Variation and impact of polygenic hematologic 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] and 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 n. 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'Etablissements (COMUE), Lyon, France and [9]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

SUPPLEMENTARY DATA

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]*

SUPPLEMENTARY METHODS

Populations

All SCD participants were of African ancestry. All patients were at > 3 months from a blood transfusion and only ten patients were taking hydroxyurea (HU) at baseline. DNA genotyping of SCD cohorts has been described elsewhere.[1, 2] We imputed missing genotypes on TOPMed freeze 5 using the Michigan Imputation Server and conserved only SNPs with sufficient imputation quality (r2 > 0.3).[3]

We used the definition from the clinical CSSCD papers to define clinical outcomes. Stroke was defined as an acute neurologic syndrome secondary to occlusion of an artery or hemorrhage with resultant ischemia and neurologic symptoms and signs and thus included transient ischemic attack, completed infarctive stroke (neurologic deficits lasting more than 24 hours), and hemorrhagic stroke.[4, 5] Vaso-occlusive crisis (VOC) was defined as the occurrence of pain in the extremities, back, abdomen, chest, or head that lasted at least two hours, led to a clinic visit, and could not be explained except by sickle cell disease.[6] Acute chest syndrome (ACS) diagnosis was retained each time a patient (1) developed a new infiltrate on chest x-ray and/or (2) had a perfusion defect demonstrable on a lung radioisotope scan.[7]

PTS for HT

PTS derived by the Blood-Cell Consortium considered the effect sizes of variants that reach genome-wide significance ($P < 5x10^{-8}$) in multi-ancestry meta-analyses of 746,667 individuals, including 15,171 African-ancestry participants.[8] We generated additive PTS for each individual

and HT by calculating the sum of HT-increasing alleles weighted by the corresponding multi-ancestry GWAS effect size. We did not test effect sizes derived exclusively from African-ancestry meta-analyses because we showed previously that they under-performed multi-ancestry effect sizes.[8]

To design a PTS for HbF, we used conditional analyses to select independent variants at *BCL11A*, *HBS1L-MYB*, and β-globin loci. We retained 2 independent variants at *BCL11A* (rs1427407, rs7606173), 3 independent variants at *HBS1L-MYB* (rs6940878, rs9389269, rs114398597), and one variant at the β-globin locus (rs10128556, **Table S2**).[9-11] We adjusted each HT for sex and age, and then applied inverse normal transformation. We tested the association between PTS z-score and normalized HT by linear regression with the 10 first principal components (PCs) as covariables.

To compare PTS performance in SCD and non-SCD cohorts, we bootstrapped 1,000 times 1,278 African-ancestry individuals from the UK Biobank cohort and compared the variance explained in each bootstrapped subset to the variance explained in the SCD cohorts (considering only cohorts with significant association for the given HT). We derived an empirical P-value which corresponds to the number of bootstrapped iterations in which the variance explained in the SCD cohort(s) was lower than in the non-SCD individuals.

We also tested the previously published 4-SNPs model for HbF (g(HbF)).[12] This model includes the rs7482144 SNP (*Xmn*I polymorphism located in the Gγ-globin gene promoter) for which the alternate allele is specific to the Senegal and the Arab-Indian sickle cell haplotypes. This SNP was not genotyped in our datasets and its imputation was not possible as there are

no linkage disequilibrium proxies in the reference TOPMed African-ancestry dataset.[3] We previously genotyped this SNP in a subset of the CSSCD,[13] and were able to test g(HbF) in 816 CSSCD participants for whom the data for the four SNPs was available.

We analyzed the respective contribution of α-thalassemia and PTS on HT variance explained in the CSSCD cohort by adding α-thalassemia as a covariate in the previous models. We reported the adjusted variance explained by each of the terms

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 for which corresponding HT was nominally associated with the outcome ($P < 0.05$). 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. To consider that death could be a competitive risk for stroke, we used Fine and Gray's method[14] to fit the subdistribution hazard model with the same covariates. Second, we repeated these analyses after replacing the HT by the corresponding PTS. To determine if the PTS improves model 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).

We attempted to replicate the association between PTS for HbF and stroke in the GEN-MOD cohort but were limited by the low cumulative incidence of stroke (3.4% in GEN-MOD,

i.e. n = 14 stroke cases) vs 8.3% in the CSSCD). Using G*Power,[15] we computed the sample size needed to replicate our PTS_{HbF}-stroke results from the CSSCD using logistic regression and the following parameters: odds ratio (OR) = 0.75, α = 0.05, power = 0.80 and raw incidence under null hypothesis (p1) = 0.034 (GEN-MOD adult data) or 0.018 (considering the current primary prevention in children using data reported elsewhere).[16] We estimated that 2,800 individuals would be required with the GEN-MOD incidence rate and 5,180 with the published cumulative incidence rate in children.

Mendelian randomization (MR)

For MR analyses, we initially focused on the following combinations of HT and complications: WBC and neutrophil counts for survival,[17, 18] and HbF for VOC rate,[19] ACS rate,[7] and stroke.[20, 21] For WBC and neutrophil counts, we used the SNPs-HT effects from 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).[22] To determine if we had sufficient power to detect causality between WBC/NEU and survival in SCD patients, we performed power calculations using the mRnd tool.[23] Under a series of realistic assumptions and parameters (N = 1,278, 44 cases, α = 0.05, odds ratio (OR) of exposure on outcome 1.15 and variance explained by the PTS ~3-3.9%), we calculated < 5% power. For this reason, we do not report MR results for survival in our study. For HbF, we selected as instruments the six SNPs used in the PTS analyses. These variants represent valid instrumental variables for MR analyses because (1) they are strongly associated ($P < 5x10^{-8}$) with the exposure (HbF) in several cohorts, and (2) are functionally implicated in HbF production through characterized molecular mechanisms, thus reducing the chance of horizontal pleiotropy.[9-11] To ensure independence between instruments, we calculated the conditional effect size of the SNPs on

normalized HbF levels in the GEN-MOD cohort using a multivariate model (linear regression adjusting for sex, age and 10 first principal components), and used these effect sizes in the CSSCD to test causality with complications using the two-sample MR framework. Association between SNPs and complications were carried out in the CSSCD cohort using logistic regression for stroke and death and linear regression for VOC and ACS rate (normalized using inverse normal transformation), adjusting for sex, age and SCD subtype in every case. We used mRnd to calculate power. For HbF and stroke, we calculated 60% power (N = 1,278, 105 cases, $\alpha = 0.05$, OR = 0.55, variance explained by PTS on HbF = 22.5%). For HbF and VOC or ACS, we calculated power that ranges from 14% (true effect size of HbF on trait = 0.05) to 92% (true effect size of HbF on trait = 0.2) using these assumptions (N = 1,278, $\alpha = 0.05$, observed effect of exposure on outcome = 0.73 for VOC and 0.1 for ACS, variance explained by PTS on HbF = 22.5%). For these power calculations, we normalized both the exposure (HbF) and the outcome (VOC or ACS) to obtain mean = 0 and variance = 1 (to be consistent with built-in functionalities of the power calculation mRnd tool).

We performed all MR analyzes in RStudio (version 1.2.5033) using the TwoSampleMR package (version 0.5.5).[24] We used the multiplicative random-effect inverse variance-weighted (IVW) approach as the main method for each MR analysis, as IVW is robust and has high statistical power in the absence of horizontal pleiotropy.[25] We also performed MR Egger and weighted median MR methods as sensitivity analyses to take into account potential horizontal pleiotropy,[26, 27] but we only considered their results if the corresponding IVW results were significant. 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). We obtained consistent results using Cox proportional-hazards (stroke and death) and quasi-Poisson (VOC and ACS rates) regression.

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),[28] 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.[29] We used the widely accepted alpha threshold of $5x10^{-8}$ to account for the number of variants tested and declare statistical significance, consistent with the GWAS literature. Because blood-cell traits are correlated, we did not correct for the number of phenotypes tested.

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 GWAS meta-analyses (above) and the published non-SCD multi-ancestry meta-analyses from the Blood-Cell Consortium.[8] 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. We corrected for multiple testing by computing a q-value for each SNP-HT association. To compare effect sizes derived from the non-SCD and SCD GWAS meta-analyses, we calculated heterogeneity *P*-values (*P-diff*) based on the following *t* statistic:[30]

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

where b_{SCD} and $b_{non\text{-}SCD}$ are the normalized effect sizes in SCD and non-SCD cohorts, respectively; SE_{SCD} and $SE_{non\text{-}SCD}$ 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. Finally, we computed a q-value from the *P-diff* obtained.

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 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.[31]

Statistical analyses

We performed all statistical analyses using RStudio (version 1.2.5033) or GraphPad Prism (version 9.2.0, GraphPad Software, LLC, CA). We performed only two-sided tests and used a P-value < 0.05 to consider statistical significance, unless a lower threshold is specified. For Cox proportional hazard models, we defined time to event as the time from inclusion to event (i.e. stroke or death), as recorded in the CSSCD database. We used Storey Tibshirani's method to obtain q-values with a 5% false discovery rate.[32]

SUPPLEMENTARY FIGURES

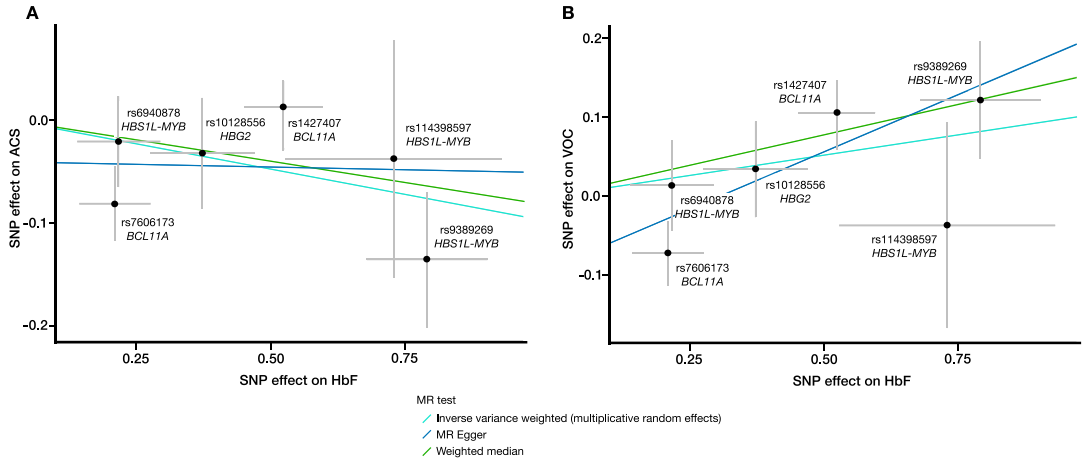

Figure S1. Mendelian randomization (MR) results for fetal hemoglobin (HbF) levels on acute chest syndrome (ACS, **A**) and vaso-occlusive crisis (VOC, **B**). Each dot represents one of the HbF-associated SNP, with its corresponding effect on normalized HbF levels (*x*-axis, standard deviation units) and normalized ACS/VOC rate (*y*-axis, standard deviation units). Horizontal pleiotropy for ACS and VOC was MR-Egger intercept = -0.04, standard error (SE) = 0.05, P = 0.44 and MR-Egger intercept = -0.09, SE = 0.05, P = 0.15, respectively. Heterogeneity for ACS and VOC was I^2 = 19% and I^2 = 34%, respectively. See **Table S5** for details.

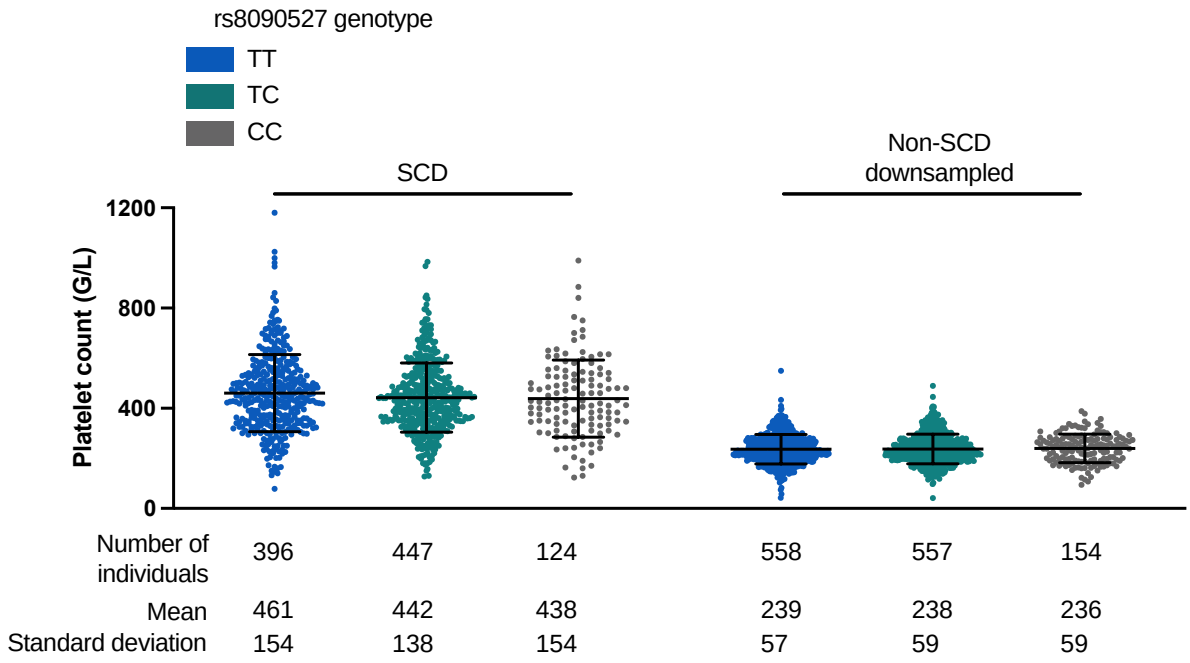

Figure S2. Platelet count distribution across rs8090527 genotypes in SCD (CSSCD cohort) and non-SCD (UK Biobank downsampled cohort) individuals using an additive model.

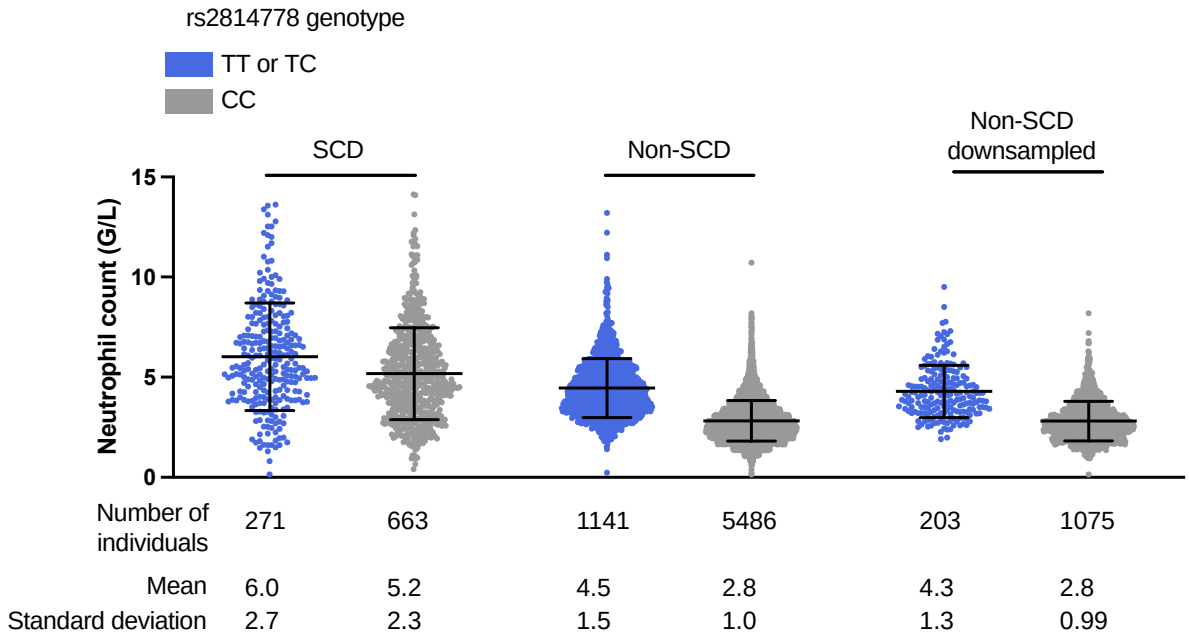

Figure S3. Neutrophil count distribution across rs2814778 genotypes in SCD (CSSCD cohort) and non-SCD (UK Biobank and UK Biobank downsampled cohorts) individuals. We used a recessive model consistent to our findings shown in **Table S7**.

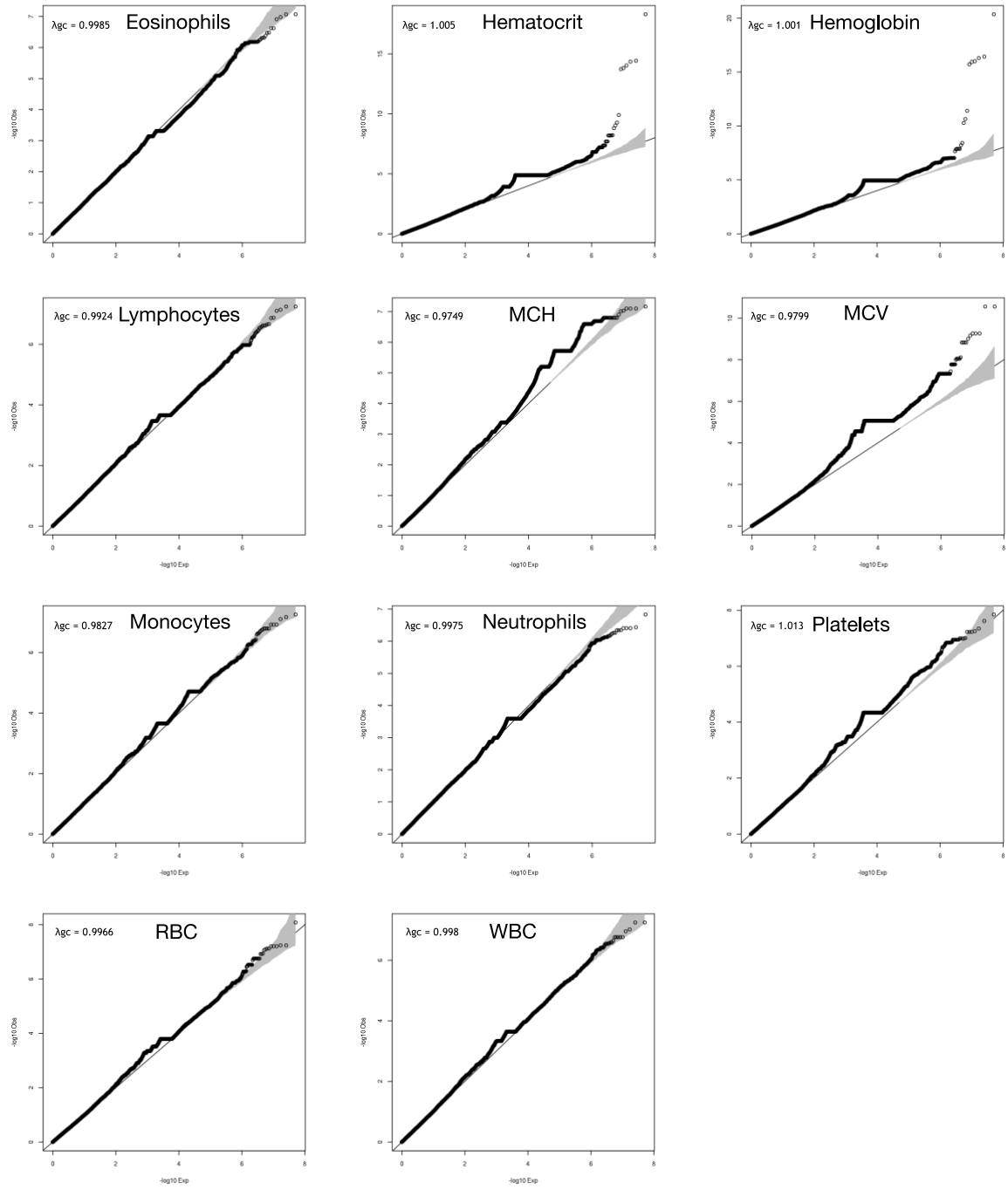

Figure S4. Quantile-quantile plots of the meta-analyses of the GWAS results for hematological traits performed in sickle cell disease patients. λ_{GC}: lambda genomic control, inflation factor. MCH: mean corpuscular hemoglobin, MCV: mean corpuscular volume, RBC: red blood cells, WBC: white blood cells.

SUPPLEMENTARY TABLES

Table S1. Population demographics. Alpha-thalassemia status is not available for the GEN-MOD and Mondor/Lyon cohorts. For vaso-occlusive crises (VOC) and acute chest syndrome (ACS), rates are defined as the number of episodes per year. NA, not available.

Characteristic	CSSCD[33]	GEN-MOD[2, 34, 35]	Mondor/Lyon[36]	OMG[37]	UK Biobank (African-ancestry)[38]
Number of individuals included	1,278	406	372	333	6,627
Sex, male/female	616/662	222/184	139/233	142/191	2,930/3,697
Age (year), mean ± SD	14 ± 12	31 ± 9	35 ± 13	35 ± 13	52 ± 8
α-thalassemia, n (%)	395 (31)	NA	NA	NA	NA
VOC rate, mean ± SD	0.81 ± 1.42	NA	NA	NA	NA
ACS rate, mean ± SD	0.13 ± 0.29	NA	NA	NA	NA
Stroke, n (%)	105 (8.3)	14 (3.4)	NA	NA	NA
Death, n (%)	44 (3.4)	19 (4.7)	NA	NA	NA
Hematocrit (%), mean ± SD	24.79 ± 3.96	25.7 ± 4.53	26.05 ± 4.24	25.11 ± 5.68	40.3 ± 3.85
Hb (g/L), mean ± SD	8.44 ± 1.25	8.76 ± 1.32	8.79 ± 1.27	8.68 ± 1.84	13.61 ± 1.35
MCH (pg), mean ± SD	30.04 ± 2.88	29.30 ± 4.14	29.25 ± 4.93	30.35 ± 4.46	29.66 ± 2.44
MCV (fL), mean ± SD	89.12 ± 8.57	86.90 ± 10.19	85.32 ± 10.81	88.22 ± 11.60	87.80 ± 6.13
RBC (10^9/L), mean ± SD	2.80 ± 0.56	3.00 ± 0.78	3.367 ± 4.68	2.93 ± 0.84	4.61 ± 0.51

HbF (%), mean ± SD	6.45 ± 4.37	6.66 ± 4.81	7.973 ± 6.37	6.40 ± 5.95	NA
WBC (10^6/L), mean ± SD	12.19 ± 3.78	10.60 ± 3.67	10.62 ± 3.01	11.85 ± 4.09	5.81 ± 1.66
Eosinophils (10^6/L), mean ± SD	0.43 ± 0.42	0.27 ± 0.31	NA	NA	0.16 ± 0.14
Basophils (10^6/L), mean ± SD	0.06 ± 0.10	0.09 ± 0.01	NA	NA	0.03 ± 0.03
Lymphocytes (10^6/L), mean ± SD	4.78 ± 2.44	3.59 ± 1.46	NA	NA	2.12 ± 0.67
Monocytes (10^6/L), mean ± SD	0.88 ± 0.56	0.81 ± 0.49	NA	NA	0.39 ± 0.16
Neutrophils (10^6/L), mean ± SD	5.42 ± 2.44	5.84 ± 2.67	NA	NA	3.10 ± 1.27
Platelets (10^9/L), mean ± SD	449.2 ± 147.00	391.0 ± 120.3	382.3 ± 131.44	390.3 ± 143.75	253.43 ± 59.34
MPV (fL), mean ± SD	NA	8.62 ± 1.02	NA	NA	9.68 ± 1.17

Table S2. SNPs selected to derive the fetal hemoglobin (HbF) polygenic trait score (PTS_{HbF}). We selected six independently HbF-associated variants and used the published effect sizes (normalized HbF betas conditioned on all other HbF variants) as weights in an additive PTS model.

SNP coordinate (hg38)	Effect allele	Effect size	SNP ID	Gene	Reference
chr2:60490908_G_T	T	0.6634	rs1427407	*BCL11A*	[9]
chr2:60498316_C_G	C	-0.2632	rs7606173	*BCL11A*	[9]
chr6:135045171_A_G	G	-0.2342	rs6940878	*HBSL1-MYB*	[10]
chr6:135106021_C_T	C	0.581725	rs9389269	*HBSL1-MYB*	[10]
chr6:135107536_A_G	G	0.677248	rs114398597	*HBSL1-MYB*	[10]
chr11:5242453_C_T	T	0.421	rs10128556	β-globin locus	[11]

Table S3. Association between sickle cell disease (SCD)-related complications and hematological traits (HT) or corresponding polygenic trait scores (PTS) in 1,278 genotyped participants from the CSSCD. We used Cox proportional-hazards models for stroke, and quasi-Poisson regression for acute chest syndrome (ACS) and vaso-occlusive crises (VOC) rates. We first tested the association between raw HT and complications, correcting for age at recruitment, sex, α-thalassemia status, and 10 first principal components. We then replaced the HT by its corresponding normalized PTS. EOS: eosinophils, LYM: lymphocytes, MCH: mean corpuscular hemoglobin, MCV: mean corpuscular volume, MON: monocytes, MPV: mean platelet volume, NEU: neutrophil count, PLT: platelet count, WBC: white blood cell count, HR: hazard ratio, SE: standard error.

	Hematological trait			**Normalized PTS for hematological trait**		
	HR or Beta	95% CI or SE	P-value	HR or Beta	95% CI or SE	P-value
Stroke, Cox proportional-hazards model						
MCH	1.03	0.92-1.14	0.62	1.00	0.83-1.22	0.97
MCV	1.03	0.99-1.06	0.11	1.02	0.84-1.23	0.84
HbF	0.89	0.83-0.95	0.0005	0.74	0.60-0.91	0.004
WBC	1.10	1.04-1.16	0.001	1.00	0.82-1.22	0.99
EOS	1.06	0.63-1.78	0.83	0.85	0.70-1.03	0.90
LYM	1.10	1.00-1.204	0.04	1.01	0.83-1.23	0.90
MONO	1.41	0.97-2.05	0.07	0.89	0.73-1.09	0.24
NEU	1.11	1.02-1.22	0.02	0.98	0.80-1.19	0.82
PLT	1.00	1.00-1.00	0.60	0.90	0.74-1.10	0.30
ACS, quasi-Poisson regression						
MCH	-0.03	0.02	0.17	0.01	0.06	0.80
MCV	-0.01	0.01	0.16	0.02	0.06	0.69
HbF	-0.07	0.02	0.00002	-0.20	0.06	0.0005
WBC	0.03	0.01	0.05	-0.04	0.06	0.51

EOS	0.25	0.12	0.04	-0.02	0.05	0.75
LYM	0.04	0.02	0.06	-0.04	0.05	0.50
MONO	0.03	0.10	0.78	-0.06	0.06	0.30
NEU	0.03	0.02	0.17	-0.01	0.06	0.83
PLT	0.00	0.00	0.44	-0.01	0.05	0.92
VOC, quasi-Poisson regression						
MCH	-0.04	0.02	0.06	0.04	0.05	0.42
MCV	-0.01	0.01	0.24	0.10	0.05	0.03
HbF	-0.03	0.01	0.036	0.06	0.05	0.21
WBC	-0.01	0.01	0.59	-0.02	0.05	0.64
EOS	-0.03	0.13	0.84	0.03	0.05	0.50
LYM	-0.04	0.03	0.10	0.02	0.05	0.68
MONO	0.03	0.10	0.72	0.03	0.05	0.47
NEU	0.01	0.02	0.70	0.01	0.05	0.82
PLT	0.00	0.00	0.15	0.03	0.05	0.60

Table S4. The polygenic trait score (PTS) for fetal hemoglobin (PTS_{HbF}) levels improves the association with vaso-occlusive crises (VOC) rates for patients with low (<10%) HbF levels. We carried out these analyses in 1,139 CSSCD participants. To compare the statistical models, we performed an analysis of deviance and compared a baseline model (HT, age, sex, α-thalassemia, and 10 first principal components) with a model that included the same predictors as well as PTS_{HbF}.

	HbF < 10% (n = 930)		**HbF ≥ 10% (n = 209)**	
	Beta (SE)	**P-value**	**Beta (SE)**	**P-value**
Baseline model (VOC association without PTS_{HbF})				
HbF	-0.002 (0.022)	0.91	-0.14 (0.046)	0.0027
Age	0.018 (0.004)	$1.9x10^{-5}$	0.031 (0.008)	0.00014
Sex	-0.009 (0.109)	0.94	-0.092 (0.24)	0.70
α-thalassemia	0.04 (0.03)	0.23	-0.0004 (0.07)	0.99
Complete model (VOC association with PTS_{HbF})				
HbF	-0.022 (0.024)	0.37	-0.14 (0.045)	0.0026
Age	0.017 (0.004)	$6.9x10^{-5}$	0.028 (0.008)	0.0007
Sex	0.001 (0.110)	1.00	-0.092 (0.23)	0.70
α-thalassemia	0.040 (0.029)	0.16	-0.006 (0.07)	0.92
PTS_{HbF}	0.129 (0.065)	0.05	0.156 (0.13)	0.25
Comparison between the two models				
	χ^2 (1 degree-of-freedom)	**P-value**	**χ^2** (1 degree-of-freedom)	**P-value**
Residual deviance difference	9.3	0.002	2.17	0.14

Table S5. Mendelian randomization (MR) results for hematological traits (HT) and sickle cell disease (SCD) complications. We selected the SNPs included in the polygenic trait scores (PTS) as instruments. We used a two-sample MR approach to test the causality of fetal hemoglobin (HbF) with acute chest syndrome (ACS) rate, vaso-occlusive crises (VOC) rate and stroke. We used inverse variance weighted as the main method and MR Egger and weighted median only as sensitivity analyses. Thus, we did not consider MR result for HbF on VOC as significant (see **Supplementary methods**). We obtained consistent results using Cox proportional-hazards (stroke and death) and quasi-Poisson (VOC and ACS rates) regression.

Outcome	Exposure (HT)	Method	Number of SNPs	Beta	Standard error	P-value
Stroke (N_{cases} = 104, $N_{controls}$= 1,168)	HbF	Inverse variance weighted (multiplicative random effects)	6	-0.499	0.147	6.50E-04
		MR Egger	6	-0.305	0.509	0.582
		Weighted median	6	-0.374	0.283	0.186
	HbF (excluding rs114398597)	Inverse variance weighted (multiplicative random effects)	5	-0.454	0.098	3.72E-06
		MR Egger	5	-0.373	0.286	0.193
		Weighted median	5	-0.191	0.518	0.737
ACS rate (N = 1,271)	HbF	Inverse variance weighted (multiplicative random effects)	6	-0.099	0.052	0.059
		MR Egger	6	-0.010	0.064	0.933
		Weighted median	6	-0.083	0.116	0.19
VOC rate (N = 1,271)	HbF	Inverse variance weighted (multiplicative random effects)	6	0.103	0.066	0.12
		MR Egger	6	0.290	0.118	0.071
		Weighted median	6	0.155	0.067	0.021

Table S6. SNPs with a significantly different effect size on hematological trait (HT) between sickle cell disease (SCD) and non-SCD individuals. For all the 4,201 SNP-HT pairs present in the 11 PTS, we compared the meta-analyzed effect sizes between SCD and non-SCD datasets using the heterogeneity *t* statistic (see **Supplementary methods**). We found only two SNPs with both a significant difference in the effect size and a significant association with HT in the SCD meta-analysis after correction for multiple testing (q-value < 0.05). Note that for the Duffy/*DARC* variant (rs2814778), the effect allele frequency (EAF) is very different between the SCD cohorts (C-allele, 85%) and the multi-ancestry meta-analyses (C-allele, 2%) because ~98% of the samples were of non-African ancestry. We confirmed the difference in Duffy/*DARC* variant effect size using a downsampled cohort of non-SCD African-ancestry individuals from the UK Biobank (n = 1,278). The association between rs8090527 and PLT count could not be replicated in downsampled UK Biobank cohort nor in 333 SCD participants from the OMG cohort (C-allele frequency = 0.34 ; Beta_C_allele = 0.0961 ; standard error (SE) = 0.0842 ; P-value = 0.25).

				SCD (CSSCD cohort, N_{max}= 1,015)				**Non-SCD (BCX, N_{max}= 716,308)**			**Non-SCD downsampled (African-ancestry, UK Biobank, N_{max}= 1,278)**			**Heterogeneity ($\beta_{SCD} = \beta_{non\text{-}SCD}$?)**		
HT	**Position (hg38)**	**SNP**	**Ref/Effect allele**	**EAF**	**Beta (SE)**	**P-value**	**q-value**	**EAF**	**Beta (SE)**	**P-value**	**EAF**	**Beta (SE)**	**P-value**	**P-value t stat**	**q-value t stat**	**P-value t stat (downsampled)**
NEU	1:159204893	rs2814778	T/C	0.85	-0.288 (0.06)	4.1E-6	0.015	0.02	-0.546 (0.012)	0	0.91	-1.050 (0.71)	1.1E-44	4.1E-6	0.031	9.1E-16
PLT	18:51257657	rs8090527	T/C	0.37	-0.144 (0.03)	1.9E-5	0.036	0.51	-0.012 (0.002)	1.1E-13	0.34	-0.001 (0.04)	0.99	1.9E-5	0.037	0.004

Table S7. Comparison of the Duffy/*DARC* null rs2814778 effect on neutrophil and white blood cell (WBC) counts between the additive and recessive genetic models. We computed the effect of rs2814779 on normalized neutrophil and WBC counts (after adjusting for age and sex) using both additive and recessive model. We used the first 10 principal components as covariates. In the UK Biobank, we only analyzed participants of African ancestry. The effect size (Beta and standard error [SE]) is for the C-allele (additive model) or the CC genotype (recessive model). We calculated the phenotypic variance explained only for nominally significant associations. Neutrophil count is not available in the Mondor/Lyon cohort. EAF: effect allele frequency, N: sample size.

			Additive			**Recessive**		
Cohort	**N**	**EAF**	**Beta (SE)**	**P-value**	**Variance (%)**	**Beta (SE)**	**P-value**	**Variance (%)**
Neutrophil count								
CSSCD	934	0.842	-0.313 (0.065)	$1.58x10^{-6}$	2.61	-0.353 (0.073)	$1.64x10^{-6}$	3.34
GEN-MOD	400	0.939	-0.083 (0.19)	0.661	-	-0.019 (0.209)	0.927	-
UK Biobank	6,564	0.906	-0.986 (0.03)	$2.03x10^{-210}$	16.51	-1.166 (0.033)	$5.96x10^{-239}$	23.12
White blood cell count								
CSSCD	1,014	0.845	-0.164 (0.056)	0.004	0.71	-0.178 (0.064)	0.005	0.84
GEN-MOD	400	0.939	-0.135 (0.189)	0.474	-	-0.092 (0.208)	0.657	-
Mondor/ Lyon	322	0.935	-0.053 (0.187)	0.775	-	-0.151 (0.214)	0.481	-
UK Biobank	6,584	0.906	-0.879 (0.032)	$1.56x10^{-150}$	13.13	-1.041 (0.036)	$4.96x10^{-171}$	18.42

Table S8. Comparison of neutrophil count between Duffy-positive individuals with a polygenic trait score (PTS) in the lowest quintile and Duffy-negative individuals with a PTS in the highest quintile from the CSSCD, African-ancestry UK Biobank individuals, and a downsampled cohort of African-ancestry UK Biobank participants (to match the number of genotyped CSSCD participants).

	Duffy+ lowest PTS quintile		**Duffy- highest PTS quintile**		
Cohort	**N**	**Neutrophil count (mean ± SD)**	**N**	**Neutrophil count (mean ± SD)**	**P-value**
CSSCD	62	5.8 ± 2.6	129	5.4 ± 2.1	0.36
UK Biobank	313	4.1 ± 1.3	1,034	3.0 ± 1.1	3.05×10^{-33}
Downsampled UK Biobank	62	4.0 ± 1.4	179	3.0 ± 1.1	5.82×10^{-9}

Table S9. Results of the GWAS meta-analyses for hematological traits (HT) in sickle cell disease patients. We only report variants that reached genome-wide significance ($P < 5x10^{-8}$). Except for the variant on chromosome (chr) 3 associated with platelet (PLT) count, all other variants associate with red blood cell (RBC) traits and are at the *BCL11A* (chr 2) and *HBS1L-MYB* (chr 6) loci. EAF: effect allele frequency, HCT: hematocrit, HGB: hemoglobin.

HT	Chr	Position (hg38)	Ref	Alt	EAF	Beta	Standard error	P-value	Direction	I^2	Hetero. P-value
HCT	2	60490908	T	G	0.2689	0.303	0.034	5.09E-19	+++	0	0.9612
HGB	2	60490908	T	G	0.2697	0.3187	0.0338	4.47E-21	+++	0	0.8823
RBC	2	60490908	T	G	0.2686	0.1918	0.0333	8.30E-09	+++	28.8	0.2453
HCT	2	60492835	C	A	0.2935	0.2512	0.0327	1.54E-14	+++	0	0.8708
HGB	2	60492835	C	A	0.2935	0.2707	0.0326	9.93E-17	+++	0	0.7614
HCT	2	60493816	A	G	0.2933	0.2504	0.0327	1.88E-14	+++	0	0.8406
HGB	2	60493816	A	G	0.2932	0.27	0.0326	1.17E-16	+++	0	0.7394
HCT	2	60494905	T	C	0.3856	-0.1801	0.031	5.94E-09	---	0	0.5633
HGB	2	60494905	T	C	0.3844	-0.1795	0.0309	6.11E-09	---	0	0.4842
HCT	2	60495961	C	CA	0.2842	0.2561	0.033	9.32E-15	+++	0	0.836
HGB	2	60495961	C	CA	0.2846	0.2705	0.0329	1.95E-16	+++	0	0.7371
HCT	2	60496951	T	C	0.2891	0.258	0.0329	4.53E-15	+++	0	0.8085
HGB	2	60496951	T	C	0.2895	0.2746	0.0327	5.05E-17	+++	0	0.6949
HCT	2	60496952	G	T	0.2902	0.2583	0.0329	3.82E-15	+++	0	0.8844
HGB	2	60496952	G	T	0.2903	0.2756	0.0327	3.73E-17	+++	0	0.735
HCT	2	60498316	C	G	0.4224	-0.183	0.0303	1.58E-09	---	0	0.7167
HGB	2	60498316	C	G	0.4213	-0.178	0.0302	3.79E-09	---	0	0.5893
PLT	3	123967670	C	T	0.0566	-0.4085	0.072	1.42E-08	---	49.3	0.1389
HCT	6	135078218	A	G	0.0267	0.5813	0.095	9.25E-10	+++	46	0.1572
HGB	6	135078218	A	G	0.0269	0.6205	0.0945	5.24E-11	+++	66.8	0.04897
HCT	6	135086355	G	A	0.0287	0.5892	0.0916	1.25E-10	+++	41.9	0.1789
HGB	6	135086355	G	A	0.0288	0.6324	0.0912	3.99E-12	+++	65	0.05735
HCT	6	135097526	C	G	0.0281	0.5744	0.0925	5.35E-10	+++	46.5	0.1544
HGB	6	135097526	C	G	0.0282	0.6157	0.0921	2.28E-11	+++	68.5	0.04185

SUPPLEMENTARY REFERENCES

1. Bae HT, Baldwin CT, Sebastiani P, et al. Meta-analysis of 2040 sickle cell anemia patients: BCL11A and HBS1L-MYB are the major modifiers of HbF in African Americans. Blood. 2012;120(9):1961-1962.
2. 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.
3. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 2021;590(7845):290-299.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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 e1114.
9. 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.
10. 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.
11. 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.
12. 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.
13. Lettre G, Sankaran VG, Bezerra MAC, et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and β-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.
14. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc. 1999;94(446):496-509.
15. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav Res Methods. 2009;41(4):1149-1160.
16. 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.
17. Nouraie M, Darbari DS, Rana S, et al. 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. 2020;95(7):766-774.
18. 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.
19. Platt OS, Thorington BD, Brambilla DJ, et al. Pain in Sickle Cell Disease. N Engl J Med. 1991;325(1):11-16.

20. 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.
21. 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.
22. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(7): s13742–015–0047–8.
23. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):1497-1501.
24. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
25. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. 2019;4:186.
26. 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.
27. Schmidt AF, Dudbridge F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. Int J Epidemiol. 2018;47(4):1217-1228.
28. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016;32(9):1423-1426.
29. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-2191.
30. Winkler TW, Justice AE, Graff M, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLoS Genet. 2015;11(10):e1005378.
31. Fisher RA. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. Trans R Soc Edinburgh. 1918;52(2):399-433.
32. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-9445.
33. 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.
34. 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.
35. Ilboudo Y, Garrett ME, Bartolucci P, et al. Potential causal role of l-glutamine in sickle cell disease painful crises: A Mendelian randomization analysis. Blood Cells Mol Dis. 2021;86:102504.
36. Pincez T, Lee SSK, Ilboudo Y, et al. Clonal hematopoiesis in sickle cell disease. Blood. 2021;138(21):2148-2152.
37. 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.
38. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-209.