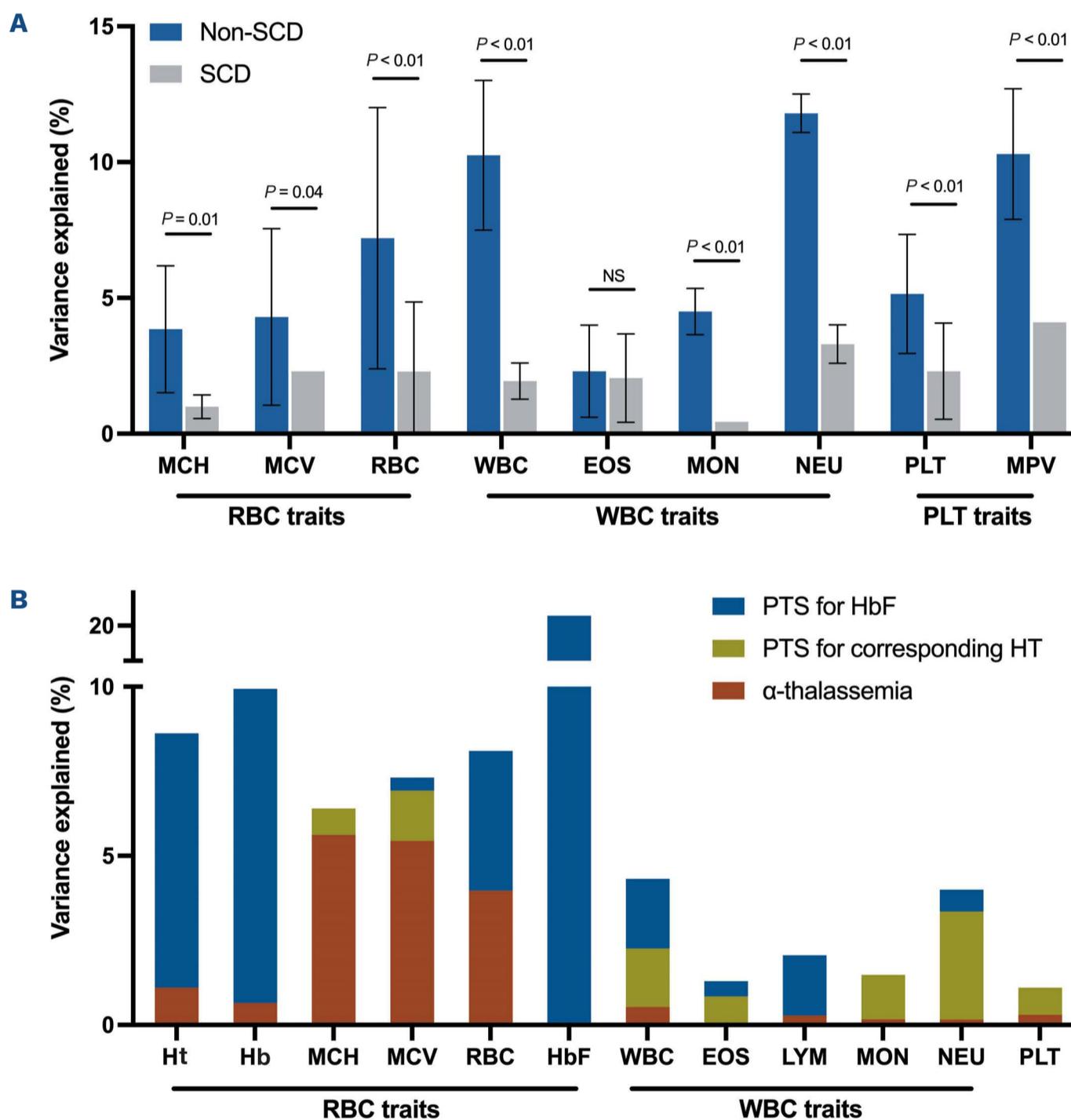
Phenotype	N of variants in the PTS <sup>a</sup>	SCD cohorts						Non-SCD cohorts					
		CSSCD (N <sub>max</sub> = 1,015)		GEN-MOD (N <sub>max</sub> = 401)		Mondor/Lyon (N <sub>max</sub> = 324)		BioME <sup>b</sup> (N <sub>max</sub> = 2,651)		UK Biobank (N <sub>max</sub> = 6,584)		Down-sampled UK Biobank (N <sub>max</sub> = 1,278)	
		Variance (%)	P	Variance (%)	P	Variance (%)	P	Variance (%)	P	Variance (%)	P	Variance (%)	P
<b>Red blood cell (RBC) traits</b>													
Hematocrit	356	-	0.70	-	0.66	-	0.14	1.4	0.0071	2.1	1.3x10 <sup>-26</sup>	1.5	9.4x10 <sup>-5</sup>
Hemoglobin	364	-	0.41	-	0.24	-	0.77	1.3	7.4x10 <sup>-9</sup>	3.0	1.9x10 <sup>-35</sup>	2.0	2.0x10 <sup>-6</sup>
MCH	384	0.69	0.014	1.3	0.021	-	0.88	2.2	5.4x10 <sup>-14</sup>	5.5	2.4x10 <sup>-42</sup>	6.9	4.6x10 <sup>-11</sup>
MCV	454	2.3	8.8x10 <sup>-7</sup>	-	0.078	-	0.55	2.0	9.1x10 <sup>-13</sup>	6.6	1.8x10 <sup>-51</sup>	10.4	6.7x10 <sup>-17</sup>
RBC	449	0.47	0.032	4.1	8.1x10 <sup>-5</sup>	-	0.16	3.8	< 2.2x10 <sup>-16</sup>	10.6	5.8x10 <sup>-95</sup>	9.3	2.2x10 <sup>-16</sup>
HbF	6	20.5	10.0x10 <sup>-59</sup>	26.5	1.2x10 <sup>-26</sup>	27.1	2.0x10 <sup>-25</sup>						
<b>White blood cell (WBC) traits</b>													
WBC	443	1.5	2.5x10 <sup>-4</sup>	2.7	0.002	1.6	0.047	12.2	< 2.2x10 <sup>-16</sup>	8.3	1.5x10 <sup>-115</sup>	7.0	2.3x10 <sup>-22</sup>
WBC (without chr1 - Duffy/DARC)	389	0.5	0.023	3.3	3.6x10 <sup>-4</sup>	2.3	0.024	1.8	< 2.2x10 <sup>-16</sup>	1.4	3.9x10 <sup>-23</sup>	1.4	6.5x10 <sup>-6</sup>
Eosinophils	346	0.9	0.0038	3.2	4.5x10 <sup>-4</sup>			1.1	6.2x10 <sup>-6</sup>	3.5	3.3x10 <sup>-42</sup>	3.7	2.8x10 <sup>-10</sup>
Lymphocytes	409	-	0.18	-	0.42			1.4	7.7x10 <sup>-9</sup>	3.6	6.9x10 <sup>-42</sup>	2.2	2.1x10 <sup>-6</sup>
Monocytes	394	0.44	0.00026	-	0.18			3.9	< 2.2x10 <sup>-16</sup>	5.1	1.9x10 <sup>-74</sup>	6.1	1.4x10 <sup>-18</sup>
Neutrophils (NEU)	352	2.8	1.4x10 <sup>-6</sup>	3.8	3.9x10 <sup>-4</sup>			12.3	< 2.2x10 <sup>-16</sup>	11.3	5.4x10 <sup>-167</sup>	10.1	8.9x10 <sup>-39</sup>
NEU (without chr1 - Duffy/DARC)	313	0.8	0.0065	3.9	1.3x10 <sup>-4</sup>			0.67	1.1x10 <sup>-4</sup>	1.2	3.6x10 <sup>-22</sup>	1.1	3.5x10 <sup>-5</sup>
<b>Platelet (PLT) traits</b>													
PLT	553	0.7	0.008	2.0	0.0045	4.2	1.9x10 <sup>-4</sup>	3.6	< 2.2x10 <sup>-16</sup>	6.7	1.9x10 <sup>-92</sup>	7.3	4.0x10 <sup>-21</sup>
MPV	392			4.1	2.4x10 <sup>-4</sup>			8.6	< 2.2x10 <sup>-16</sup>	12.0	1.7x10 <sup>-140</sup>	12.3	3.4x10 <sup>-30</sup>

Only the phenotypic variance explained for polygenic trait scores (PTS) that were nominally significant were calculated ( $P < 0.05$ ). Empty cells indicate phenotypes that were not available in the corresponding study. All  $P$ -values  $< 0.05$  remain statistically significant after multiple testing correction except for WBC and WBC (without chr1 - Duffy/DARC) phenotypes in Mondor/Lyon cohort. The <sup>a</sup>PTS models (except for fetal hemoglobin [HbF]) and the results from the <sup>b</sup>BioMe African American participants have been reported previously.<sup>19</sup> For HbF, we selected six independently associated variants.<sup>21-23</sup> SCD: sickle cell disease; CSSCD: Cooperative Study of SCD; GEN-MOD: Genetic Modifier; N: number; MCH: mean corpuscular hemoglobin; MCV: mean corpuscular volume; MPV: mean platelet volume.

### Quantifying the causal impact of HbF levels on SCD complications by Mendelian randomization

Using MR, we sought to confirm the causal protective effect of high HbF levels on stroke,<sup>27,28</sup> ACS,<sup>4</sup> and VOC.<sup>3</sup> Although the literature suggests links between high WBC or neutrophil counts and SCD survival,<sup>2</sup> we did not test these potential causal relationships by MR because of limited statistical power (see *Online Supplementary Appendix*). We performed two-sample MR using the pseudo-independent SNP from  $PTS_{HbF}$ . We found a causal association between HbF and stroke: a one standard deviation increase in genetically-determined HbF levels (corresponding to 4.8% of HbF) de-

creases the risk of stroke by 39% (odds ratio [95% confidence interval]=0.61 [0.46-0.81];  $P=0.0006$ ) (Figure 3). In this analysis, the low-frequency variant rs114398597 located between *HBS1L* and *MYB* on chromosome 6 appears as an outlier (Figure 3), but we confirmed that the MR result remained significant after its exclusion ( $P=3.7 \times 10^{-6}$ ) (*Online Supplementary Table S5*). Although the direction of the effect was similar using the sensitivity MR-Egger and weighted median methods, these analyzes were not statistically significant, suggesting insufficient power (*Online Supplementary Table S5*). We found no heterogeneity in the effect ( $I^2 = 0\%$ ) and confirmed the absence of horizontal



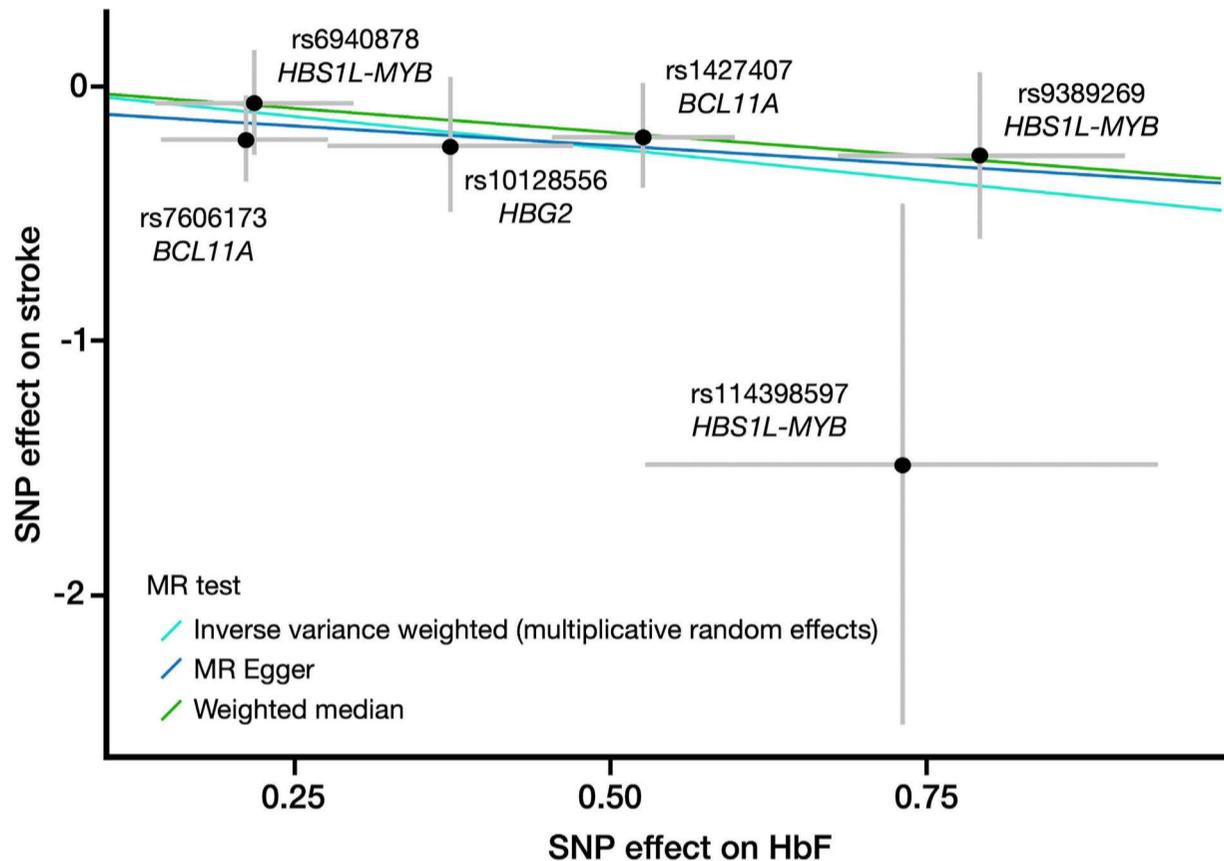
**Figure 2. Variance explained by polygenic trait scores for hematologic traits.** (A) Comparison of variance explained by polygenic trait scores (PTS) for hematologic traits (HT) in African-ancestry non-sickle cell disease (SCD) individuals from BioMe and the UK Biobank, and participants from three SCD cohorts (CSSCD, GEN-MOD, Mondor/Lyon). We only present the variance explained by nominally significant PTS in at least one SCD cohort (Table 1). When the PTS for a given HT was significant in more than one SCD cohort, we calculated the mean and standard deviation (error bars) of the variance explained. We show the empirical  $P$ -value for comparison between SCD and non-SCD individuals. (B) Respective contribution of the PTS for corresponding HT,  $PTS_{HbF}$  and  $\alpha$ -thalassemia status on the HT variance explained. Analyses performed in the CSSCD cohort. Only the variance explained if the PTS or  $\alpha$ -thalassemia was nominally significant is presented ( $P<0.05$ ). EOS: eosinophils; MCH: mean corpuscular hemoglobin; MCV: mean corpuscular volume; MON: monocytes; MPV: mean platelet volume; NEU: neutrophil count; PLT: platelet count; RBC: red blood cell count; WBC: white blood cell count; Ht: hematocrit; Hb: hemoglobin; HbF: fetal hemoglobin.

pleiotropy (Egger intercept, -0.09, standard error = 0.21;  $P=0.69$ ). We did not find causal association between HbF levels and ACS/VOC rates (*Online Supplementary Table S5, Online Supplementary Figure S1*); this may have been due to limited statistical power (*Online Supplementary Appendix*).

**SCD partially masks the genetic effect of the Duffy/DARC null variant on white blood cell and neutrophil counts**

To understand why the PTS under-performed in SCD patients (excluding  $PTS_{HbF}$ ), we carried out meta-analyses of

GWAS results for the 11 HT in the three SCD cohorts, and compared effect sizes ( $\beta_{SCD}$ ) for the SNP found in the PTS (4,201 SNP-HT pairs) with the effect sizes from multi-ancestry meta-analyses ( $\beta_{non-SCD}$ ).<sup>19</sup> Across all SNP and HT, normalized effect sizes were weakly correlated when considering the same effect alleles (Pearson's  $r = 0.09$ ;  $P=2.4 \times 10^{-10}$ ) (Figure 4A). Among the 273 variant-HT pairs that were nominally associated in SCD meta-analyses ( $P < 0.05$ ), 162 (59.3%) had a concordant direction of effect in non-SCD meta-analyses ( $P=0.002$ , binomial test).



**Figure 3. Mendelian randomization results for fetal hemoglobin levels on stroke.** Each dot represents one of the fetal hemoglobin (HbF)-associated SNP, with its corresponding effect on normalized HbF levels (x-axis, standard deviation units) and stroke risk (y-axis, logistic regression beta). The analysis without rs114398597 is significant; see *Online Supplementary Table S5* for details. MR: Mendelian randomization.

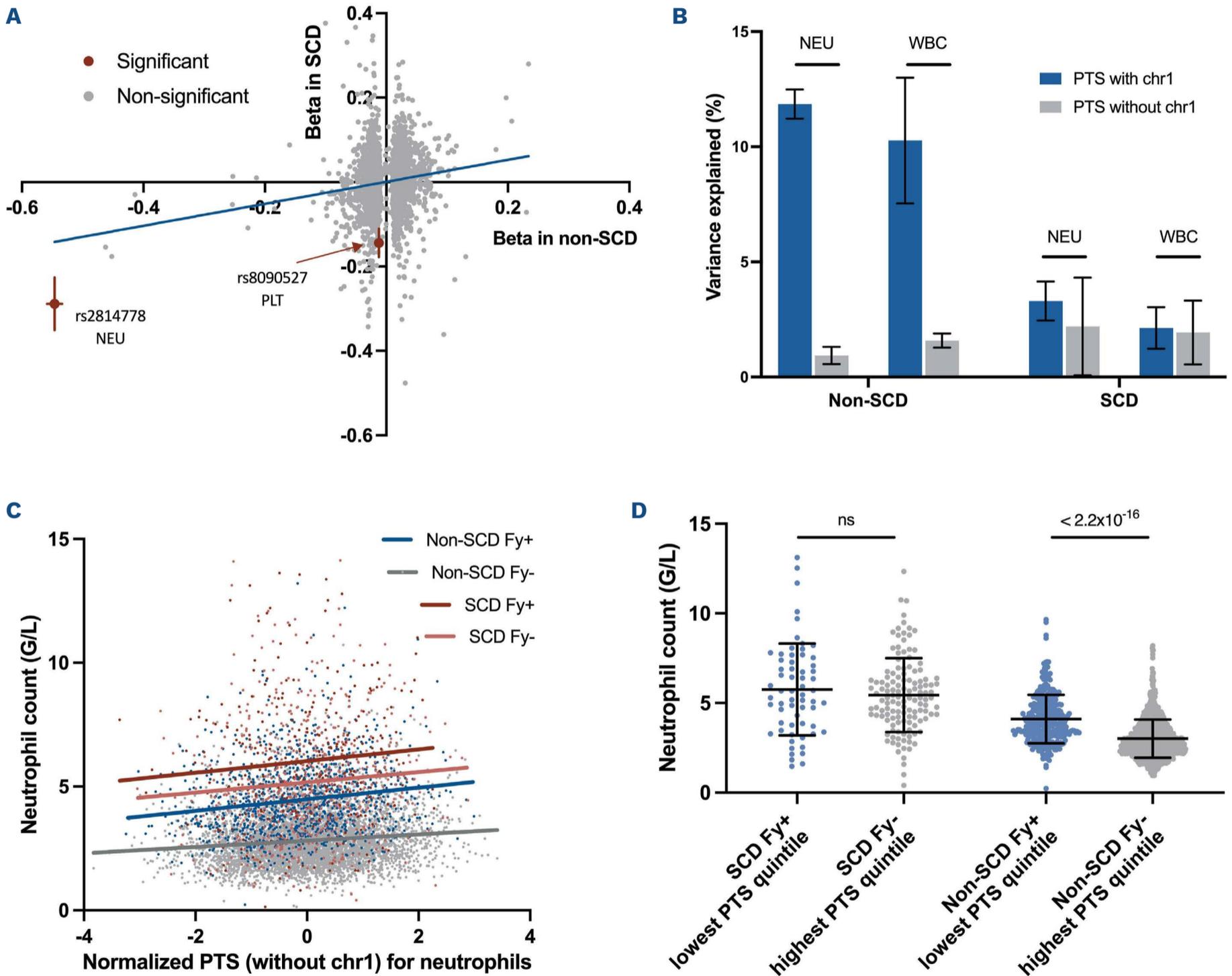
**Table 2.** Polygenic trait score for fetal hemoglobin levels is associated with stroke and acute chest syndrome and improves a statistical model of vaso-occlusive crisis rate.

Outcome	PTS	Association of PTS with outcome		Comparison of models with and without PTS	
		HR [95%CI] or Beta (SE)	P	$\chi^2$ (1 df)	P
Stroke (N cases = 104) (N controls = 1,168)	HbF*	0.72 [0.59-0.88]	0.002	0.23	0.63
	WBC	1.00 [0.82-1.22]	0.99	1.42	0.23
	NEU	0.98 [0.81-1.19]	0.82	2.43	0.12
	LYM	1.01 [0.83-1.23]	0.90	0.25	0.62
ACS rate (N = 1,271)	HbF	-0.20 (0.06)	0.0005	0.91	0.34
	EOS	-0.02 (0.05)	0.75	0.48	0.49
VOC rate (N = 1,271)	HbF	0.06 (0.05)	0.21	16.1	$6.0 \times 10^{-5}$

Analyses in 1,278 genotyped CSSCD participants. To contrast models, we performed an analysis of deviance and compared a baseline model (hematologic trait [HT], age, sex,  $\alpha$ -thalassemia, and 10 first principal components) with a model that included the same predictors as well as the polygenic trait score (PTS). All  $P$ -values  $< 0.05$  remain statistically significant after multiple testing correction. \*Fine and Gray's method was used to consider death as a competitive risk of stroke that may have influenced the result.  $PTS_{HbF}$  was still associated with stroke in the subdistribution hazard model: hazard ratio (HR) 0.75, 95% confidence interval (CI) (0.61-0.94),  $P=0.01$ . ACS: acute chest syndrome; VOC: vaso-occlusive crisis; Beta: effect size in standard deviation units; SE: standard error; df: degree-of-freedom; HbF: fetal hemoglobin; WBC: white blood cell count; EOS: eosinophils; NEU: neutrophil count; LYM: lymphocyte count.

After correction for multiple testing ( $q$ -value  $<0.05$ ), we found two variants significantly associated with HT in SCD patients, but with significantly different effect size when comparing  $\beta_{\text{SCD}}$  and  $\beta_{\text{non-SCD}}$ : rs8090527 and rs2814778 (*Online Supplementary Table S6, Online Supplementary Figures S2, S3*). We confirmed that the effect sizes were still different when comparing  $\beta_{\text{SCD}}$  to  $\beta_{\text{non-SCD}}$  derived from a down-sampled subset of African-ancestry UK Biobank individuals ( $P < 0.05$ ) (*Online Supplementary Table S6*). We did not explore the association between the intergenic rs8090527

variant and PLT count further as we were unable to replicate it in an independent SCD cohort (*Online Supplementary Table S6*). The second variant is the previously described Duffy/DARC null variant (rs2814778),<sup>31</sup> which had a two-fold weaker effect on neutrophil count in SCD when compared to non-SCD individuals. Based on this observation, we wondered if the apparent poor performance of the WBC and neutrophil count PTS in SCD participants was due to the lower impact of the Duffy/DARC null variant in this patient population. In non-SCD individuals, removing the



**Figure 4. Sickle cell disease partially masks the genetic impact of the Duffy/DARC null variant (rs2814778) on neutrophil count.**

(A) Comparison of effect sizes (Beta) for 3,917 SNP-hematologic trait pairs in sickle cell disease (SCD) patients (x-axis) and non-SCD participants (y-axis). Only variants with a minor allele frequency  $>1\%$  in the SCD meta-analyses were considered. We highlight two variants with significantly different effect sizes between SCD and non-SCD individuals. (See text for details). The blue line represents the best-fit linear regression line. (B) Variance explained (mean and standard deviation) by polygenic trait scores (PTS) for neutrophil (NEU) and white blood cell (WBC) counts with and without chromosome (chr) 1 variants in SCD (three cohorts) and non-SCD (BioMe and UK Biobank cohorts) individuals. (C) Raw neutrophil count (y-axis) as a function of the multi-ancestry neutrophil PTS (without chr 1 variants, x-axis) in Duffy-positive (Fy+, T/T or C/T genotypes at rs2814778) and Duffy-negative (Fy-, C/C genotype) SCD (CSSCD) or non-SCD (UK Biobank) individuals. Regression lines for each of the four subgroups are shown. (D) Comparison of raw neutrophil count between Fy+ individuals with a PTS within the lowest quintile and Fy- individuals with a PTS within the highest quintile, for both SCD (CSSCD) and non-SCD (UK Biobank) individuals. ns: non-significant.

PTS variants on chromosome 1 (to ensure that the large admixture signal due to the Duffy/DARC locus does not impact the analysis) reduced the mean variance explained for WBC from 10.3% to 1.6%, and for neutrophils from 11.9% to 0.9% (Table 1, Figure 4B). When we repeated this analysis in SCD participants, the PTS for WBC was not affected (from 2.1% to 1.9%), whereas the variance explained by the neutrophil count PTS changed slightly from 3.3% to 2.2% (Figure 4B).

We next specifically focused on the association between Duffy/DARC rs2814778 and WBC or neutrophil counts in SCD and non-SCD individuals. First, consistent with the recessive inheritance of the Duffy-negative blood group, we showed that a recessive genetic model provided a better fit with the data than the standard additive model (Online Supplementary Table S7). Thus, we used a recessive model for all subsequent genetic analyses of this variant. For non-SCD individuals, the single Duffy/DARC variant explained 18.4–24.2% of the phenotypic variance (Table 3). In contrast, the associations between rs2814778 and WBC or neutrophil counts were either weak or non-significant in SCD participants, with this variant contributing only 0.9–3.3% of the variance in these HT (Table 3). To quantify the magnitude of the difference in effect sizes and provide meaningful clinical estimates, we calculated that the Duffy null genotype (homozygosity for the C-allele at rs2814778) was associated with a mean reduction of  $0.76 \times 10^9$  WBC/L ( $P=0.004$ ) and  $0.84 \times 10^9$  neutrophils/L ( $P=1.6 \times 10^{-6}$ ) in the CSSCD, and of  $1.9 \times 10^9$  WBC/L ( $P=3.3 \times 10^{-164}$ ) and  $1.6 \times 10^9$  neutrophils/L ( $P=4 \times 10^{-199}$ ) in the UK Biobank. When we considered both SCD status and the Duffy blood group, we found that: 1) SCD has the strongest impact on neutrophil count; 2) Duffy has a weaker effect on neutrophil count in SCD patients;

and 3) the neutrophil PTS (without chromosome 1 variants) remains associated with neutrophil count in all groups (Figure 4C). To further illustrate how SCD modifies the effect of Duffy, we considered the neutrophil PTS (without chromosome 1 variants) quintiles and compared neutrophil count in Duffy-positive individuals with a PTS in the lowest quintile with Duffy-negative individuals with a PTS in the highest quintile (Figure 4D). Whereas in non-SCD UK Biobank participants Duffy outweighs the PTS effect, it is equivalent in SCD patients. We observed a similar effect when analyzing a subset of the non-SCD African-ancestry cohort from the UK Biobank with a sample size similar to the CSSCD, suggesting that the difference in effect size is not due to a difference in sample size (Online Supplementary Table S8, Online Supplementary Figure S3). Finally, we investigated whether sickle cell trait (heterozygosity for the hemoglobin S allele) also modifies the Duffy/DARC effect on neutrophil count. We did not find a significant interaction between these two genotypes in the UK Biobank ( $P=0.33$ ), indicating that the epistatic effect is specific to SCD.

Taken together, our data suggest that SCD partially masks the strong effect of the Duffy/DARC null variant (rs2814778) on neutrophil count. In non-SCD individuals, the Duffy null variant is the main determinant of neutrophil count, whereas in SCD individuals, its effect is equivalent to the effect of other neutrophil count-associated common variants (as captured by the neutrophil PTS without chromosome 1).

### Genome-wide association studies of hematologic traits in SCD patients

To determine if new genetic variation could specifically modulate HT variation in SCD, we carried out GWAS for 11

**Table 3.** Genetic association results between the Duffy/DARC null variant (rs2814778) and white blood cell and neutrophil counts.

	N	Mean (SD) 10 <sup>9</sup> /L	EAF (C-allele)	Beta (SE)	P	Variance (%)
<b>Neutrophils</b>						
CSSCD	934	5.42 (2.44)	0.842	-0.353 (0.073)	$1.6 \times 10^{-6}$	3.3
GEN-MOD	400	5.84 (2.67)	0.939	-0.019 (0.209)	0.93	-
UK Biobank	6,564	3.10 (1.27)	0.906	-1.166 (0.033)	$6.0 \times 10^{-239}$	23.1
Down-sampled UK Biobank	1,278	3.04 (1.17)	0.912	-1.228 (0.080)	$7.0 \times 10^{-49}$	24.2
<b>WBC</b>						
CSSCD	1,014	12.19 (3.78)	0.845	-0.178 (0.064)	0.005	0.9
GEN-MOD	400	10.60 (3.67)	0.939	-0.092 (0.208)	0.66	-
Mondor/Lyon	322	10.62 (3.01)	0.935	-0.151 (0.214)	0.48	-
UK Biobank	6,584	5.81 (1.66)	0.906	-1.041 (0.036)	$5.0 \times 10^{-171}$	18.4
Down-sampled UK Biobank	1,278	5.75 (1.56)	0.912	-1.038 (0.083)	$4.8 \times 10^{-34}$	17.3

The direction of the effect is given for the CC genotype (in a recessive genetic model). Neutrophil count is not available in Mondor/Lyon. Effect sizes (Beta) and standard errors (SE) are in standard deviation units. SD: standard deviation; CSSCD: Cooperative Study of Sickle Cell Disease; GEN-MOD: Genetic Modifier; WBC: white blood cell count; EAF: effect allele frequency.

blood-cell traits across all three SCD cohorts available. Given the relatively small sample size of the dataset, we restricted our analyses to variants with a minor allele frequency (MAF) >1%. We found little evidence of association, except for Ht and Hb levels (*Online Supplementary Figure S4*). We found 24 genome-wide significant ( $P < 5 \times 10^{-8}$ ) SNP-HT associations, including 23 at the known HbF loci and associated with Hb levels, Ht or RBC count (*Online Supplementary Table S9*). The last variant, rs113819343, was associated with PLT count in the SCD meta-analyses ( $P = 1.4 \times 10^{-8}$ ). This variant is common in African-ancestry individuals (MAF=5.5%, gnomAD) but rarer in European-ancestry populations (MAF=0.098%). This variant is not associated with PLT count in the multi-ancestry ( $n=473,895$ ;  $P=0.72$ ) nor in the African-specific ( $n=15,171$ ;  $P=0.27$ ) meta-analyses from the Blood-Cell Consortium.<sup>19</sup> Our attempt to replicate this association with PLT count in 333 SCD patients from the OMG cohort was unsuccessful ( $P=0.69$ ), so it is not possible to conclude if this association is real or a false-positive result.

## Discussion

In this study, we explored the utility of PTS for HT in SCD patients. Our analyses highlight several important results, including some with clinical implications. First, we found that PTS for HT derived in non-SCD individuals largely underperformed when tested in SCD patients. This emphasizes the need to derive polygenic predictors directly in SCD patients before trying to implement them into precision medicine initiatives for this patient population. This will require larger SCD cohorts with comprehensive clinical and genetic information. Second, our characterization of the weak association between the Duffy/DARC null rs2814778 variant and neutrophil count in SCD patients suggested that the weaker impact of PTS may partly be due to an epistatic effect of SCD with other HT-associated variants. These results explain the unsuccessful attempts to use the Duffy/DARC null variant as an SCD genetic modifier,<sup>32,33</sup> and stress the importance to use all neutrophil-associated variants (ideally identified by GWAS carried out in SCD patients) as potential predictors of SCD complications. Third, a small set of six HbF-associated variants (PTS<sub>HbF</sub>) were associated with stroke and ACS rate, and improved the statistical modeling of VOC rate. These results support the addition of PTS<sub>HbF</sub> in clinical efforts to stratify SCD patients based on risk of developing complications. Finally, PTS<sub>HbF</sub> allowed us to confirm and quantify the causal protective impact of HbF increase on stroke risk reduction, a controversial point in SCD literature (see below). Several non-mutually exclusive factors could explain why PTS for HT were not as strongly associated with HT variation in SCD patients. These patients have different base-

line levels and distribution values for several HT such as Hb and WBC count. This is in part due to the direct hemolytic effect of hemoglobin S, but also to the broad consequences of SCD, such as the induction of a chronic inflammatory state that can lead to cytokine-driven higher WBC count.<sup>34,35</sup> Moreover, the frequent intercurrent complications (e.g., VOC) experienced throughout the natural course of SCD could result in greater variability in HT values. Finally, SNP genotyping arrays do not capture all structural variants, which are the main alterations in  $\alpha$ -thalassemia, a major determinant MCH and MCV. We showed here that adding  $\alpha$ -thalassemia and PTS<sub>HbF</sub> to PTS for corresponding HT greatly increased the variance explained. Thus, the performance of HT PTS will improve once comprehensive whole-genome sequencing of large SCD cohorts becomes available.

The C-allele of the Duffy/DARC null variant (rs2814778) results in erythroid-specific loss of expression of the Duffy/DARC chemokine reservoir expression gene.<sup>36</sup> It has been known for some time that the Duffy/DARC null variant is strongly associated with lower circulating neutrophil and WBC counts in SCD and non-SCD individuals due to extravasation to tissues,<sup>31,37,38</sup> and in particular the spleen.<sup>39</sup> Although it is unclear why SCD partially masks the effect of this variant on neutrophil count, we may speculate that the precocious and pervasive splenic atrophy observed in SCD patients could lead to a reduced reservoir size. In addition, various Duffy/DARC variants in non-SCD and SCD individuals have been shown to affect the binding of DARC with inflammatory markers such as interleukin 8.<sup>40-42</sup> Definitive data are still lacking, but the proinflammatory state of SCD patients may contribute to the observed discrepancy in effect.<sup>43</sup>

Whether or not the Duffy phenotype is associated with complications in SCD patients remains unclear.<sup>32,33,37,44-46</sup> The potential consequences of the Duffy/DARC status in SCD would be linked to circulating proinflammatory cytokines and, in particular, the resulting effect on neutrophil and WBC counts, which has been implicated in several SCD complications.<sup>2,6,7</sup> However, our data showed that the Duffy/DARC negative phenotype is not a good proxy for neutrophil count in SCD patients, in contrast to non-SCD individuals in whom it accounts for a large fraction of the phenotypic variance.

Although complex multi-ancestry PTS for general HT that include hundreds of variants performed poorly in SCD patients, we showed that a simple PTS for HbF made of six variants at three loci can capture a large fraction of the phenotypic variance in this important SCD modifier, consistent with previous reports.<sup>23,24</sup> One major difference between the general HT and HbF is that we developed PTS<sub>HbF</sub> using GWAS data from SCD patients. Interestingly, we observed an association between PTS<sub>HbF</sub> and ACS or stroke in the large CSSCD. We also validated previous observations

that a PTS for HbF can improve a statistical model for VOC rate,<sup>29,30</sup> and further discovered that this  $PTS_{HbF}$  was useful in the subset of patients with HbF levels <10%.

Whether the increase in HbF levels has a beneficial effect on stroke has long been unclear due to data discrepancies. The pre-hydroxyurea (HU) CSSCD (still the largest SCD cohort to date) did not find that high HbF reduces ischemic stroke nor risk of silent infarct (changes in white matter).<sup>26,47</sup> Of note, in the CSSCD, results for hemorrhagic stroke and “overall” stroke (considering both ischemic and hemorrhagic stroke) were not reported. A follow-up study of SCD patients identified through screening of newborn did not find a protective effect of HbF.<sup>48</sup> In contrast, historical,<sup>49</sup> as well as more recent data, found a protective effect of elevated HbF levels on “overall” stroke.<sup>28</sup> This protective effect was also reported for incidence of silent infarct.<sup>27,50</sup> Furthermore, some studies investigating the association between HbF-associated SNP and stroke reported an association,<sup>51,52</sup> while another did not.<sup>53</sup> Several factors could have led to such inconsistent results: 1) different studies used different definitions of stroke; 2) only the CSSCD separated ischemic and hemorrhagic stroke based on computed tomography findings (with the limitations associated with this technology to define stroke sub-type); 3) retrospective studies used history of stroke as endpoint; 4) small sample sizes with limited power to detect an effect of HbF on stroke; and 5) differences between SCD studies regarding the age at which HbF levels were measured. (HbF is physiologically higher in younger patients<sup>54</sup>). Finally, HU has clearly been shown to reduce stroke risk in prospective studies,<sup>55</sup> but given the multiple mechanisms by which the drug is beneficial in SCD,<sup>56</sup> its protective effect cannot be solely linked to HU-mediated HbF increase.

In this study, we used the large genotyped subset of the CSSCD and combined genetic information in a PTS to show that HbF associates with stroke in SCD patients. Our MR analyses revealed that HbF levels have a causal and protective effect on stroke, although we also acknowledge that the Winner’s curse could have artificially increased our estimate. In effect, while the MR Egger and weighted median MR approaches are sensitive methods to detect horizontal pleiotropy, they are also known to have less statistical power than the multiplicative random-effect IVW method used as the primary discovery method in our analyses.<sup>57,58</sup> Thus, the non-significant results for the  $PTS_{HbF}$ -stroke using MR-Egger and weighted median approaches should not be interpreted as a lack of validation. What this negative result means is that we cannot exclude the possibility that horizontal pleiotropy could have impacted our HbF-stroke MR result. However, because the MR instruments were selected at clear HbF loci with known regulatory mechanisms (*BCL11A*, *HBS1L-MYB*, beta-globin), we suggest that the risk of confound-

ing due to horizontal pleiotropy is minimal. Our MR analysis suggests that the impact of HU on stroke is,<sup>55</sup> at least in part, mediated by HbF and distinguishes the effect of HU by specifically quantifying how genetically-determined (lifelong) HbF levels modulate stroke risk.

Additional large replication cohorts were not available to our group to validate our findings and we acknowledge that this is an important limitation of our study. Specifically, replication of the association and MR results for  $PTS_{HbF}$  and stroke would require large SCD cohorts (approx. 2,800-5,180 SCD patients; see *Online Supplementary Appendix*) given the reduction in stroke incidence following the implementation of successful stroke primary prevention programs.<sup>48,59</sup> Therefore, we encourage investigators to test our models in their own SCD cohorts, and applaud efforts to create new large collaborative studies of SCD to energize the field of SCD modifier genetics.

Our findings have implications beyond SCD. While PTS have been shown to modulate the penetrance of monogenic mutations in diseases such as coronary artery disease and familial breast and colorectal cancers,<sup>60</sup> much less is known about their effect on expressivity (or disease severity).<sup>13</sup> This distinction is important because, although they may not cause the disease, several clinical variables and other endophenotypes that are captured by PTS can strongly modify disease severity (e.g., PTS for kidney functions in the context of hypertension, PTS for retinopathy/cataract in diabetic patients). Our analyses indicate that simply translating the genetics of polygenic traits formulated in healthy individuals to diseased populations may not provide the expected gain in risk stratification in the context of precision medicine. Fortunately, large biobanks and other cohorts should soon be able to use powerful GWAS for genetic modifiers in >10,000 patients who all suffer from the same disease.

### Disclosures

*No conflicts of interest to disclose.*

### Contributions

*TP and GL designed the study. TP, KSL and MEG performed the analyses. TP, ALPHAO, MEG, CB, AEA-K, MJT, FG, PJ, PB and GL collected the clinical and genetic data. TP and GL drafted the paper. GL supervised the study. All authors contributed to data interpretation, revised the manuscript for critical content and approved the final manuscript.*

### Acknowledgments

*We thank all participants for their contribution to this project. We thank Gabrielle Boucher for statistical support. TP is a recipient of a Charles Bruneau Foundation fellowship award and merit scholarship program for foreign students from the Ministry of Education and Higher Education of Quebec.*

**Funding**

This work was funded by the Canadian Institutes of Health Research (PJT #156248), Bioverativ, a Sanofi Company, and the Canada Research Chair Program (to GL). GEN-MOD samples and data collection were supported by NIH grant HL-68922. AA-K, MJT and establishment and analysis of the OMG cohort have been funded by NHLBI (R01HL68959, HL79915, HL70769, HL87681). This research has been conducted using the UK Biobank Resource under Application Number 11707.

**Data-sharing statement**

The CSSCD genetic dataset is available on the database of Genotypes and Phenotypes (dbGaP: <https://www.ncbi.nlm.nih.gov/gap/>), accession phs000366.v1.p1. The UK Biobank dataset is publicly available (<https://www.ukbiobank.ac.uk/>). The GEN-MOD, Mondor/Lyon and BioMe datasets and code supporting the current study have not been deposited in a public repository because data are not public but are available from the corresponding author on request.

**References**

- Kato GJ, Piel FB, Reid CD, et al. Sickle cell disease. *Nat Rev Dis Primers*. 2018;4:18010.
- Platt OS, Brambilla DJ, Rosse WF, et al. Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med*. 1994;330(23):1639-1644.
- Platt OS, Thorington BD, Brambilla DJ, et al. Pain in sickle cell disease. Rates and risk factors. *N Engl J Med*. 1991;325(1):11-16.
- Castro O, Brambilla DJ, Thorington B, et al. The acute chest syndrome in sickle cell disease: incidence and risk factors. *The Cooperative Study of Sickle Cell Disease. Blood*. 1994;84(2):643-649.
- Steinberg MH, Chui DH, Dover GJ, Sebastiani P, Alsultan A. Fetal hemoglobin in sickle cell anemia: a glass half full? *Blood*. 2014;123(4):481-485.
- Sebastiani P, Nolan VG, Baldwin CT, et al. A network model to predict the risk of death in sickle cell disease. *Blood*. 2007;110(7):2727-2735.
- Elmariah H, Garrett ME, De Castro LM, et al. Factors associated with survival in a contemporary adult sickle cell disease cohort. *Am J Hematol*. 2014;89(5):530-535.
- Miller ST, Sleeper LA, Pegelow CH, et al. Prediction of adverse outcomes in children with sickle cell disease. *N Engl J Med*. 2000;342(2):83-89.
- Vichinsky EP, Neumayr LD, Earles AN, et al. Causes and outcomes of the acute chest syndrome in sickle cell disease. National Acute Chest Syndrome Study Group. *N Engl J Med*. 2000;342(25):1855-1865.
- Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12(1):44.
- Sun J, Wang Y, Folkersen L, et al. Translating polygenic risk scores for clinical use by estimating the confidence bounds of risk prediction. *Nat Commun*. 2021;12:5276.
- Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584-591.
- Oetjens MT, Kelly MA, Sturm AC, Martin CL, Ledbetter DH. Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nat Commun*. 2019;10:4897.
- Farber MD, Koshy M, Kinney TR. Cooperative Study of Sickle Cell Disease: demographic and socioeconomic characteristics of patients and families with sickle cell disease. *J Chronic Dis*. 1985;38(6):495-505.
- Bartolucci P, Brugnara C, Teixeira-Pinto A, et al. Erythrocyte density in sickle cell syndromes is associated with specific clinical manifestations and hemolysis. *Blood*. 2012;120(15):3136-3141.
- Ilboudo Y, Bartolucci P, Rivera A, et al. Genome-wide association study of erythrocyte density in sickle cell disease patients. *Blood Cells Mol Dis*. 2017;65:60-65.
- Pincez T, Lee SSK, Ilboudo Y, et al. Clonal hematopoiesis in sickle cell disease. *Blood*. 2021;138(21):2148-2152.
- Xu JZ, Garrett ME, Soldano KL, et al. Clinical and metabolomic risk factors associated with rapid renal function decline in sickle cell disease. *Am J Hematol*. 2018;93(12):1451-1460.
- Chen MH, Raffield LM, Mousas A, et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell*. 2020;182(5):1198-1213.e14.
- Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
- Canver MC, Lessard S, Pinello L, et al. Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat Genet*. 2017;49(4):625-634.
- Bauer DE, Kamran SC, Lessard S, et al. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science*. 2013;342(6155):253-257.
- 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(12):1049-1051.
- Gardner K, Fulford T, Silver N, et al. g(HbF): a genetic model of fetal hemoglobin in sickle cell disease. *Blood Adv*. 2018;2(3):235-239.
- Lettre G, Bauer DE. Fetal haemoglobin in sickle-cell disease: from genetic epidemiology to new therapeutic strategies. *Lancet*. 2016;387(10037):2554-2564.
- Ohene-Frempong K, Weiner SJ, Sleeper LA, et al. Cerebrovascular accidents in sickle cell disease: rates and risk factors. *Blood*. 1998;91(1):288-294.
- Calvet D, Tuilier T, Mélé N, et al. Low fetal hemoglobin percentage is associated with silent brain lesions in adults with homozygous sickle cell disease. *Blood Adv*. 2017;1(26):2503-2509.
- Sommet J, Alberti C, Couque N, et al. Clinical and haematological risk factors for cerebral macrovasculopathy in a sickle cell disease newborn cohort: a prospective study. *Br J Haematol*. 2016;172(6):966-977.
- Rampersaud E, Kang G, Palmer LE, et al. A polygenic score for acute vaso-occlusive pain in pediatric sickle cell disease. *Blood Adv*. 2021;5(14):2839-2851.
- Lettre G, Sankaran VG, Bezerra MAC, et al. 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):11869-11874.
- Visscher PM, Reich D, Nalls MA, et al. Reduced neutrophil count

- in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genetics*. 2009;5(1):e1000360.
32. Farawela HM, El-Ghamrawy M, Farhan MS, Soliman R, Yousry SM, AbdelRahman HA. Association between Duffy antigen receptor expression and disease severity in sickle cell disease patients. *Hematology*. 2016;21(8):474-479.
  33. Schnog JB, Keli SO, Pieters RA, Rojer RA, Duits AJ. Duffy phenotype does not influence the clinical severity of sickle cell disease. *Clin Immunol*. 2000;96(3):264-268.
  34. Hermand P, Azouzi S, Gautier EF, et al. The proteome of neutrophils in sickle cell disease reveals an unexpected activation of interferon alpha signaling pathway. *Haematologica*. 2020;105(12):2851-2854.
  35. Nader E, Romana M, Connes P. The red blood cell-inflammation vicious circle in sickle cell disease. *Front Immunol*. 2020;11:454.
  36. Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet*. 1995;10(2):224-228.
  37. Afenyi-Annan A, Kail M, Combs MR, Orringer EP, Ashley-Koch A, Telen MJ. Lack of Duffy antigen expression is associated with organ damage in patients with sickle cell disease. *Transfusion*. 2008;48(5):917-924.
  38. Schaefer BA, Flanagan JM, Alvarez OA, et al. Genetic modifiers of white blood cell count, albuminuria and glomerular filtration rate in children with sickle cell anemia. *PLoS One*. 2016;11(10):e0164364.
  39. Duchene J, Novitzky-Basso I, Thiriot A, et al. Atypical chemokine receptor 1 on nucleated erythroid cells regulates hematopoiesis. *Nat Immunol*. 2017;18(7):753-761.
  40. Nebor D, Durpes MC, Mougengel D, et al. Association between Duffy antigen receptor for chemokines expression and levels of inflammation markers in sickle cell anemia patients. *Clin Immunol*. 2010;136(1):116-122.
  41. Moreno Velasquez I, Kumar J, Bjorkbacka H, et al. Duffy antigen receptor genetic variant and the association with Interleukin 8 levels. *Cytokine*. 2015;72(2):178-184.
  42. Schnabel RB, Baumert J, Barbalic M, et al. Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. *Blood*. 2010;115(26):5289-5299.
  43. van Beers EJ, Yang Y, Raghavachari N, et al. Iron, inflammation, and early death in adults with sickle cell disease. *Circ Res*. 2015;116(2):298-306.
  44. Elliott L, Ashley-Koch AE, De Castro L, et al. Genetic polymorphisms associated with priapism in sickle cell disease. *Br J Haematol*. 2007;137(3):262-267.
  45. Nolan VG, Adewoye A, Baldwin C, et al. Sickle cell leg ulcers: associations with haemolysis and SNPs in Klotho, TEK and genes of the TGF-beta/BMP pathway. *Br J Haematol*. 2006;133(5):570-578.
  46. Drasar ER, Menzel S, Fulford T, Thein SL. The effect of Duffy antigen receptor for chemokines on severity in sickle cell disease. *Haematologica*. 2013;98(8):e87-89.
  47. Kinney TR, Sleeper LA, Wang WC, et al. Silent cerebral infarcts in sickle cell anemia: a risk factor analysis. *Pediatrics*. 1999;103(3):640-645.
  48. Bernaudin F, Verlhac S, Arnaud C, et al. Impact of early transcranial Doppler screening and intensive therapy on cerebral vasculopathy outcome in a newborn sickle cell anemia cohort. *Blood*. 2011;117(4):1130-1140.
  49. Powars DR, Schroeder WA, Weiss JN, Chan LS, Azen SP. Lack of influence of fetal hemoglobin levels or erythrocyte indices on the severity of sickle cell anemia. *J Clin Invest*. 1980;65(3):732-740.
  50. van der Land V, Mutsaerts HJMM, Engelen M, et al. Risk factor analysis of cerebral white matter hyperintensities in children with sickle cell disease. *Br J Haematol*. 2016;172(2):274-284.
  51. Leonardo FC, Brugnerotto AF, Domingos IF, et al. Reduced rate of sickle-related complications in Brazilian patients carrying HbF-promoting alleles at the BCL11A and HMIP-2 loci. *Br J Haematol*. 2016;173(3):456-460.
  52. Saraf SL, Akingbola TS, Shah BN, et al. Associations of  $\alpha$ -thalassemia and BCL11A with stroke in Nigerian, United States, and United Kingdom sickle cell anemia cohorts. *Blood Adv*. 2017;1(11):693-698.
  53. Arez AP, Wonkam A, Ngo Bitoungui VJ, et al. 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.
  54. Maier-Redelsperger M, Noguchi CT, de Montalembert M, et al. Variation in fetal hemoglobin parameters and predicted hemoglobin S polymerization in sickle cell children in the first two years of life: Parisian Prospective Study on Sickle Cell Disease. *Blood*. 1994;84(9):3182-3188.
  55. Ware RE, Davis BR, Schultz WH, et al. Hydroxycarbamide versus chronic transfusion for maintenance of transcranial doppler flow velocities in children with sickle cell anaemia—TCD With Transfusions Changing to Hydroxyurea (TWITCH): a multicentre, open-label, phase 3, non-inferiority trial. *Lancet*. 2016;387(10019):661-670.
  56. Platt OS. Hydroxyurea for the treatment of sickle cell anemia. *N Engl J Med*. 2008;358(13):1362-1369.
  57. Schmidt AF, Dudbridge F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. *Int J Epidemiol*. 2018;47(4):1217-1228.
  58. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304-314.
  59. Fullerton HJ, Adams RJ, Zhao S, Johnston SC. Declining stroke rates in Californian children with sickle cell disease. *Blood*. 2004;104(2):336-339.
  60. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun*. 2020;11(1):3635.