

Variation and impact of polygenic hematological traits in monogenic sickle cell disease

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

1 Montreal Heart Institute, Montréal, Québec, Canada

2 Department of Pediatrics, Division of Pediatric Hematology-Oncology, Charles-Bruneau Cancer Center, CHU Sainte-Justine, Université de Montréal, Montréal, Québec, Canada

3 Red Cell Genetic Disease Unit, Hôpital Henri-Mondor, Assistance Publique-Hôpitaux de Paris (AP-HP), Université Paris Est, IMRB - U955 - Équipe no 2, Créteil, France

4 Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA

5 Department of Laboratory Medicine, Boston Children's Hospital, Boston, MA, USA

6 Department of Medicine, Division of Hematology, Duke University Medical Center, Durham, NC, USA

7 Unité Fonctionnelle 34445 'Biochimie des Pathologies Érythrocytaires', Laboratoire de Biochimie et Biologie Moléculaire Grand-Est, Groupement Hospitalier Est, Hospices Civils de Lyon, Bron, France

8 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

9 Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, Canada.

Corresponding author:

Guillaume Lettre

Montreal Heart Institute, 5000 Belanger Street, Montreal, Quebec, H1T 1C8, Canada

514-376-3330

guillaume.lettre@umontreal.ca

ABSTRACT

Several complications observed in sickle cell disease (SCD) are influenced by variation in hematological 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 1000s of variants associated with HT, which have then been used to develop multi-ancestry polygenic trait scores (PTS). Here, we tested if these PTS associate with HT in SCD patients and can improve the prediction of SCD-related complications. In 2,056 SCD patients, we found that the PTS predicted less HT variance than in non-SCD African-ancestry individuals. 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 routinely measured HT 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 (17.1-26.4%), associated with acute chest syndrome and stroke risk, and improved the prediction of vaso-occlusive crises. Using Mendelian randomization, we found that increasing HbF by 4.8% reduces stroke risk by 36% ($P = 0.0008$). 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

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 already support their value in the context of precision medicine initiatives.^{1,2} One abundantly discussed limitation of PTS is their poor performance when tested in populations that have different ancestral backgrounds than the populations in which they were optimized.³ 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.⁴ This is important because PTS could, in theory, be useful to stratify patients into mild or severe categories. For instance, a PTS for estimated glomerular filtration rate (eGFR) could help identify hypertensive patients more likely to suffer from kidney failure, but because this PTS_{eGFR} would have been developed in largely normotensive individuals, it is unclear how useful it would be in patients with hypertension.

To address this question, we took advantage of the classic monogenic disorder sickle cell disease (SCD), and explored how this pathology impacts the performance of PTS for hematological traits (HT). SCD, the most frequent monogenic disease worldwide, is caused by mutations in the β -globin gene.⁵ Despite being primarily a mature red blood cell disorder, SCD is associated with a broad range of consequences, both hematological (*e.g.* hemolysis, ineffective erythropoiesis) and extra-hematological (*e.g.* inflammatory state, endothelial cell activation).^{5,6} 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.⁵ Critically, the causes of this clinical heterogeneity are only poorly understood.

HT are among the main factors known to be associated with clinical outcomes in SCD. fetal hemoglobin (HbF) is a major disease modifier, associated with reduction in the occurrence of several complications such as VOC, ACS and death.⁷⁻⁹ A percentage of HbF > 30% is associated with an almost complete absence of complications in SCD patients.^{10,11} However, most SCD patients have lower HbF levels while not receiving disease-modifying therapy and the risk reduction associated with HbF has not been quantified in large cohorts for some complications such as stroke. Several other HT have been associated with SCD-related complications, notably elevated white blood cell (WBC) and neutrophil counts with survival,^{7,12,13} low hemoglobin levels with composite severe outcomes and death,^{13,14} and platelet count with ACS.¹⁵

Here, we investigated if PTS for HT developed in largely healthy cohorts also associate with blood-cell phenotypes in SCD patients. Furthermore, we evaluated the clinical relevance of these PTS in terms of predicting complications in this patient population.

SUBJECTS AND METHODS

Populations

We collected data from three SCD cohorts with genome-wide genotype information available: the Cooperative Study of Sickle Cell Disease (CSSCD, n = 1278),¹⁶ Genetic Modifier (GEN-MOD, n = 406),¹⁷⁻¹⁹ and Mondor/Lyon (n = 372)²⁰ (**Table S1**). Data collection was made according to the Helsinki declaration and the study was approved by the institutional ethics committees. DNA genotyping has been described elsewhere^{18,21} and we used reference haplotypes from the TOPMed project²² to impute missing genotypes. Eight to 14 hematological traits were measured at steady state and available in these SCD cohorts. All patients were at > 3 months from a blood transfusion and only ten patients were taking hydroxyurea at baseline. For comparison, we also accessed data from African-ancestry individuals from the BioMe cohort^{23,24} and the UK Biobank.²⁵

PTS for HT

For all HT except HbF, we used the multi-ancestry PTS derived by the Blood-Cell Consortium to test for association with HT.²³ Briefly, these PTS considered the effect sizes of variants that reach genome-wide significance in multi-ancestry meta-analyses of 746,667 individuals, including 15,171 African-ancestry participants. We generated additive PTS for each individual and HT by calculating the sum of HT-increasing alleles weighted by the corresponding genome-wide association study (GWAS) effect size. 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.^{26–28} We tested the association between normalized PTS and HT by linear regression with the four first principal components as covariables.

GWAS of HT in SCD patients

We adjusted HT for sex and age, and then applied inverse normal transformation. We performed GWAS for each HT available in the three SCD cohorts separately using RvTests (v20190205),²⁹ testing an additive genetic model and correcting for the 10 first principal components. We then performed a meta-analysis of the GWAS results using METAL.³⁰

For the Duffy/DARC null variant (rs2814778) association with neutrophil and WBC counts, we compared the additive and recessive models in each cohort, correcting each model for the 10 first principal components. We calculated the effect of rs2814778 on raw neutrophil and WBC counts using a recessive model with age, sex and the 10 first principal components as covariates. We computed the variance explained in each cohort by rs2814778 using the following formula: $2pq\beta^2$, where p is the frequency of the effect allele, q is $1-p$ and β is the normalized effect size of the effect allele on the HT.

To replicate the association between rs113819343, rs8090527 and platelet count, we analyzed 333 SCD patients from the Duke University Outcome Modifying Genes (OMG) cohort with whole-genome sequencing data available. We performed the association test as described above.

Comparing effect sizes of HT-associated SNPs in SCD patients and non-SCD individuals

For each SNP-HT pair considered in the multi-ancestry PTS models, we retrieved association results from the SCD meta-analyses (above) and the published non-SCD multi-ancestry meta-analyses from the Blood-Cell Consortium.²³ While there are 4,502 SNP-HT pairs in the PTS, we could recover results for 4,201 (93%) of them in the SCD meta-analyses. From this dataset, we corrected for multiple testing by computing a q-value for each SNP-HT association. Finally, we computed a q-value from the *P-diff* obtained (see *statistical analyses* section below).

Association between HT or PTS with SCD-related clinical outcomes

We limited our analyses of the association between PTS and outcomes (VOC rate, ACS rate, stroke) to the large CSSCD. Further, we only considered PTS that were nominally associated with the corresponding HT ($P < 0.05$). First, we fitted Cox proportional hazard ratio models (for stroke) or quasi-Poisson regression models (for VOC and ACS rates) for the outcome on each HT (measured value), adjusting for age, sex and SCD subtype. Second, we repeated these analyses after replacing the HT by the corresponding PTS. To determine if the PTS improves model prediction beyond the measured HT, we performed an analysis of deviance. The difference between the residual deviances of the two models follows a χ^2 distribution with n degrees of freedom, where n corresponds to the difference in the number of degrees of freedom of the two models (*i.e.* one degree of freedom when adding the PTS).

Mendelian Randomization (MR)

We used a two-sample MR approach to test if HT captured by PTS causally impact SCD-related complications. We focused on the following combinations of HT and complications: WBC and neutrophil counts for survival, and HbF for VOC, ACS and stroke (for VOC and ACS, we dichotomized the data as no event vs. at least one event).

Association between SNPs and HT were performed in the GEN-MOD cohort (linear regression adjusting for sex and age), whereas association between SNPs and complications were carried out in the CSSCD (logistic regression, adjusting for sex, age and SCD subtype). For WBC and neutrophil counts, we used as instruments genome-wide significant SNPs reported by the Blood-Cell Consortium²³ and used in the PTS analyses described above. To ensure that these variants were independent, we further pruned them using PLINK1.9b6.10 ($r^2 > 0.01$ within 5-Mb windows).³¹ For HbF, we selected as instruments the six SNPs used in the PTS analyses. Although there is residual linkage disequilibrium between some of these HbF variants, we performed conditional analyses to obtain adjusted effects sizes and standard errors.

We performed all MR analyzes in RStudio (version 1.2.5033) using the TwoSampleMR package (version 0.5.5).³² We used the multiplicative random-effect inverse variance-weighted (IVW) approach as the main method for each MR analysis, but we also performed MR-Egger and weighted-median methods as sensitivity analyses. We assessed the validity of our statistically significant results by testing for horizontal pleiotropy (using the MR-Egger intercept test) and heterogeneity (using Cochran's Q and I^2 statistics).

Statistical analyses

To compare effect sizes derived from the meta-analyses, we calculated heterogeneity P -values (P -diff) based on the following t statistic:³³

$$= \frac{b_{SCD} - b_{non-SCD}}{\sqrt{\frac{SE_{SCD}^2 + SE_{non-SCD}^2}{2} - 2 \cdot r \cdot SE_{SCD} \cdot SE_{non-SCD}}}$$

where b_{SCD} and $b_{non-SCD}$ are the normalized effect sizes in SCD and non-SCD cohorts, respectively; SE_{SCD} and $SE_{non-SCD}$ are the standard errors in SCD and non-SCD cohorts respectively, and r is the Spearman rank correlation coefficient computed using effect sizes (for the same effect allele) of all SNPs available in the meta-analyses. In our datasets, r ranged from -0.0016 to 0.0007 across different HT. From the t statistic, we can calculate a P -value using the normal distribution.

All statistical analyses were performed using RStudio (version 1.2.5033) or GraphPad Prism (version 9.2.0, GraphPad Software, LLC, CA).

RESULTS

Performance of hematological PTS in SCD patients

We investigated the phenotypic variance explained by PTS of HT in African-ancestry SCD patients 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 African-ancestry individuals from the UK Biobank. These PTS were derived from large multi-ancestry GWAS meta-analyses involving > 700,000 participants.²³ In SCD participants, the PTS were significant (P -value < 0.05) in at least one cohort for nine of the 12 HT tested (**Table 1**). The non-significant PTS were for hematocrit, hemoglobin concentration and lymphocyte count. The variance explained by significant PTS ranged from 1.0 to 4.0% for RBC traits, 0.9 to 3.9% for WBC traits, and 0.6 to 4.0% for PLT traits. When we compared the performance of these PTS in SCD participants and non-SCD African-ancestry individuals, we found that all the 9 PTS with significant association explained less phenotypic variance in SCD participants (**Table 1** and **Figure 1**). One of the most striking differences was for WBC and neutrophil counts: the mean variance explained was 2.2% and 3.3% in SCD participants and 10.3% and 11.9% in non-SCD individuals, respectively (**Figure 1**).

HbF levels are an important modifier of severity in SCD. Because it is rarely measured in large non-SCD cohorts, the genetics of this trait has 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 β -globin locus (reviewed in ref. ³⁴). With this information, we derived an HbF PTS that includes six

conditionally independent variants (**Methods** and **Table S2**). This HbF PTS was strongly associated with HbF levels in all three SCD cohorts, and explained 17.1-26.4% of the variance (**Table 1**).

SCD partially masks the genetic effect of the Duffy/DARC null variant on WBC and neutrophil counts

To understand why the PTS under-performed in SCD patients, we carried out meta-analyses of GWAS results for the 11 HT in the three SCD cohorts, and compared effect sizes (β_{SCD}) for the SNPs found in the PTS (4,201 SNP-HT pairs) with the effect sizes from multi-ancestry meta-analyses ($\beta_{\text{non-SCD}}$, ref. ²³). Across all SNPs 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 2A**). Among the 273 variant-HT pairs that were nominally associated in SCD meta-analyses (P -value < 0.05), 162 (59.3%) had a concordant direction of effect in non-SCD meta-analyses ($P = 0.002$, binomial test).

After correction for multiple testing (q -value < 0.05), we found two HT-associated variants significantly associated with HT in SCD patients, but that also had a significantly different effect size when comparing β_{SCD} and $\beta_{\text{non-SCD}}$: rs8090527 and rs2814778 (**Table S3**). We did not explore the association between the intergenic rs8090517 variant and platelet count further as we could not replicate it in an independent SCD cohort (**Table S3**). The second variant is the previously described Duffy/DARC null variant (rs2814778),³⁵ and 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 counts 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 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** and **Figure 2B**). 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 2B**).

Next, we specifically focused on the association between Duffy/DARC rs2814778 and WBC or neutrophil counts in SCD and non-SCD individuals. For these analyses, we used 6,627 African-ancestry participants from the UK Biobank. 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 (**Table S4**). Thus, we used a recessive model for all subsequent genetic analyses of this variant. For non-SCD individuals, the single Duffy/DARC variant explained 18.4-23.1% of the phenotypic variance (**Table 2**). 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 2**). 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 0.76×10^9 WBC/L ($P = 0.004$) and 0.84×10^9 neutrophils/L ($P = 1.6 \times 10^{-6}$) in the CSSCD, and of 1.9×10^9 WBC/L

($P = 3.3 \times 10^{-164}$) and 1.6×10^9 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 2C**). 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 2D**). Whereas in non-SCD UK Biobank participants Duffy outweighs the PTS effect, it is equivalent in SCD patients. Put together, our data suggest that SCD partially masks the strong effect of the *Duffy*/DARC null variant (*rs2814778*) on neutrophil count. Finally, we investigated whether sickle cell trait (heterozygosity for the HbS allele) also modified the Duffy/DARC effect on neutrophil count. We did not find a significant interaction between these two genotypes in the UK Biobank ($P = 0.327$), suggesting that the epistatic effect is specific to SCD (i.e. homozygous HbS).

GWAS of HT in SCD patients

To determine if new genetic variation could specifically modulate HT variation in SCD, we carried out GWAS for 11 blood-cell traits across all three SCD cohorts available (**Methods**). 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 hematocrit and hemoglobin levels (**Figure S1**). We found 24 genome-wide significant ($P < 5 \times 10^{-8}$) SNP-HT associations, including 23 at the known HbF loci and associated with hemoglobin

levels, hematocrit, or RBC count (**Table S5**). The last variant, rs113819343, was associated with platelet (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 ($P = 0.72$, $N = 473,895$) nor African-specific ($P = 0.27$, $N = 15,171$) meta-analyses from the Blood-Cell Consortium.²³ 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.

Associations between hematological PTS and SCD-related complications

Variation in HT has been associated (prospectively) with several clinical outcomes observed in SCD patients such as death,⁷ stroke,³⁶ VOC,⁸ and ACS.⁹ We could reproduce most of those results in the genotyped subset of the CSSCD (**Table S6**). Having established that multi-ancestry PTS are associated with HT in SCD participants (although to a lesser extent than in non-SCD individuals), we asked if they were also associated with these complications in the CSSCD. For these analyses, we used nominally significant PTS (**Table 1**) for which the corresponding HT was also significantly associated with the complication (**Table S6**). In total, we tested six PTS-complication combinations, and found significant associations for PTS_{HbF}-stroke and PTS_{HbF}-ACS (**Table 3**).

We extended these analyses to determine if the PTS could improve the predictive performance of the models beyond the baseline HT measures. We reasoned that because the

PTS capture HT heritable variation, they would more faithfully represent “life-long exposure” and add information that is independent from HT lab measure imprecisions. The models did not improve for stroke and ACS rate, but we found that PTS_{HbF} improved the prediction of VOC, consistent with previous reports (**Table 3**).^{37,38} Interestingly, additional analyses revealed that PTS_{HbF} improved the prediction of VOC in patients with low HbF lab values (< 10%), levels at which HbF is not associated with VOC (**Table S7**). For high HbF values (> 10%), HbF is strongly associated with VOC and adding PTS_{HbF} did not improve the predictive model.

Quantifying the causal impact of HbF levels on SCD complications by MR

Finally, we sought to confirm some of the causal effects of HT on SCD complications as reported in the literature: the protection that high HbF levels confer against stroke,^{39,40} ACS⁹ and VOC,⁸ as well as the increased risk of death associated with high WBC⁷ and neutrophil counts.⁴¹ To this end, we performed two-sample MR using the pseudo-independent SNPs from the PTS of HbF, WBC and neutrophils (**Methods**). 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) decreases by 36% the risk of stroke (odds ratio [95% confidence interval] = 0.64 [0.49-0.83], $P = 0.0008$, **Figure 3**). Although the direction of the effect was similar using MR-Egger and weighted median, these analyzes were not statistically significant, suggesting insufficient power (**Table S8**). We found no heterogeneity in the effect ($I^2 = 0\%$) and confirmed the absence of horizontal pleiotropy (Egger intercept, -0.07, $P = 0.75$). All other MR tests were non-significant (**Table S8**).

DISCUSSION

In this study, we showed that PTS for HT derived in non-SCD individuals largely underperformed when tested in SCD patients. We hypothesize that this is due to an epistatic effect of the SCD genotype on the large number of variants that contribute to polygenic variation in HT. In particular, we found a dramatic impact of SCD on the magnitude of the effect of the Duffy/*DARC* null rs2814778 variant on neutrophil count. Potentially because the phenotypic variance explained by these PTS for HT remain low, we could not detect an effect of these PTS on disease expressivity when considering different SCD-related complications. However, a small set of six HbF-associated variants was useful to predict VOC when considered as a PTS, and to quantify the causal protective impact of HbF increase on stroke risk reduction.

Several non-mutually exclusive factors could explain why PTS for HT were not as predictive in SCD patients. These patients have different baseline levels and distribution values for several HT such as hemoglobin and WBC counts. 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 cytokines-driven higher WBC count.^{42,43} Moreover, the frequent intercurrent complications (*e.g.* VOC) experienced throughout SCD natural history could result in greater variability in HT values. Finally, SNP genotyping arrays do not capture all structural variants, which are the main alterations in α -thalassemia, a major determinant of mean corpuscular hemoglobin and mean corpuscular volume. Thus, the performance of HT PTS will improve once comprehensive whole-genome sequencing of 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* (also known as *ACKR1*) chemokine receptor expression gene.^{44,45} It has been known 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,⁴⁶⁻⁵⁰ and in particular the spleen.⁵¹ 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. Additionally, 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 8.⁵²⁻⁵⁴ Definitive data are still lacking, but the proinflammatory state of SCD patients may contribute to the discrepancy in effect observed.⁵⁵

Whether or not the Duffy phenotype is associated with complications in SCD patients remains unclear.⁵⁶⁻⁶¹ The potential consequences of the Duffy/*DARC* status in SCD would be linked to circulating proinflammatory cytokines and especially the resulting effect on neutrophil and WBC counts, which has been implicated in several SCD complications.^{7,12,13} 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. Thus, while it remains possible that neutrophils causally contribute to SCD clinical heterogeneity, we have not been able to use genetics, including MR, to confirm this

epidemiological observation owing to the relatively weak effect of the neutrophil-associated variants in SCD patients.

Although complex multi-ancestry PTS for general HT that included 100s 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. One major difference between the general HT and HbF is that we developed PTS_{HbF} using GWAS data from SCD patients. Interestingly, we observed an association between PTS_{HbF} and ACS or stroke in the large CSSCD. We validated previous reports that a PTS for HbF can improve the prediction of VOC,^{37,38} and further discovered that this PTS_{HbF} was useful in the subset of patients with HbF levels < 10%. Our limited sample size prevented powerful analyses of causality, yet we could use MR to quantify the protective effect of HbF on stroke risk in SCD patients, although we acknowledge that the Winner's curse could have biased upward our estimate. This result is consistent with clinical trials that have shown that the HbF-inducing drug hydroxyurea (HU) can reduce stroke risk in primary prevention in SCD.^{62,63} Furthermore, although HU can improve SCD through different mechanisms,⁶⁴ our MR analysis is useful to partition its effect and specifically quantify how genetically-determined (lifelong) HbF levels modulate stroke risk. Given its hematological rationale, simplicity, and performance, we believe that the time is right to formally test in a pragmatic randomized clinical trial whether PTS_{HbF} is useful to manage clinical heterogeneity in SCD patients.

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 (CAD) and familial breast and colorectal cancers,⁶⁵ much less is known about their effect on expressivity (or disease severity).⁴ 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 or CAD, PTS for retinopathy/cataract in diabetic patients). Our analyses indicate that simply translating the genetics of polygenic traits learned 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 enable powerful GWAS for genetic modifiers in > 10,000 patients who all suffer from the same disease.

SUPPLEMENTAL DATA

Supplemental data include one figure and eight tables.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENTS

We thank all participants for their contribution to this project. We thank Gabrielle Boucher for statistical support. T.P. 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. This work was funded by the Canadian Institutes of Health Research (PJT #156248), Bioverativ, a Sanofi Company, and the Canada Research Chair Program (to G.L.). GEN-MOD samples and data collection were supported by NIH grant HL-68922. A.A-K., M.J.T. 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 AND CODE AVAILABILITY

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.

AUTHOR CONTRIBUTIONS

T.P. and G.L. designed the study. T.P., K.S.L. and M.E.G. performed the analyses. T.P., A.L.P.H.A.O., M.E.G, C.B., A.E.A.-K., M.J.T., F.G., P.J., P.B and G.L. collected the clinical and genetic data. T.P. and G.L. drafted the paper. G.L. supervised the study. All authors participated in data interpretation, revised the manuscript for critical content and approved the final manuscript.

REFERENCES

1. Lewis, C.M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* *12*, 44.
2. Sun, J., Wang, Y., Folkersen, L., Borné, Y., Amlie, I., Buil, A., Orho-Melander, M., Børghlum, A.D., Hougaard, D.M., Regeneron Genetics Center, et al. (2021). Translating polygenic risk scores for clinical use by estimating the confidence bounds of risk prediction. *Nat. Commun.* *12*, 5276.
3. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* *51*, 584–591.
4. Oetjens, M.T., Kelly, M.A., Sturm, A.C., Martin, C.L., and Ledbetter, D.H. (2019). Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nat. Commun.* *10*, 4897.
5. Kato, G.J., Piel, F.B., Reid, C.D., Gaston, M.H., Ohene-Frempong, K., Krishnamurti, L., Smith, W.R., Panepinto, J.A., Weatherall, D.J., Costa, F.F., et al. (2018). Sickle cell disease. *Nat Rev Dis Primers* *4*, 18010.
6. Ware, R.E., de Montalembert, M., Tshilolo, L., and Abboud, M.R. (2017). Sickle cell disease. *Lancet* *390*, 311–323.

7. Platt, O.S., Brambilla, D.J., Rosse, W.F., Milner, P.F., Castro, O., Steinberg, M.H., and Klug, P.P. (1994). Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N. Engl. J. Med.* *330*, 1639–1644.
8. Platt, O.S., Thorington, B.D., Brambilla, D.J., Milner, P.F., Rosse, W.F., Vichinsky, E., and Kinney, T.R. (1991). Pain in sickle cell disease. Rates and risk factors. *N. Engl. J. Med.* *325*, 11–16.
9. Castro, O., Brambilla, D.J., Thorington, B., Reindorf, C.A., Scott, R.B., Gillette, P., Vera, J.C., and Levy, P.S. (1994). The acute chest syndrome in sickle cell disease: incidence and risk factors. The Cooperative Study of Sickle Cell Disease. *Blood* *84*, 643–649.
10. Ngo, D.A., Aygun, B., Akinsheye, I., Hankins, J.S., Bhan, I., Luo, H.Y., Steinberg, M.H., and Chui, D.H.K. (2012). 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.
11. Steinberg, M.H., Chui, D.H.K., Dover, G.J., Sebastiani, P., and Al Sultan, A. (2014). Fetal hemoglobin in sickle cell anemia: a glass half full? *Blood* *123*, 481–485.
12. Sebastiani, P., Nolan, V.G., Baldwin, C.T., Abad-Grau, M.M., Wang, L., Adewoye, A.H., McMahan, L.C., Farrer, L.A., Taylor, J.G., 6th, Kato, G.J., et al. (2007). A network model to predict the risk of death in sickle cell disease. *Blood* *110*, 2727–2735.
13. Elmariah, H., Garrett, M.E., De Castro, L.M., Jonassaint, J.C., Ataga, K.I., Eckman, J.R., Ashley-Koch, A.E., and Telen, M.J. (2014). Factors associated with survival in a contemporary adult sickle cell disease cohort. *Am. J. Hematol.* *89*, 530–535.
14. Miller, S.T., Sleeper, L.A., Pegelow, C.H., Enos, L.E., Wang, W.C., Weiner, S.J., Wethers, D.L., Smith, J., and Kinney, T.R. (2000). Prediction of adverse outcomes in children with sickle cell disease. *N. Engl. J. Med.* *342*, 83–89.
15. Vichinsky, E.P., Neumayr, L.D., Earles, A.N., Williams, R., Lennette, E.T., Dean, D., Nickerson, B., Orringer, E., McKie, V., Bellevue, R., et al. (2000). Causes and outcomes of the acute chest syndrome in sickle cell disease. National Acute Chest Syndrome Study Group. *N. Engl. J. Med.* *342*, 1855–1865.
16. Farber, M.D., Koshy, M., and Kinney, T.R. (1985). Cooperative Study of Sickle Cell Disease: Demographic and socioeconomic characteristics of patients and families with sickle cell disease. *J. Chronic Dis.* *38*, 495–505.
17. Bartolucci, P., Brugnara, C., Teixeira-Pinto, A., Pissard, S., Moradkhani, K., Jouault, H., and Galacteros, F. (2012). Erythrocyte density in sickle cell syndromes is associated with specific clinical manifestations and hemolysis. *Blood* *120*, 3136–3141.
18. Ilboudo, Y., Bartolucci, P., Rivera, A., Sedzro, J.-C., Beaudoin, M., Trudel, M., Alper, S.L.,

Brugnara, C., Galactéros, F., and Lettre, G. (2017). Genome-wide association study of erythrocyte density in sickle cell disease patients. *Blood Cells Mol. Dis.* 65, 60–65.

19. Ilboudo, Y., Garrett, M.E., Bartolucci, P., Brugnara, C., Clish, C.B., Hirschhorn, J.N., Galactéros, F., Ashley-Koch, A.E., Telen, M.J., and Lettre, G. (2021). Potential causal role of l-glutamine in sickle cell disease painful crises: A Mendelian randomization analysis. *Blood Cells Mol. Dis.* 86, 102504.

20. Pincez, T., Lee, S.S.K., Ilboudo, Y., Preuss, M.H., Pham Hung d’Alexandry d’Orengiani, A.-L., Bartolucci, P., Galacteros, F., Joly, P., Bauer, D.E., Loos, R., et al. (2021). Clonal hematopoiesis in sickle cell disease. *Blood* 138, 2148–2152.

21. Bae, H.T., Baldwin, C.T., Sebastiani, P., Telen, M.J., Ashley-Koch, A., Garrett, M., Hooper, W.C., Bean, C.J., Debaun, M.R., Arking, D.E., et al. (2012). Meta-analysis of 2040 sickle cell anemia patients: BCL11A and HBS1L-MYB are the major modifiers of HbF in African Americans. *Blood* 120, 1961–1962.

22. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299.

23. Chen, M.-H., Raffield, L.M., Mousas, A., Sakaue, S., Huffman, J.E., Moscati, A., Trivedi, B., Jiang, T., Akbari, P., Vuckovic, D., et al. (2020). Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* 182, 1198–1213.e14.

24. Belbin, G.M., Cullina, S., Wenric, S., Soper, E.R., Glicksberg, B.S., Torre, D., Moscati, A., Wojcik, G.L., Shemirani, R., Beckmann, N.D., et al. (2021). Toward a fine-scale population health monitoring system. *Cell* 184, 2068–2083.e11.

25. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O’Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209.

26. Galarneau, G., Palmer, C.D., Sankaran, V.G., Orkin, S.H., Hirschhorn, J.N., and Lettre, G. (2010). Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat. Genet.* 42, 1049–1051.

27. Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., et al. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* 342, 253–257.

28. Canver, M.C., Lessard, S., Pinello, L., Wu, Y., Ilboudo, Y., Stern, E.N., Needleman, A.J., Galactéros, F., Brugnara, C., Kutlar, A., et al. (2017). Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat. Genet.* 49, 625–634.

29. Zhan, X., Hu, Y., Li, B., Abecasis, G.R., and Liu, D.J. (2016). RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 32, 1423–1426.
30. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191.
31. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
32. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 7,.
33. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpeläinen, T.O., et al. (2015). The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet.* 11, e1005378.
34. Lettre, G., and Bauer, D.E. (2016). Fetal haemoglobin in sickle-cell disease: from genetic epidemiology to new therapeutic strategies. *Lancet* 387, 2554–2564.
35. Reich, D., Nalls, M.A., Kao, W.H.L., Akylbekova, E.L., Tandon, A., Patterson, N., Mullikin, J., Hsueh, W.-C., Cheng, C.-Y., Coresh, J., et al. (2009). Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet.* 5, e1000360.
36. Ohene-Frempong, K., Weiner, S.J., Sleeper, L.A., Miller, S.T., Embury, S., Moohr, J.W., Wethers, D.L., Pegelow, C.H., and Gill, F.M. (1998). Cerebrovascular accidents in sickle cell disease: rates and risk factors. *Blood* 91, 288–294.
37. Lettre, G., Sankaran, V.G., Bezerra, M.A.C., Araújo, A.S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F.F., Hirschhorn, J.N., et al. (2008). 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.* 105, 11869–11874.
38. Rampersaud, E., Kang, G., Palmer, L.E., Rashkin, S.R., Wang, S., Bi, W., Alberts, N.M., Angheliescu, D., Barton, M., Birch, K., et al. (2021). A polygenic score for acute vaso-occlusive pain in pediatric sickle cell disease. *Blood Adv* 5, 2839–2851.
39. Sommet, J., Alberti, C., Couque, N., Verlhac, S., Haouari, Z., Mohamed, D., François, M., Missud, F., Holvoet, L., Elmaleh, M., et al. (2016). Clinical and haematological risk factors for cerebral macrovasculopathy in a sickle cell disease newborn cohort: a prospective study. *Br. J. Haematol.* 172, 966–977.
40. Calvet, D., Tuilier, T., Mélé, N., Turc, G., Habibi, A., Abdallah, N.A., Majhadi, L., Hemery, F.,

Edjlali, M., Galacteros, F., et al. (2017). Low fetal hemoglobin percentage is associated with silent brain lesions in adults with homozygous sickle cell disease. *Blood Adv* *1*, 2503–2509.

41. Nouraie, M., Darbari, D.S., Rana, S., Minniti, C.P., Castro, O.L., Luchtman-Jones, L., Sable, C., Dham, N., Kato, G.J., Gladwin, M.T., et al. (2020). Tricuspid regurgitation velocity and other biomarkers of mortality in children, adolescents and young adults with sickle cell disease in the United States: The PUSH study. *Am. J. Hematol.* *95*, 766–774.

42. Nader, E., Romana, M., and Connes, P. (2020). The Red Blood Cell-Inflammation Vicious Circle in Sickle Cell Disease. *Front. Immunol.* *11*, 454.

43. Hermand, P., Azouzi, S., Gautier, E.-F., Guillonneau, F., Bondet, V., Duffy, D., Dechavanne, S., Tharaux, P.-L., Mayeux, P., Le Van Kim, C., et al. (2020). The proteome of neutrophils in sickle cell disease reveals an unexpected activation of interferon alpha signaling pathway. *Haematologica* *105*, 2851–2854.

44. Peiper, S.C., Wang, Z.X., Neote, K., Martin, A.W., Showell, H.J., Conklyn, M.J., Ogborne, K., Hadley, T.J., Lu, Z.H., Hesselgesser, J., et al. (1995). The Duffy antigen/receptor for chemokines (DARC) is expressed in endothelial cells of Duffy negative individuals who lack the erythrocyte receptor. *J. Exp. Med.* *181*, 1311–1317.

45. Tournamille, C., Colin, Y., Cartron, J.P., and Le Van Kim, C. (1995). Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat. Genet.* *10*, 224–228.

46. Grann, V.R., Ziv, E., Joseph, C.K., Neugut, A.I., Wei, Y., Jacobson, J.S., Horwitz, M.S., Bowman, N., Beckmann, K., and Hershman, D.L. (2008). Duffy (Fy), DARC, and neutropenia among women from the United States, Europe and the Caribbean. *Br. J. Haematol.* *143*, 288–293.

47. Reich, D., Nalls, M.A., Kao, W.H.L., Akylbekova, E.L., Tandon, A., Patterson, N., Mullikin, J., Hsueh, W.-C., Cheng, C.-Y., Coresh, J., et al. (2009). Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet.* *5*, e1000360.

48. Nalls, M.A., Wilson, J.G., Patterson, N.J., Tandon, A., Zmuda, J.M., Huntsman, S., Garcia, M., Hu, D., Li, R., Beamer, B.A., et al. (2008). Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am. J. Hum. Genet.* *82*, 81–87.

49. Afenyi-Annan, A., Kail, M., Combs, M.R., Orringer, E.P., Ashley-Koch, A., and Telen, M.J. (2008). Lack of Duffy antigen expression is associated with organ damage in patients with sickle cell disease. *Transfusion* *48*, 917–924.

50. Schaefer, B.A., Flanagan, J.M., Alvarez, O.A., Nelson, S.C., Aygun, B., Nottage, K.A., George, A., Roberts, C.W., Piccone, C.M., Howard, T.A., et al. (2016). Genetic Modifiers of White Blood

Cell Count, Albuminuria and Glomerular Filtration Rate in Children with Sickle Cell Anemia. *PLoS One* *11*, e0164364.

51. Duchene, J., Novitzky-Basso, I., Thiriot, A., Casanova-Acebes, M., Bianchini, M., Etheridge, S.L., Hub, E., Nitz, K., Artinger, K., Eller, K., et al. (2017). Atypical chemokine receptor 1 on nucleated erythroid cells regulates hematopoiesis. *Nat. Immunol.* *18*, 753–761.

52. Schnabel, R.B., Baumert, J., Barbalic, M., Dupuis, J., Ellinor, P.T., Durda, P., Dehghan, A., Bis, J.C., Illig, T., Morrison, A.C., et al. (2010). Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. *Blood* *115*, 5289–5299.

53. Moreno Velásquez, I., Kumar, J., Björkbacka, H., Nilsson, J., Silveira, A., Leander, K., Berglund, A., Strawbridge, R.J., Ärnlov, J., Melander, O., et al. (2015). Duffy antigen receptor genetic variant and the association with Interleukin 8 levels. *Cytokine* *72*, 178–184.

54. Nebor, D., Durpes, M.C., Mouguel, D., Mukisi-Mukaza, M., Elion, J., Hardy-Dessources, M.-D., and Romana, M. (2010). Association between Duffy antigen receptor for chemokines expression and levels of inflammation markers in sickle cell anemia patients. *Clin. Immunol.* *136*, 116–122.

55. van Beers, E.J., Yang, Y., Raghavachari, N., Tian, X., Allen, D.T., Nichols, J.S., Mendelsohn, L., Nekhai, S., Gordeuk, V.R., Taylor, J.G., 6th, et al. (2015). Iron, inflammation, and early death in adults with sickle cell disease. *Circ. Res.* *116*, 298–306.

56. Drasar, E.R., Menzel, S., Fulford, T., and Thein, S.L. (2013). The effect of Duffy antigen receptor for chemokines on severity in sickle cell disease. *Haematologica* *98*, e87–e89.

57. Afenyi-Annan, A., Kail, M., Combs, M.R., Orringer, E.P., Ashley-Koch, A., and Telen, M.J. (2008). Lack of Duffy antigen expression is associated with organ damage in patients with sickle cell disease. *Transfusion* *48*, 917–924.

58. Nolan, V.G., Adewoye, A., Baldwin, C., Wang, L., Ma, Q., Wyszynski, D.F., Farrell, J.J., Sebastiani, P., Farrer, L.A., and Steinberg, M.H. (2006). Sickle cell leg ulcers: associations with haemolysis and SNPs in Klotho, TEK and genes of the TGF-beta/BMP pathway. *Br. J. Haematol.* *133*, 570–578.

59. Elliott, L., Ashley-Koch, A.E., De Castro, L., Jonassaint, J., Price, J., Ataga, K.I., Levesque, M.C., Brice Weinberg, J., Eckman, J.R., Orringer, E.P., et al. (2007). Genetic polymorphisms associated with priapism in sickle cell disease. *Br. J. Haematol.* *137*, 262–267.

60. Schnog, J.B., Keli, S.O., Pieters, R.A., Rojer, R.A., and Duits, A.J. (2000). Duffy phenotype does not influence the clinical severity of sickle cell disease. *Clin. Immunol.* *96*, 264–268.

61. Farawela, H.M., El-Ghamrawy, M., Farhan, M.S., Soliman, R., Yousry, S.M., and AbdelRahman, H.A. (2016). Association between Duffy antigen receptor expression and disease

severity in sickle cell disease patients. *Hematology* 21, 474–479.

62. Bernaudin, F., Verlhac, S., Arnaud, C., Kamdem, A., Hau, I., Leveillé, E., Vasile, M., Kasbi, F., Madhi, F., Fourmaux, C., et al. (2016). Long-term treatment follow-up of children with sickle cell disease monitored with abnormal transcranial Doppler velocities. *Blood* 127, 1814–1822.

63. Ware, R.E., Davis, B.R., Schultz, W.H., Brown, R.C., Aygun, B., Sarnaik, S., Odame, I., Fuh, B., George, A., Owen, W., et al. (2016). 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* 387, 661–670.

64. Platt, O.S. (2008). Hydroxyurea for the treatment of sickle cell anemia. *N. Engl. J. Med.* 358, 1362–1369.

65. Fahed, A.C., Wang, M., Homburger, J.R., Patel, A.P., Bick, A.G., Neben, C.L., Lai, C., Brockman, D., Philippakis, A., Ellinor, P.T., et al. (2020). Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat. Commun.* 11, 3635.

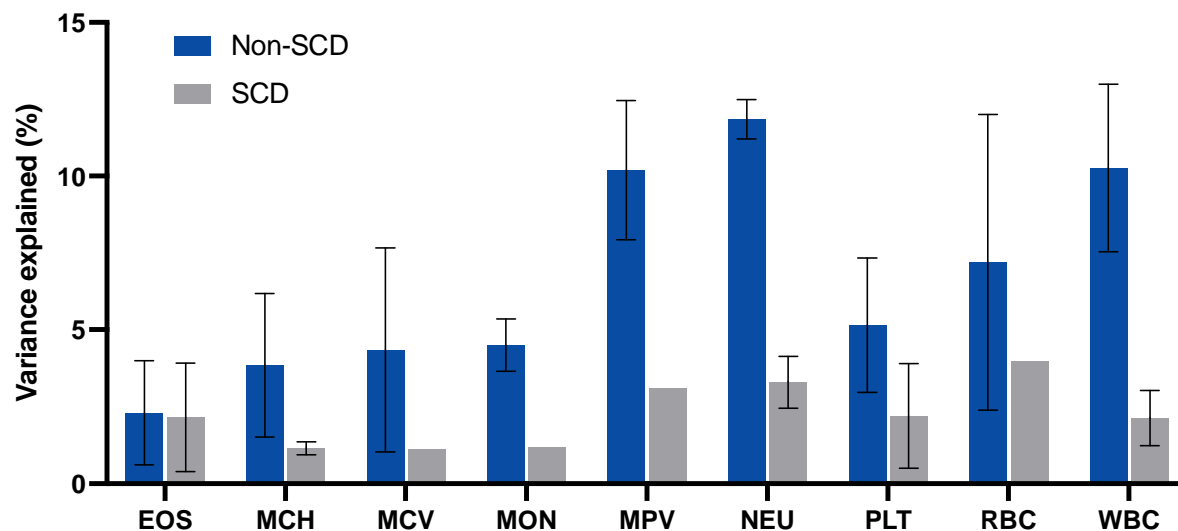


Figure 1. Variance explained by polygenic trait scores (PTS) for hematological traits (HT) in African-ancestry non-sickle cell disease (SCD) individuals from BioMe and the UK Biobank, and participants from three SCD cohorts (CSSCD, GEN-MOD, Mondor/Lyon). We only present the variance explained by nominally 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. 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.

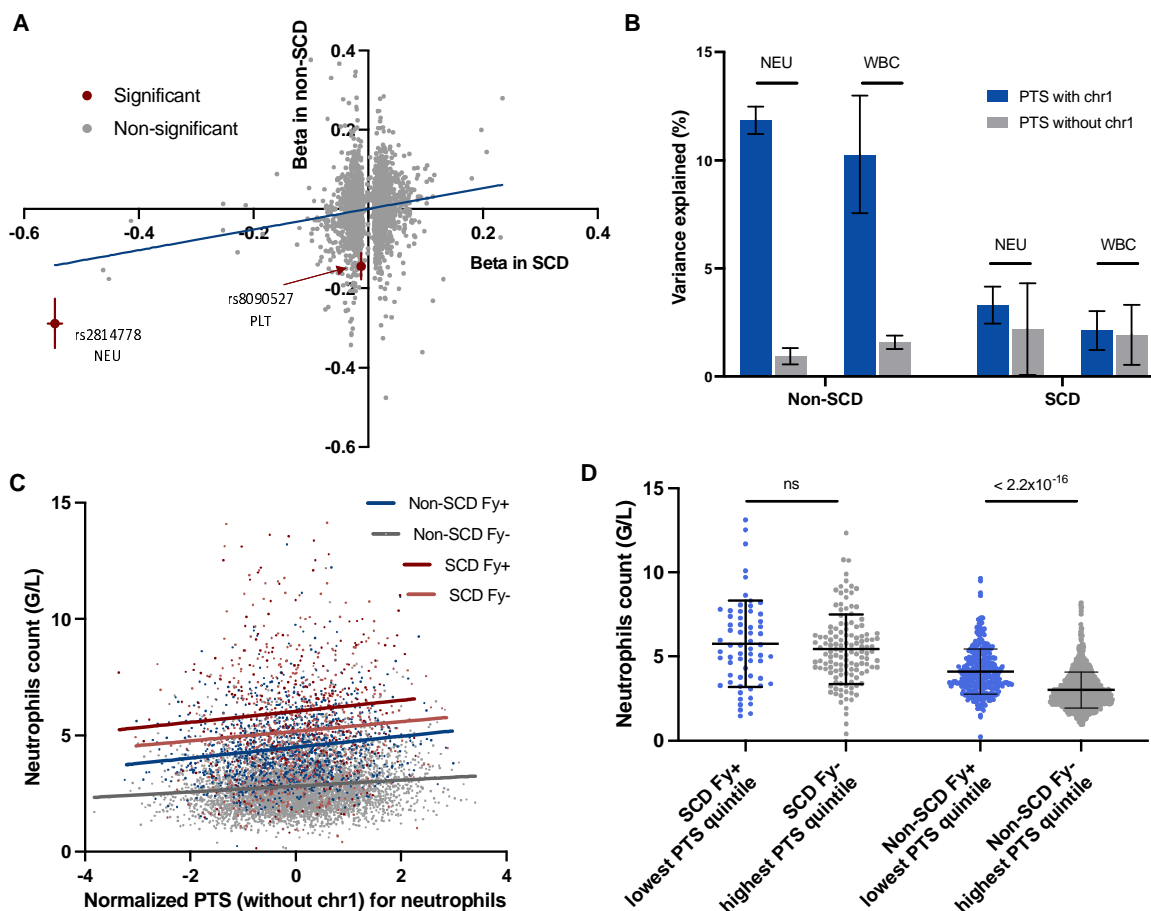


Figure 2. Sickle cell disease (SCD) partially masks the genetic impact of the Duffy/*DARC* null variant (rs2814778) on neutrophil count. **(A)** Comparison of effect sizes (Beta) for 3,917 SNP-hematological trait pairs in SCD patients (x-axis) and non-SCD participants (y-axis). We only considered variants with a minor allele frequency > 1% in the SCD meta-analyses. 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 counts (WBC) with and without chromosome (chr) 1 in SCD and non-SCD individuals. **(C)** Raw neutrophil count (y-axis) as a function of the multi-ancestry NEU 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 or non-SCD 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 and non-SCD participants. ns: non-significant.

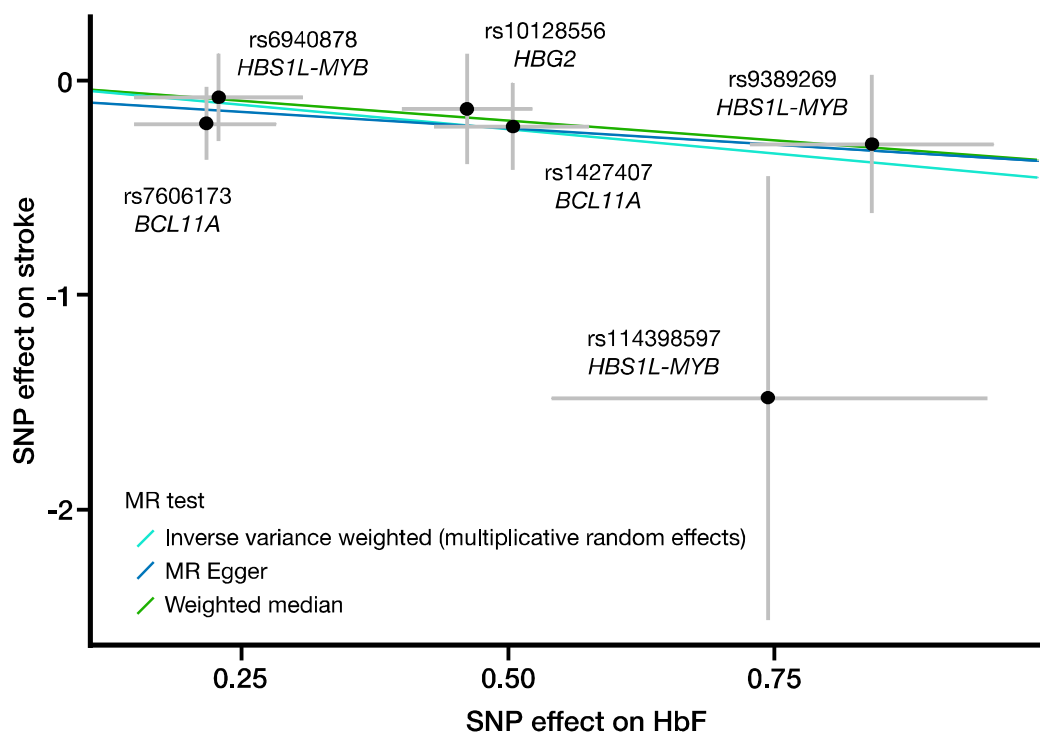


Figure 3. Mendelian randomization (MR) results for fetal hemoglobin (HbF) levels on stroke. Each dot represents one of the HbF-associated SNP, with its corresponding effect on normalized HbF levels (x-axis, standard deviation units) and stroke risk (y-axis, logistic regression beta). The analysis without rs114398597 is significant; see **Table S8** for details.

Table 1. Hematological trait (HT) variance explained by polygenic trait scores (PTS) in sickle cell disease (SCD) patients. We only calculated the phenotypic variance explained for PTS that were nominally significant (P -value < 0.05). Grey shaded cells indicate phenotypes that were not available in the corresponding study. The ^aPTS models (except for fetal hemoglobin [HbF]) and the results from the ^bBioMe African American participants were reported previously.²³ For HbF, we selected six independently associated variants.^{26–28} MCH: mean corpuscular hemoglobin, MCV: mean corpuscular volume, MPV: mean platelet volume.

Phenotype	Number of variants in the PTS ^a	CSSCD (N _{max} = 1015)		GEN-MOD (N _{max} = 401)		Mondor/Lyon (N _{max} = 324)		BioME ^b (non-SCD) (N _{max} = 2,651)		UK Biobank (non-SCD) (N _{max} = 6,584)	
		Variance (%)	P-value	Variance (%)	P-value	Variance (%)	P-value	Variance (%)	P-value	Variance (%)	P-value
<i>Red blood cell (RBC) traits</i>											
Hematocrit	356	-	0.77	-	0.78	-	0.10	1.4	0.0071	2.1	< 2.2x10 ⁻¹⁶
Hemoglobin	364	-	1.0	-	0.23	-	0.48	1.3	7.4x10 ⁻⁹	3.0	< 2.2x10 ⁻¹⁶
MCH	384	1.0	6.1x10 ⁻⁴	1.3	0.021	-	0.95	2.2	5.4x10 ⁻¹⁴	5.5	< 2.2x10 ⁻¹⁶
MCV	454	1.1	4.4x10 ⁻⁵	-	0.054	-	0.35	2.0	9.1x10 ⁻¹³	6.7	< 2.2x10 ⁻¹⁶
RBC	449	-	0.093	4.0	7.1x10 ⁻⁵	-	0.15	3.8	< 2.2x10 ⁻¹⁶	10.6	< 2.2x10 ⁻¹⁶
HbF	6	17.1	< 2.2x10 ⁻¹⁶	22.5	< 2.2x10 ⁻¹⁶	26.4	< 2.2x10 ⁻¹⁶				
<i>White blood cell (WBC) traits</i>											
WBC	443	1.2	7.2x10 ⁻⁵	3.0	0.0010	2.2	0.019	12.2	< 2.2x10 ⁻¹⁶	8.3	< 2.2x10 ⁻¹⁶
WBC (without chr1 - Duffy/DARC)	389	0.4	0.013	3.1	3.8x10 ⁻⁴	2.3	0.0067	1.8	< 2.2x10 ⁻¹⁶	1.4	< 2.2x10 ⁻¹⁶
Eosinophils	346	0.9	0.0045	3.4	2.7x10 ⁻⁴			1.1	6.2x10 ⁻⁶	3.5	< 2.2x10 ⁻¹⁶
Lymphocytes	409	-	0.21	-	0.38			1.4	7.7x10 ⁻⁹	3.6	< 2.2x10 ⁻¹⁶
Monocytes	394	1.2	0.00041	-	0.15			3.9	< 2.2x10 ⁻¹⁶	5.1	< 2.2x10 ⁻¹⁶
Neutrophils	352	2.7	2.2x10 ⁻⁷	3.9	2.7x10 ⁻⁴			12.3	< 2.2x10 ⁻¹⁶	11.4	< 2.2x10 ⁻¹⁶
NEU (without chr1 - Duffy/DARC)	313	0.7	0.0059	3.7	1.4x10 ⁻⁴			0.67	1.1x10 ⁻⁴	1.2	< 2.2x10 ⁻¹⁶
<i>Platelet (PLT) traits</i>											
PLT	553	0.6	0.0058	2.0	0.0046	4.0	2.8x10 ⁻⁴	3.6	< 2.2x10 ⁻¹⁶	6.7	< 2.2x10 ⁻¹⁶
MPV	392			3.1	9.9x10 ⁻⁴			8.6	< 2.2x10 ⁻¹⁶	11.8	< 2.2x10 ⁻¹⁶

Table 2. Genetic association results between the Duffy/DARC null variant (rs2814778) and white blood cell (WBC) and neutrophil counts. 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. EAF, effect allele frequency.

	N	EAF (C-allele)	Beta (SE)	P-value	Variance (%)
<i>Neutrophils</i>					
CSSCD	934	0.842	-0.353 (0.073)	1.6×10^{-6}	3.3
GEN-MOD	400	0.939	-0.019 (0.209)	0.93	-
UK Biobank	6,564	0.906	-1.166 (0.033)	6.0×10^{-239}	23.1
<i>White blood cells</i>					
CSSCD	1014	0.845	-0.178 (0.064)	0.005	0.9
GEN-MOD	400	0.939	-0.092 (0.208)	0.66	-
Mondor/Lyon	322	0.935	-0.151 (0.214)	0.48	-
UK Biobank	6,584	0.906	-1.041 (0.036)	5.0×10^{-171}	18.4

Table 3. The polygenic trait score (PTS) for fetal hemoglobin (HbF) levels is associated with stroke and acute chest syndrome (ACS), and improves the prediction of vaso-occlusive crises (VOC) rates. We carried out these analyses in 1,278 genotyped CSSCD participants. To compare predictive models, we performed an analysis of deviance and compared a baseline model (HT, age, sex, α -thalassemia) with a model that included the same predictors as well as the PTS (Methods).

Outcome	PTS	Association of the PTS with the outcome		Comparison of predictive models with and without PTS	
		HR [95% CI] or beta (SE)	P-value	χ^2	P-value
Stroke (N _{cases} = 104, N _{controls} = 1,168)	HbF	0.75 [0.61-0.91]	0.004	0.14	0.70
	WBC	1.01 [0.84-1.22]	0.92	2.07	0.15
	NEU	0.98 [0.81-1.19]	0.86	3.21	0.07
ACS rate (N = 1,271)	HbF	-0.21 (0.06)	0.0003	0.231	0.63
	EOS	-0.02 (0.05)	0.78	0.68	0.41
VOC rate (N = 1,271)	HbF	0.05 (0.05)	0.26	14.2	0.0002